Journal of Engineering and Sustainable Development
Vol. 24 No.02, March 2020
ISSN 2520-0917
https://doi.org/10.31272/jeasd.24.2.1

# OUTLIER DETECTION TECHNIQUE USING CT-OCSVM AND FUZZY RULE-BASED SYSTEM IN WIRELESS SENSOR NETWORKS

Hussein Hassan Shia[1], Dr. Mohammed Ali Tawfeeq[2], *Dr. Sawsan Mousa Mahmoud[3]

1) M.Sc. Student, Computer Engineering Department, Mustansiriyah University, Baghdad, Iraq.
2) Assistant Prof., Computer Engineering Department, Mustansiriyah University, Baghdad, Iraq.
3) Assistant Prof., Computer Engineering Department, Mustansiriyah University, Baghdad, Iraq.

Received 13/ 1/2019 Accepted in revised form 25/2/ 2019 Published 1/3/2020

**Abstract:** The development of Wireless Sensor Networks (WSNs) has been attained in the past few years due to its important using in wide range of application. The readings of data derived from WSN nodes are not always accurate and may contain abnormal data. This paper proposed an anomaly detection and classification algorithm in WSNs. At first, an integration of Contourlet Transform (CT) algorithm and One Class Support Vector Machine (OCSVM) algorithm (CT-OCSVM) is utilized to detect outliers then Fuzzy Inference System (FIS) is used to identify the source of these outliers. The underlying aim of this paper focuses on treating the collected streams of data as raw datum of an image, which is then passed through some filters using CT to get compressed size of directional subbands coefficients. The coefficients of CT are examined by OCSVM algorithm to detect anomalies. Finally the source of anomalies is identified based on using FIS and by exploiting the spatial temporal correlation existing between the sensed data. The integrated algorithm is tested using different types of filters. Real datasets collected from a small WSN constructed in a local lab are used for testing the integrated algorithms. The simulation results have shown a high rate of accurate classification with high detection rate and low false alarm rate.

**Keywords**: *Contourlet Transform; Fuzzy Inference System; One class support vector machine; Outlier; Wireless Sensor Networks.*

# اكتشاف القيم المتطرفة باستخدام التحويلات الكنتورية ونظام الاستدلال الضبابي في الشبكات اللاسلكية

**الخلاصة:** لقد تم تطوير شبكات الاستشعار اللاسلكية (WSNs) في السنوات القليلة الماضية بسبب استخدامها المهم في نطاق واسع من التطبيقات. لا تكون قراءات البيانات المشتقة من عقد WSN دقيقة دائمًا وقد تحتوي على بيانات غير طبيعية. في هذا المقال تم تقديم خوارزمية كشف وتصنيف البيانات غير الطبيعية في WSNs. في البداية ، يتم استخدام خوارزمية التحويلات الكونتورية (CT) وخوارزمية One Class Support Vector Machine معا لتشكيل (CT-OCSVM) لاكتشاف القيم المتطرفة ، ثم يستخدم نظام الاستدلال المشوش (FIS) لتحديد مصدر هذه القيم المتطرفة. ويركز الهدف الأساسي من هذه المقالة على معالجة شريط البيانات المجمعة كمسند خام لصورة ، والتي يتم تمريرها بعد ذلك من خلال بعض المرشحات باستخدام CT للحصول على حجم مضغوط لمعاملات فرعي الاتجاه. يتم فحص معاملات CT من خوارزمية OCSVM للكشف عن الحالات الشاذة. وأخيرًا ، يتم تحديد مصدر الانحرافات بناءً على استخدام FIS واستغلال الارتباط الزماني المكاني الموجود بين البيانات المستشعرة. تم اختبار الخوارزمية المتكاملة باستخدام أنواع مختلفة من الفلاتر. تم استخدام مجموعات البيانات الحقيقية التي تم جمعها من WSN صغيرة شيدت في مختبر محلي لاختبار الخوارزميات المتكاملة. أظهرت نتائج المحاكاة نسبة عالية من التصنيف الدقيق مع معدل اكتشاف عالي ومعدل إنذار زائف منخفض.

**Corresponding Author*: sawsan.mahmoud@uomustansiriyah.edu.iq

## 1. Introduction

Wireless Sensor Networks (WSNs) have received much attention in the recent and past few years. They are composed of a large number of low energy, low cost and multifunctional sensors nodes that are networked together. WSNs are widely used and have gained attention in various fields including: personal application, health care application, military and surveillance application which they used to improve the life of humans [1]. The purpose of using WSNs is not only to collect raw deployment data, but essentially to decompose these data and for decision making [2]. The collected raw sensor data especially in large WSNs may be inaccurate and of low quality [3]. The inaccuracy in gathering data may be made due to one or combination of three main reasons [4]. First, the resource constraints in the sensor device itself such as energy, storage and processing. Second, WSNs may be deployed in harsh and unintended environments. While the third reason is due to malicious attack such as selective forwarding, sinkhole, wormhole, denial of service and blackhole attack which may also yield inaccurate and low quality data [5]. This inaccurate data is called anomalies. Another type of outliers causes due to natural phenomena in the environments such as earthquake, rainfall, and forest fire. Therefore, three source of outlier that can be noticed includes: malicious attacks, event, and noise and error. The algorithms of anomaly detection allow cleaning, refining, and improving the robustness of the collected raw data under the presence of faulty and malicious sensors node [2]. In this paper Contourlet Transform (CT) technique is used because of its ability to extract relevant information about the outliers. To the best of our knowledge there is no study to use and compare the effectiveness of an integration of Contourlet Transform (CT) and One Class Support Vector Machine (OCSVM) for anomaly detection in WSNs. Afterward, FIS is utilized to identify the source of these outlier by exploiting the temporal spatial correlation existing between the sensed data The aim of this paper is to propose an effective anomaly detection and classification algorithm in WSNs by using the mentioned techniques.

This paper is organized as follows. Section II reviews some related works. Section III presents the proposed anomaly detection algorithm. Section IV introduces the results and evaluation of the proposed algorithm. Finally, the conclusion of this work is presented in section V.

## 2. Related Works

Many researchers have introduced algorithms and techniques to detect and classify outliers in WSNs. For example, Zhang et al. [6] have presented three techniques for anomaly detection in WSNs based on OCSVM algorithm. The outliers were identified by using spatial and temporal correlations existed between sensors data. These techniques can achieve high detection rate with low false alarm rate but suffer from the high computational capacity. Chitradevi et al. [7] have proposed an outlier detection technique in WSNs by using both of Principle Component Analysis (PCA) and Mahalanobis Distance (MD) techniques. PCA technique is used to predict an accurate model of the normal pattern of the sensed data, while the MD technique is used to detect abnormal data. Siripanadorn et al. [8] have presented an integration of Discrete Wavelet Transform DWT and Self-Organizing Map (SOM) for outlier detection in WSNs. The raw data are encoded by utilizing DWT technique afterward this encoded data infused to SOM for outlier detection. Takianngan and Usaha [9] have proposed an integration of Discrete Wavelet Transform (DWT) algorithm and OCSVM algorithm for outlier detection in WSNs in which, the testing data is encoded by using DWT algorithm then this encoded data injected to OCSVM algorithm for anomaly detection. Zhang et al. [10] have presented five techniques for anomaly detection in WSNs using spatial temporal correlation. The proposed techniques are able to detect outliers and recognize between error and event. Zhang et al. [11] have proposed two online and distributed methods for anomaly detection in WSNs. These methods are based on hyper-ellipsoid OCSVM algorithm to increase true alarm rate and reduce false alarm rate. Salem et al. [12] have used linear regression to predict the data reading of the sensor from its neighbors reading. The proposed technique can detect outlier and recognize between error and unusual event in medical WSN. Mahmoud et al. [13] have presented an outlier detection technique for smart home WSN using a combination of PCA and fuzzy rule based system. According to the authors the technique can successfully detect the outlier and recognize the data as normal or abnormal and determined the degree of abnormality. Kamal et al. [14] have proposed a technique for outlier detection and identification in WSNs by using Euclidean distance and fuzzy rule based approach. The proposed technique is able to recognize the outlier as error or event. Garcia et al. [15] presented comparison of four anomaly detection methods in WSNs including MD, OCSVM, Local Outlier Factor (LOF) and hierarchical clustering. The data is collected from smart city of Barcelona and were examined with different type of attack. The results show that OCSVM algorithm is the most appropriate technique by achieving the highest detection rate and the lowest false alarm rate. Meng et al. [16] proposed an anomaly detection model based on PCA which is able to detect normal and abnormal user behavior accurately and efficiently. The user behavior and web browsing behavior are decomposed for such characteristic.

## 3. Proposed Technique

This section introduces the proposed algorithm in detecting and identifying anomalies in WSNs which includes the integration of CT algorithm and OCSVM algorithm to construct CT-OCSVM algorithm in addition to the classification process which is based on the use of FIS technique.

### *3.1. Contourlet Transform*

Pyramidal Direction Filter Bank (PDFB) or Contourlet transform is proposed by Do and Vetterli [17] as directional multi-resolution image representation. It is a two dimensional transform developed in discrete domain. PDFB is a combination of Laplacian Pyramid (LP) and Directional Filter Bank (DFB). At each level the LP down sampled and decomposed the input data into low pass and high pass subbands to determine point discontinuities in the data. Fig. 1 shows the decomposition process, in which H and G are analysis and synthetic filter while, M represents the sampling matrix. DFB are applied to the high pass subbands of the LP to link discontinuities to linear structure. CT decomposes the input data into different directional subbands at multiple scale to achieve critical sampling and it will decompose the 2-D frequency spectrum into wedge-shaped region as depicts in Fig. 2 [17]. Contourlet transform provides set of features which are, multiresolution analysis, anisotropy, localization, directionality and critical sampling. The general Contourlet filter bank is shown in Fig. 3.

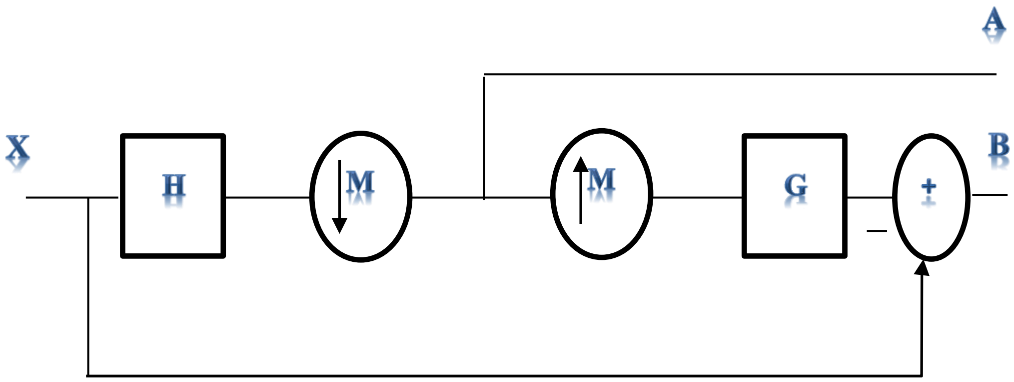



Figure 1. LP decomposition

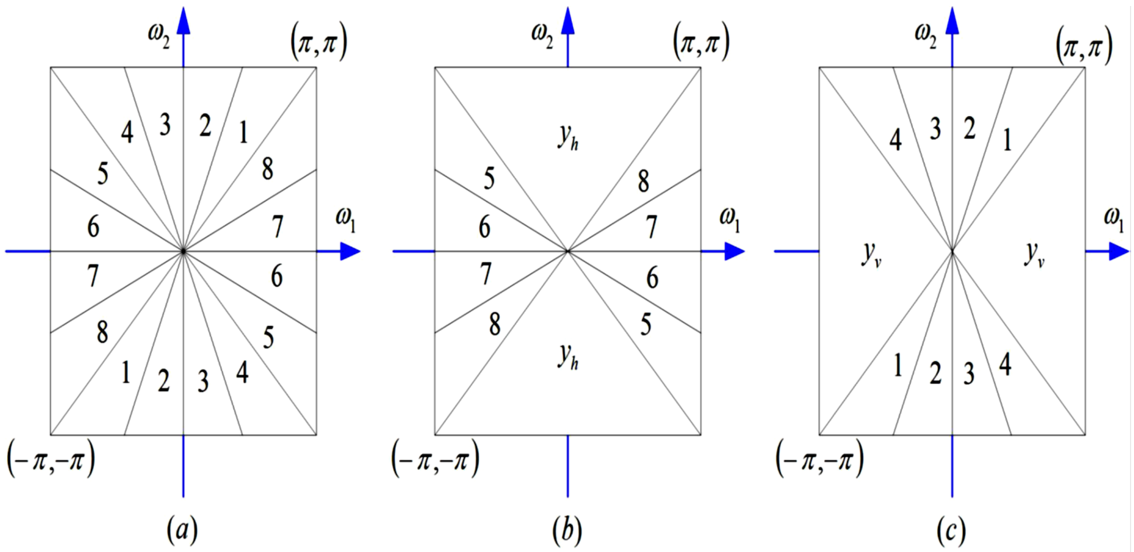



Figure 2. Frequency division of DFB (a) three level frequency division (b) vertical DFB (c) horizontal DFB

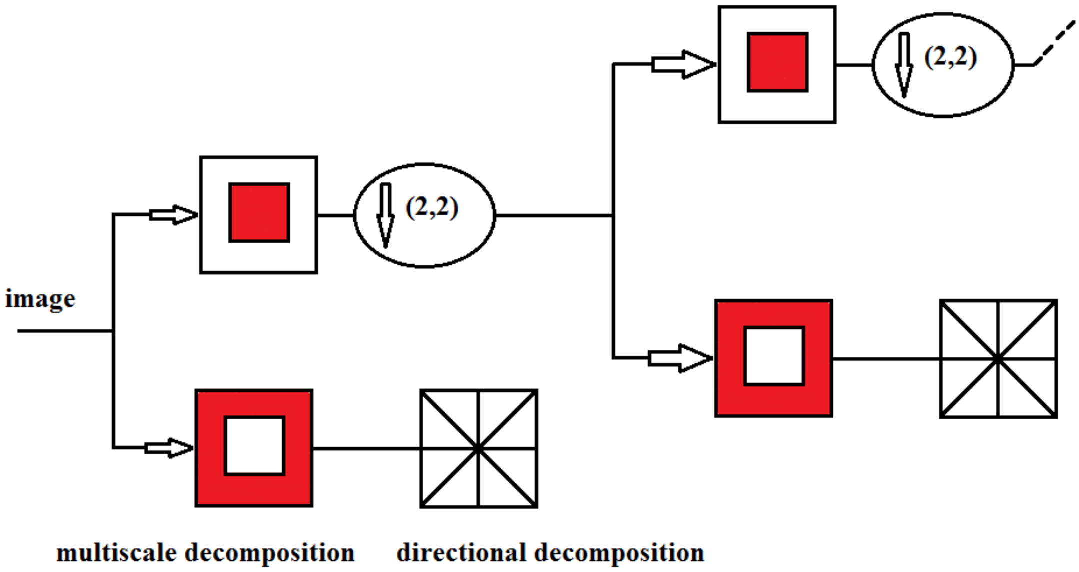



Figure 3. Original Contourlet transform with two decomposition level

### *3.2. One Class Support Vector Machine*

OCSVM is a machine learning classification model which is vastly applied in the context of intrusion detection. OCSVM requires two steps. First, training the classification model by using instances set of data. Then, the trained model is used to categorize instances into one of the learned (normal/abnormal) class. It defines hyperplane to separate between different classes which maximize the margin. The data is mapped into higher dimension feature space which can be easily separated [18].

### *3.3. Integration of CT and OCSVM*

In this work, CT algorithm is integrated with the OCSVM algorithm to construct the proposed anomaly detection CT-OCSVM algorithm. The operation of the proposed algorithm is started by converting the gathered raw data of the sensor nodes to 2-D image matrix. Then, the CT part of the algorithm encoded the data of the 2-D image by making dimension reduction without losing any significant information. Reducing the dimensionality of the data helps in prolonging the lifetime of the nodes, as well as increasing the performance of the network by reducing the communication overhead in WSNs. CT encoded coefficients are examined by OCSVM part of the integrated proposed algorithm to detect anomalies. Since OCSVM works with the principle of unary-classification, it will classify the testing data as either normal or abnormal. Fig. 4 shows a block diagram of the proposed algorithm.

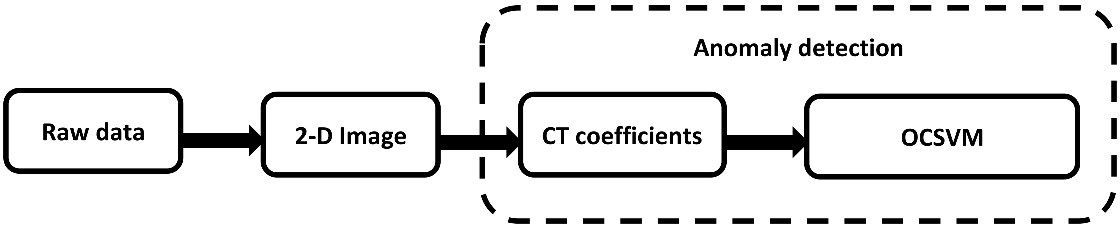



Figure 4. Block diagram of the proposed algorithm

### *3.4. Outlier Classification using FIS*

One of the most important issues in making the appropriate decision in WSNs is to classify the identity of the detected outlier. These outliers may be due to a real event which required some actions, or it may be as a result of some error (due to noise, malfunction …etc.). In this work, FIS is utilized for this purpose and by exploiting the spatial temporal correlation existing between the sensed data. Fuzzy logic is one of the most essential methodologies that are widely utilized in machine control and in manipulating information to model different engineering issues. It was introduced by Lotfi A. zaden in 1960 [19].

Three inputs are utilized including:

- $R_i(t)$: current CT-OCSVM value of node i which is detected as outlier.
- $R_i(t-1)$: previous CT-OCSVM value of node i.
- $R_N(t)$: current CT-OCSVM values of the neighbor nodes that shared the same sensing area with node i.

According to the output of the proposed CT-OCSVM technique, the detected outlier values are either negative or positive. This fact can be used as an additional fuzzy parameter that helps in the classification process. As a result, the fuzzy set terms for each of the three input variable are considered to be: Negative outlier (Ng), Normal Data (Nr), and Positive outlier (Ps).

In OCSVM algorithm, the input data is mapped to two categories separated by hyperplane with a clear gap of maximum possible wide, therefore in node i, the fuzzy term Nr in the variable Ri(t) will not be taken into account as its result is definitely determined as normal data.

To overcome the problem of divergence in the scale of the input data, the input to the FIS is normalized to be within the range [0, 1]. The membership functions are chosen to be trapezoidal functions as shown in Fig. 5 to accommodate the output from CT-OCSVM algorithm.

Three fuzzy terms with triangular membership function are used to distinguish the output of the proposed FIS; Event (E), Instantaneous-error (Ie), and Continuous-error (Ce) as demonstrated in Fig. 6. Event refers to the case of detecting actual phenomena or real event that requires to raise alarm or to take the appropriate action. Instantaneous-error refers to the case in which the algorithm categorizes the current data of node i at the instant t as outlier, while the previous time step categorization for this node is normal. This will leads to uncertainty, which requires checking the status of the neighbor nodes at the instantaneous time and the status of node i at the next instant (t+1). Finally, Continues-error refers to the case in which the error is happened in both current and previous measurements and it may also exist in the neighbor reading in the case of different error levels. Similarly, a wait for checking must be done if this error continues for the next reading and to do an appropriate decision. In such a case, the error can be occurred due to node malfunction or high noise which required taking the right decision such as exchange the erroneous node, removing it from the network, or repairing it.

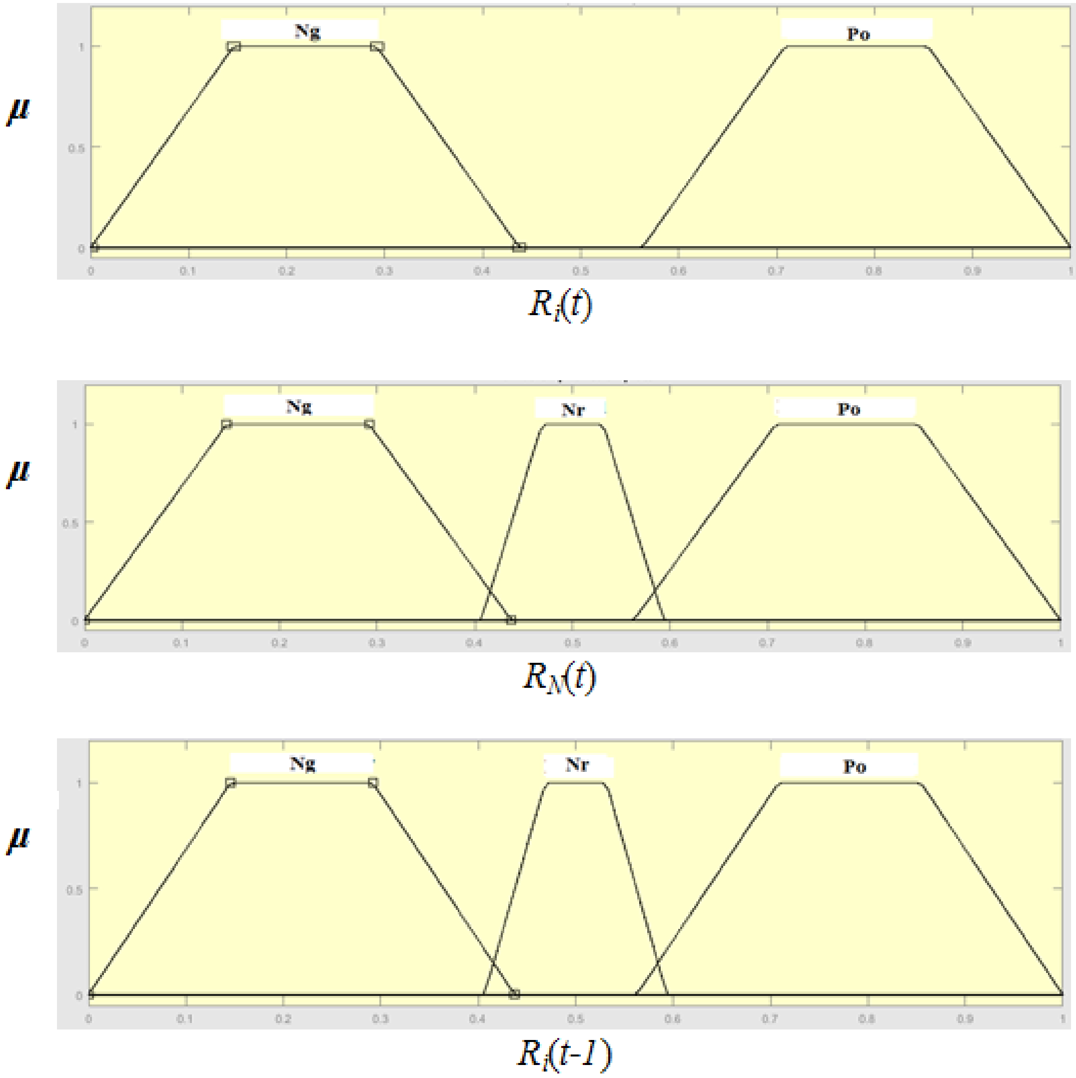



Figure 5. Membership functions of the input variables

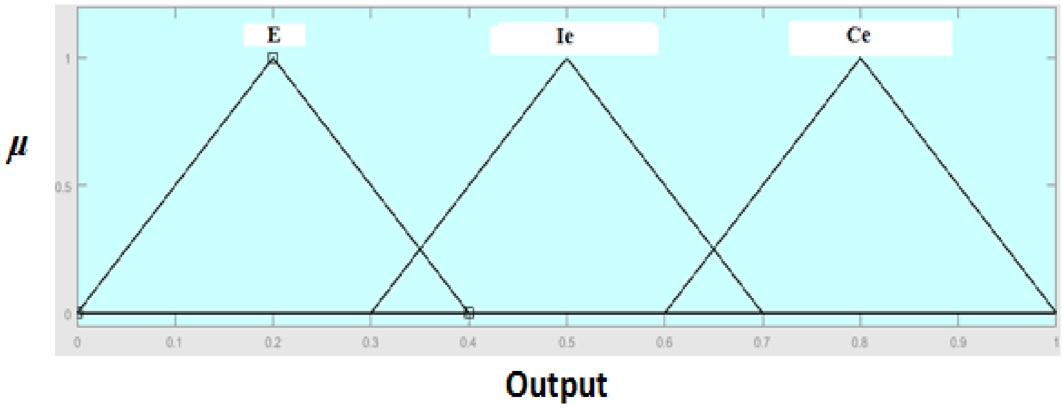



Figure 6. Membership functions of the output variables.

The generic form of the fuzzy rule is as follows:
If $R_i(t)$ is (A) and $R_i(t-1)$ is B and $R_N(t)$ is C THEN an outlier is D.
Where A, B represent the linguistic values (fuzzy labels as is mentioned before) for the current and previous CT-OCSVM values of node i respectively, and C is the neighbor nodes linguistic value that shared the same sensing area with node i. The prescribe output value is represented by the linguistic value D. In this work, Fuzzy rules for outlier classification with the following configuration are used:

R1: IF $R_i(t)$ is Ps and $R_N(t)$ is Ps THEN an outlier is E
R2: IF $R_i(t)$ is Ps and $R_i(t-1)$ is Ps and $R_N(t)$ is NR THEN an outlier is Ce
R3: IF $R_i(t)$ is Ps and $R_i(t-1)$ is Ps and $R_N(t)$ is Ng THEN an outlier is Ce
R4: IF $R_i(t)$ is Ps and $R_i(t-1)$ is Ng and $R_N(t)$ is Nr THEN an outlier is Ce
R5: IF $R_i(t)$ is Ps and $R_i(t-1)$ is Ng and $R_N(t)$ is Ng THEN an outlier is Ce
R6: IF $R_i(t)$ is Ps and $R_i(t-1)$ is Nr and $R_N(t)$ is Nr THEN an outlier is Ie
R7: IF $R_i(t)$ is Ps and $R_i(t-1)$ is Nr and $R_N(t)$ is Ng THEN an outlier is Ie
R8: IF $R_i(t)$ is Ng and $R_i(t-1)$ is Ps and $R_N(t)$ is Ps THEN an outlier is Ce
R9: IF $R_i(t)$ is Ng and $R_i(t-1)$ is Ps and $R_N(t)$ is Nr THEN an outlier is Ce
R10: IF $R_i(t)$ is Ng and $R_i(t-1)$ is Ng and $R_N(t)$ is Ps THEN an outlier is Ce
R11: IF $R_i(t)$ is Ng and $R_i(t-1)$ is Ng and $R_N(t)$ is Nr THEN an outlier is Ce
R12: IF $R_i(t)$ is Ng and $R_i(t-1)$ is Nr and $R_N(t)$ is Ps THEN an outlier is Ie
R13: IF $R_i(t)$) is Ng and $R_i(t-1)$ is Nr and $R_N(t)$ is Nr THEN an outlier is Ie
R14: IF $R_i(t)$ is Ng and $R_N(t)$ is Ng THEN an outlier is E

The above fuzzy rules are used to identify or classify the outliers resulted from an event or an instance or continues error. The fuzzy rule base model is shown in Table 1.

Table 1. Fuzzy rule base model

| | $R_i(t)$ = Ng | | | $R_i(t)$ = Ps | | |
|---|---|---|---|---|---|---|
| $R_i(t-1)$ / $R_N(t)$ | Ng | Nr | Ps | Ng | Nr | Ps |
| Ng | E | E | E | Ce | Ie | Ce |
| Nr | Ce | Ie | Ce | Ce | Ie | Ce |
| Ps | Ce | Ie | Ce | E | E | E |

The final classification outlier degree denoted as D* is based on determining the Center of Gravity (CoG) for the aggregate area of the membership rules' results [20]:

$$D^* = \frac{\int x\mu_A(x)dx}{\int \mu_A(x)} \quad (1)$$

Where x is the range value of the universe and $\mu_A(x)$ is the membership degree.

## 3. Experimental results

In this work, real data collected from a small WSN constructed in a local lab to monitor temperature, humidity and light are used to examine the proposed techniques. Table 2 shows a sample of the sensory data collected from the local WSN environment for temperature (T), humidity (H), and light (L) in different packet sequences (S) of four nodes (1, 2, 3, and 4).

Table 2. Sample of collected raw data from the local wireless

| Date | Time | Node Id. | S | T | H | L |
|---|---|---|---|---|---|---|
| 18/12/2018 | 15:33:52 | 4 | 84 | 23 | 44 | 77 |
| 18/12/2018 | 15:33:57 | 1 | 96 | 21 | 45 | 2 |
| 18/12/2018 | 15:34:03 | 2 | 87 | 19 | 49 | 12 |
| 18/12/2018 | 15:34:16 | 3 | 100 | 18 | 49 | 27 |
| 18/12/2018 | 15:34:22 | 4 | 85 | 23 | 43 | 76 |
| 18/12/2018 | 15:34:27 | 1 | 97 | 21 | 45 | 2 |
| 18/12/2018 | 15:34:33 | 2 | 88 | 19 | 49 | 9 |
| 18/12/2018 | 15:34:46 | 3 | 1 | 16 | 49 | 26 |

A) Outlier Detection using CT-OCSVM

This section presents the results of anomaly detection using CT-OCSVM algorithm based on the temperature data gathered from the node of the locally constructed WSN. Fig.7 shows the recorded temperature for the period from 18$^{th}$ to 20$^{th}$ December 2018 for node 2, 3, and 4 respectively. In these experiments, the data of node 4 (the maximum recorded degree is 32 °C, while the minimum degree is 16 °C) is used for testing the capability of the proposed algorithm in detecting anomalies in which its data showed high nonlinear behavior comparing to the data of the other nodes as seen in Fig. 7. The collected data is firstly converted to images to be applied to the CT technique as it can deal with 2D image. For instance, Fig. 8 (a) shows an image of normal data, while Fig. 8 (b) shows the same image after imposing abnormal data to it.

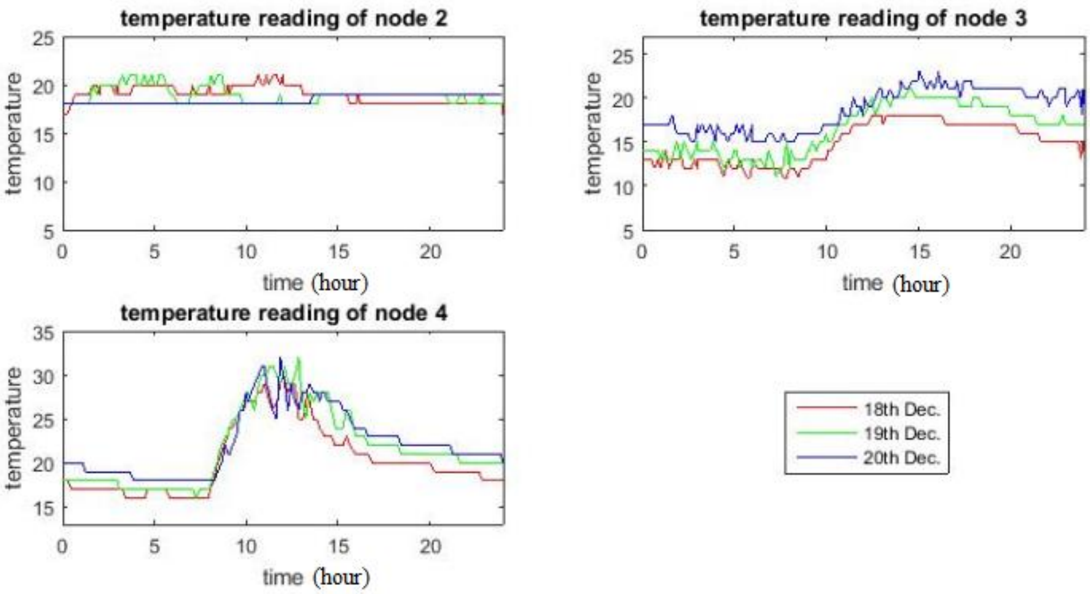



Figure 7. Sample of three successive days of the temperature values collected by three nodes in the local WSN environment

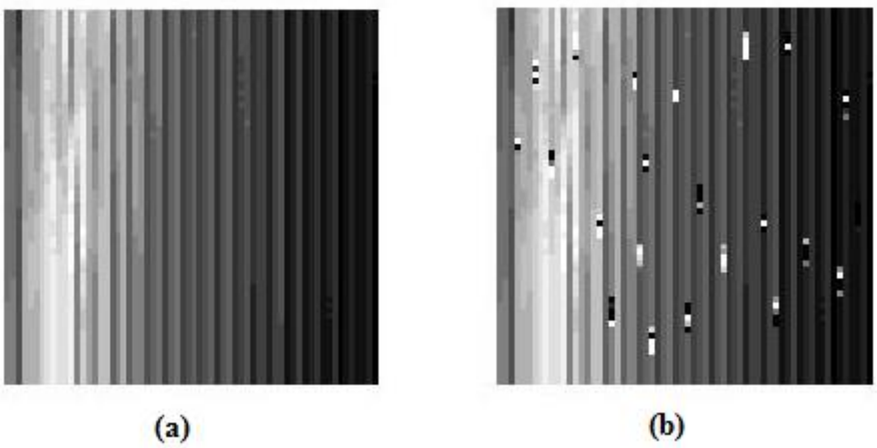



Figure 8. (a) Image of normal data, (b) Image of data with outliers

In the CT part of CT-OCSVM algorithm, two combinations of (LP : DFT) filters are proposed. The first combination uses 9-7 filter as LP filter [21] [22], and PKVA filter [23] as DFT filter, while the second combination proposed Haar filter as LP filter with PKVA filter. After many test it is noticed that only one decomposition level is suitable to be used in CT technique and to implement the DFB filter via 3-level binary tree decomposition which leads to $2^3 = 8$ subbands (frequency division) in order to reduce the time required to complete detection. Empirically, among these subbands, it is found that the second direction of CT (d2) obtains the best performance.

Three performance metrics are considered; the first one is the Detection Rate (DR) or true alarm rate, the second metric is the False Positive Rate (FPR), while the last one is the Accuracy.

$$\mathrm{TPR} = \frac{\mathrm{TP}}{\mathrm{TP+FN}} \tag{2}$$

$$\mathrm{FPR} = \frac{\mathrm{FP}}{\mathrm{FP+TN}} \tag{3}$$

$$\mathrm{ACC} = \frac{\mathrm{TP+TN}}{\mathrm{TP+TN+FP+FN}} \tag{4}$$

Two synthetic types of faults are used as outliers, these are:

- Additive White Gaussian Noise (AWGN) with 25dBw ranged from short faults (15 outliers) to noisy faults (300 outliers).
- Synthetic faults with high amplitude (from 50 to 100) added to the image matrix ranged from short faults (15 outliers) to noisy faults (300 outliers).

Precise locations of the added noise were predetermined which are later used to identify DR and FPR. When an outlier is detected then its location is compared with the known faults locations, if it is existed, then the detected outlier is considered as true alarm otherwise it is missed. In addition, if an outlier is detected but the fault location did not exist, then the outlier is considered as false alarm. The proposed technique without outlier gives 100% DR with 0% FPR. Fig. 9 shows the Contourlet coefficients with [9-7 + PKVA] filters and 80 outliers. Fig. 10 shows the confusion matrix results after applying OCSVM on d2 direction of the CT coefficients. It gives 93.75% DR, 0%FPR, and 99.9% ACC.

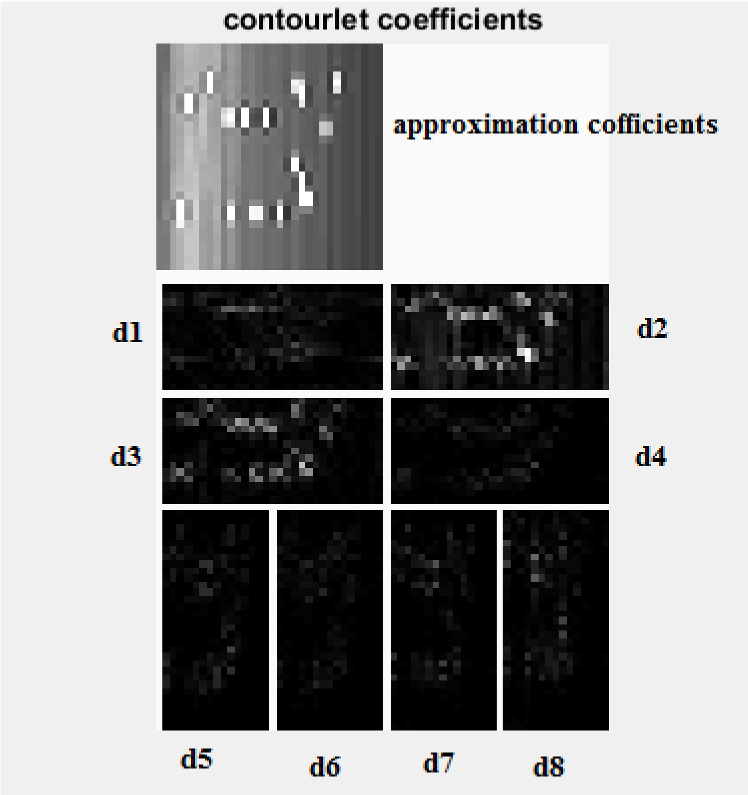



Figure 9. Contourlet coefficients

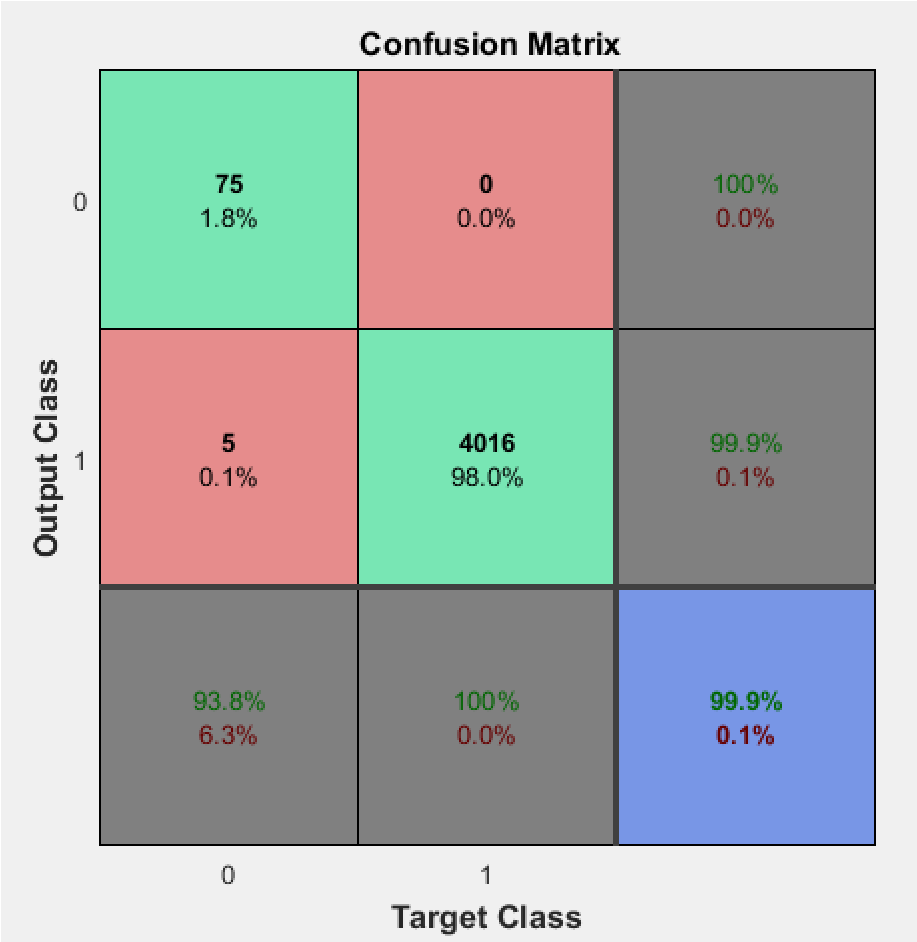



Figure 10. Confusion matrix of CT-OCSVM

Table 3 presents the performance of CT-OCSVM using different percentage of AWGN. In most cases the proposed technique gives high DR, and ACC with low FPR.

Table 3. Anomaly detection rates with different AWGN values

| Outliers | CT-OCSVM [9-7 + PKVA] | | | | | | CT-OCSVM [Haar + PKVA] | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | [LL] | | | d2 | | | [LL] | | | d2 | | |
| | DR % | FP R% | Acc. % | DR % | FPR % | Acc. % | DR% | FPR % | Acc.% | DR % | FPR % | Acc. % |
| (0.4%) | 100 | 0.8 | 99.2 | 73.7 | 0 | 99.9 | 80 | 0 | 99.9 | 80 | 0 | 99.9 |
| (0.8%) | 100 | 0.5 | 99.5 | 100 | 0.1 | 99.9 | 67 | 0 | 99.8 | 93 | 0 | 99.9 |
| (1.25%) | 100 | 0.5 | 99.5 | 74 | 0 | 99.7 | 64 | 0 | 99.6 | 80 | 0 | 99.8 |
| (2%) | 100 | 0.5 | 99.5 | 80 | 0 | 99.6 | 60 | 0 | 99.2 | 85 | 0 | 99.7 |
| (2.5%) | 100 | 0.7 | 99.3 | 93 | 0 | 99.8 | 62 | 0 | 99.1 | 94 | 0 | 99.9 |
| (5%) | 100 | 0.4 | 99.6 | 84.5 | 0 | 99.2 | 65 | 0 | 98.2 | 81 | 0 | 99 |
| (7.5%) | 100 | 0.2 | 99.8 | 76.3 | 0 | 98.3 | 67 | 0 | 97.5 | 78.3 | 0 | 98.4 |

Table 4 presents the performance of CT-OCSVM using different percentage of high amplitude faults (from 50 to 100). For all of the tested cases the proposed algorithm can achieve high performance in term of TPR, FPR and Accuracy. On the other hand, when the amount of outlier is increased (for noisy faults) the CT algorithm with [Haar + PKVA] filters outperforms the CT with [9-7 + PKVA] filters in the detection rate and Accuracy. This is because Haar filter utilizes only two adjacent data to calculate the coefficients while the 9-7 filter use four.

Table 4. Anomaly detection rates with different values of synthetic faults

| Outliers | CT-OCSVM [9-7 + PKVA] | | | | | | CT-OCSVM [Haar + PKVA] | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | [LL] | | | d2 | | | [LL] | | | d2 | | |
| | DR % | FPR % | Acc. % | DR % | FPR % | Acc. % | DR% | FPR % | Acc.% | DR % | FPR % | Acc. % |
| 0.4% | 100 | 0.8 | 99.2 | 100 | 0 | 100 | 67 | 0 | 99.9 | 100 | 0 | 100 |
| 0.8% | 100 | 0.8 | 99.2 | 100 | 0 | 100 | 63 | 0 | 99.7 | 100 | 0 | 100 |
| 1.25% | 100 | 0.7 | 99.3 | 100 | 0 | 100 | 58 | 0 | 99.5 | 100 | 0.2 | 99.8 |
| 2% | 100 | 0.4 | 99.6 | 93.8 | 0 | 99.9 | 60 | 0 | 99.2 | 100 | 0.1 | 99.9 |
| 2.5% | 100 | 0.6 | 99.4 | 91 | 0 | 99.8 | 54 | 0 | 98.9 | 94 | 0 | 99.9 |
| 5% | 100 | 0.8 | 99.2 | 90.5 | 0 | 99.5 | 65 | 0 | 98.2 | 95.1 | 0 | 99.8 |

As it can be seen from the above tables the output of d2 subband direction outperforms the results obtained from low coefficients LL of the LP filter. This is due to the ability of CT in extracting high level information from the image data by using its important features specially directionality and anisotropy. In these tests, CT made reduction to one eighth (1:8) (for example from 4096 to 512) of the total number of the input data where, CT is able to extract relevant feature without losing any significant information. Data dimensionality reduction helps in reducing consumed energy of transmitting all data to the base station where this is another advantage of CT.

Finally, for CT-OCSVM the binary level for DFB can be increased but the complexity and time will be increased, for example if the binary level increased from 3 to 4 the time will increased from about 0.17 second to 0.33 second, while the reduction in data will be (1:16). But this reduction in the data may cause some of important outlier information to be lost which leads to a reduction in the performance of detecting anomalies (DR and ACC will be reduced if outlier percentage $\geq$ 6.25%). Therefore, DFB with 3 binary level is appropriate to be used in the detection of outlier.

B) Outlier Classification results based on FIS

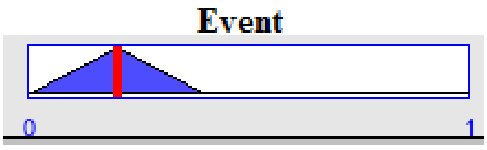
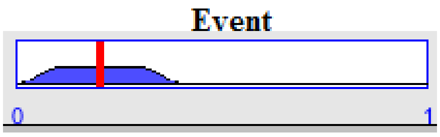
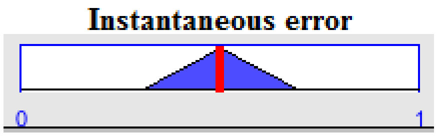
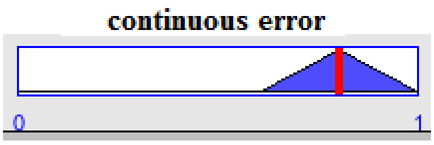
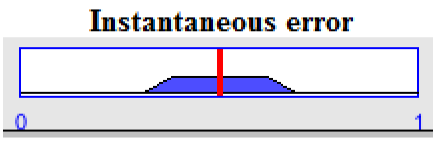
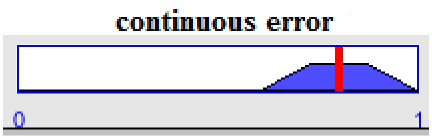
In this section, the results of the outlier classification on the temperature data are presented. FIS technique is used to identify the source of anomalies after extracted it by the utilizing CT-OCSVM technique as mentioned in section 3. To deal with a coherent and consistent data set, each input to the FIS is normalized to be within the range [0, 1] based on using equation (5).

$$\check{c} = (c - \min(c))/(\max(c) - \min(c)) \qquad (5)$$

Where č is the normalized value of coefficient c.

Table (5) shows sample of the results after applying the classification process on some sample of the temperature data.

Table 5. Outlier Classification results synthetic faults

| $R_i(t)$ (č) | $R_N(t)$ (č) | $R_i(t-1)$ (č) | output |
|---|---|---|---|
| 0.8158 | 0.811 | 0.4933 |  |
| 0.3749 | 0.373 | 0.4947 |  |
| 0.21752 | 0.52424 | 0.524702 |  |
| 0.76534 | 0.514663 | 0.292326 |  |
| 0.751738 | 0.377476 | 0.523989 |  |
| 0.84366 | 0.279948 | 0.659553 |  |

## 5. Conclusions

This paper proposed techniques for outlier detection and classification in WSNs. Where, first an integration of CT and OCSVM algorithms are utilized for anomaly detection then FIS is used to identify the source of these outliers. The proposed algorithm was examined using real world datasets collected from local Lab. The algorithm can attain high detection rate, accuracy with low FPR. Furthermore, CT-OCSVM algorithm compressed the input data to one eighth of its original size where this reduction in the data can reduce the consumed power which leads to prolonging the lifetime of the nodes, at the same time this will save the bandwidth and reduce the lost packets in WSNs. This makes CT-OCSVM for anomaly detection in WSNs. Finally, Fuzzy Inference System (FIS) can be used for outlier classification as event, instantaneous error, and continuous error by exploiting the spatial temporal relationships existing between sensor data. Different types of attacks can be tested using CT-OCSVM for future work for example sinkhole, selective forwarding, and denial of services attack etc.

**Abbreviations**

| | |
|---|---|
| $\mu_A(x)$ | Membership degree |
| č | Normalized value of C |
| FN | False negative |
| FP | False positive |
| Ng | Negative Outlier |
| Nr | Normal Data |
| Ps | Positive Outlier |
| $R_i(t)$ | current CT-OCSVM value of node i which is detected as outlier |
| $R_i(t-1)$ | previous CT-OCSVM value of node i |
| $R_N(t)$ | current CT-OCSVM values of the neighbor nodes that shared the same sensing area with node i |
| TN | True negative |
| TP | True positive |

## 4. References


1. Yick, j., Mukherjee, B., and Ghosal, D. (2008). "*Wireless sensor network survey*". Computer Networks, Vol. 52, No. 12, pp. 2992-2805.
2. Zhang, Y., Meratnia, N., and Havinga, P. (2010). "*Outlier detection techniques for wireless sensor networks: A survey*". IEEE Communication Survey & Tutorials, Vol. 12, No. 2, pp. 159-170.
3. Krontiris, I., Benenson, Z., Giannetsos, T., Freiling, F., and Dimitriou, T. (2010). "*Cooperative intrusion detection in wireless sensor networks*". In: European Conference on Wireless Sensor Networks, Springer, pp. 263-278.
4. Jurdak, R., Wang, X.R. , Obst, O., and Valencia, P. (2011). "*Wireless sensor Network anomalies: Diagnosis and Detection Strategies*". In: Intelligence-Based Systems Engineering., Springer, pp. 309-325.
5. Farooqi, A.H., and Khan, F.A. (2009). "*Intrusion Detection Systems for Wireless Sensor Networks: A Survey*". In Communication and Networking, Springer, pp. 234–241.
6. Zhang, Y., Meratnia, N., and Havinga, P.J.M. (2009). "*Adaptive and online one-class support vector machine-based outlier detection techniques for wireless sensor networks*". in proceedings of IEEE International Conference on Advanced Information Networking and Applications Workshops, pp. 990-995.
7. Chitradevi, N., Palanisamy, V., Baskaran, K. , and Barakkath, U. (2010). "*Outlier aware Data Aggregation in Distributed Wireless Sensor Network Using Robust*

*Principal Component Analysis*". in International Conference on Computing Communication and Networking Technologies, pp. 1-9.

8. Siripanadorn, S., Hattagam, W., and Teaumroong, N. (2010). "*Anomaly Detection in Wireless Sensor Networks using Self-Organizing Map and Wavelets*". International Journal of Communications, vol. 4, No. 3,pp. 74–83.
9. Takianngam, S., and Usaha, W. (2011). "*Discrete Wavelet Transform and One-Class Support Vector Machine for Anomaly Detection in Wireless Sensor Network*". In: Intelligent Signal Processing and Communication Systems (ISPACS), International Symposium on IEEE, pp. 1-6.
10. Zhang, Y., Hamm, N.A.S., Meratnia, N., Stein, A., van de Voort, M., and Havinga, P.J.M. (2012) "*Statistics-Based Outlier Detection for Wireless Sensor Networks*". in International Journal of Geographical Information Science, vol. 26, No. 8, pp. 1373-1392.
11. Zhang, Y., Meratnia, N., and Havinga, P.J. (2012). "*Distributed online outlier detection in wireless sensor Networks using ellipsoidal support vector machine*". Ad Hoc Networking, Vol. 11, pp. 1062-1074.
12. Salem, O., Guerassimov, A., Mehaoua, A., Marcus, A., and Furht, B. (2013). "*Sensor fault and patient anomaly detection and classification in medical wireless sensor networks*". In: communications (ICC), 2013 IEEE International Conference, pp. 4373–4378.
13. Mahmoud, S., Lotfi, A., and Langensiepen, C. (2014). "*User activities outliers detection; integration of statistical and computational intelligence techniques*". Computational Intelligence, Vol. 32, No. 1, pp. 49-71.
14. Kamal, S., Ramadan, R.A., and Fawzy, E.R. (2016). "*Smart outlier detection of wireless sensor network*". Facta Universitatis, series: Electronic and Energetics, Vol. 29, No. 3, pp. 383–393.
15. Garcia-Font, V., Garrigues, C., and Rif-Pous, H.A. (2016). "*A Comparative study of anomaly detection techniques for smart city wireless sensor networks*". sensors, Vol. 16, No. 16.
16. Meng, Bi., Jian, Xu., Wang, Mo., and Fucai Zhou. (2016). "*Anomaly detection model of user behavior based on principal component analysis*". Journal of Ambient Intelligence Humanized Computing, Springer, Vol. 7, pp. 547-554.
17. Do, M.N., and Vetterli, M. (2005). "*The Contourlet Transform: An Efficient Directional Multiresolution Image Representation*". IEEE Transactions On Image Processing, Vol 14, pp. 2091-2106.
18. Tax, D.M.J., and Duin, R.P.W. (2004). "*Support vector description*". Machine Learning, Vol. 54, No. 1, pp. 45-66.
19. Zadeh, L. A. (1995). "*Fuzzy sets, Information and Control*". direct science, vol. 8, pp. 338-353.
20. Zhuang, H., Wang, D., and Bai, Y. (2006). "*Advanced Fuzzy Logic Technologies in Industrial Applications*". Springer-Verlag London Limited.
21. Shapiro, J.M. (1993). "*Embedded image coding using zerotrees of wavelet coefficients*". IEEE Transactions on Signal Processing. , vol. 41, No. 12, pp. 3445–3462.

22. Cohen, I., Daubechies, I., and Feauveau, J.C. (1992). "*Biorthogonal bases of compactly supported Wavelets*". Communication on Pure and Applied Mathematics, vol. 45, No. 5, pp. 485–560.
23. Phoong, S.M, Kim, C.W, Vaidyanathan, P.P, and Ansari, R. (1995). "*A new class of two-channel Biorthogonal filter banks and wavelet base*". IEEE Transactions on Signal Processing, vol. 43, No. 3, pp. 649–665.