Int. J. Advance Soft Compu. Appl, Vol. 10, No. 2, July 2018
ISSN 2074-2827

Unsupervised Anomaly Detection for Unlabelled Wireless Sensor Networks Data

Nurfazrina Mohd Zamry[1], Anazida Zainal[1], Murad A. Rassam[2]

[1] Information Assurance and Security Research Group, Faculty of Computing, Universiti Teknologi Malaysia,
81310 Skudai, Malaysia
[2] Faculty of Engineering and Information Technology, Taiz University, Taiz 6803, Yemen
e-mail: nurfazrina.mohdzamry@gmail.com, and anazida@utm.my

Abstract

With the advances in sensor technology, sensor nodes, the tiny yet powerful device are used to collect data from the various domain. As the sensor nodes communicate continuously from the target areas to base station, hundreds of thousands of data are collected to be used for the decision making. Unfortunately, the big amount of unlabeled data collected and stored at the base station. In most cases, data are not reliable due to several reasons. Therefore, this paper will use the unsupervised one-class SVM (OCSVM) to build the anomaly detection schemes for better decision making. Unsupervised OCSVM is preferable to be used in WSNs domain due to the one class of data training is used to build normal reference model. Furthermore, the dimension reduction is used to minimize the resources usage due to resource constraint incurred in WSNs domain. Therefore one of the OCSVM variants namely Centered Hyper-ellipsoidal Support Vector Machine (CESVM) is used as classifier while Candid-Covariance Free Incremental Principal Component Analysis (CCIPCA) algorithm is served as dimension reduction for proposed anomaly detection scheme. Environmental dataset collected from available WSNs data is used to evaluate the performance measures of the proposed scheme. As the results, the proposed scheme shows comparable results for all datasets in term of detection rate, detection accuracy and false alarm rate as compared with other related methods.

Keywords: *Support Vector Machine, Unsupervised SVM, the Wireless Sensor Network, Dimension Reduction, Unlabeled Data.*

1 Introduction

Wireless Sensor Networks (WSNs) are formed by deployed a large number of sensor nodes in large areas to collect the desired data from the target phenomenal. WSNs have been used in many domains due to the tiny features of sensor nodes are favored to capture the needed data. For instance, the sensor deployed in 1) the mountain, dessert or urban area to collect the environmental data like ambient temperature, relative humidity, soil moisture, wind speed; 2) industrial and agricultural application for tracking and control purposed 3) military area or danger zone for alert system and monitoring purposed. Sensing unit, processing, unit, radio unit, and power unit are the basic unit equipped with sensor nodes as shown in Figure 1 while can be added to other unit depending on the requirement. Unfortunately, sensor node has limited resource constraint in term of energy, computation, and storage. Basically, wireless sensor data are communicated continuously via wireless channel followed the network architecture designed, based on flat or hierarchical network architecture. As sensor nodes are deployed in the critical area, in most situations, it will be utilized until the battery is depleted.

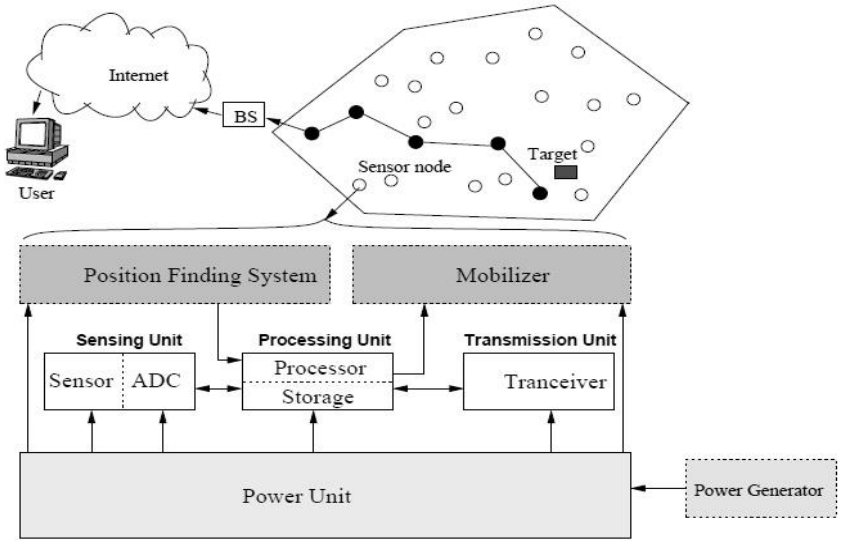


Figure 1: Basic Component of Sensor Nodes [1]

On the other hand, the raw data collected from the phenomenon are usually inaccurate and unreliable due to some reason. For instance, due to the nature of sensor node is tiny in size and limited resources in certain point sensor will fail to send data to the base station when energy is depleted. Moreover, as sensors are randomly deployed in the critical area, sensor nodes are prone to the malicious attack. Nevertheless, due to the unintended environment like the dynamic climate changing or harsh phenomenon in wildlife area sensor might be reported with unstable data. Therefore, to ensure the reliable and accurate data collected for decision making at the base station, anomaly detection is a possible solution to

detect anomalous or outlier from the raw data. Hodge and Austin, (2004) [1] have stated that outlier detection is the closest task to the initial motivation behind the data mining.

Anomaly detection is the process to detect the data which significantly deviate from the rest of the normal data. Generally, anomaly detection model built the normal reference model using normal data which contrast to the misused detection that used both normal and anomalous data as a reference. By taking only normal data as references, thus anomaly detection is capable of detecting new types of security attacks or intrusions that emerge in the system [2]. Furthermore, sensor nodes may potentially collect anomalous data which come from the noise and error, actual event as well as a malicious attack. In the first case, the noise and erroneous data are needed to be eliminated, however, the other sources of anomalous data need to be carefully analyzed as it may give the meaningful results in the decision making at the base station. The generic framework of anomaly detection has been illustrated by [3] as shown in Figure 2. The generic anomaly detection illustrated in Figure 2 is composed of input, data processing, analysis and decision, and output derived from [2].

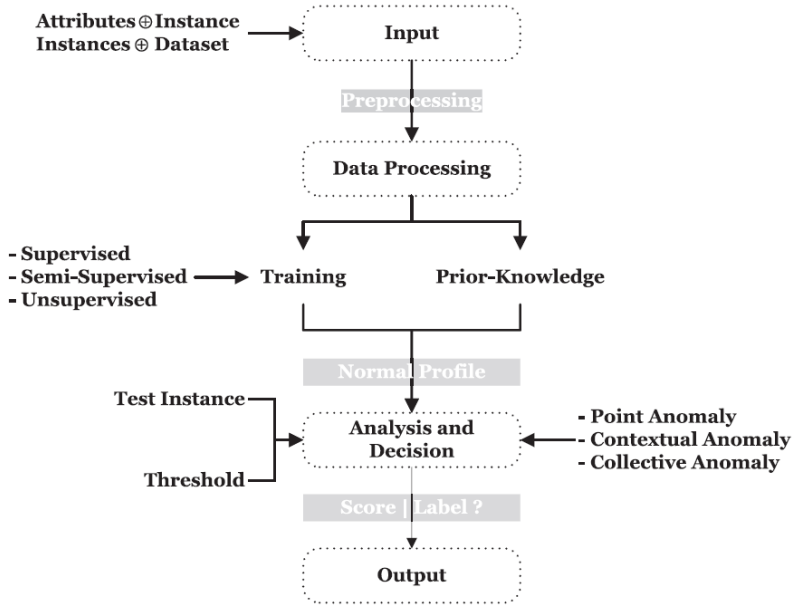


Figure 2. Generic framework of anomaly detection[3].

On the other hand, there are lots of taxonomies of anomaly detection have been discussed for WSNs domain like in [2], [4], [5]. Figure 3 shows the taxonomy of anomaly detection in WSNs as reviewed in [4] and have categorized anomaly detection approach as Statistics-based, Nearest Neighbor-based, Clustering-based, Classification-based as well as Spectral Decomposition-based. Each of the categories has different algorithms to detect anomalous data measurements.

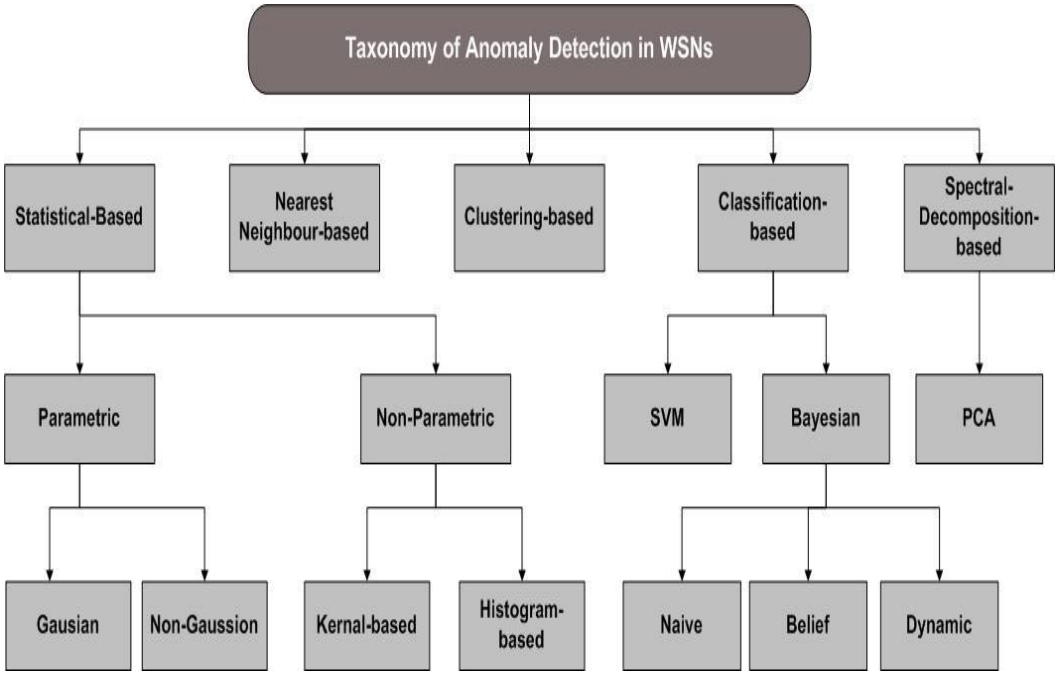


Figure 3: Taxonomy of Anomaly Detection in WSNs by [4]

Based on Figure 3, classification-based generally comes from data mining and machine learning community. Generally, classification-based anomaly detection approach learns the normal features of data measurements which known as a training set to classify the new data measurement which known as a testing set as anomalous or normal instances. Support Vector Machine (SVM) which is one of classification-based technique has widely used in WSNs dataset included in [6]-[9].

One of the challenged faced by SVM-based anomaly detection is to obtain error-free and labeled data for training [10]. Unfortunately, hundreds of thousand raw data collected from the sensor nodes are normally unlabeled. Furthermore, obtaining such clean and labeled data is often an expensive or manually intensive exercise [7]. The solution is to implement the unsupervised anomaly detection approach which suitable for unlabeled data collection. One-class SVM, on the other hand, learned the one-class normal data technique such as One-class Support Vector Machine (OCSVM) and Unsupervised Principle Component Analysis (UNPCA) has been widely used in the machine learning environment. Moreover, OCSVM has been widely studied to suit the sensor nodes limitation.

As mention earlier, sensor nodes are limited to resources constraint, thus anomaly detection must be carefully modeled in order to minimize the energy and computational restriction. Therefore, dimension reduction approach can be considered to incorporate in the anomaly detection process to reduce the computational overhead as well as the data communication. There are many dimensions reduction techniques like Principal component analysis (PCA),

Discrete Wavelet Transform (DWT) and Deep Belief Network (DBN) have been proposed by hybridizing with anomaly detection for more resource minimization.

In this paper, we implemented unsupervised anomaly detection with the dimension reduction technique for unlabeled WSNs data as well as to suit the resources constraints incurred in WSNs. The rest of the paper is organized as follows. The related work on the SVM-based anomaly detection and dimension reduction are discussed in Section 2. The methodology of anomaly detection technique is presented in section 3 including the dataset, data pre-processing and techniques used in the anomaly detection scheme. The experimental results and performance comparison with other algorithms are presented in section 4. Finally, our work of this paper is summarized in the last section.

2 Related Work

Classification-based anomaly detection technique for SVM-based classifier will be discussed in this section. Generally, classification-based anomaly detection is performed in two basic phases namely, namely training and testing phase. Compared to Bayesian-based classification-based techniques, SVM-based have much better generalization ability because they tend to minimize the separation between different classes by making used Mercer Kernels [11]. In the early version SVM technique introduces by Scholkopf *et. al*.,(2001)[12] the maximum margin hyperplane is used to separate the normal class from outlier classes. The data in input space are mapped to a high dimensional space called features space using Mercer Kernel to separate normal from anomalous as mention before. Again, as data are mapped to features space which minimized the computational overhead, thus SVM-based is preferable to be used in WSNs rather than Bayesian-based technique. SVM-based have also been classified based on the target class number included Multi-class SVM, Binary-class SVM and One-Class SVM (OCSVM). In this paper, OCSVM will be used as classifier due to the nature of OCSVM used normal data to model anomaly detection. Moreover, as the absence of ground truth labeled data in WSNs dataset, thus OCSVM is suited for modeling the anomaly detection for WSNs dataset. OCSVM is classified as unsupervised classification technique as no prior labeled data are required to learn the normal model.

Shahid *et. al.* (2013)[11] have reviewed the various One-class SVM-based techniques formulations. The first variant of one-class SVM called hyperplane-SVM have proposed by Scholkopf *et. al*.,(2001)[12]. In this variant hyperplane margin is used to separate the anomalous data from normal data measurement. Meanwhile, Hypersphere-SVM has been proposed by Tax and Duin (1999) [13] by calculating minimum radius as a decision boundary to identify anomalous data. Wang et al. (2006) [14] have modeled Hyper-Ellipsoidal (TOCC) to separate the normal and anomalous data based on minimum effective radii as a decision boundary. Both formulations are differentiated based on their shape of decision boundary which is spherical and ellipse shape respectively. Due to the quadratic

optimization used to calculate the decision boundary for both techniques incurred computational complexity. Therefore, the other researcher has proposed linear optimization to mitigate the computational overhead. Laskov et al. (2004) [15] have made an alteration in calculating the decision boundary by adopting linear optimization to proposed quarter-sphere based one-class SVM (QSSVM). On the other hand, Centered Hyperellipsoidal Support Vector Machine (CESVM) based anomaly detection has been proposed by Rajasegarar et al. (2008) [16] by combining the idea of linear optimization in QSSVM with the Hyper-Ellipsoidal-SVM. As reported in [11] the classification performance and generalization ability of one-class formulations can be arranged in the following increasing order: hyperplane < hypersphere ≈ quarter-sphere < hyperelliptic ≈ centered ellipsoid. The formulation of OCSVM variants is presented in [11].

In WSNs anomaly detection domains, SVM techniques have been used as a classifier in various anomaly detection schemes. Takianngam and Usaha(2011) [17] used Quarter-Sphere OCSVM as a classifier to detect anomalous data measurements while incorporating Discrete Wavelet Transform (DWT) as data compression for data pre-processing. Researcher in [7], [16], [18] have used Centered Hyperspherical One-Class SVM and Hyperellipsoidal One-Class SVM and Quarter-Sphere SVM algorithm to perform anomaly detection in WSNs data. Meanwhile, Researcher in [19] used Quarter-Sphere SVM to proposed Spatio-Temporal-Attribute Quarter-sphere SVM (STA-QS-SVM) formulation by considering attribute correlations between the sensor nodes to model their anomaly detection schemes. Meanwhile, research in [20] has proposed two unsupervised methods to estimating the optimal setting for hyper-plane based One-Class Support Vector Machine (OCSVM) and Hypersphere-SVM faster parameter estimation. In other domain, researchers in [21], one-class SVM have been combined with deep belief networks (DBNs) which DBNs is used to extract the features from the input data.

3 Research Methodologies

The flow chart of the research methodologies is depicted in Figure 4 below is applied to this research paper. ~~As~~ compared to generic anomaly detection presented in Figure 2, the component of dimension reduction is added in the component to reduce the data dimension. Thus result reduces the detection efficiency in term of memory and energy usage. Meanwhile, the procedure to identify the normal profile (normal reference model) is performed using CESVM classifier based on unsupervised anomaly detection scheme. The testing data instances are then can be classified as anomalous or normal data based on the established normal reference model. In CESVM classifier for anomaly detection scheme phase, analysis and decision making is taking place which corresponds to analysis and decision, and output components in Figure 2. The detail of each component will be elaborated in each sub-section.

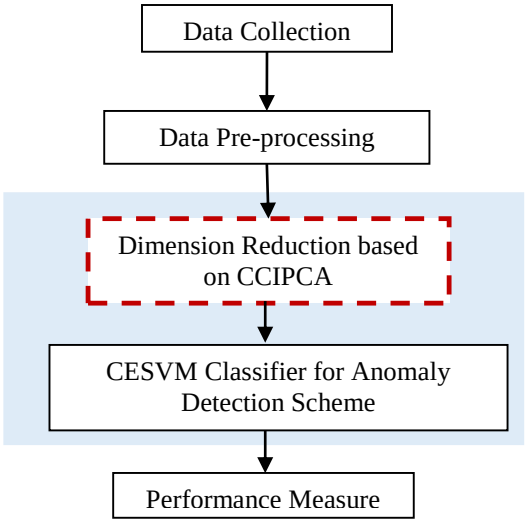


Figure 4: Flowchart of Proposed Methodology using OCSVM Classifier and CCIPCA Dimension Reduction algorithm.

3.1 Data Collection

Data collection phase is the equivalent to the input component from Figure 2 which represent the dataset collected from the sensor nodes. The environmental datasets are extracted from Intel Berkeley Research Laboratory (IBRL), Lausanne Urban Canopy Experiment (LUCE), PDG and Networked Aquatic Microbial Observing System (NAMOS) which widely used in anomaly detection schemes to detect anomalies in WSNs data. In the experiment, both IBRL, LUCE and NAMOS datasets will be tested on univariate dataset while PDG dataset will be tested on multivariate data. Temperature and the ambient temperature is selected from IBRL and LUCE data sample respectively. Two variables are used in *SensorScope* PDG 2008, which is ambient and surface temperature is selected from the sensor in station pdg2008-metro-1. Lastly, one variable which is Chlorophyll concentration is extracted from NAMOS dataset located at buoy no. 103.

3.2 Data Pre-processing

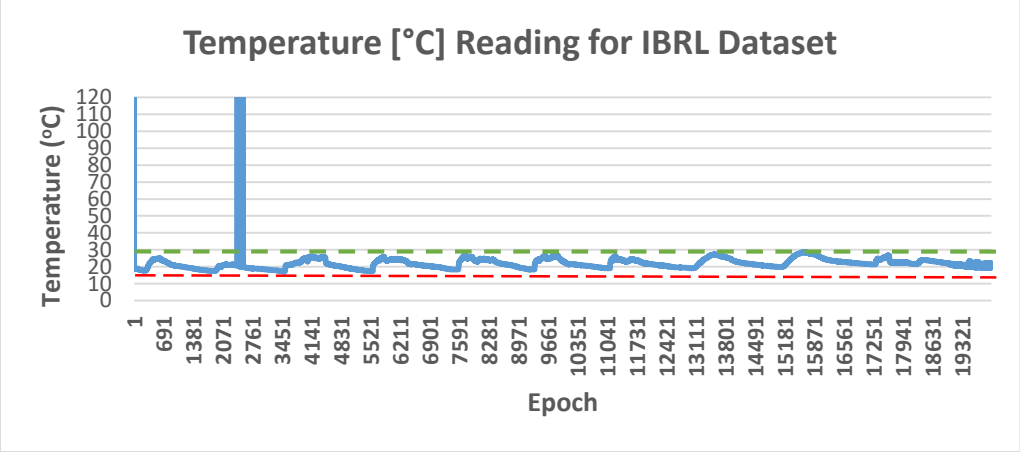
In data pre-processing, histogram-based data labeling will be used to label the normal and anomalous data using visual inspection. This data labeling has been used in [8], [17], [22] to evaluate the effectiveness of their proposed detection schemes. From histogram plot observation, the normality region is obtained to label the collected sensor data. Moreover, the patterns of anomalies are found from the

observation as stated in [8]. The normality region for all the datasets are presented in Table 1.

Table 1: Normality regions for histogram-based datasets [17]

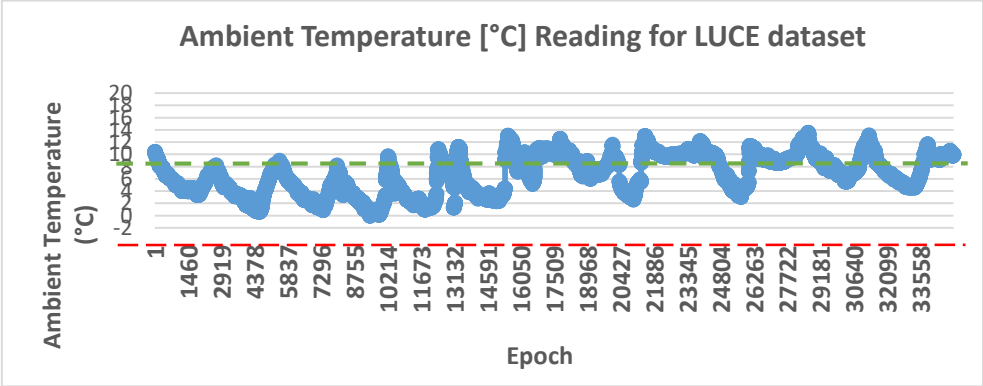
Dataset	Normal region lower bound	Normal region upper bound
IBRL	16	30
LUCE	1.5	9
PDG (ambient temp.)	4	-12
PDG (surface temp.)	4	-14
NAMOS	0	500

Meanwhile the histogram plots for all datasets are shown in Figure 5(a)-(d). The green and red dash line indicates the maximum and minimum normality region respectively.

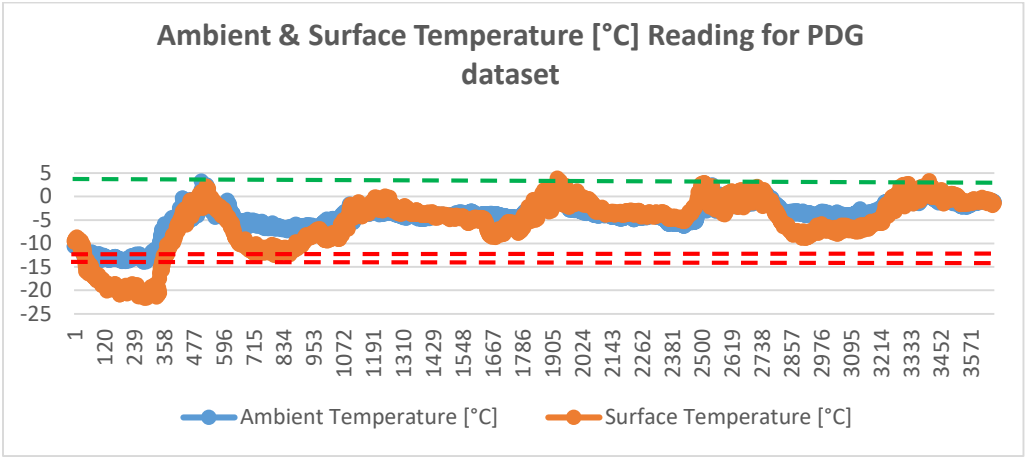


(a) IBRL Dataset

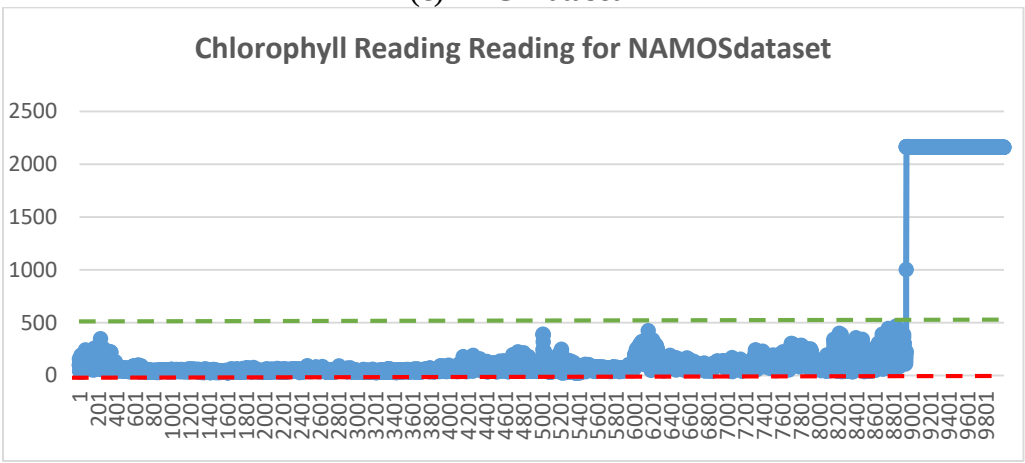
Figure 5. Histogram for (a) IBRL, (b) LUCE, (c) PDG and (d) NAMOS Dataset



(b)LUCE Dataset



(c) PDG Dataset



(d) NAMOS Dataset

Figure 5. Histogram for (a) IBRL, (b) LUCE, (c) PDG and (d) NAMOS Dataset(cont.)

From the histogram plot, short anomalies have been observed in IBRL dataset. The sharp and short plot are spotted around epoch 1 and 2500 which the reading reported are more than the normality region as stated in Table 1. Anomalous data reported in LUCE dataset can be categorized into noise and short anomalies as there is increasing the value of the variance of sensed data and some sharp plot are also presented in Figure 5(b). The data measurements also show there are some data plotted over maximum and below the minimum normality region. The same noise pattern is also observed in PDG dataset which the inconsistent shape presented in the histogram plot. As the dataset is using multivariate features, two minimum normality regions are marked in the histogram plot. The last dataset, on the other hand, presented the constant plot at the last 9000 epochs which indicated the presence of constant anomalies. The value reported is also above the maximum normality region which is labeled as anomalous data.

3.3 Dimension Reduction based on Candid-Covariance free Incremental Principal Component Analysis (CCIPCA)

As explained before, this paper will use dimension reduction to minimize the resource constraint incurred in sensor nodes. CCIPCA which have successfully used dimension reduction in anomaly detection scheme proposed by [23] will be used in this paper. Weng *et. al*.,(2003) [24] proposed CCIPCA to reduce the computational complexity of the original PCA. The pseudocode for CCIPCA is shown in Figure 6.

Algorithm: CCIPCA
1 Input: $S_{m\times n} = [s(1), s(2), \dots s(m)]$
2 Output: V, D //equivalent to Λ, P
3 For m=1,2,..,m, do the following steps
4 $s_1(m) \leftarrow s(m)$
5 For $i = 1,2.., \min\{k, m\}$ do// k is first dominant V
If i= m do
6 Initialize the i^{th} eigenvector as $f_i(m) = s_i(m)$
7 Else do
8 $f_i(m) = \frac{m-1-l}{m} f_i(m-1) + \frac{1+l}{m} s_i(m) s_i^T(m) \frac{f_i(m-1)}{\|f_i(m-1)\|}; m > 0$
9 End if
10 End For
For $i = 1,2.., \min\{k, m\}$ do
11 $f_i = \frac{f_i(m)}{\|f_i(m)\|}; \lambda_i = \|f_i(m)\|$
// $Normalize\ f_i\ to\ get\ the\ V, D$
12 End For

Figure 6. The algorithm of CCIPCA proposed in Weng *et. al*.,(2003)[24]

Basically, CCIPCA can be used either in a batch or incremental learning mode. For batch modes, principal component is generated using data collected in a specific period of time. Meanwhile, in incremental mode, the principal components are updated for each of data observation.

3.4 CESVM Classifier for Anomaly Detection scheme

CESVM will be used in this paper, to model the anomaly detection scheme for detecting anomalous data instance from WSNs data. This is due to the ability of ellipsoidal-based SVM techniques to capture multivariate data as well as the significant reduction of computational complexity compared to hyper-ellipsoidal by using linear optimization problem for decision boundary calculation. CESVM aims to place the majority of the of image vector within the minimum effective radii and centered at the origin in the feature space.

Firstly, consider $X = \{\phi(x_i) : i = 1, \dots, n\}$ as the dataset with d variate data vectors in the input space, $x_i : i = (x^1, x^2, \dots, x^d)$, $i = 1,2, \dots n$ and n is the number of data vectors. Then in features space, the input vector are mapped into image vector $X = \{\phi(x_i) : i = 1, \dots, n\}$ using non-linear function $\phi(.)$ and produce $\phi(.) : \Re^d \rightarrow \Re^p$. The optimization problem is calculated based from [14] as shown in Equation (1).

$$\min_{R \in R, \xi \epsilon R} \; R^2 + \frac{1}{vm}\sum_{i=1}^{m} \xi_i \tag{1}$$

$$\text{s.t: } (x_i)\Sigma^{-1}\phi(x_i)^T \leq R^2 + \xi_i, \xi_i \geq 0, i = 1,2, \dots m$$

CESVM is then transformed to the linear optimization problem formulated based on equation (1) as shown in equation (2).

$$\min_{\alpha \in \Re} -\sum_{i=1}^{m} \alpha_i \; \left\| \sqrt{m}\, \Lambda^{-1} P^T K_c^i \right\|^2 \tag{2}$$

$$\text{s.t: } \sum_{i=1}^{m} \alpha_{i=} 1 \,, 0 \leq \alpha_i \leq \frac{1}{vm}, i = 1,2, \dots m$$

Calculation of centered kernel matrix, K_c obtained from kernel matrix, K is equal to $K_c = \mathrm{K} - 1_m \mathrm{K} - \mathrm{K}\, 1_m + 1_m \mathrm{K}\, 1_m$ where 1_m is the $m \times m$ matrix and all value equal to $\frac{1}{m}$. In OCSVM, Mercer Kernel is used to computation the dot product of image vector in the features space which can be computed in input space. The linear optimization technique such as simplex or interior point method can be used to obtained the value of α_i. This α_i value is used to classify the data vector as 1) $\alpha_i = 0$ data vector is classify as normal data; 2) $\alpha_i > 0$ data vector is classify as support vector and 3) $\alpha_i = \frac{1}{vm}$ data vector is classify as border support vector. Finally, effective radii, R are computed using any border support vector as in equation (3).

$$R = \left\| \sqrt{m}\, \Lambda^{-1} P^T K_c^b \right\| \tag{3}$$

Lastly, the decision function used to classify the new data measurement is calculated by equation (4). Table 1 shows the explanation for notation used in equation (1) - (3).

$$f(x) = sgn(R^2 - \left\| \sqrt{m}\, \Lambda^{-1} P^T K_c^i \right\|^2 \tag{4}$$

Table 2 shows the explanation for notation used in equation (1) - (3).

Table 2: Explanation of Notation used in equation (1) - (3).

Notation	Explanation
m, n	Size of input data input data
$\phi(.)$	non-linear function
p	The dimension of features space

v	Regularization parameter
Σ^{-1}	The inverse of the covariance matrix
α_i	Lagrange multipliers
P	Positive eigenvector matrix
Λ	Positive diagonal eigenvalue
K	Kernel Matrix
K_c^i	Centered Kernel Matrix
R	Effective Radii

3.5 Performance Measure

The common evaluation metric used to measure the effectiveness of anomaly detection performance are detection rate, detection accuracy and false alarm rate which calculated as Equation (5)-(8).

$$\text{Detection Rate (DR)} = \frac{TN}{FN + TN} \tag{5}$$

$$\text{Accuracy} = \frac{TP + TN}{TP + FP + FN + TN} \tag{6}$$

$$\text{False Positive Rate (FPR)} = \frac{FP}{TP + FP} \tag{7}$$

$$\text{False Negative Rate (FNR)} = \frac{FN}{FN + TN} \tag{8}$$

The result of performance measure will be reported in term of percentage. The best detection rate and accuracy when results are reported 100% reading while the false alarm rate (FPR and FNR) shows the best results when 0% are reported.

4 Experimental Results

The effectiveness performance result of the WSNs dataset included IBRL, LUCE, PDG, and NAMOS above are discussed in this section. Datasets included mixed of univariate and multivariate datasets with a different type of anomalies will be discussed based on the performance measure discussed in the previous section. The linear kernel presented as $k_{linear} = x_1.x_2$ is used in the experiments as follow the experiment setup in [17]. Meanwhile the v value varied from each dataset since each dataset reporting different the value of outliers. This is due the v represents the maximum outlier can be in the dataset. The results reported are taken from v which gave the best performance measure for all the datasets.

4.1 The result of Dataset IBRL

In IBRL dataset, two scenarios are examined for effectiveness performance as stated in section 4.5. Temperature is selected as a variable in two scenarios. In scenario 1, 1000 data instances are used to build normal reference model and 2000 instances are used in scenario 2 to build normal reference model using CESVM technique. These two scenarios are chosen as based on histogram plot in Figure 5(a), the short anomalies are exhibited in the first 1000 data the data instance while the next 1000 data instances are free from anomalous data.

Table 3. Effectiveness Results for IBRL Dataset with Different Scenario

Scenario	***Training Set Size and Position***	***DR (%)***	***ACC (%)***	***FPR (%)***	***FNR (%)***
1	1-1000 (1000)	100	99.8	0.02	0.02
2	1-2000 (2000)	100	98.4	1.6	0

Both scenarios reported 100% detection rate that indicates the proposed scheme can clearly detect anomalous data as shown in Figure 5. Since short anomalies are reported in IBRL which is represented by the sudden sharp changes around epoch 1, therefore, anomalous data can be detected easily. However, the false alarm rate is reported to be more than 0% for both scenarios. This shows that the training set does not well represent the current data situations. Therefore, as the false alarm increases, the detection accuracy is decreased.

4.2 The result of Dataset LUCE

As same as IBRL dataset, the LUCE dataset is examined using univariate data by selecting ambient temperature as variable. The performance of LUCE dataset also tested in two scenarios. The first scenario, 1000 data instances are selected to use as a training set, meanwhile, 4000 instances are selected for scenario 2. The results of the detection effectiveness are shown in Table 4.

Table 4. Effectiveness Results for LUCE Dataset with Different Scenario

Scenario	***Training Set Size and Position***	***DR (%)***	***ACC (%)***	***FPR (%)***	***FNR (%)***
1	1-1000 (1000)	100	98.0	2.0	0
2	1-4000 (4000)	100	98.0	2.0	0

Based on Table 4, 100% detection rate is reported for both scenarios which indicated the proposed scheme can clearly detect anomalous data. Since short and noise anomalies are reported in LUCE dataset, the anomalous data are more difficult to differentiated compared to the previous dataset. Therefore, the false positive rate is reported to be more than 0% for both scenarios. This result indicate

that some normal data is classified as an anomalous data instance. This shows that the training set does not well represent the current data situations. Therefore, as the false alarm increases while the detection accuracy is decreased.

4.3 The result of Dataset PDG

The multivariate dataset is presented in PDG dataset by using two variables. Ambient and surface temperature variables are selected to evaluate the effectiveness performance. In scenario 1 and scenario 2, 700 and 2000 data instances are used respectively as training set and to build normal reference model.

Table 5. Effectiveness Results for PDG Dataset with Different Scenario

Scenario	*Training Set Size and Position*	*DR (%)*	*ACC (%)*	*FPR (%)*	*FNR (%)*
1	701-1400 (700)	100	81.6	32.7	0
2	1667-3666 (2000)	99.4	75.6	29.3	0.01

For PDG dataset, Scenario 1 presented 100% detection rate which as same as previous datasets. Meanwhile, 99.4% of detection rate the second scenario. Low detection accuracy is reported as reflected by the high false positive rate in both scenarios. This is due to PDG dataset is tested with multivariate data as well as the noise anomalies is reported in PDG. Noise anomalies represent the increases in value in the dataset shows fluctuating value in the histogram. However, the false negative rate shows a good result for both scenarios which is 0% and 0.01% for scenario 1 and scenario 2 respectively. This indicated the classifier successfully differentiates the anomalous from the normal data instances.

4.4 The result of Dataset NAMOS

The last dataset used chlorophyll concentration from NAMOS dataset and tested and evaluated in two scenarios as shown in Table 6. From the histogram plot in Figure 1(d), constant anomalies are demonstrated in NAMOS dataset where a long period of anomalies shown around epochs 9001.

Table 6. Effectiveness Results for NAMOS Dataset with Different Scenario

Scenario	*Training Set Size and Position*	*DR (%)*	*ACC (%)*	*FPR (%)*	*FNR (%)*
1	1-3000(3000)	100	100	0	0
2	3001-6000 (3000)	100	99.9	0	0.001

NAMOS dataset demonstrates good results for all performance measures in both scenarios as compared to previous datasets. Furthermore, the false negative rate is 0.001% which affected the detection accuracy in scenario 2. However, the result is better than PDG dataset. This is due to that PDG dataset is tested with multivariate variables while in NAMOS dataset is tested using a univariate variable. Moreover, the constant anomalies reported in the dataset make the classifier easier to distinguish the normal and anomalous data.

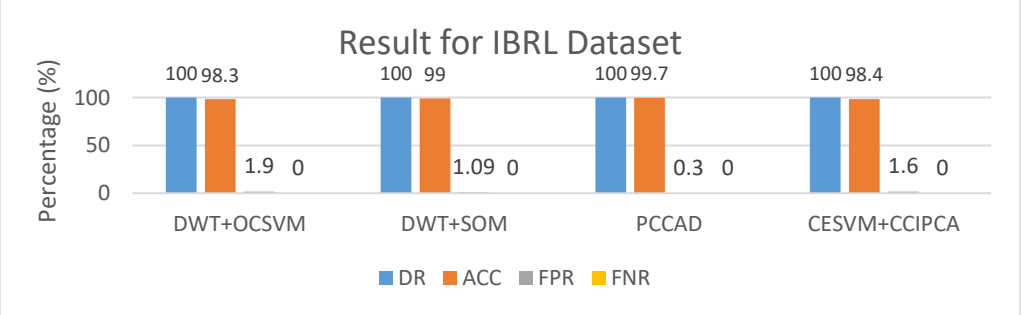
As the research is using the same histogram-based data samples, the results using CESVM classifier with CCIPCA dimension reduction (CESVM+CCIPCA) is then compared with other PCCAD [8], DWT+SOM [22] and DWT+OCSVM [17] anomaly detection schemes using WSNs datasets. Table 7 presented the effectiveness performance measured with other schemes and the graphical result is presented in Figure 7 (a) – (d).

Table 7. The Comparison of Effectiveness Evaluation with Other Related Anomaly Detection Schemes Using Histogram-Based Labelling

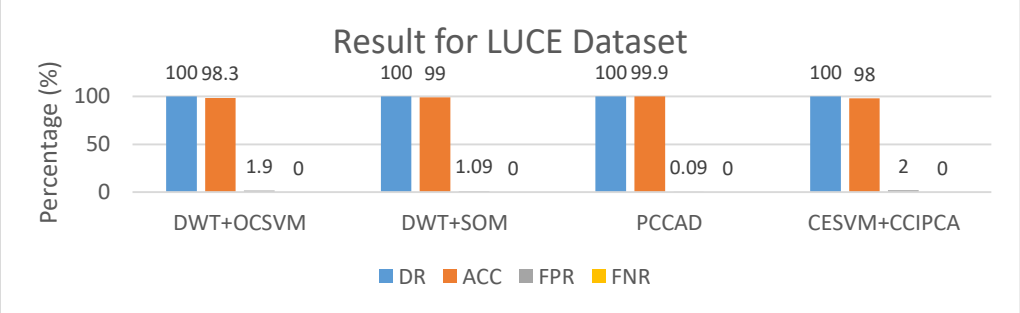
Dataset	Model	DR	ACC	FPR	FNR
IBRL	DWT+OCSVM	100	98.3	1.9	0
	DWT+SOM	100	99	1.09	0
	PCCAD	100	99.7	0.3	0
	CESVM+CCIPCA	100	98.4	1.6	0
LUCE	DWT+OCSVM	100	98.3	1.9	0
	DWT+SOM	100	99	1.09	0
	PCCAD	100	99.9	0.09	0
	CESVM+CCIPCA	100	98	2	0
PDG	DWT+OCSVM	99.7	97.6	2.6	0.3
	DWT+SOM	83	97.8	0.5	16.5
	PCCAD	97.9	96.7	3.5	2.1

	CESVM+CCIPCA	99.1	78.6	25.8	0.01
NAMOS	DWT+OCSVM	100	88.6	12.8	0
	DWT+SOM	100	99.4	0.5	0
	PCCAD	100	90.2	11.5	0
	CESVM+CCIPCA	100	100	0	0

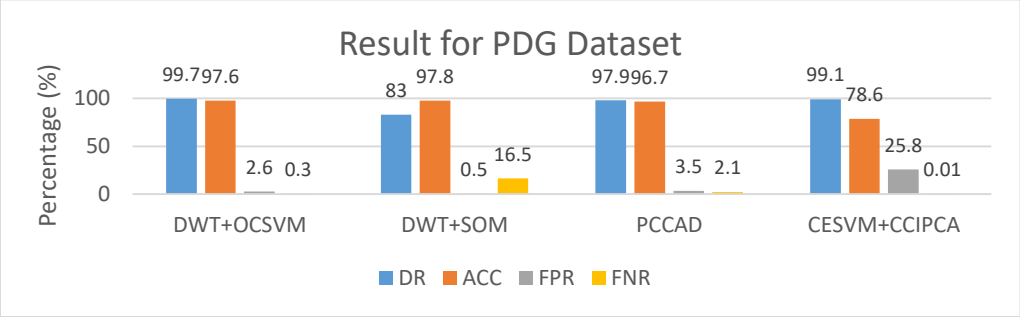
From Table 7 and Figure 7, the experiments using CESVM + CCIPCA demonstrated comparable performance measure results for all the anomaly detection schemes using the same WSNs datasets. In terms of detection rate and false negative rate, CESVM+CCIPCA shows almost the higher detection value and lower false negative rate. Meanwhile, detection accuracy shows the lowest value in PDG dataset which reflecting the high false positive results. The detection accuracy and false negative rate show the comparable result in IBRL and LUCE datasets, especially with the DWT+OCSVM anomaly detection scheme due to the SVM-based classifier used for both schemes. Overall the CESVM+CCIPCA schemes show the best detection accuracy and with the lower false negative rate for NAMOS datasets compared to all schemes.



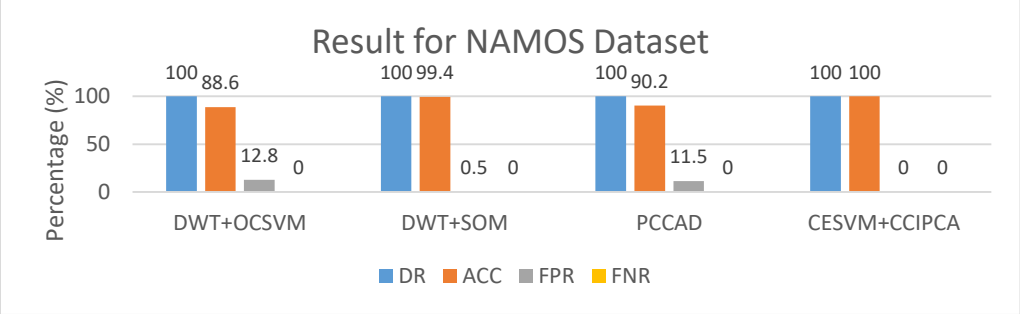
(a) IBRL Dataset



(b) LUCE Dataset



(c) PDG Dataset



(d) NAMOS Dataset

Figure 7: Effectiveness Evaluation with Other Related Anomaly Detection Schemes Using Histogram-Based Labelling

5 Conclusion

In this paper, unsupervised classification-based anomaly detection scheme based on OCSVM is used to evaluate the effectiveness performance in term of detection rate, detection accuracy, false positive rate and false negative rate. Unsupervised OCSVM is suggested to be used in anomaly detection schemes in WSNs data as the absence of ground truth labeling data collected from the sensor nodes. To reduce the recourse constraint incurred in sensor nodes, one PCA variant known as CCIPCA dimension reduction is used to minimize the computational complexity of the CESVM classifier. Meanwhile, histogram-based data labeling technique is used to label the dataset to use as the training set. Environmental sensor dataset collected from IBRL, LUCE, PDG, and NAMOS are used in the experiments and the effectiveness performance is compared with the anomaly detection schemes that incorporated dimension reduction technique. The results show the CESVM+CCIPCA anomaly detection scheme is comparable in most of the WSNs dataset in term of performance measured mentioned above.

ACKNOWLEDGEMENTS.
This work is supported by the Ministry of Higher Education (MOHE) and Research Management Centre (RMC) at the Universiti Teknologi Malaysia (UTM) under UTM GUP Grant 16H56.

References

[1] Rassam, M. A., Zainal, A., & Maarof, M. A. (2013). Advancements of data anomaly detection research in wireless sensor networks: a survey and open issues. *Sensors (Basel).*, 13, 10087–10122.

[2] Chandola, V., Banerjee, A., & Kumar, V. (2009). Anomaly detection: A Survey. *ACM Comput. Surv.*, 41(3), 1–58.

[3] Xie, M., Han, S., Tian, B., & Parvin S. (2011). Anomaly detection in wireless sensor networks : A survey. *J. Netw. Comput. Appl.*, 34(4), 1302–1325.

[4] Zhang, Y., Meratnia, N., & Havinga, P. (2010). Outlier detection techniques for wireless sensor networks : A survey. *IEEE Communications Surveys & Tutorials* 12(2), 159–170.

[5] Rajasegarar, S., Leckie, C., & Palaniswami, M. (2008). Anomaly detection in wireless sensor networks. *IEEE Wirel. Commun.*, 15(4), 34–40.

[6] Zhang, Y., Meratnia, N., & Havinga, P. (2009). Hyperellipsoidal SVM-based outlier detection technique for geosensor networks. I*n n GeoSensor Networks, 2009*. *International conference o on (*pp. 31–41). Springer.

[7] Rajasegarar, S., Leckie, C., Bezdek, J. C., & Palaniswami, M. (2010). Centered hyperspherical and hyperellipsoidal one-class Support Vector Machines for anomaly detection in sensor networks. *IEEE Transactions on Information Forensics and Security*, 5(3), 518–533.

[8] Rassam, M. A., Maarof, M. A., & Zainal, A. (2014). Adaptive and online data anomaly detection for wireless sensor systems. *Knowledge-Based Syst.*, 60, 44–57.

[9] Moshtaghi, M., Leckie, C., Karunasekera, S., & Rajasegarar, S. (2014). An adaptive elliptical anomaly detection model for wireless sensor networks. *Comput. Networks*, (64), 195–207.

[10] Zhang, Y., Meratnia, N., & Havinga, P. (2009). Adaptive and online one-class support vector machine-based outlier detection techniques for wireless sensor networks. In *Advanced Information Networking and Applications Workshops, 2009. AINA '09. 23rd IEEE International Conference on.* (pp. 990–995). IEEE.

[11] Shahid, N., Naqvi, I. H., & Bin Qaisar, S. (2015). One-class Support Vector Machines: Analysis of outlier detection for wireless sensor networks in harsh environments. *Artif. Intell. Rev.*, 43(4), 515-563.

[12] Scholkopf, B., Platt, J. C., Shawe-Taylor, J. C., Smola, A. J., & Williamson,

R. C. (2001). Estimating the support of a high dimensional distribution. *J. Neural Comput.*, 13(7), 1443–1471.

[13] Tax D. M. J., & Duin, R. P. W. (1999). Support vector domain description. *Pattern Recognit. Lett.*, 20(11-13), 1191–1199.

[14] Wang, D., Yeung, D. S., & Tsang, E. C. C., (2006). Structured one-class classification. *IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics)* 36(6), 1283–1295.

[15] Laskov, P., Schäfer, C., Kotenko, I., &Müller, K. R. (2004). Intrusion detection in unlabeled data with quarter-sphere support vector machines. In *Detection of Intrusion and Malware & Vulnerability Assessment, Proceedings. DIMVA Conference on* (pp. 71-82).

[16] Rajasegarar, S., Leckie, C., & Palaniswami, M. (2008). CESVM : Centered Hyperellipsoidal Support Vector Machine based anomaly detection. In *Communications, 2008. ICC '08. IEEE International Conference on* (pp. 1610–1614). IEEE.

[17] Takianngam, S., & Usaha, W. (2011). Discrete wavelet transform and one-class support vector machines for anomaly detection in wireless sensor networks. In *Intelligent Signal Processing and Communications Systems (ISPACS), 2011. International Symposium on* (*pp. 1–6).*IEEE.

[18] Rajasegarar, S., Leckie, C., Palaniswami, M., & Bezdek, J. C. (2007). Quarter sphere based distributed anomaly detection in wireless sensor networks. In *Communications, 2007. ICC '07. IEEE International Conference on* (pp. 3864–3869). IEEE.

[19] Shahid, N., Naqvi, I. H., & Bin Qaisar, S. (2012). Quarter-Sphere SVM : attribute and spatio-temporal correlations based outlier & event detection in wireless sensor networks. In *Wireless Communications and Networking Conference (WCNC), 2012. WCNC 2012 IEEE on* (pp. 2048–2053). IEEE.

[20] Ghafoori, Z., Rajasegarar, S., Erfani, S. M., Shanika, K., & Leckie, C. A. (2016). Unsupervised parameter estimation for one-class support vector machines. In *Lecture Notes in Computer Science,2016. Pacific-Asia Conference on Knowledge Discovery and Data Mining, 2016 PAKDD on*, (pp. 183–195). Springer.

[21] Erfani, S. M., Rajasegarar, S., Karunasekera, S., & Leckie, C. (2016). High-dimensional and large-scale anomaly detection using a linear one-class SVM with deep learning. *Pattern Recognit.*, 58, 121–134.

[22] Siripanadorn, S., Hattagam, W., & Teaumroong, N. (2010). Anomaly detection using self-organizing map and wavelets in wireless sensor networks. In *Applied Computer Science, 2010. Proceedings. 10th Scientific and Engineering Academy and Society (WSEAS) International Conference on* (pp. 291–297). ACM.

[23] Rassam, M. A., Zainal, A., & Maarof, M. A. (2013). An adaptive and efficient dimension reduction model for multivariate wireless sensor networks applications. *Appl. Soft Comput. J.*, 13(4), 10087–10122.

[24] Weng, J., Zhang, Y., & Hwang, W.S. (2003). Candid covariance-free

incremental principal component analysis. *IEEE Trans. Pattern Anal. Mach. Intell.*, 25(8), 1034–1040.