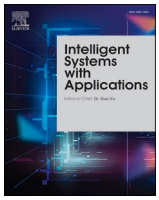
Intelligent Systems with Applications 22 (2024) 200381

Contents lists available at ScienceDirect

Intelligent Systems with Applications

journal homepage: www.journals.elsevier.com/intelligent-systems-with-applications



Cluster-based wireless sensor network framework for denial-of-service attack detection based on variable selection ensemble machine learning algorithms

Ayuba John [a,*], Ismail Fauzi Bin Isnin [b], Syed Hamid Hussain Madni [c], Muhammed Faheem [d]

[a] Faculty of Computing, Federal University Dutse, Jigawa State, Nigeria
[b] Faculty of Computing, Universiti Teknologi Malaysia (UTM), Johor Bahru, Malaysia
[c] School of Electronic & Comp. Sc, University of Southampton, Johor Bahru, Malaysia
[d] School of Technology and Innovations, University of Vaasa, 65200, Vaasa, Finland

ARTICLE INFO

Keywords:
Cluster-based
Wireless sensor network
Machine learning
DoS attacks
Detection system

ABSTRACT

A Cluster-Based Wireless Sensor Network (CBWSN) is a system designed to remotely control and monitor specific events or phenomena in areas such as smart grids, intelligent healthcare, circular economies in smart cities, and underwater surveillance. The wide range of applications of technology in almost every field of human activity exposes it to various security threats from cybercriminals. One of the pressing concerns that requires immediate attention is the risk of security breaches, such as intrusions in wireless sensor network traffic. Poor detection of denial-of-service (DoS) attacks, such as Grayhole, Blackhole, Flooding, and Scheduling attacks, can deplete the energy of sensor nodes. This can cause certain sensor nodes to fail, leading to a degradation in network coverage or lifetime. The detection of such attacks has resulted in significant computational complexity in the related works. As new threats arise, security attacks get more sophisticated, focusing on the target system's vulnerabilities. This paper proposed the development of Cluster-Based Wireless Sensor Network and Variable Selection Ensemble Machine Learning Algorithms (CBWSN_VSEMLA) as a security threats detection system framework for DoS attack detection. The CBWSN model is designed using a Fuzzy C-Means (FCM) clustering technique, whereas VSEMLA is a detection system comprised of Principal Component Analysis (PCA) for feature selection and various ensemble machine learning algorithms (Bagging, LogitBoost, and RandomForest) for the detection of grayhole attacks, blackhole attacks, flooding attacks, and scheduling attacks. The experimental results of the model performance and complexity comparison for DoS attack evaluation using the WSN-DS dataset show that the PCA_RandomForest IDS model outperforms with 99.999 % accuracy, followed by the PCA_Bagging IDS model with 99.78 % accuracy and the PCA_LogitBoost model with 98.88 % accuracy. However, the PCA_RandomForest model has a high computational complexity, taking 231.64 s to train, followed by the PCA_LogitBoost model, which takes 57.44 s to train, and the PCA_Bagging model, which takes 0.91 s to train to be the best in terms of model computational complexity. Thus, the models surpassed all baseline models in terms of model detection accuracy on flooding, scheduling, grayhole, and blackhole attacks.

1. Introduction

In a wireless sensor network, the clustering architecture provides various advantages. It can reduce the size of inter-node communication, for example, by focusing on data transmission inside defined clusters and minimizing the number of transmissions to the base station (Tirani et al., 2020). However, security attacks have been a major concern in both homogeneous and heterogeneous wireless sensor networks according to John and Igimoh (2017). Attackers may purposefully exploit the target system's vulnerabilities and launch various threats to gain access to the system, some of which may divulge sensitive information (Kocher et al., 2020). Unfortunately, new threats and vulnerabilities arise at a rapid pace as attackers become more expert (Zhou et al., 2020). In wireless sensor networks (WSNs), an intrusion detection system (IDS) is a tried-and-true approach for dealing with hostile threats, Saranya et al. (2020) An intrusion detection system (IDS) is a

* Corresponding author.
E-mail addresses: john@graduate.utm.my, ayuba.john@fud.edu.ng (A. John).

https://doi.org/10.1016/j.iswa.2024.200381
Received 4 December 2023; Received in revised form 5 April 2024; Accepted 25 April 2024
Available online 30 April 2024

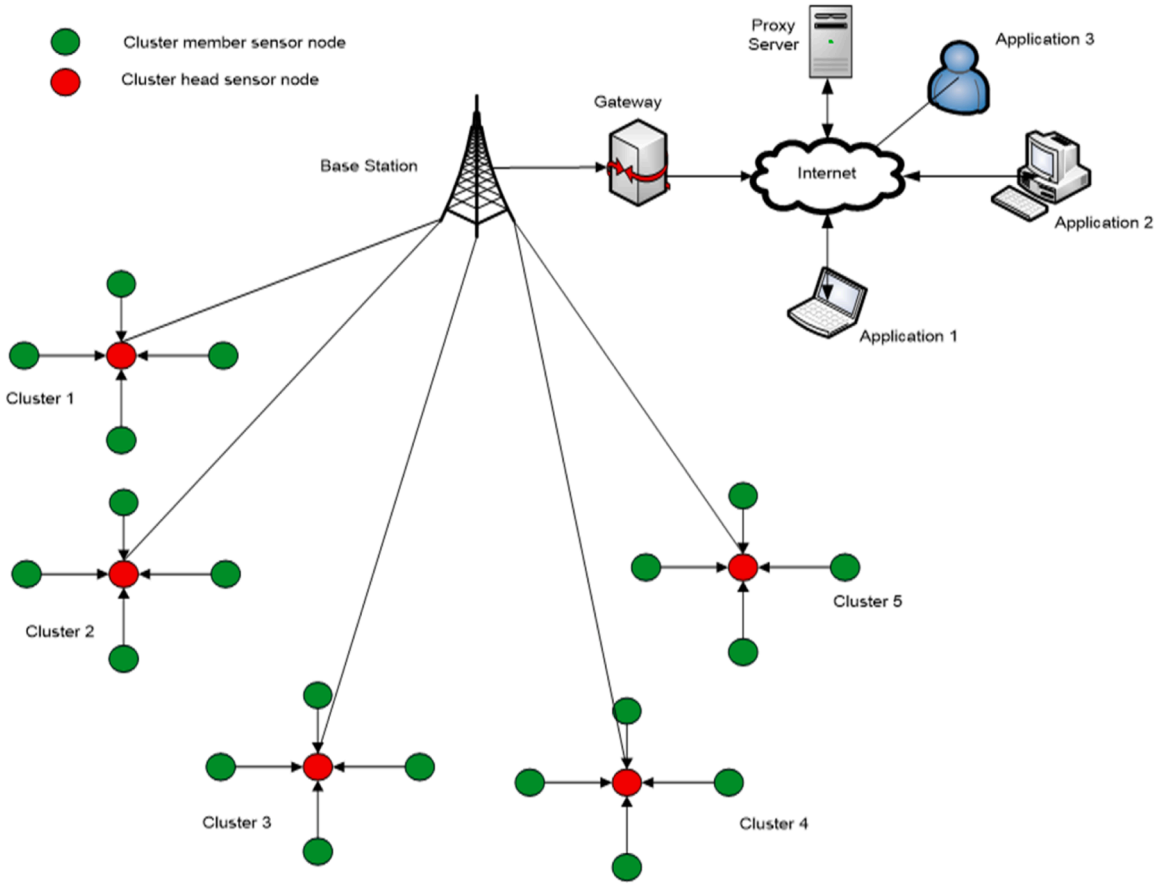


Fig. 1. Cluster-based wireless sensor networks communication system.

software or hardware device that regularly manages and recognizes intrusions and warns the computer or system administrator to perform a specified action. This alert report assists the system administrator or operator in finding and resolving vulnerabilities in the system (Mutlag et al., 2019). According to Li et al. (2022), Anderson pioneered the concept of intrusion detection systems (IDS) in the 1980s, and they can protect a network or host from attacks; however, their design and construction differ. Because WSN is vulnerable to a wide range of attacks, most of the proposed IDS systems now demonstrate some strength, but unfortunately, some of the systems generate computational overhead and consume sensor node resources (Elsaid & Albatati, 2020). Connections in cluster-based wireless sensor networks include sensor node to sensor node, sensor node as a member of the cluster head, cluster to cluster, and lastly cluster head to base station (Sarkar & Senthil Murugan, 2019). Fig. 1 depicts a Cluster-Based Wireless Sensor Network (CBWSN) Communication System. Several ensemble machine learning algorithms have been used to develop intrusion detection systems for DoS attacks, but most researchers have found it difficult to select the appropriate algorithm that could be used to develop a DoS attack detection system model for better performance in detecting several DoS attack classes.

The primary contributions of this paper are the development of Cluster-Based Wireless Sensor Network and Variable Selection Ensemble Machine Learning Algorithms (CBWSN_VSEMLA) as a security framework comprised of intrusion detection models with low computational complexity for DoS attacks detection, which can detect by comparing the performance of several ensemble machine learning algorithms.

The remaining parts of this paper are organized as follows: Section 2 discusses the related research works on the detection system of DoS attacks in cluster-based wireless sensor networks, Section 3 describes the clustering wireless sensor network using Fuzzy C-Means, Section 4 provides the design implementation of the CBWSN-VSEMLA framework, Section 5 describes the design development of cluster-based wireless sensor network, Section 6 discussed on the design development of the variable selection ensemble machine learning algorithm, Section 7 described the specialized cluster-based WSN-DS dataset, Section 8 gives the performance evaluation metrics used for experimental measures, Section 9 discussed the experimental results, and the remainder of the section presented the conclusion, acknowledgment and references.

2. Related works

DoS attacks of various forms have a substantial impact on the security vulnerability of a cluster-based wireless sensor network (John et al., 2023), especially in the communication system illustrated in Fig. 1. Zhiqiang et al. (2022) proposed an enhanced empirical-based component analysis for relevant feature selection techniques and used a short-term memory classifier to develop an intrusion detection technique that has achieved high detection accuracy with no misclassification error, though the accuracy is low on attacks in some datasets. To solve the problem of injection of false data in a cluster wireless sensor network, Lai et al. (2022) proposed a time-spatial and event correlation to develop a detection system that has a high detection rate for malicious nodes but has instability in the data collection generated from the nodes. Khan et al. (2023) present an auto-machine learning framework voting technique to detect a high-speed attack by selecting an optimal classifier to maximize the accuracy, which results in a high delay time for efficient intrusion detection. Ramana et al. (2022) achieved a low computational overhead and improved detection by using the whale optimization gate recurrent unit, though it failed to access changes in sensor data trends.

Premkumar and Sundararajan (2020) achieved high detection of DoS attacks due to a lack of synchronization among nodes by using a deep

learning-based defense mechanism, which is only applicable to nodes with low mobility. A hybrid whale optimization algorithm with an artificial bee colony feature selection technique is proposed by Hussain et al. (2022) which improves the detection rate with less execution time but significantly lower accuracy. Patil and Chaudhari (2016) developed an intrusion detection system using an immune system with fuzzy logic, which has improved the learning ability and has high detection accuracy on DoS attacks, though it has not been experimentally implemented. Gandhimathi and Murugaboopathi (2021) presented a defective mechanism with mobile agents that uses a single mobile agent performance verification at the cluster head that is efficient to verify all nodes to detect malicious nodes and traffic overhead, though it requires more detection time. Ganeshkumar et al. (2016) proposed a jammer detection framework for high detection of jamming attacks but could not determine the position of the jammer node in the cluster.

A Bayes theorem with direct trust value and centralized trust value was used by Basan et al. (2016) to detect node failure malicious attacks and block their activities, but it excluded the significance of unsuccessful events in the network. Kishore and Pappa (2015) proposed elliptic curve cryptography and a Bayesian probabilistic model that were resilient to intruder attacks and logically disconnected insider attackers from the network, though they could not detect a novelty attack in the network. Jianjian et al. (2018) proposed the AdaBoost-RBFSVM algorithm for effective detection and removal of malicious nodes, though it was not able to detect other types of attacks. (Kalnoor & Agarkhed, 2018) proposed an intrusion detection system using the KMP pattern matching technique to detect DoS attacks, but it is limited to some DoS attack detections such as SYN flood, Smurf, and land attacks. A machine learning technique was also introduced by Quincozes et al. (2023) for the detection of DoS attacks, though it had good accuracy on grayhole, blackhole, and flooding attacks but failed to detect scheduling attacks, which were also among the normal traffic flows in the dataset used.

Yu et al. (2021) proposed an artificial neural network with a support vector machine technique for DoS attack detection; it was effective in media access control DoS attacks but had a high time delay for detection. A deep learning technique proposed by Premkumar and Sundararajan (2020) was able to detect and isolate DoS attacks in the data forwarding phase but was not able to eliminate unidentified multiple adversary attacks. NG and Selvakumar (2019) developed an intrusion detection model by using deep radial intelligence and cumulative incarnation, which have improved the detection of DoS attacks and reduced the rate of false alarms, but make it difficult to extract knowledge from the training model. A bio-inspired bat algorithm was used by Sreeram and Vuppala (2019) for fast and early detection of HTTP flood attacks, though it was very effective but could not detect any unknown attack. Cheng et al. (2023) proposed a comparator based on Lyapunov stability theory to develop an intrusion detection for DoS attacks; it effectively resists the influence of intermittent DoS attacks but could not synchronize under a deception attack in a complex network. A finite-time fuzzy technique and Takagi-Sugeno fuzzy model were used by Huang et al. (2022) for the detection of false data injection attacks; they achieved fast and accurate detection but were not effective on a linear complex network.

Poor detection of DoS attacks can deplete sensor node energy and degrade network coverage or lifetime by causing certain sensor nodes to fail. The detection of such attacks has resulted in significant computational complexity in the related works. Implementing a Cluster-Based Wireless Sensor Network with Variable Selection Ensemble Machine Learning Algorithms (CBWSN_VSEMLA) as a security framework will lead to a significant improvement in the detection accuracy of various Denial of Service (DoS) attacks, including grayhole attacks, blackhole attacks, scheduling attacks, and flooding attacks, when compared to related research works. this improvement is expected to be accompanied by a reduction in computational complexity and false alarm rates, thereby enhancing the overall performance and reliability of the CBWSN, mitigating the risk of network coverage failure, and extending the network lifetime.

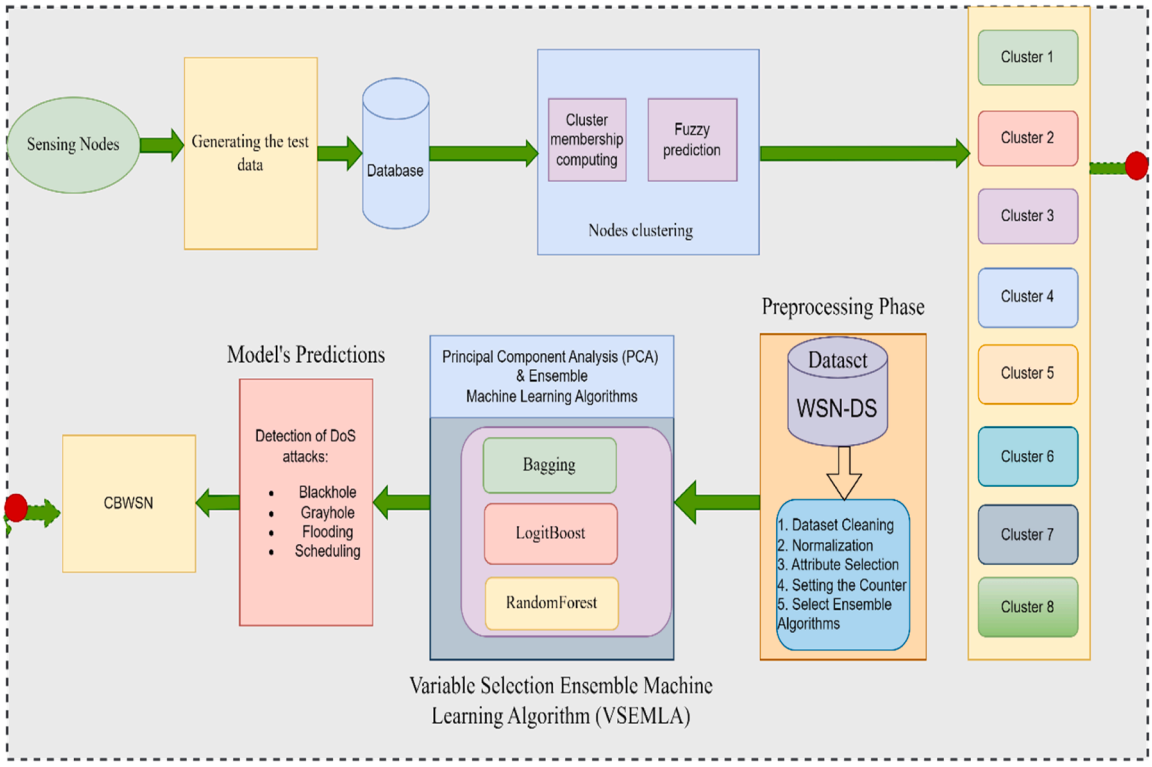


Fig. 2. CBWSN_VSEMLA Security Threats Detection System Framework.

Algorithm 1
The algorithm of the fuzzy C-means for *i*th node.

1. **Input data:** C = [(c_1, c_2, c_3, ..., c_j), a = (a_1, a_2, a_3, ..., a_k), $\boldsymbol{\mu} = \boldsymbol{\mu_{k,j}} \epsilon [\mathbf{0}, \mathbf{1}]$;
2. **Output result:** c, $\boldsymbol{\mu_{k,j}}$ (k = 1, …, m);
3. /* setting the range to a specific number of cluster */
4. In the range of $\mathbf{2} \leq \boldsymbol{n} < \boldsymbol{m}$ and $\mathbf{1} < \boldsymbol{R} < \infty$,
5. Generating the test data for the nodes
6. Randomly initializes the Fuzzy C-Means partition matrix $\boldsymbol{\mu}^{(\mathbf{0})}$,
7. /* Choose the initial centers for the cluster to build the fuzzy model*/
8. Find the cluster nodes i = (0,1, 2, ...) using $\boldsymbol{\mu}^{(i)}$. The $\boldsymbol{j^{th}}$ cluster centre is provided by equation 1.

$$\boldsymbol{C_j} = \frac{\sum_{k=1}^{m} \boldsymbol{\mu}_{kj}^{R} . a_k}{\sum_{k=1}^{m} \boldsymbol{\mu}_{kj}^{R}} \quad \mathbf{1}$$

9. i=1;
10. **While true do**
11. /*Calculate membership matrix */
12. **for** j=1; j≤n; j=++ do

$$\boldsymbol{\mu_{kj}} = \frac{1}{\sum_{i=1}^{n} \left\{ \frac{\|a_k - c_j\|}{\|a_k - c_i\|} \right\}^{\frac{2}{R-1}}} \quad \mathbf{2}$$

13. **End**
14. /* update the cluster centre to μ^((i+1)) to get new centers */
15. i=i+1;
16. **for** j=1; j≤n; j=++ do

$$\boldsymbol{\mu_{kj}} = \frac{1}{\sum_{i=i+1}^{n} \left\{ \frac{\|a_k - c_j\|}{\|a_k - c_i\|} \right\}^{\frac{2}{R-1}}} \quad \mathbf{3}$$

17. **End**
18. /*Use the Euclidean distance between k^{th} data and j^{th} to compare c_j*/
19. **If** $\left\|\boldsymbol{\mu}^{(i+1)} - \boldsymbol{\mu}^{(i)}\right\| < \boldsymbol{\varepsilon}$, then
20. **Break;**
21. **End**
22. **End**
23. /* Showing the fuzzy cluster model built*/
24. $\mathbf{C} = \boldsymbol{\mu}^{(i+1)}$

3. Clustering WSN using fuzzy c-means

In this phase, a Fuzzy logic principle is used to build a cluster-based wireless sensor network model by assigning each sensor node membership in each cluster and classifying new data based on the cluster prediction using a Fuzzy C-Means clustering algorithm. The performance of the cluster-based model can then be evaluated using a Fuzzy Partition Coefficient (Madni et al., 2017). Using a Fuzzy C-Means clustering algorithm may be capable of classifying all nodes into clusters and reducing node isolation during clustering (Rajput & Kumaravelu, 2021).

Clustering is the process of forming groups from a given dataset using specified preconditions (Grachev et al., 2020). Each data point is considered to belong to two or more clusters in Fuzzy C-Means clustering, with membership samples ranged from [0, 1], and the maximum value is known for all clusters equal '1' (Pantula et al., 2020). That is, a set of clusters $c = (c_1, c_2, c_3, \ldots c_j)$ was generated for a set of members with data points $a = (a_1, a_2, a_3, \ldots, a_k)$ and the Fuzzy partition matrix $\mu = \mu_{k,j} \epsilon [0, 1]$, where k = 1 … m, j = 1 … n, and $\mu_{k,j}$ indicates the membership degree of data point a_k in cluster c_j

The Fuzzy C-Means clustering is based on iterative minimization giving as in Eq. (1).

$$jR = \sum_{k=1}^{m} \sum_{j=1}^{n} \mu_{kj}^{R} a_k - c_j^2 1 \leq R < \infty \quad (1)$$

Where "i^{th}" is the iteration step, "j^{th}" is the objective function, "n" is the number of clusters, "R" is any real number, "$\|*\|$" is the inner product norm, "$\mu^{(0)}$" is the initial cluster centre, "a_k" is k^{th} point of the given dataset, "c_j" is the centre of the cluster, "$\| a_k - c_j \|^2$" is the Fuzzy weighting exponent and "μ_{kj}^{R}" is the membership value of any real number of the j^{th} data point in k^{th} cluster. The following are the steps involved in a Fuzzy C-Means clustering algorithm:

Step 1: In the range of $2 \leq n < m$ and $1 < R < \infty$, specify the number of clusters (n) and the real number of parameters (R)
Step 2: Randomly initializes the Fuzzy C-Means partition matrix $\mu^{(0)}$.
Step 3: Find the cluster nodes $i = (0, 1, 2, \ldots)$ using $\mu^{(i)}$. The *j*th cluster centre is provided by Eq. (2).

$$C_j = \frac{\sum_{k=1}^{m} \mu_{kj}^{R} . a_k}{\sum_{k=1}^{m} \mu_{kj}^{R}} \quad (2)$$

Step 4: Update the cluster centre to $\mu^{(i+1)}$ which is provided by Eq. (3).

$$\mu_{kj} = \frac{1}{\sum_{i=1}^{n} \left\{ \frac{\|a_k - c_j\|}{\|a_k - c_i\|} \right\}^{\frac{2}{R-1}}} \quad (3)$$

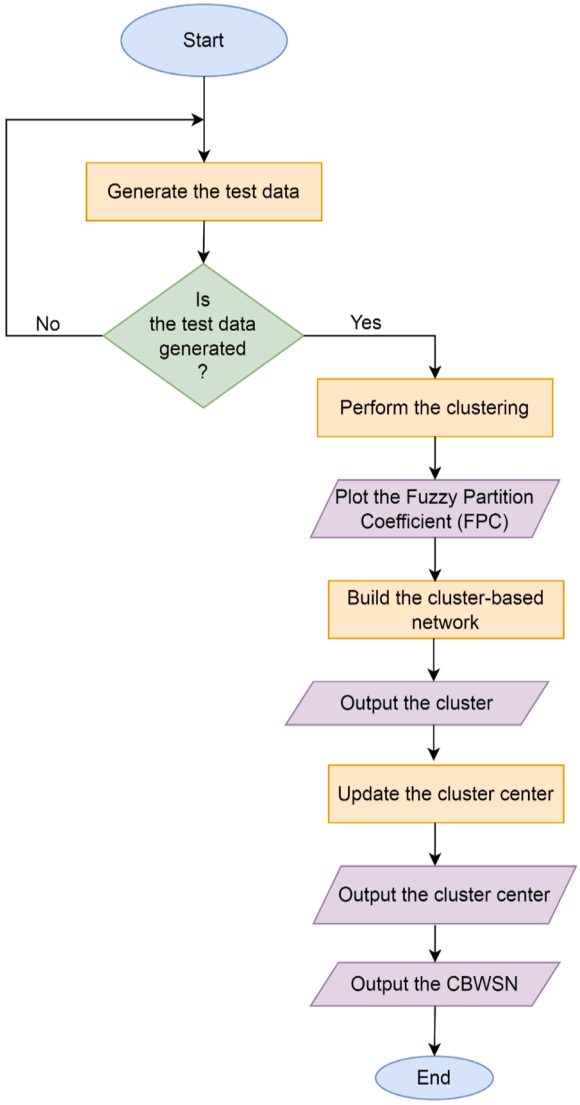


Fig. 3. Flowchart Diagram for the CBWSN Development.

where $\| a_k - c_i \|$ is the Euclidean distance between kth data and jth.

Step 5: Using the following formula, compare the cluster centre to the updated one:

If $\| \mu^{(i+1)} - \mu^{(i)} \| < \varepsilon$, then the iterations will end. Where "ε" is the termination criterion ($0 \leq \varepsilon \leq 1$).

4. The design implementation of CBWSN_VSEMLA framework

Fig. 2 shows the proposed security framework for detecting and evaluating DoS attacks due to node isolation in order to secure the cluster-based wireless sensor network against such attacks. This design merges the Cluster-Based Wireless Sensor Network (CBWSN) with a Variable Selection Ensemble Machine Learning Algorithm (VSEMLA). The VSEMLA is designed to accurately detect various types of DoS attacks, such as blackhole attack, grayhole attack, flooding attack, and scheduling attack on the CBWSN.

5. The design development of the CBWSN

This section proposes the use of the Fuzzy C-Means algorithm. The algorithm can be employed to represent and approximate human reasoning. Here, it is utilized to create a Cluster-Based Wireless Sensor Network (CBWSN). The objective is to minimize the Fuzzy C-Means objective function and generate a CBWSN with fewer isolated nodes, which represents the underlying topology of the network. According to the flowchart for the proposed Algorithm 1 as shown in Fig. 3, all nodes generated in the network tend to execute the Fuzzy C-Means algorithm synchronously distributed in order to obtain the final centres and the membership degree value of the Fuzzy Partition Matrix, which describes the degree of membership of "k" nodes that belong to the cluster "j" obtained by each node.

The algorithm of the Fuzzy C-Means for ith node above consists of the following steps:

I. Defining the cluster centres and limiting the range to a certain number of clusters so that the number of clusters "n" is not larger than two but less than "m" membership degree of a real integer less than infinity, and then generating test data for the nodes.
II. To build the Fuzzy model, choose the initial cluster centres by randomly initializing the Fuzzy C-Means partition matrix.
III. To build the cluster network, find the cluster nodes by calculating the membership matrix.
IV. Update the cluster centres of the original Fuzzy C-Means partition matrix to get new centres for the next iteration and continuing comparing it with the Euclidean distance between each cluster and its membership until the Euclidean distance achieves the termination condition.
V. To gain the new membership for each cluster centre for the CBWSN of the approximation imagination, use the newly obtained Fuzzy C-Means model to detect node's isolation by generating uniformly sampled data dispersed across the range of the network's "X" and "Y" planes.

6. The design development of VSEMLA

The operational flowchart of the Variable Selection Ensemble Machine Learning Algorithm (VSEMLA) module for the proposed security framework of CBWSN is described in Fig. 4. The WSN-DS dataset is a specialized network traffic dataset for CBWSN, which includes normal traffic flows and four types of DoS attacks: flooding, scheduling, grayhole, and blackhole attacks. The data was initially uploaded into the WEKA simulation environment and cleaned. The attack attribute was then selected. To ensure that the data was cleaned, null values were checked for before normalization. If any duplicates or null values were found, the cleaning process was repeated until they were all removed. Finally, the data was visualized to show the number of data points per class distribution or the relationship among the attributes. A counter was set for the machine learning algorithms used in the ensemble method. Principal Component Analysis (PCA) was chosen for each algorithm to detect and evaluate Denial of Service (DoS) attacks present in the network traffic flows. The system checks the number of algorithms selected for each alteration, and when it is less than three, it saves the detected attacks and informs the counter to select the next algorithm for attack detection. Otherwise, the system outputs the detected attack classes for the entire iteration and ends the process. Algorithm 2 depicts the entire process involved.

7. The WSN-DS dataset

The WSN-DS is a specialized dataset for detecting four types of DoS attacks in a wireless sensor networks, specifically Cluster-Based Wireless Sensor Networks (CBWSN): blackhole, flooding, grayhole, and scheduling attacks, all of which are referred to as DoS attacks. Almomani et al.

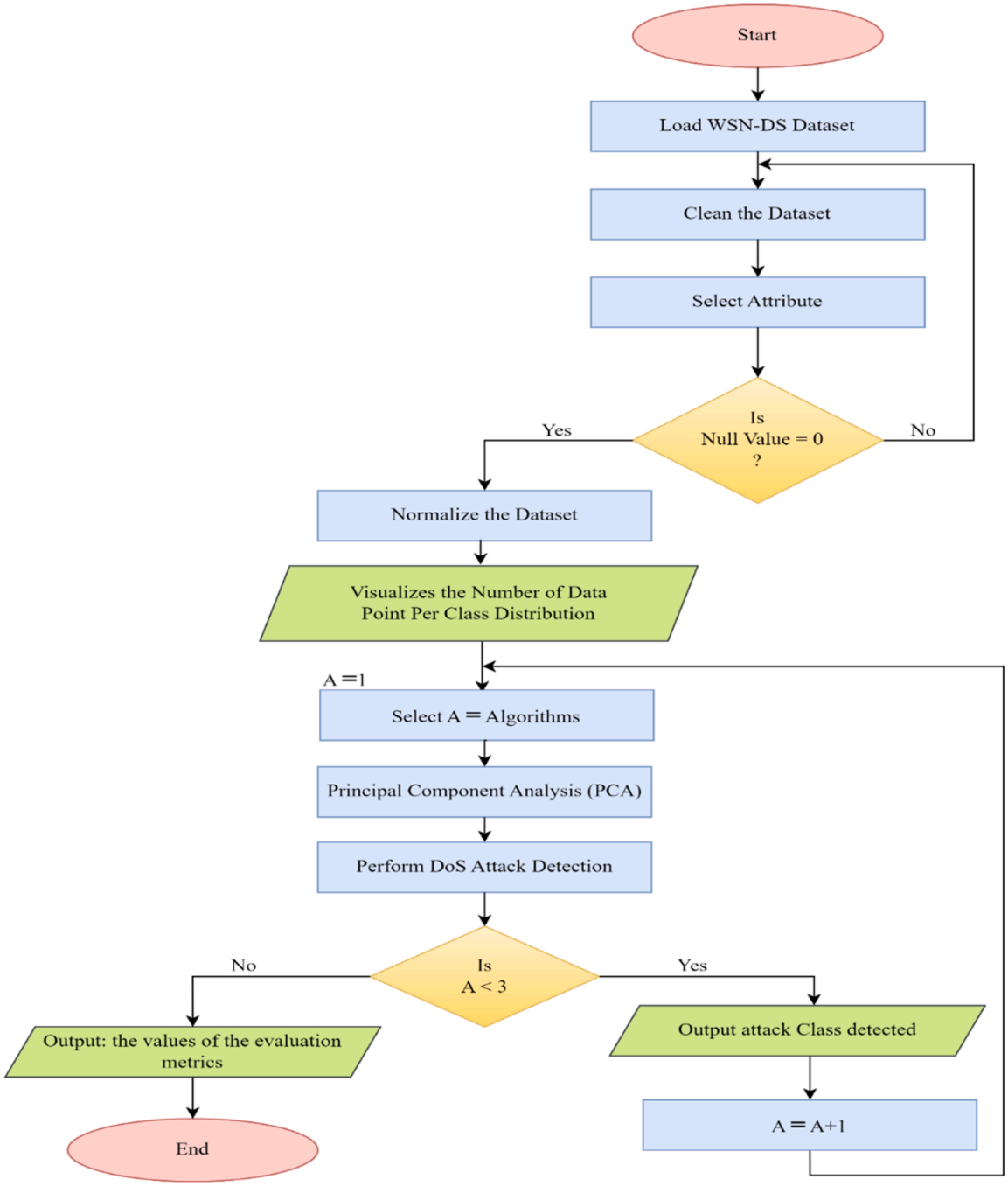


Fig. 4. Flowchart diagram for VSEMLA.

(2016) created the dataset in a Network Simulation Two (NS2) environment with 100 nodes in a 10,000 m^2 region, resulting in eighteen different attributes of a class label of around 374,661 data records for intrusion detection systems in wireless sensor networks. The dataset can be used to prevent infiltration by prohibiting malicious nodes from entering the network, with DoS attacks being the most hazardous and damaging to WSNs due to vulnerabilities to security threats. Table 1 gives the attack labels in the WSN-DS dataset and the amount together with the relative weight of each attack label, while Table 2 shows the relationship between the labels based on the number of attributes, instances, and the sum of the weight.

Gray hole attack

The attacker established a link with each node as shown in Fig. 5; in order to disrupt the network and communicate with other nodes at high speeds over the networks in order to breach routing in sensor networks, due to the fact that routing protocols have no measures in place to prevent attacks (Boubiche et al., 2021; Fang et al., 2020). This malicious node functions as a normal node, discarding selective packets that pass through it and sending malicious packets to the next node. They disrupt the normal operation of the CBWSN by packet creation and can act as isolated nodes that are not part of the network but place hostile nodes to carry out the network attack (Khan et al., 2021; Younas et al., 2022).

Algorithm 2
Algorithm for detecting DoS (Flooding, Scheduling, Grayhole, and Blackhole) attacks.

1. **Input:** WSN-DS dataset for training and testing;
2. **Output:** Classification, TP Rate, FP Rate, Precision, Recall, F1-Scotre, ROC Area, Attack Class;
3. **Begin:** Data preprocessing;
4. Cleaning;
5. Normalization;
6. **End;**
7. **Begin:** Feature extraction;
8. Use Principal Component Analysis (PCA) to select the feature;
9. **End;**
10. **Begin:** Classification;
11. Train the Classifiers (Bagging, LogitBoost, and RandomForest);
12. Test evaluation on the WSN-DS test dataset by fitting into the
13. Classifiers (Bagging, LogitBoost, and RandomForest)
14. To detect attacks (Flooding, Scheduling, Grayhole, Blackhole);
15. **End;**
16. **Return:** the classification results;

Table 1
Description of WSN-DS dataset.

Attack Labels	Count	Weight
Normal	340,066	340,066.0
Flooding	3312	3312.0
Scheduling	6638	6638.0
Grayhole	14,596	14,596
Blackhole	10,049	10,049.0

Table 2
The Relation in WSN-DS dataset.

Relation	WSD-DS
Attributes	19
Instances	374,661
Sum of weight	374,661

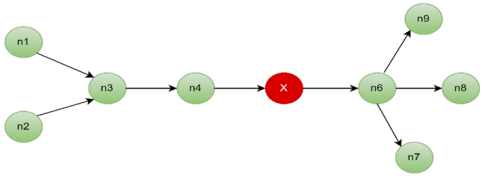


Fig. 5. Gray hole attack.

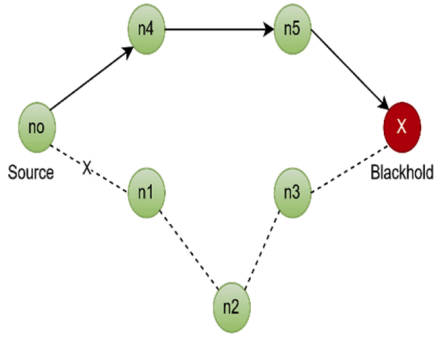


Fig. 6a. Black hole attack.

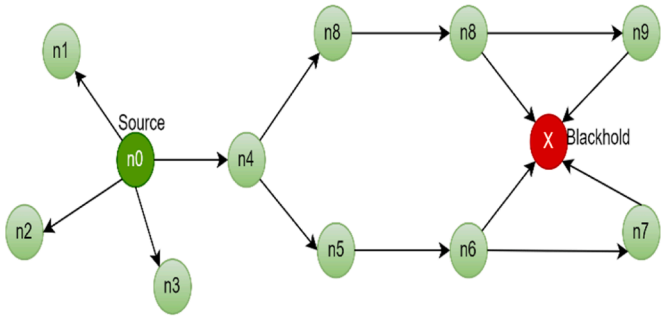


Fig. 6b. Black hole attack.

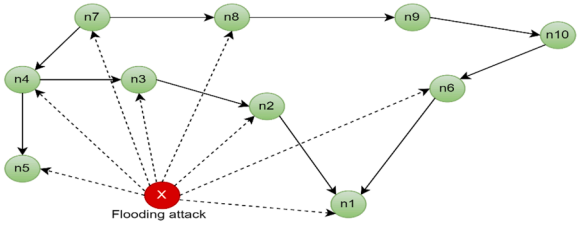


Fig. 7. Flooding attack.

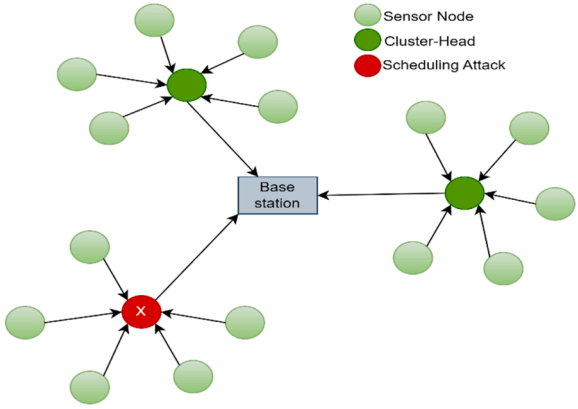


Fig. 8. Scheduling attack.

Black hole attack

This malicious node will establish connections with the normal nodes and advertise itself as the shortest path as shown in Fig. 6a, with the goal of absorbing all packets from the traffic flow that are drawn to it as their final destination as shown in Fig. 6b, preventing them from being transmitted any further (Pullagura & Dhulipalla, 2023; Reddy & Dhananjaya, 2022).

Flooding attack

The attacker broadcasts a HELLO packet to all nodes in the network within radio range as shown in Fig. 7, and when the nodes receive the packets, they assume it is a message from normal nodes and acknowledge the transmission from the attacker, while nodes further away from the radio range send a packet into oblivion, keeping the network confused (Maurya & Kushwaha, 2022; Srinivas & Manivannan, 2020). The attacker broadcasts the packets to the nodes with the highest transmitting power, depleting the sensor's energy by receiving a large number of broadcasting packets. This might cause the failure of some sensor nodes, limiting network coverage and hence the sensor network's lifetime (Islam et al., 2021; Radhika et al., 2022).

Scheduling attack

This form of DoS attack typically occurs when the cluster head is configured for data transfer time (Eliyan & Di Pietro, 2021; Sharathkumar & Sreenath, 2023). The attacker acts as the cluster head as shown in Fig. 8, designating some sensor nodes as cluster members with a set time slot for transmitting packets to the cluster head. The attack is carried out by changing the broadcast mode to unicast, resulting in packet collisions and packet loss during transmission (Jayabalan & Pugazendi, 2022; Yoon & Kim, 2021).

8. Performance evaluation metrics

The performance of a classifier is evaluated using certain parameters such as Recall Score, Precision Score, F1 Score, Detection Rate (DR), False-alarm Rate (FR), Accuracy, and the Area Under the Curve (AUC)-Receiver Operation Characteristic (ROC), which helps in assessing the performance of the classifier. Additionally, the fuzzy partition coefficient (FPC) is used to evaluate the performance of the cluster-based wireless sensor network or to determine the optimal number of clusters. The evaluation of performance indicators consists of several component matrices, which are identified below.

TN = True negative that signify correctly predicted as normal.
FN = False negative which signify mis-predicted as normal.
TP = True positive which signify correctly predicted as abnormal.
FP = False positive which signify mis-predicted as abnormal.

The aforementioned parameters are combined to form various equations which serve as the metrics predominantly used in research-related works presented in the literature review. These equations are the evaluation indicators that was utilized during the experiment to determine the accuracy, precision, recall score, F1 score, area under the curve (AUC) of the receive operational characteristics (ROC), and algorithm running time. The accuracy is defined as the percentage of sample data that has been correctly identified as normal or abnormal data, as demonstrated in Eq. (4).

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

Precision refers to the percentage of accurately predicted data out of the total data predicted to exhibit abnormal behavior. A high precision value indicates a lower error rate of the algorithm used in the model for normal behavior of the data, as shown in Eq. (5).

$$Precision = \frac{TP}{TP + FP} \tag{5}$$

The recall score is a percentage that represents the number of accurately predicted abnormal behaviors out of the total abnormal data, as shown in Eq. (6). A higher recall score indicates that the model has a lower mis-detection rate for abnormal behavior.

$$Recall\ Score = \frac{TP}{TP + FN} \tag{6}$$

The F1 Score is the harmonic multiplication of precision and recall, indicating model performance quality, as shown in Eq. (7).

$$F1\ Score = \frac{2 * Precision * Recall\ score}{Precision + Recall\ Score} \tag{7}$$

9. Experiment and results discussion

The experiments are performed on a Python environment and a Weka software environment concurrently due to a lack of adequate resources for the implementation of wireless sensor networks (Bhushan & Sahoo, 2020), which is one of the major challenges of the simulation of the security framework designed for CBWSN. The system configurations involved in the implementation are as follows: processor 13th Gen Intel (R) Core (TM) i7-1355U 1.70 GHz, Installed RAM of 8.00 GB (7.72 GB usable) on a system type 64-bit operating system, x64-based processor, Windows 11 Home Single Language.

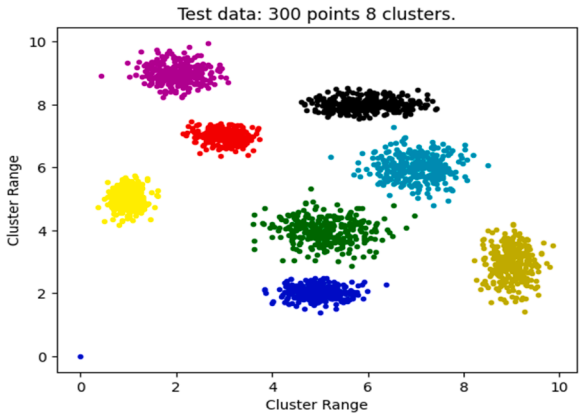


Fig. 9. Visualization of the test data 300 points 8 clusters.

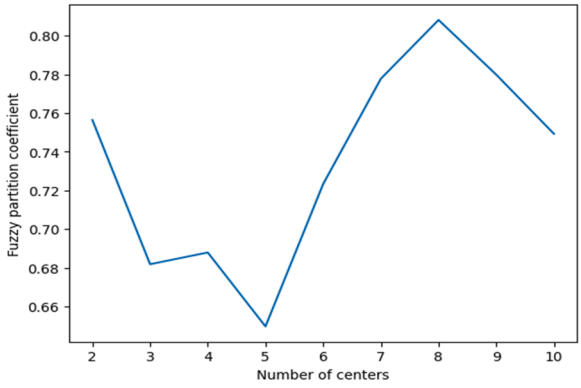


Fig. 10. Graph of The Fuzzy Partition Coefficient (FPC).

Fig. 11. Showing the 8-clusters of the trained model of CBWSN.

An experiment was conducted to exhibit the significance of isolated nodes in establishing cluster-based wireless sensor networks. To form eight clusters, 300 points were generated as test data in a Python environment for each cluster, as shown in Fig. 9.

The data was processed using the fuzzy C-means approach using Skfuzzy. The C-Means function library is used to create the cluster-based model, and a Fuzzy Partition Coefficient (FPC) is used as the metric to measure the cluster points, which are in the range between 0 and 1 on

Table 3
PCA_Bagging model.

Class Attacks	TP Rate	FP Rate	Precision	Recall	F-Measure	ROC Area
Normal	0.999	0.016	0.998	0.999	0.999	0.999
Flooding	0.993	0.000	0.968	0.993	0.980	1.000
Scheduling	0.927	0.000	0.999	0.927	0.962	0.997
Grayhole	0.992	0.000	0.994	0.992	0.992	1.000
Blackhole	0.998	0.000	0.991	0.998	0.994	1.000

Table 4
PCA_LogitBoost model.

Class Attacks	TP Rate	FP Rate	Precision	Recall	F-Measure	ROC Area
Normal	0.998	0.069	0.993	0.998	0.995	0.999
Flooding	0.999	0.001	0.904	0.999	0.949	0.967
Scheduling	0.926	0.001	0.962	0.926	0.943	0.937
Grayhole	0.849	0.002	0.934	0.849	0.890	0.969
Blackhole	0.932	0.001	0.964	0.932	0.948	0.970

Table 5
PCA_RandomForest model.

Class Attacks	TP Rate	FP Rate	Precision	Recall	F-Measure	ROC Area
Normal	1.000	0.000	1.000	1.000	1.000	1.000
Flooding	1.000	0.000	1.000	1.000	1.000	1.000
Scheduling	1.000	0.000	1.000	1.000	1.000	1.000
Grayhole	1.000	0.000	1.000	1.000	1.000	1.000
Blackhole	1.000	0.000	1.000	1.000	1.000	1.000

the vertical axis, with 1 being the best, and the horizontal axis between 2 and 10, as shown in the graph in Fig. 10, with 8 clusters being the best to perform the model clustering; that is to say, the best model has eight output clusters by maximizing the value of the FPC.

The cluster-trained model shows that certain nodes scatter away from their cluster, as seen in Fig. 11. The model provided an approximate human reason for the formation of cluster-based wireless sensor networks (CBWSN). The isolated nodes can be used by an attacker to launch DoS attacks on the network. The targets will be either to destroy the network or to reduce the lifetime of network coverage.

The second experiment involves the use of a WSN-DS dataset. This dataset represents a simulated network traffic in a real-world environment of a cluster-based wireless sensor network. The main goal of this experiment is to evaluate the network's ability to detect four different types of DoS attacks amidst normal traffic. These attacks are known as the flooding attack, the scheduling attack, the grayhole attack, and the blackhole attack. By using PCA to select the feature that was used on each of the following machine learning algorithm ensembles: Bagging, LogitBoost, and RandomForest Algorithms, Tables 3–5 shows the accuracy of each DoS class attack recognized by the built models, as well as the values of the evaluation metrics utilized on each attack (flooding, scheduling, grayhole, and blackhole).

Table 6 compares the three proposed models (PCA_Bagging, PCA_LogitBoost, and PCA_RandomForest) to the baseline IDS models, and Table 7 compares the proposed model's performance and complexity, as measured by training time.

The experimental results of a model performance and complexity comparison for DoS attack evaluation using the WSN-DS dataset show that the PCA_RandomForest IDS model outperforms with 99.999 % accuracy, followed by the PCA_Bagging IDS model with 99.78 % accuracy and the PCA_LogitBoost model with 98.88 % accuracy, as shown in Fig. 12.

Fig. 13 depicted that, in terms of model computational complexity, the PCA_RandomForest model is the most complicated, taking 231.64 s to train, followed by the PCA_LogitBoost model, which takes 57.44 s to train, and the PCA_Bagging model, which takes 0.91 s to train.

The models surpassed all baseline models in terms of model detection accuracy on flooding, scheduling, grayhole, and blackhole attacks. The prediction of the individual model is shown in the confusion matrix in Figs. 14–16. Subsequently, the ROC_AUC curve of the VSEMLA is depicted in Fig. 17 which shows high values of the indicator metric as a better model.

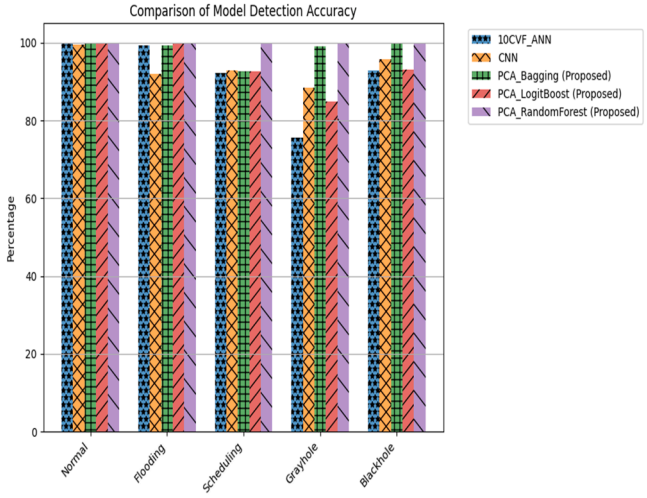


Fig. 12. Bar chart showing the comparison of model detection accuracy.

Table 6
Comparison of model detection accuracy.

Class Attacks	10CVF_ANN (Almomani et al., 2016)	CNN (Salmi & Oughdir, 2023)	PCA_Bagging (Proposed)	PCA_LogitBoost (Proposed)	PCA_RandomForest (Proposed)
Normal	99.8 %	99.49 %	99.9 %	99.8 %	100 %
Flooding	99.4 %	92.06 %	99.3 %	99.9 %	100 %
Scheduling	92.2 %	92.90 %	92.7 %	92.6 %	100 %
Grayhole	75.6 %	88.35 %	99.2 %	84.9 %	100 %
Blackhole	92.8 %	95.78 %	99.8 %	93.2 %	100 %

Table 7
Comparison of model performance and complexity.

Model	Accuracy	Precision	Recall	F1 Score	ROC Area	Training Time
PCA_Bagging	99.78 %	99.8 %	99.80 %	99.80 %	99.9 %	0.91 s
PCA_RandomForest	99.999 %	100 %	100 %	100 %	100 %	231.64 s
PCA_LogitBoost	98.88 %	98.90 %	98.9 %	98.9 %	99.40 %	57.44 s

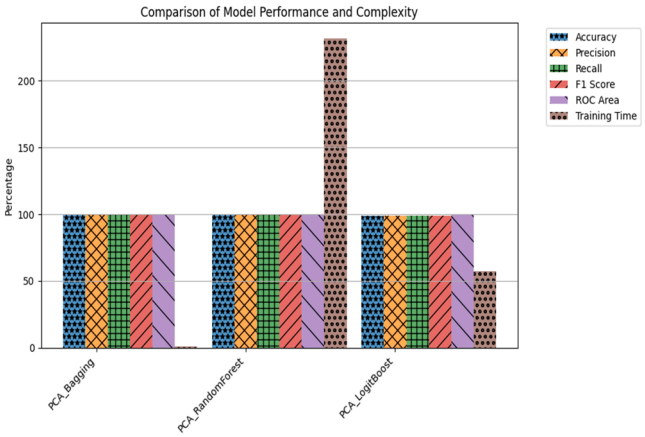


Fig. 13. Bar chart showing comparison of model performance.

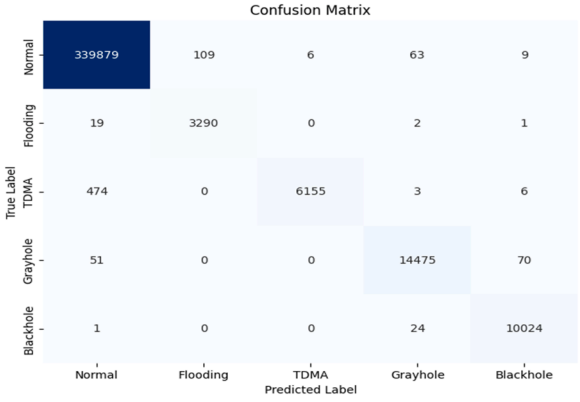


Fig. 14. PCA_ Bagging model prediction.

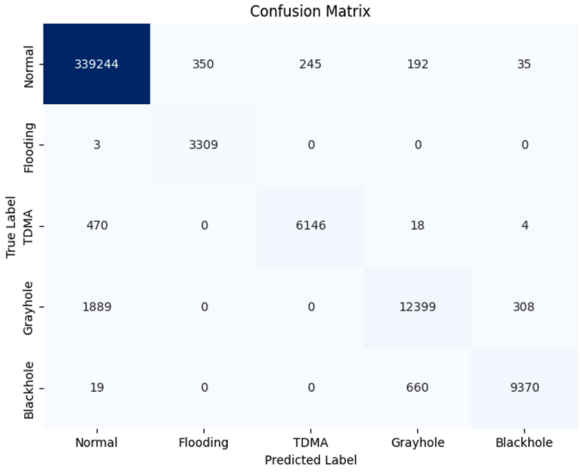


Fig. 15. PCA_ Logit-boost model prediction.

Conclusion

This paper proposes a Cluster-Based Wireless Sensor Networks Variable Selection Ensemble Machine Learning Algorithm (CBWSN_VSEMLA) security threats detection system framework for detecting Denial of Service (DoS) attacks in cluster-based wireless sensor networks, such as grayhole attacks, blackhole attacks, scheduling attacks, flooding attacks, and so on, among normal network traffic flows.

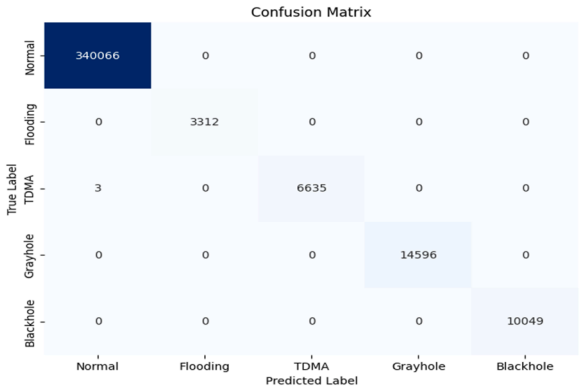


Fig. 16. PCA_ Random-forest model prediction.

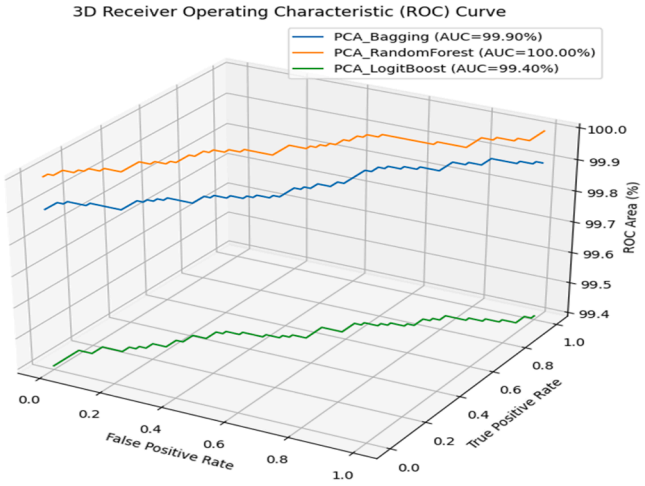


Fig. 17. ROC_AUC curve of VSEMLA.

The CBWSN training model was initially created using the Fuzzy C-Means clustering algorithm to demonstrate the effects of isolated nodes in a clustering wireless sensor network, and it is measured using Fuzzy Coefficient Partition (FCP) to determine the best number of clusters that could be used for the effective model designed. The VSEMLA module is then designed as a detection system for DoS attacks on a cluster-based wireless sensor network by using a Principal Component Analysis (PCA) as the feature selection technique, with numerous ensemble machine learning algorithms (Bagging, LogitBoost, and RandomForest algorithms) as the classifiers; operating in variable selection modes to build an intrusion detection system model. The proposed IDS models (PCA_Bagging, PCA_LogitBoost, and PCA_RandomForest) produce a low false positive rate and high detection accuracy on flooding attacks, scheduling attacks, grayhole attacks, and blackhole attacks, with a low computational complexity in comparison to the baseline IDS models (Almomani et al., 2016; Salmi & Oughdir, 2023).

Further research work will be conducted to incorporate a different security mechanism into the CBWSN_VSEMLA security threats detection system framework that can be used as an intrusion response system (IRS), which will respond to detect attacks by isolating and disconnecting any compromised nodes from the network.

Credit author statement

All authors contributed to this research work.

Declaration of competing interest

The authors declare no conflict of interest.

Data availability

Data will be made available on request.

Acknowledgements

The authors thank the Nigerian Petroleum Technology Development Fund (PTDF) for providing the scholar with a fully funded scholarship to pursue his studies in this field at the Universiti Technologi Malaysia.

This research is supported by the Universiti Teknologi Malaysia (UTM), Malaysia and the University of Vaasa, Finland under project no. WP3-Profi6(2708102611). The authors also would like to thank both universities for providing resources to accomplish this research.

References

Almomani, I., Al-Kasasbeh, B., & Al-Akhras, M. (2016). WSN-DS: A dataset for intrusion detection systems in wireless sensor networks. *Journal of Sensors, 2016*.

Basan, A., Basan, E., & Makarevich, O. (2016). Development of the hierarchal trust management system for mobile cluster-based wireless sensor network. In *Proceedings of the 9th international conference on security of information and networks*.

Bhushan, B., & Sahoo, G. (2020). Requirements, protocols, and security challenges in wireless sensor networks: An industrial perspective. *Handbook of computer networks and cyber security: Principles and paradigms* (pp. 683–713).

Boubiche, D. E., Athmani, S., Boubiche, S., & Toral-Cruz, H (2021). Cybersecurity issues in wireless sensor networks: Current challenges and solutions. *Wireless Personal Communications, 117*, 177–213.

Cheng, Y., Zhang, R., Liu, Y., & Xiao, J. (2023). Secure synchronization control for a class of complex time-delay dynamic networks against denial-of-service attacks. *Journal of The Franklin Institute*.

Eliyan, L. F., & Di Pietro, R. (2021). DoS and DDoS attacks in Software Defined Networks: A survey of existing solutions and research challenges. *Future Generation Computer Systems, 122*, 149–171.

Elsaid, S. A., & Albatati, N. S. (2020). An optimized collaborative intrusion detection system for wireless sensor networks. *Soft Computing*, 1–15.

Fang, W., Zhang, W., Chen, W., Liu, Y., & Tang, C. (2020). TMSRS: Trust management-based secure routing scheme in industrial wireless sensor network with fog computing. *Wireless Networks, 26*(5), 3169–3182.

Gandhimathi, L., & Murugaboopathi, G. (2021). Mobile malicious node detection using mobile agent in cluster-based wireless sensor networks. *Wireless Personal Communications, 117*, 1209–1222.

Ganeshkumar, P., Vijayakumar, K., & Anandaraj, M. (2016). A novel jammer detection framework for cluster-based wireless sensor networks. *EURASIP Journal on Wireless Communications and Networking, 2016*, 1–25.

Grachev, S., Skobelev, P., Mayorov, I., & Simonova, E. (2020). Adaptive clustering through multi-agent technology: Development and perspectives. *Mathematics, 8*(10), 1664.

Huang, X., Li, Z., & Ding, D.-W. (2022). Finite-time attack detection for nonlinear complex cyber-physical networks under false data injection attacks. *Journal of The Franklin Institute, 359*(18), 10510–10524.

Hussain, K., Xia, Y., Onaizah, A. N., Manzoor, T., & Jalil, K. (2022). Hybrid of WOA-ABC and proposed CNN for intrusion detection system in wireless sensor networks. *Optik, 271*, Article 170145.

Islam, M. N. U., Fahmin, A., Hossain, M. S., & Atiquzzaman, M. (2021). Denial-of-service attacks on wireless sensor network and defense techniques. *Wireless Personal Communications, 116*, 1993–2021.

Jayabalan, E., & Pugazendi, R. (2022). Deep learning model-based detection of jamming attacks in low-power and lossy wireless networks. *Soft Computing, 26*(23), 12893–12914.

Jianjian, D., Yang, T., & Feiyue, Y. (2018). A novel intrusion detection system based on IABRBFSVM for wireless sensor networks. *Procedia Computer Science, 131*, 1113–1121.

John, A., & Igimoh, J. A. (2017). Implementation of wireless sensor networks for real time monitoring of oil and gas flow rate metering infrastructure.

John, A., Isnin, I. F., & Madni, S. H. H. (2023). Current security threats in applications of wireless sensor network. *International Journal on Engineering, Science and Technology, 5*(3), 255–272.

Kalnoor, G., & Agarkhed, J. (2018). Detection of intruder using KMP pattern matching technique in wireless sensor networks. *Procedia Computer Science, 125*, 187–193.

Khan, M. A., Iqbal, N., Jamil, H., & Kim, D.-H. (2023). An optimized ensemble prediction model using AutoML based on soft voting classifier for network intrusion detection. *Journal of Network and Computer Applications, 212*, Article 103560.

Khan, T., Singh, K., Hasan, M. H., Ahmad, K., Reddy, G. T., Mohan, S., & Ahmadian, A. (2021). ETERS: A comprehensive energy aware trust-based efficient routing scheme for adversarial WSNs. *Future Generation Computer Systems, 125*, 921–943.

Kishore, R., & Pappa, A. C. (2015). Light weight security architecture for cluster based wireless sensor networks. In *2015 IEEE international conference on ubiquitous wireless broadband (ICUWB)*.

Kocher, P., Horn, J., Fogh, A., Genkin, D., Gruss, D., Haas, W., Hamburg, M., Lipp, M., Mangard, S., & Prescher, T. (2020). Spectre attacks: Exploiting speculative execution. *Communications of the ACM, 63*(7), 93–101.

Lai, Y., Tong, L., Liu, J., Wang, Y., Tang, T., Zhao, Z., & Qin, H. (2022). Identifying malicious nodes in wireless sensor networks based on correlation detection. *Computers & Security, 113*, Article 102540.

Li, Y., Qin, T., Huang, Y., Lan, J., Liang, Z., & Geng, T. (2022). HDFEF: A hierarchical and dynamic feature extraction framework for intrusion detection systems. *Computers & Security*, Article 102842.

Madni, S. H. H., Abd Latiff, M. S., Abdullahi, M., Abdulhamid, S.i. M., & Usman, M. J (2017). Performance comparison of heuristic algorithms for task scheduling in IaaS cloud computing environment. *PLoS ONE, 12*(5), Article e0176321.

Maurya, P., & Kushwaha, V. (2022). Impact analysis of hello flood attack on RPL. In *International conference on advanced network technologies and intelligent computing*.

Mutlag, A. A., Abd Ghani, M. K., Arunkumar, N.a., Mohammed, M. A., & Mohd, O (2019). Enabling technologies for fog computing in healthcare IoT systems. *Future Generation Computer Systems, 90*, 62–78.

NG, B. A., & Selvakumar, S. (2019). Deep radial intelligence with cumulative incarnation approach for detecting denial of service attacks. *Neurocomputing, 340*, 294–308.

Pantula, P. D., Miriyala, S. S., & Mitra, K. (2020). An evolutionary neuro-fuzzy C-means clustering technique. *Engineering Applications of Artificial Intelligence, 89*, Article 103435.

Patil, S., & Chaudhari, S. (2016). DoS attack prevention technique in wireless sensor networks. *Procedia Computer Science, 79*, 715–721.

Premkumar, M., & Sundararajan, T. (2020). DLDM: Deep learning-based defense mechanism for denial of service attacks in wireless sensor networks. *Microprocessors and Microsystems, 79*, Article 103278.

Pullagura, J. R., & Dhulipalla, V. R. (2023). Black-hole attack and counter measure in ad hoc networks using traditional routing optimization. *Concurrency and Computation: Practice and Experience, 35*(9), e7643.

Quincozes, S. E., Kazienko, J. F., & Quincozes, V. E. (2023). An extended evaluation on machine learning techniques for Denial-of-Service detection in Wireless Sensor Networks. *Internet of Things, 22*, Article 100684.

Radhika, S., Anitha, K., Kavitha, C., Lai, W.-C., & Srividhya, S. (2022). Detection of hello flood attacks using fuzzy-based energy-efficient clustering algorithm for Wireless Sensor Networks. *Electronics, 12*(1), 123.

Rajput, A., & Kumaravelu, V. B. (2021). FCM clustering and FLS based CH selection to enhance sustainability of wireless sensor networks for environmental monitoring applications. *Journal of Ambient Intelligence and Humanized Computing, 12*(1), 1139–1159.

Ramana, K., Revathi, A., Gayathri, A., Jhaveri, R. H., Narayana, C. L., & Kumar, B. N. (2022). WOGRU-IDS—An intelligent intrusion detection system for IoT assisted Wireless Sensor Networks. *Computer Communications, 196*, 195–206.

Reddy, B., & Dhananjaya, B. (2022). The AODV routing protocol with built-in security to counter blackhole attack in MANET. *Materials Today: Proceedings, 50*, 1152–1158.

Salmi, S., & Oughdir, L. (2023). Performance evaluation of deep learning techniques for DoS attacks detection in wireless sensor network. *Journal of Big Data, 10*(1), 1–25.

Saranya, T., Sridevi, S., Deisy, C., Chung, T. D., & Khan, M. A. (2020). Performance analysis of machine learning algorithms in intrusion detection system: A review. *Procedia Computer Science, 171*, 1251–1260.

Sarkar, A., & Senthil Murugan, T. (2019). Cluster head selection for energy efficient and delay-less routing in wireless sensor network. *Wireless Networks, 25*, 303–320.

Sharathkumar, S., & Sreenath, N. (2023). Distributed Clustering based Denial of Service Attack Prevention Mechanism using a Fault Tolerant Self Configured Controller in a Software Defined Network.

Sreeram, I., & Vuppala, V. P. K. (2019). HTTP flood attack detection in application layer using machine learning metrics and bio inspired bat algorithm. *Applied Computing and Informatics, 15*(1), 59–66.

Srinivas, T. A. S., & Manivannan, S. (2020). Prevention of hello flood attack in IoT using combination of deep learning with improved rider optimization algorithm. *Computer Communications, 163*, 162–175.

Tirani, S. P., Avokh, A., & Azar, S. (2020). WDAT-OMS: A two-level scheme for efficient data gathering in mobile-sink wireless sensor networks using compressive sensing theory. *IET Communications, 14*(11), 1826–1837.

Yoon, Y., & Kim, H. (2021). Resolving persistent packet collisions through broadcast feedback in cellular V2X communication. *Future Internet, 13*(8), 211.

Younas, S., Rehman, F., Maqsood, T., Mustafa, S., Akhunzada, A., & Gani, A. (2022). Collaborative detection of black hole and gray hole attacks for secure data communication in VANETs. *Applied Sciences, 12*(23), 12448.

Yu, D., Kang, J., & Dong, J. (2021). Service attack improvement in wireless sensor network based on machine learning. *Microprocessors and Microsystems, 80*, Article 103637.

Zhiqiang, L., Mohiuddin, G., Jiangbin, Z., Asim, M., & Sifei, W. (2022). Intrusion detection in wireless sensor network using enhanced empirical based component analysis. *Future Generation Computer Systems, 135*, 181–193.

Zhou, Y., Cheng, G., Jiang, S., & Dai, M. (2020). Building an efficient intrusion detection system based on feature selection and ensemble classifier. *Computer Networks, 174*, Article 107247.

Ayuba John is a PhD candidate in Computer Science at Universiti Teknologi Malaysia (UTM). In 2010, he received his B. Eng. Engineering Degree in Computer Engineering from the University of Maiduguri, Nigeria, and in 2017, he received his M.Eng. Computer Engineering Degree from the University of Benin, Nigeria. He worked for the National Control Centre (NCC) in the Transmission Company of Nigeria (TCN) as a transmission engineer in 2014, and he is a member of the Nigerian Society of Engineers (NSE). He is now a lecturer at Nigeria's Federal University Dutse. His research areas of interest include cybersecurity, machine learning and wireless sensor networks.

E-mail: ayuba.john@fud.edu.ng or john@graduate.utm.my.
https://orcid.org/0000-0003-0496-765x

Ismail Fauzi Bin Isnin received a Ph.D. degree and M.S. degree in Network System Engineering from the University of Plymouth U.K., in the year 2011 and 2004, respectively. Currently, he is a Senior Lecturer in School of Computing, Faculty of Engineering, Universiti Teknologi Malaysia, Malaysia. He is a member of Pervasive Computing Research Group, School of Computing. His research interests are in wired and wireless computers network and communication, mobile ad-hoc network and communication, high performance, and parallel computing. He can be contacted at email: ismailfauzi@utm.my.

https://orcid.org/0000-0002-9765-3491

Syed Hamid Hussain Madni is a currently based in Malaysia and working as a Senior Lecturer at School of Electronic and Computer Science, University of Southampton of Engineering, Malaysia. He received his PhD Degree in 2020 from Universiti Teknologi Malaysia (UTM) and worked as a senior lecturer before he moved to his current station. His area of research is "Optimal Resource Scheduling for Infrastructure as a Service in Cloud Computing based on Cuckoo Search". He has received MS (CS) degree in 2009 from Federal Urdu University Arts, Science and Technology, Islamabad, Pakistan. His areas of interest are cloud computing, analysis of algorithm, network security, e-commerce, web development and Internet of Things. He has published about 18 research papers in in High Impact Journals. He is also conducting trainings on research publications with different International Universities in various countries.

Email: s.h.h.madni@soton.sc.uk or madni4all@yahoo.com
https://orcid.org/0000-0002-3816-1382

Muhammad Faheem received the B.Sc. Computer Engineering degree in 2010 from the Department of Computer Engineering at the University College of Engineering &Technology, Bahauddin Zakariya University Multan, Pakistan. In 2012, he received an MS degree in Computer Science from the school of Computing at Universiti Teknologi Malaysia. He served as a lecturer at Comsats Institute of Information & Technology from 2011 to 2012, Pakistan. Since 2013 he is working as a Sr. Network Researcher at Abdullah Gul University, Kayseri, Turkey. He received his Ph.D. in Computer Science Universiti Teknologi Malaysia. His research interest includes the areas of smart grid communications, energy harvesting, underwater acoustic communications, cognitive radio sensor networks, and information storage and retrieval architecture from the sensor memory. Mr. Faheem has authored several papers in refereed journals and has been serving as a reviewer for numerous Journals, such as Journal of network and computer applications, Ad-hoc networks, Australian journal of electrical engineering, International Journal of Computer Communication & Control, Computer standards and interfaces, IEEE Access, IEEE Transaction on Vehicular Technology Communication magazine and Future Generation Computer Systems. Email: muhammad.faheem@uwasa.fi

https://orcid.org/0000-0003-4628-4486