I.J. Wireless and Microwave Technologies, **2024, 3, 14-30**
Published Online on June 8, 2024 by MECS Press (http://www.mecs-press.org/)
DOI: 10.5815/ijwmt.2024.03.02

Detection of Threats in Wireless Sensor Network Based on Optics Clustering With DE-BiLSTM Classifier

R. Preethi*
Bishop Heber College, Tiruchirappalli - 620017, TN, India
Email: preethihcc@gmail.com
ORCID iD: https://orcid.org/0009-0004-7989-3450
*Corresponding Author

Received: 03 January, 2024; Revised: 31 January, 2024; Accepted: 20 March, 2024; Published: 08 June, 2024

Abstract: An intelligent distributed network system is the Wireless Sensor Network (WSN), which is a strategy required to address security threats as well as energy consumption that has a direct impact on a network's lifetime. Thus, attempting to identify malicious attacks with a low consumption of data transmission makes a lot of sense. The high energy consumption of nodes due to the transmission of data shortens the lifetime of the network. To overcome these issues, the proposed method is based on the Ordering Points to Identify Cluster Structure (OPTICS) with Bi-directional Long Short Term Memory using Differential evolution (DE-BiLSTM) classifier to detect the threats in WSN for smart building. Initial deployment of the sensor nodes (SN) and formation of the cluster nodes (CN) by employing the OPTICS density-based clustering approach that partitions clusters with different densities. In order to transport data to the base station, the cluster head (CH) nodes are chosen from the CN according to their more energy as well as shorter distance. Then, in order to forecast the threats, the size of the batch and hidden layers are set using the differential evolution method (DE) and the classification of the data is performed using BiLSTM to detect as attack or non-attack. Performance for predicting an attack is measured by network and classification parameters such as Packet Delivery Ratio (PDR), Average Residual Energy (ARE), Throughput, Accuracy and Precision. The results of the performance obtained are 91.78% for PDR, 8.56J for ARE, 2.52mbps for throughput with 100 nodes, then 93.78% for accuracy and 93.04% for precision. Thus, the designed detection of threats in WSN based on OPTICS clustering with DE-BILSTM classifier performs better for malicious attack prediction with low energy consumption sensor nodes.

Index Terms: WSN; OPTICS; DE-BiLSTM; SB; Cluster Head.

1. Introduction

In shopping malls and smart buildings (SB), WSN have been crucial for monitoring smoke, fire alarms, security, surroundings and security threats [1]. A WSN is a type of network comprised of several sensor nodes and can serve as an essential component of the Internet of Things (IoT). To gather the necessary data and transmit it to a centralized node, or sink node also called a base station (BS), which is a more potent, capable node utilized in many real-time applications, these nodes are dispersed across a wide variety of different regions [2]. Nowadays, the term "smart" is a catchphrase and is used to describe an improvement in our way of living. Automation is required in this busy society to make daily life simpler. People of today seek simple control and monitoring of their devices. Cities all over the world are starting to see an increase in the number of smart buildings (SB) [3]. Smart buildings are not a new concept, however, new technological advancements have been largely responsible for its evolution resulting in permits resources and processes to be more intelligent, which enhances the building's ability to function in a more effective, adaptable, interactive, and sustainable manner.

A smart building is a building that is equipped with numerous electrical and electronic components which can be controlled and monitored by PC or smartphones [4]. With the help of the IoT, the idea of a smart building has been developed. Essentially, the control and monitoring of home appliances are linked via a complicated network. By interacting with numerous electronic gadgets across wireless means, the IoT provides users with sufficient information [5]. Building management is now a cost-effective and effective solution thanks to the IoT. Sensors, communication buses, actuators, human machine interfaces, controllers, application servers, databases, and cloud components are the primary elements of a building control network, according to a standard smart building automated system design. In large commercial buildings, such components are organized either functionally or according to the module or task to that they

belong in order to handle certain tasks in a particular section of the buildings. The HVAC (Heating, Ventilation, and Air Conditioning) subsystem, security and access control, lighting, networking, elevators and electrical infrastructure are some of the major modules in a modern building [6]. As they are employed for apartment complexes, offices, shopping centers, or other retail establishments, these buildings serve as the fundamental building elements of modern society. Making them intelligent is an excellent first step toward a smart city.

Smart buildings further decrease residential waste of energy with intelligently managing high-energy-demanding building devices. It further introduces some security problems that involve network hacking and attacks in the real worl implementations [7]. The sensor node is susceptible to many malicious attacks, namely Sinkhole, Wormhole, Flooding and Greyhole, among others, because of its limited energy, storage capacity, and computational power due to the nature of wireless transmission and the uncontrolled features of WSN. Such common attacks lead to traffic on the network to divert from normal network traffic, causing significant damage to the WSN in a short period of time [8]. For military applications such as smart logistics, perimeter monitoring, and infiltration detection, wireless sensors are essential. But in order to execute them, limitations like hardware, power consumption, cost, scalability, and fault tolerance must be taken into considerations. Thus, significant technology for identifying network intrusions and transfer data energy efficiently in WSN network is inspired a lot of attention for the efficient and secure networks in recent years. Recent classifications of WSN attack detection include abuse detection, anomaly detection, hybrid system detection and specification-based detection. The most common approaches for detecting malware presently in utilize are Naive Bayes, support vector machines (SVM), artificial neural networks (ANN), Bayesian networks (BN), random forests (RF), decision trees (DT), random weight neural networks, artificial immunity and other approaches. However, the main issues these systems are struggling with computational complexity, a higher false alarm rate, a lower detection accuracy, and a longer training and testing period and also detecting unknown attack is ineffective. To overcome the drawbacks, the designed model proposed to detect threats in WSN based on OPTICS clustering with DE-BILSTM classifier. Major contributions to the proposed method are mentioned as follows:

- Threat detection in wireless sensor network based on OPTICS clustering with DE-BILSTM classifier for smart building.
- Nodes are deployed and clustered using OPTICS, which is a density based clustering to cluster the sensor nodes for reducing energy consumption in WSN.
- The cluster head (CH) is selected based on distance and higher energy from the cluster node and transmits data to the base station.
- Hyperparameter tuned Bi-directional Long Short Term Memory (Bi-LSTM) with differential evolution (DE) model called DE-BiLSTM classifier is used to detect the network as attack or non-attack.

Each section of the article is described in the list that follows: The second sector investigates the numerous articles on attack forecasting in WSN. The third sector provides a brief description of the proposed method. The fourth sector presents experimental results for the proposed approach, and the final section of the research draws a conclusion.

2. Literature Review

Attacks are frequent threats to WSN security because the attacks are extremely simple to launch. Over several decades, numerous research have presented various techniques for attack detection. In this section, some of this past works are reviewed below based on recent machine learning techniques.

Dong et al. [9] demonstrated a WSN intrusion detection model utilizing the bagging algorithm and information gain ratio was presented employing the ensemble learning technique's basic principles. First, in this model, the characteristic of sensor node info about traffic was chosen utilizing the information gain ratio approach. Second, an ensemble classifier was built using the Bagging approach to train several enhanced C4.5 decision trees. The dynamic pruning method was implemented after ten iterations of parameter optimization for the ensemble classifier. Finally, a majority voting strategy was used to classify and detect the classification outcomes of the C4.5 decision tree. The experimental results demonstrated that the introduced model for detection in Blackhole, Flooding, Grayhole, Scheduling, and other attacks.

Biswas and Samanta [10] presented a method for detecting anomalies in WSN using ensemble random forests (ERF). As the ensembles base learners, select Naive Bayes, Decision Tree, and K-Nearest Neighbor. Additionally, bootstrap sampling was employed to build the random forest. Here, the learning algorithm was implemented using the sci-kit learn 0.23.1 machine learning module and Python 3.7.7. Using activity recognition on the basis of a multi-sensor data fusion (AReM) dataset, which was a real-world sensor dataset, researchers evaluated the ERF method.

Ifzarne et al. [11] provided a model for intrusion detection that was appropriate with WSN features. This model was depended on the online Passive aggressive classifier and information gain ratio. In the beginning of the process, the information gain ratio is used to choose the key components of the sensor data. The second step was to train the online passive aggressive algorithm for categorized and recognized various Deny of Service attacks. A dataset from a WSN detection system (WSN-DS) was applied for the studies. The presented model ID-GOPA achieved 96% detection rate when identifying if the network was functioning normally or vulnerable to attacks.

Alqahtani et al. [12] introduced an XGBoot classifier and a genetic algorithm were used to create the GXGBoost model, a new method of detecting intrusion assaults. The current version represents a gradient boosting model created to enhance the capability of conventional models to recognize attack types in attack minorities in the extremely unbalanced data traffic of WSN. On a dataset for the WSN-detection system (WSN-DS), a series of tests were carried out utilizing the holdout and tenfold cross-validation procedures. The presented approach beat previous ensemble learning classifiers for a high rate of detection and state-of-the-art approaches, according to the findings of ten fold cross validation testing.

Hemanand et al. [13] developed a combine the Likelihood Support Vector Machine (LSVM) and Cuckoo Search Greedy Optimization (CSGO) models creates an intelligent IDS system for increasing WSN security. The most widely used network datasets, including UNSW-NB15 and NSL-KDD, are taken into consideration for this model's validation. Initial dataset preprocessing involves the removal of unimportant data, the prediction of missing values, and filtering in order to normalize the attributes. Following preprocessing, the ideal number of features is chosen and inputted into the CSGO algorithm. This method then calculates the ideal fitness function for choosing the best features. Finally, the classification method based on LSVM is used to forecast whether the data would be classed as normal or abnormal.

Most articles are examined above concern the detection of threats in WSN. Some of the limitations are found in the investigated articles are causes a loss in models interpretability [9], ensembling is difficult to learn, each incorrect choice might result in poorer predicted accuracy [10], Lack of data [11], struggles with sparse and unstructured data [12], target classes are overlapping [13]. Thus, the proposed model overcomes the above mentioned problem by using the OPTICS clustering algorithm and DE-Bi-LSTM for detecting attack. Table 1 provides a comparative study of advantages and drawbacks of the existing techniques.

Table 1. A comparative analysis for threat detection with existing methodologies

Author name	Title	Methods	Advantage	Limitation
Shuai Jiang [14]	SLGBM: An Intrusion Detection Mechanism for Wireless Sensor Networks in Smart Environments	SBS was utilized to select sensor node traffic data, while LightGBM was employed to detect traffic attacks under WSN.	Enhances network robustness, provides higher detection rates.	Challenges such as low detection rate, high false alarm rate, and high redundancy.
Thi-Kien Dao [15]	Identification Failure Data for Cluster Heads Aggregation in WSN Based on Improving Classification of SVM	Data aggregation in Hierarchical WSNs by Support Vector Machines (SVM).	Reduced latency, scalability, energy efficiency, and lifetime extension are an advantages of clustering-based WSN.	Network reliability of monitoring and predicting applications at CHs was reduced due to an aggregated data fault problem.
Mnahi Alqahtani [16]	A Genetic-Based Extreme Gradient Boosting Model for Detecting Intrusions in Wireless Sensor Networks	GXGBoost model was developed to detect minority attacks in the imbalanced data traffic of wireless sensor networks.	Increase the effectiveness of WSN intrusion detection.	Imbalanced data problems in network traffic.
Neelakandan Subramani [17]	An Efficient Metaheuristic-Based Clustering with Routing Protocol for Underwater Wireless Sensor Networks	MCR-UWSN technique focused on selecting an efficient set of cluster heads (CHs) and directing them to their destination.	Significantly improve the lifespan of a network.	Because of their low energy capacity and the laborious procedure of battery replacement or recharging, underwater sensors have energy efficiency issues.
Jin Wang [18]	An Improved Routing Schema with Special Clustering Using PSO Algorithm for Heterogeneous Wireless Sensor Network	Energy Centers Searching with Particle Swarm Optimization (EC-PSO) was developed to efficiently discover energy centers for CH selection.	Enhances the scalability of the network	Average hop distance was large, it may result in long-distance communication and a significant use of energy.

3. Proposed Methodology

Internet-based communication is permitted between humans, machines, and every non-living things in the WSN. WSN devices are widely available these days. All of the devices are linked together in a WSN so that data can be sent immediately over an Internet connection. An open connectivity to the internet emerges with linked devices that are vulnerable to security attacks. In the proposed design, the attacker mainly affects the WSN systems that is addressed by the DE-BiLSTM techniques to identify the attack.

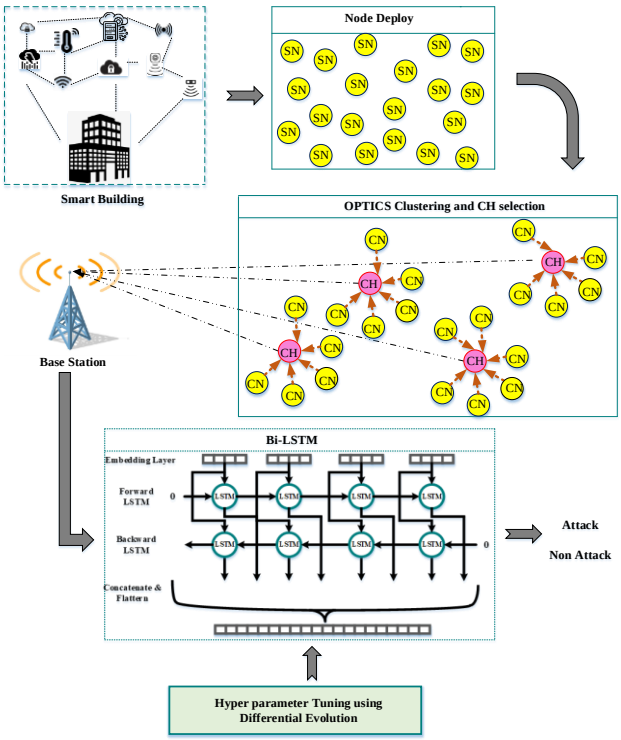

Fig. 1. Schematic representation of a proposed method.

Fig. 1 depicts a flow diagram representation of the proposed method. In this proposed model, Sensor Nodes (SN) are temperature sensor, humidity sensor, motion sensor, safety sensor, light sensor, and smoke sensors are considered to collect the data. Initially, the SN is deployed and cluster nodes (CN) are formed using the OPTICS clustering method based on density, which separates clusters with similar characteristics and extracts clusters with various densities and shapes. Then, the cluster head (CH) nodes are chosen from the CN based on higher energy and minimum distance and the data is transmitted to the base station. Afterwards, the data are classified using BiLSTM based on tuning the hyperparameters, namely the number of hidden layers and batch size are evaluated by means of differential evolution algorithm (DE) to predict as attack or non-attack in the WSN.

3.1. Node Deployment

Sensor nodes considered for this article are 100 nodes which have been distributed in a $1000\,m \times 1000\,m$ region in the network and the activities performed such as processing, information gathering, monitoring the environment, and dispatch.

3.2. OPTICS clustering and CH selection

Ordering Points to Identify Cluster Structure (OPTICS) technique provides data with an ordered list instead of grouping, with the closer-together data on this list being in the denser regions because it uses a density-based clustering method [19]. Simply expressed, the nodes in the aforementioned list that have higher densities and are more closely spaced out are arranged in order. In this structure, it is simple to retrieve data about cluster centres. The OPTICS approach received much of interest in the current investigation as a result of these benefits.

The OPTICS density-based clustering environment, a powerful algorithm capable of producing clusters with various dimensions, served as the model for the proposed method. The clusters are initially created using this procedure, and following this, the cluster-head nodes are chosen. The nodes which don't become a part of any clusters in the following stage join the cluster that is closest to their cluster head node. The sensors finally gather the necessary data and send it to the cluster head nodes.

On the basis of the Euclidean distance among the samples, the OPTICS approach determines whether a new cluster will develop when a specific number of samples come together. The Optics approach is actually a DBSCAN-based pre-process that orders the points according to how easily the core points can reach them. The ordered points are next subjected to the DBSCAN algorithm, after which the points are clustered. This technique defines 2 novel concepts: core distance and reachability distance.

Core-distance: "Core Distance" or "CD" refers to the shortest neighbourhood radius, which may be taken into account for a specific spot or point p, given which the spot p remains the core spot in the specified density.

Fig. 2 demonstrates the distance between the core point p and its fourth neighbour, designated as ε', is the smallest distance that may be taken into account under the present circumstances to maintain the point p status as the core point.

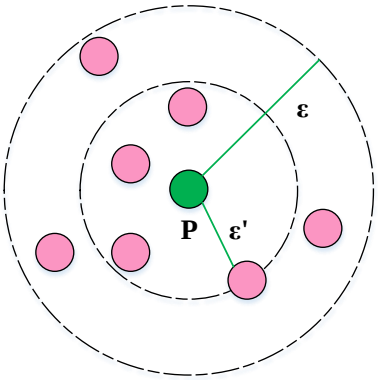

Fig. 2. Concept of core distance.

Reachability-distance: The point core distance is the similar as ε′When it is inside the neighbourhood radius of ε′. The distance between the point and the core is known as RD or the reachability distance of the core to that point when the point is exterior of the radius.

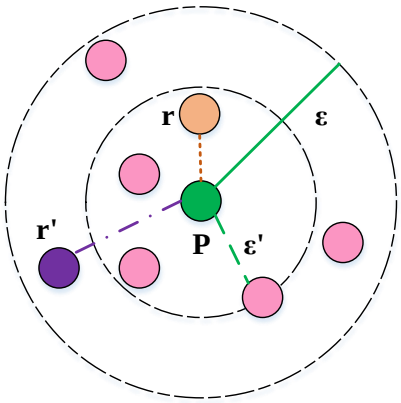

Fig. 3. Concept of Reachability-distance.

As shown in Fig. 3, since the separation between the core point p and the point r is within the neighbor radius ε', the p reaching to the point r is taken into consideration to be equal to ε', however if the point r' is exterior of this neighbourhood radius, the distance between the core point p and the point r' is taken into consideration to be RD.

The proposed approach uses two parameters are $Minpts$ and $Eps.Minpts$, which determines the lowest number of points required to create a cluster, and Eps, which determines the highest neighbour radius and increases the number of clusters required to start the process. Along with the development of the algorithm, a list of CD and RD is created. The mentioned algorithm starts off by ordering the points according to RD, or how reachable they are to one another, and then shows the results as a Dendrogram.

This method CH selection procedure is defined by the interaction of 2 factors: its separation from the base station as well as the node's remaining energy in the present round. In the same manner, a cost function that includes the two parameters discussed above is computed for each node following the clustering procedure, as can be seen in Equation (1).

$$cost(m) = \frac{1}{RemE(m)} + D\ to\ BS(m) \tag{1}$$

Each node's cost values are compared with one another, and the lowest value of the cluster node is chosen as the cluster head since it includes the maximum remaining energy and is closest to the base station. In Equation (1), $D\ to\ BS(m)$ indicates the distance between the base station and the node m and $RemE(m)$ denotes the node m remaining energy in the present round.

3.3. DE-Bi-LSTM

After clustering the nodes by using the OPTICS method and data transmission is performed in the base station. Then, the data from the base station is classified using Bi-LSTM to classify the attack and non-attack in WSN.

3.3.1. Bi-LSTM

LSTM is trustworthy when using a standard RNN approach, but it uses a variety of techniques to determine the hidden state, which fixes the RNN's problem but cannot handle dependence over a long distance [20]. An array of identical memory units utilizing three gates makes up the LSTM technology. The following list contains the specific estimate functions, where $\odot$ denotes the dot multiplication σ indicates a sigmoid function. A forget gate is defined by the f_t:

$$f_t = \sigma(W_f w_r + u_f H_{r-1} + B_f) \tag{2}$$

The input gate is referred to as i_t and stated as follows,

$$i_t = \sigma(W_i w_t + u_i H_{t-1} + B_i) \tag{3}$$

When $tanh$ is used, it refers to the tangent hyperbolic function and the $\tilde{m}_t$ stands for the candidate memory state of the cell at a recent time step;

$$\tilde{C} = \tanh(W_c w_t + u_i H_{t-1} + B_c) \tag{4}$$

The value of f_t and i_t is between (0, 1), while C_t corresponds to the state value. The estimation of $i_t \odot \tilde{C}_t$ implies that new data collected from the candidate unit $\widetilde{C}_t$, is recorded in C_t. When $f_t \odot C_{t-1}$ is calculated, it indicates that data has been preserved and kept in the current memories C_{t-1}.

$$C_t = i_r \odot \widehat{C_t} + f_t \odot C_{t-1} \tag{5}$$

The output gate is represented as O_t and defined as

$$O_t = \sigma(W_0 w_t + u_0 H_{t-1} + B_o) \tag{6}$$

At time t, H_t represents the hidden layer state referred as

$$H_t = O_t \odot \tanh(C_t) \tag{7}$$

Periodical sequence information makes up the LSTM, which is insufficient. When this data is accessible in a later stage, it might be highly useful for a series of actions. Layers of the backward and forward LSTM comprise the structure of bidirectional LSTM. The technique for this procedure is as follows: the forward layer of LSTM retains the sequence's previous data, whereas the backward layer holds the new sequence data. An identical output layer was connected to these layers. The use of the sequence content information to its greatest degree is another important aspect of this model.

Assume that the word embedding w_t is the input for time t at time $t-1$, and the forward hidden unit produces the result $\vec{H}_t$ and the backward hidden unit produces the simulation outcome $\overrightarrow{H}_{t+1}$, respectively.

$$\vec{H}_t = \text{▯}(w_t, \vec{H}_{t-1}, C_{t-1}) \tag{8}$$

$$\overleftarrow{H}_t = \text{▯}(w_t, \overleftarrow{H}_{t+1}, C_{t+1}) \tag{9}$$

Where the LSTM hidden layer task is represented by ▯(·). In order to obtain the text feature, the forward output vector is represented as $\vec{H}_t \in \mathcal{R}^{1\times H}$ and the backward output vector is denoted as $\overleftarrow{H}_t \epsilon \leftarrow \mathcal{R}^{1\times H}$ must be integrated. The entire set of cells in the hidden layer is obviously represented by H_t:

$$H_t = \vec{H}_t || \overleftarrow{H}_t \leftarrow \tag{10}$$

3.3.2. DE-based Bi-LSTM Model

Utilizing the DE model is the key objective to improve the hyper-parameters of the Bi-LSTM classifier and produce the best classification outcome. The parameters that are used in this case are number of hidden neurons and batch size. The DE approach starts with the main result that is generated at random and works to increase the classification model accuracy until it satisfies the termination requirement. The Bi-LSTM system's fitness function (FF) is in charge of estimating the attack categorization accuracy.

Storn developed the DE model for real parameter and real value function optimization [21]. This approach, which is a population-based method, is widely used for problems involving frequent searching. DE uses the crossover and mutation models in a manner similar to the Genetic Algorithm (GA) but regrettably without an explicit upgrade function. Initialization, Mutation, Crossover, and Selection make up the first four steps of the optimization problem in DE. The final three stages are carried out till the termination requirements are met.

Step 1: Initialization

NP initial parameter vector values are produced during the initialization phase. The bottom and upper boundaries of every parameter are fixed, and each vector contains G parameters. Each parameter is a number at random that falls inside the given range. $v_{p,q}(R)$ stands for the starting p^{th}vector of the q^{th} parameter in creation r. It is provided in the form of:

$$v_{p,q}(R) = rand_q(0,1).(R_u - R_l), \tag{11}$$

Where upper and bottom boundaries are denoted by R_u and R_l.

Step 2: Differential Mutation

Take into consideration the first population vector as the target vector. Three random vectors $(v_{d1,q}, v_{d2,q}, v_{d3,q})$ are selected with regard to this target vector. Then, every element of the distinct vector gets multiplied via the scaling factor S. The variance between the elements that correspond to each of the final two vectors is then calculated. The resulting vector transforms into the initial target vector's mutant vector. This procedure will continue till the final population estimate. As a result, the appropriate mutant vectors $m_{p,q}(R+1)$ are created for every target vector number. The formula for the equation that produced the mutant vector is as follows:

$$m_{p,q}(R+1) = v_{d1,q}(R) + S.(v_{d2,q}(R) - v_{d3,q}(R)) \tag{12}$$

DE/rand/1 indicates this differential mutation variant. The additional DE variations are DE/rand/2, DE/best/1, and DE/best/2, based on various mutation techniques. The mutation processes used in distinct DE variations are

$$m_{p,q}(R+1) = v_{d5,q}(R) + S.(v_{d1,q}(R) + v_{d2,q}(R) - v_{d3,q}(R) - v_{d4,q}(R)) \tag{13}$$

$$m_{p,q}(R+1) = v_{best,q}(R) + S.(v_{d2,q}(R) - v_{d3,q}(R)) \tag{14}$$

$$m_{p,q}(R+1) = v_{best,q}(R) + S.(v_{d1,q}(R) + v_{d2,q}(R) - v_{d3,q}(R) - v_{d4,q}(R)) \tag{15}$$

Where S represents a real constant in the range$[0, 2]$. The best population member, $v_{best,q}$, is amplified by this control parameter, which also increases the differential variation.

Step 3: Cross over

Based on the likelihood of the crossover ratio CR, the crossover process includes modifying the parameters for the mutant vector and the initially selected target vector. The trial vector, or resultant vector,$u_{p,q}(R+1)$, can be calculated as

$$u_{p,q}(R+1) = \begin{cases} m_{p,q}(R+1), & if\ rand(0,1) \le CR \quad or\ q = q_{rand} \\ v_{p,q}(R) & if\ rand(0,1) > CR \quad or\ q \ne q_{rand} \end{cases} \tag{16}$$

Where CR is a number ranging from zero to one that determines the likelihood that certain mutant vector parameters can be passed over to the trail vector. $rand(0,1)$ is used to produce a random number and .

A random number $rand(0,1)$ is generated, the parameter of the mutant vector is passed on to the trial vector when its value is lower or equal to CR. If not, the trial vector inherits the parameter of the target vector. For every pair of mutant and target vectors, this procedure is repeated.

Step 4: Selection

The resultant trial vector, $v_{p,q}(R)$, is evaluated using the cost function. The trial vector remains and substitutes the target vector in the following generation if its cost is lower than the target vector's; if not the target vector is kept for a further generation. The mathematical formula for the selection procedure is

$$v_{p,q}(R+1) = \begin{cases} u_{p,q}(R), & if\ f(u_{p,q}) \le f(v_{p,q}) \\ v_{p,q}(R), & otherwise \end{cases} \tag{17}$$

Where target and trial vector costs are denoted by $f(v_{p,q})$ and $f(u_{p,q})$, respectively. Repeat steps 2, 3, and 4 till the termination requirement is satisfied.

3.3.3. Softmax Classifier

For attack detection, it gives the resultant vector v_{gap} directly given to a softmax layer. The results of the analysis are as follows:

$$\hat{z} = softmax(wv_{gap} + B) \tag{18}$$

The cross-entropy primary objective is to determine a model which simulates the difference between the examined attack classes $\hat{z}$ and the actual classes z.

$$loss = -\sum_j z_j \log(\hat{z}) \tag{19}$$

Where j represents the value of the index. Thus, the proposed method is predicted as attack and non-attack using DE-based BiLSTM.

Algorithm 1: Pseudocode for attack prediction using OPTICS clustering with DE-BiLSTM techniques

```
# Node deployment
n = n1, n2, n3 .... nn
M= Node deployment (n)
# Clustering and CH selection
Ch=OPTICS(M)            // clustering the nodes and select CH by OPTICS method using equation (1)
datatx = Collected data from nodes
BS= datatx (Ch)         // Data transmission from CH to BS
#Classification
H = DE-BiLSTM (BS)      // Attack prediction based on the DE using BiLSTM algorithm using the equation (2)-(19)
   If (H=0)
   Print (attack)
   Else
  Print(non-attack)
end
```

Output: *prediction of attack in WSN*

4. Results And Discussions

A wireless sensor network using OPTICS clustering and a DE-BiLSTM classifier is used to implement the proposed method in order to identify attacks. The model that has been created is evaluated using the program called matlabR2021b, which is running using the following components: an Intel Core i7 CPU, NVIDIA GeForce RTX 3070 GPU, and an operating system with 64-bit support. Initial deployment of SN and these nodes are clustered using OPTICS clustering then from the cluster nodes CH nodes are selected. Following that data transmission is done in BS, an attack is predicted using the BiLSTM classifier with DE to optimally select the hyperparameter.

4.1. Dataset

Dataset [22] is used for training the DE-BiLSTM for predicting the attack and non-attack from the collected data. The parameters considered for training the classifier are Time, id, Is_CH, who CH, , ADV_S, Dist_To_CH, ADV_R, JOIN_R, JOIN_S, SCH_S, Rank, SCH_R, DATA_S, Data_Sent_To_BS, DATA_R, dist_CH_To_BS, Expaned Energy and send_code. Based on this parameters, the classifier is trained. In total 80% of the data are trained and 20% of the data are tested.

Table 2. Simulation parameters of the proposed method.

Parameter	Range
No of Nodes in WSN	100
Network Area	1km
Initial Energy (Mega Joule)	2.6 J
Packet size	512 bytes
No of Cluster	5
Transmission range	250m
Routing protocol	AODV

Table 3. Simulation parameters of Bi-LSTM

Parameter	Value
Optimizer	Adam
Loss Function	Cross Entropy
State Activation Function	Tanh
Gate Activation Function	Sigmoid
Initial Learning Rate	0.005
Dropout	0.5
Maximum Epochs	800
Maximum Batch size	100

Table 2 represents the simulation parameters of the proposed method. Simulation parameters are the number of nodes considered here as 100 nodes, network area, initial energy, size of the packet, number of clusters, and range of transmission are provided. Table 3 represents the simulation parameters for Bi-LSTM. The parameters for simulation are optimized by adam, function of loss, function of activation, dropout and by max.batch size.

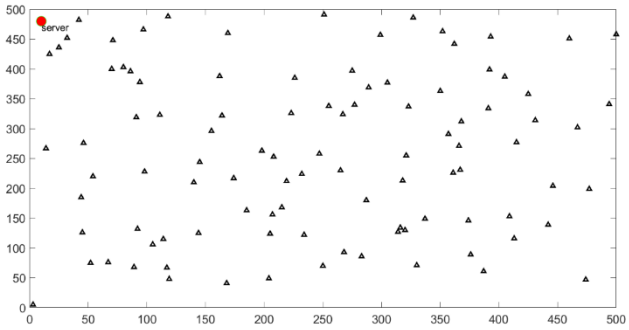

Fig. 4. Setup of a node in 1000 x 1000m.

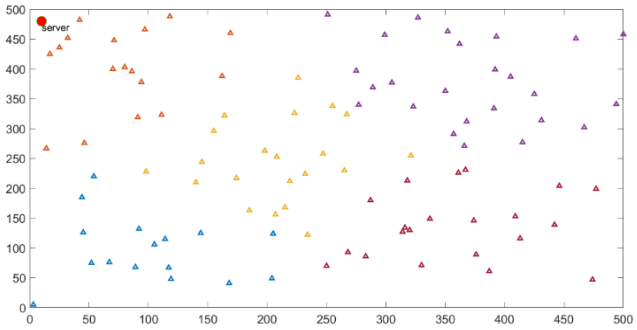

Fig. 5. Clustering nodes using the OPTICS algorithm.

Fig. 4 and Fig. 5 depict node deployment and clustering using the OPTICS method. For the purpose of assessing the data transmission, 100 nodes have been deployed over a 1000 x 1000 m area. The OPTICS algorithm is used to cluster the deployed nodes in the square region, which clusters the WSN nodes based on density. By using the OPTICS algorithm five clusters are formed from the deployed nodes.

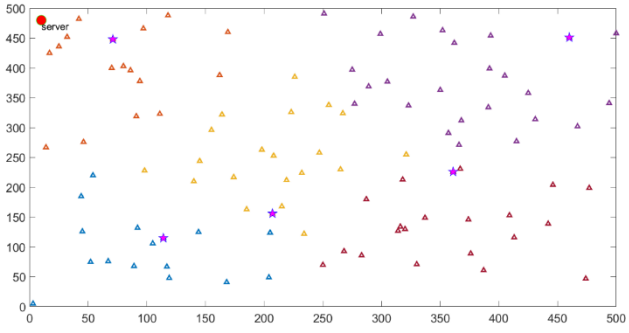

Fig. 6. Representation of Cluster head selection.

The graph in Fig. 6 shows the cluster head selection utilizing OPTICS and data transmission from BS to CH. Based on distance minimization and energy maximization, the cluster head is chosen. The data from the cluster head (CH) is transmitted to the BS for more energy-efficient data transfer.

4.2. Network performance metrics

The implementation of the proposed models are compared with existing models evaluates the network performance parameters that involve PDR, Average Residual energy (ARE), Throughput, Packet Loss Ratio (PLR), and delay. Performance metrics are evaluated for the proposed OPTICS method and existing models such as Power Efficient Gathering in Sensor Information Systems (PEGASIS), Low-energy adaptive clustering hierarchy (LEACH) and Threshold Sensitive Energy Efficient Network Protocol (TEEN).

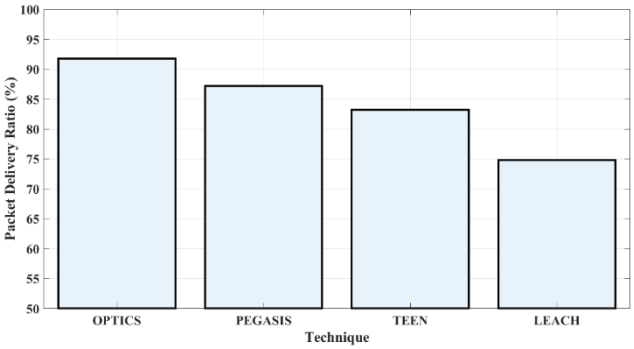

Fig. 7. Packet Delivery Ratio (PDR) analysis for both the proposed and existing model.

The analysis of the PDR for the proposed OPTICS algorithm and existing models are PEGASIS, TEEN and LEACH is presented in Fig. 7. The ratio of the number of packets received to all packets generated is measured by the Packet Delivery Ratio (PDR). Packet Loss Ratio and PDR are inversely related. A higher packet delivery ratio leads to lower packet loss. The value obtained for the proposed OPTICS method is 91.78% and the existing methods are PEGASIS, TEEN, and LEACH are attained values as 87.17%, 83.2% and 74.79%. Thus, the PDR value obtained for the proposed method is higher than the other methods. In regarding data transmission efficiency, the proposed OPTICS method performs better than other methods.

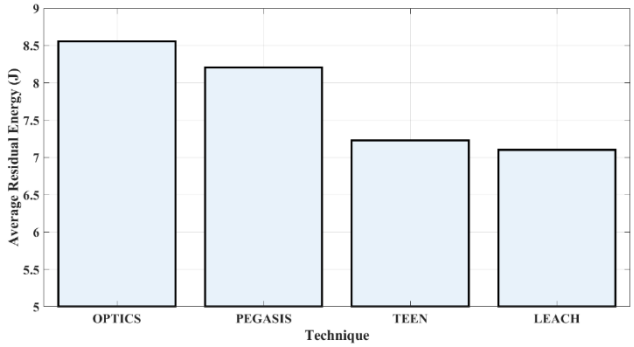

Fig. 8. Average Residual Energy (ARE) analysis for both the proposed and existing model.

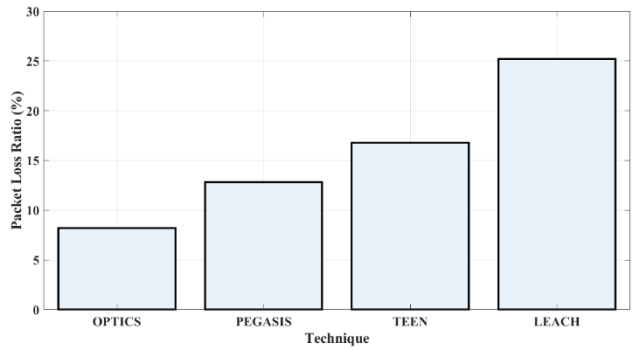

Fig. 9. Packet loss Ratio (PLR) analysis for both proposed and existing model.

Analyses of the Average Residual Energy (ARE) for the proposed and current models are portrayed in Fig. 8. The energy that was consumed while the sensor node was in each condition can be added up to determine the remaining energy of the node. The proposed OPTICS approach provided a value of 8.56 J, whereas the PEGASIS, TEEN, and LEACH methods results in values of 8.21 J, 7.23 J, and 7.1 J, respectively. The proposed method thus yielded a greater ARE value compared to the earlier methods. Fig. 9 demonstrates an examination of both the proposed and the existing model's ratio of packet loss (PLR). PLR, which measures how many data packets are dropped while transmission, can significantly affect performance. The result for the proposed OPTICS approach is 8.22%, while the values for the PEGASIS, TEEN, and LEACH methods are 12.83%, 16.8%, and 25.21%, respectively. Compared to the other models OPTICS has a lesser packet loss rate.

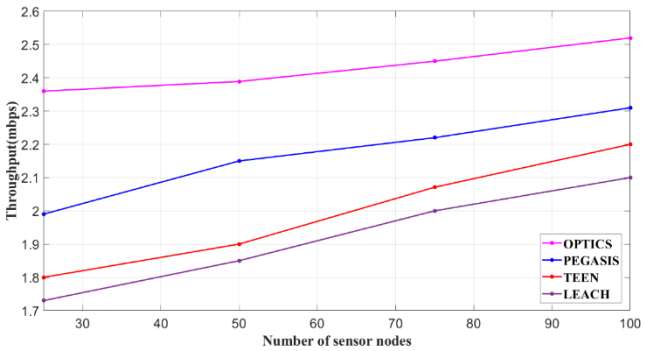

Fig. 10. Analysis of throughput for both proposed and existing model.

Fig. 10 depicts an analysis of throughput for proposed OPTICS and existing approaches such as PEGASIS, TEEN, and LEACH. Network throughput is the total amount of data that is successfully transmitted via a network in a specific amount of time, frequently expressed in megabits per second (mbps). Throughput measures the average rate at which bits of information effectively arrive at their final destinations. Optics has a throughput of 2.36 for 25 nodes, 2.38 for 50 nodes, 2.45 for 75 nodes, and 2.52 for 100 nodes. Similarly, PEGASIS achieves throughput values of 1.99, 2.15, 2.22, and 2.31 for 25, 50, 75, and 100 nodes, respectively. TEEN yields throughput results of 1.8 for 25 nodes, 1.9 for 50 nodes, 2.07 for 75 nodes, and 2.2 on 100 nodes. The LEACH throughput estimates for 25, 50, 75, and 100 nodes are 1.73, 1.85, and 2.1, respectively. The graph shows that OPTICS performs better than other approaches in terms of throughput.

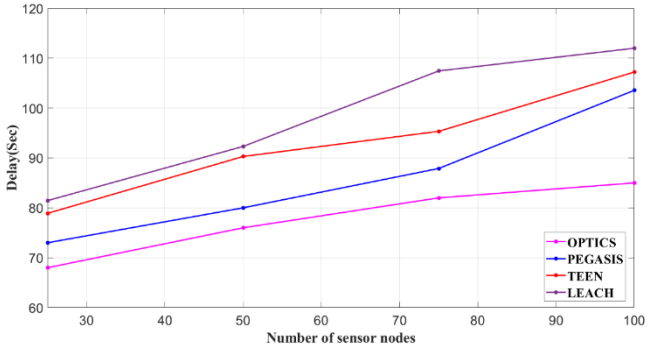

Fig. 11. Delay analysis for both proposed and existing model.

Fig. 11 illustrates a delay comparison between the proposed OPTICS approach and existing models like PEGASIS, TEEN, and LEACH. The network's efficiency is characterized by the delay, which is expressed in milliseconds. Delay measurements used to gauge network efficiency also show that delays are important in the system despite the fact that the sources of delay have a considerable impact on system performance. In a network for communication, the amount of time it takes a packet to get from its source to its endpoint is known as delay. The results of the OPTICS method for determining delay are 68ms for 25 nodes, 76ms for 50 nodes, 82ms for 75 nodes, and 85ms for 100 nodes. When using PEGASIS, the delay is 73ms for 25 nodes, 80ms for 50 nodes, 87.89ms for 75 nodes, and 103.56 milliseconds for 100 nodes. If TEEN is applied, there exists a delay of 78.9ms for 25 nodes, 90.3ms for 50 nodes, 95.34ms for 75 nodes, and 107.23ms for 100 nodes. The delay for LEACH is 81.45ms at 25 nodes, 92.31ms at 50 nodes, 107.45ms at 75 nodes, and 112ms at 100 nodes. The graphical representation makes it evident that the average delay measured in OPTICS is lower than that of the other existing approaches.

4.3. Classification performance metrics

The proposed model's confusion metrics are depicted in Fig. 12. It is used to assess the ability of a classification model to distinguish between correct and wrong predictions. A good approach provides high TP and TN rates while having low FP and FN rates. The target variable's true labels as positive and negative are represented in the rows and columns, respectively. Data for classes 0 and 1 are predicted to be 2024 and 2208, respectively. The total amount of data applied to testing is 4509, of which 4232 are predictions made in accordance with the actual class and the remaining 277 are predictions that are wrong.

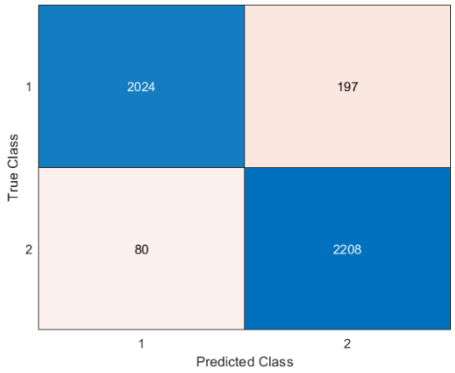

Fig. 12. Measures of confusion for the proposed technique.

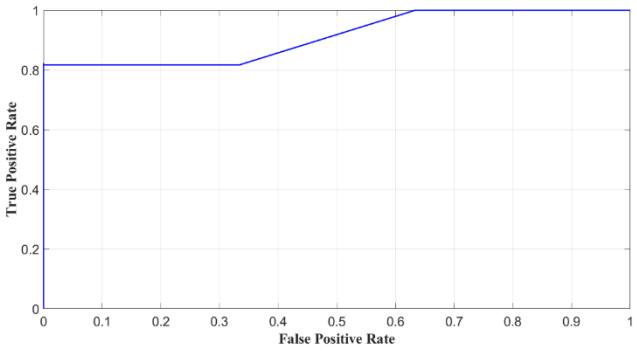

Fig. 13. ROC plot for the proposed method.

Fig. 13 depicts the ROC curve of the proposed model. A graph that displays the effectiveness of a classification model is called a ROC curve or receiver operating characteristic curve. When a predictive model is used with various likelihood thresholds, ROC Curves describe the relation between the rate of true positives and false positive.

Table 4. Analysis of variance (ANOVA) test table for feature set

Source	Features	SS	df	MS	F	Pro> F
Columns	10	0.8695	1	0.8695	3.4776	0.0622
	20	4.2549	1	4.2549	17.0060	0.0037
	30	0.1215	1	0.1215	0.4857	0.4859
	40	0.1419	1	0.1419	0.5674	0.4513

ANOVA table 4 shows the source as between-groups variation (Columns) for different features as 10, 20, 30 and 40. SS represents the sum of squares, df represents the degrees of freedom, and MS as the mean squared error, which is SS/df for each source of variation. F represents the F-statistic the ratio of the mean squared errors. Pro> F represent the p-value probability that the test statistic can take a value greater than the value of the computed test statistic. The small p-value of 0.0037 indicates that differences between column means are significant. Performance indicators involving accuracy, F1-score, precision, specificity, FPR, error, FNR and sensitivity are all evaluated between the proposed method and existing methods.

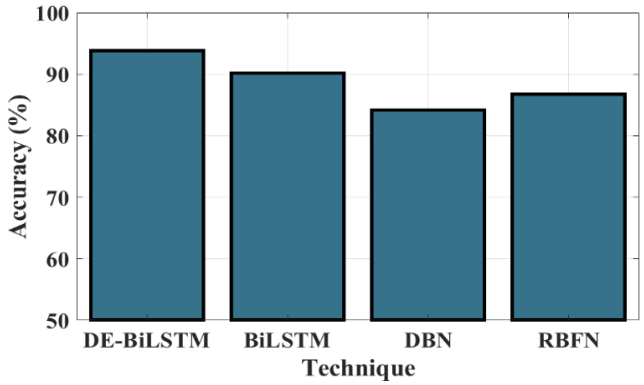

Fig. 14. Accuracy analysis for both proposed and existing model.

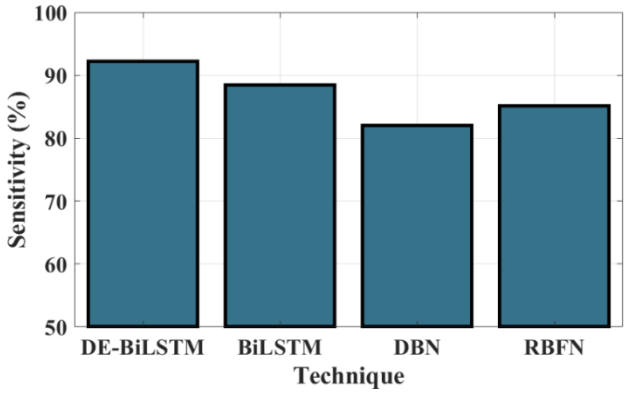

Fig. 15. Sensitivity estimation for both proposed and existing model.

The proposed DE-BiLSTM strategy's accuracy is measured and contrasted to that of existing methods in the graph presented in Fig. 14. Accuracy is the proportion of accurately categorized subjects in the evaluation dataset. The proposed approach DE-BiLSTM is evaluated for accuracy by comparing it to the existing techniques BiLSTM, DBN, and RBFN. For DE-BiLSTM, BiLSTM, DBN, and RBFN, the equivalent values are 93.78%, 90.17%, 84.2%, and 86.79%. In comparison to past methods, the proposed model is more accurate and effective at detecting attack in wireless sensor networks. The graph in Fig. 15 displays the results and compares the sensitivity of the proposed DE-BiLSTM strategy to that of existing techniques. Comparable results for DE-BiLSTM, BiLSTM, DBN, and RBFN are 92.23%, 88.45%, 82%, and 85.12%. As a result, the sensitivity value of the proposed approach is higher than that of the previously employed methods.

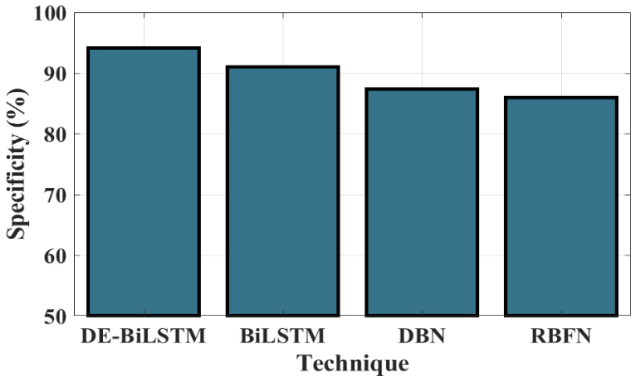

Fig. 16. Specificity estimation for both proposed and existing model.

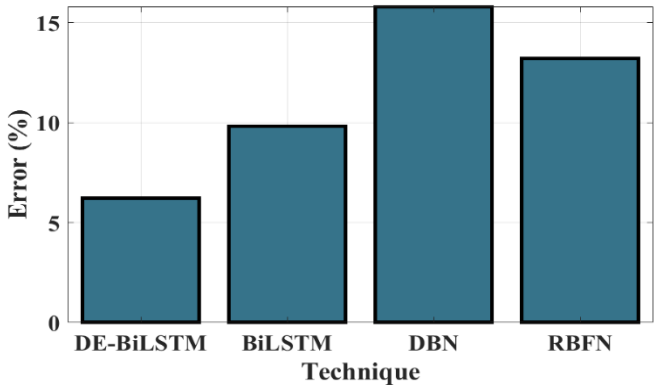

Fig. 17. Error estimation for both proposed and existing model.

The graph in Fig. 16 demonstrates how the DE-BiLSTM strategy's specificity compares against the performance of existing methods. To assess specificity, it is necessary to contrast the proposed DE-BiLSTM model with the existing techniques such as BiLSTM, DBN, and RBFN. The final results for DE-BiLSTM, BiLSTM, DBN, and RBFN are 94.12%, 91%, 87.34%, and 86%, respectively. As a result, the specificity value of the proposed technique increases in comparison to each of the existing approaches. The results of the proposed DE-BiLSTM strategy's error and that of existing strategies are evaluated in the graph illustrated in Fig. 17. The results are 6.22%, 9.83%, 15.8%, 13.21% for DE-BiLSTM, BiLSTM, DBN, and RBFN. The proposed framework shows lesser inaccuracy than earlier approaches.

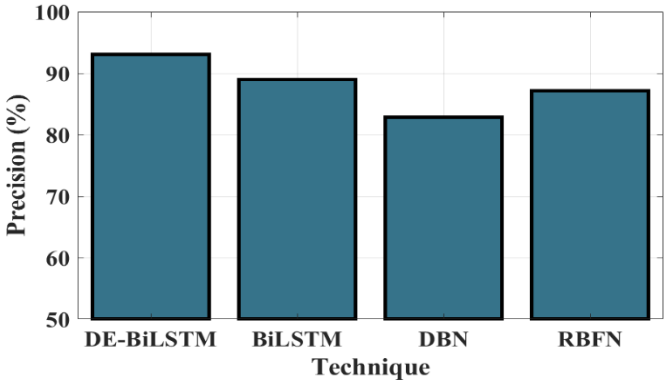

Fig. 18. Precision analysis for both proposed and existing model.

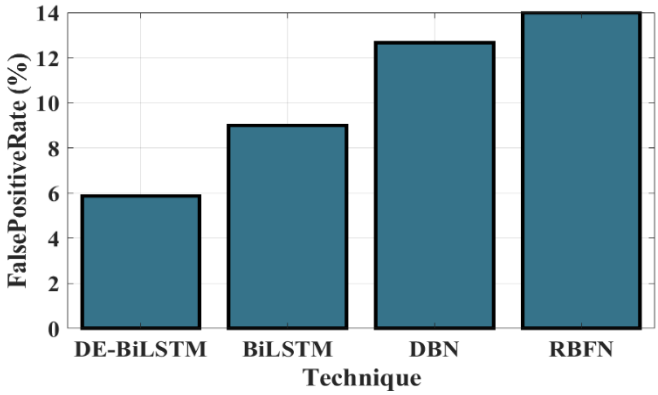

Fig. 19. FPR analysis for both proposed and existing model.

Fig. 18 describes the examination of the proposed DE-BiLSTM precision to various existing techniques. The proposed DE-BiLSTM is contrasted with BiLSTM, DBN, and RBFN. Precision values of 1.02%, 3.85%, 7.54%, and 11.48% are obtained by DE-BiLSTM, BiLSTM, DBN, and RBFN, respectively. The proposed method's precision value is thus higher than that of existing approaches. In the graph depicted in Fig. 19, the FPR of the proposed DE-BiLSTM strategy is contrasted with that of the existing approach. DE-BiLSTM, BiLSTM, DBN, and RBFN have FPR values of 5.26%, 8.5%, 12.84%, and 17.48%, respectively. Compared to more traditional approaches, the proposed model exhibits lower FPR values.

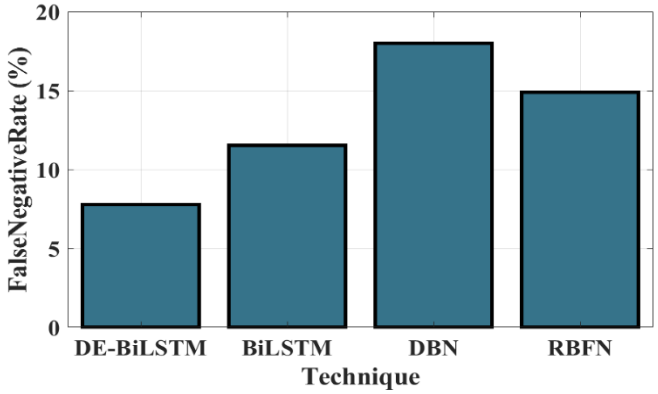

Fig. 20. FNR analysis for both proposed and existing model.

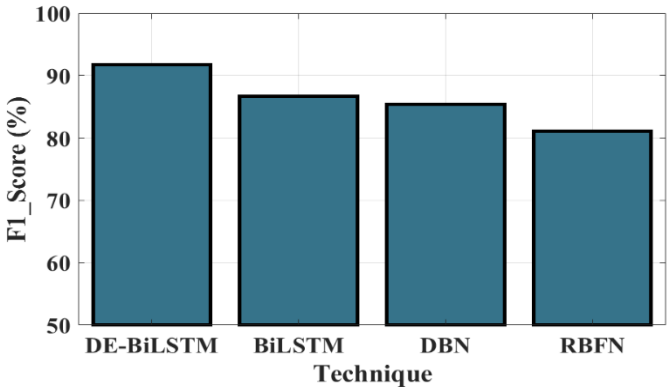

Fig. 21. F1-Score analysis for both proposed and existing model.

In regard to FNR values, Fig. 20 analyses the proposed DE-BiLSTM method with the already used techniques. The respective FNR rates for DE-BiLSTM, BiLSTM, DBN, and RBFN are 7.77%, 11.55%, 18%, and 14.88%. This result indicates the proposed method's FNR value is lower as compared with the existing model. Fig. 21 compares the proposed DE-BiLSTM method to existing approaches based on F1_score. The proposed approach is compared to already-used approaches like BiLSTM, DBN, and RBFN. F1_score of 91.78%, 86.7%, 85.4%, and 81.1%, respectively, are obtained using the proposed DE-BiLSTM approach and the previously employed methods. With regard to existing approaches, the F1_score rate is higher score in the proposed method.

Table 5. Comparative analysis of the proposed and exiting methodologies

Algorithms	Performance metrics			
	Accuracy	Precision	Recall	F1_Score
LSVM-CSGO	86	84	86	89
SLGBM	91.7	90.3	88.7	90.5
SVM	85	82	84	87
GXGBoost	90	91	87	89
Proposed	93.78	93.04	92.23	91.78

Table 5 illustrates the comparative analysis of the proposed and existing methodologies. The existing models such as LSVM-CSGO [13], SLGBM [14], SVM [15] and GXGBoost [16]. These existing models are designed in the proposed model scenario and the performance metrics results that the proposed algorithm attains better performance metrics values than the existing models.

5. Conclusion

A wireless sensor network is a network composed of numerous sensor nodes that gather and send data to a centralized location. It faces a number of security issues caused by resource-constrained nodes, deployment approaches, and communication routes. Therefore, detecting illegal access is crucial to enhancing the security aspects of wireless sensor networks. Such services are provided by network security systems, which are a must for every form of communication network. To address these problems, a proposed method to identify threats in wireless sensor networks for smart buildings is based on the OPTICS clustering with DE-BiLSTM) classifier. Initially, the SN are deployed, and CN nodes are generated using the OPTICS density-based clustering algorithm that divides clusters with similar properties

and densities. Following that, the CH nodes are chosen within the CN on the basis of the greater energy and less distance, and they send data to the base station. After that, the batch size and hidden layers are selected by DE in order to classify the data using BiLSTM to predict attack and non-attack. Performance to predict attack is assessed by network parameters are PDR, ARE, PLR, throughput, delay with proposed OPTICS and existing techniques like OPTICS, PEGASIS, TEEN, and LEACH and classification parameters such as accuracy, specificity, F1_score, error, FNR, sensitivity, FPR, and precision are contrasted with proposed DE-BiLSTM and existing methods namely BiLSTM, DBN, and RBFN. The outcomes achieved for network performance attained value of PDR are 91.78%, 8.56J for ARE, 8.22% for PLR, throughput and delay for 100 nodes are 2.52mbps and 85sec and the classification parameters evaluated values are 93.78% for accuracy, 94.12% for specificity, 91.78% for F1_score, 6.22% for error, 7.77% for FNR, 92.23% for sensitivity, 5.88% for FPR, and 93.04% for precision of the proposed method. It demonstrates that the proposed method is superior to the existing techniques. Thus, this designed model based on threat detection in the WSN can influence the WSN security protocols, which can protect the data by detecting the attacked nodes and the security protocol or authentication scheme for such nodes can be changed or modified. In future, the proposed model based on IoT enabled devices for routing and threat detection will be essential for the development of a secure smart city, smart farming, smart homes and smart grids, which can detect threats automatically and secure the device and user data from an intruder without being manipulated. Also, the device data can be transferred or routed energy efficiently from various IoT devices to the cloud.

Acknowledgement

Funding: The authors declare that no funds, grants, or other support were received during the preparation of this manuscript.

Conflict of Interest: The authors declared that they have no conflicts of interest to this work. We declare that we do not have any commercial or associative interest that represents a conflict of interest in connection with the work submitted.

Availability of data and material: Not applicable

Code availability: Not applicable

Author contributions: The corresponding author claims the major contribution of the paper including formulation, analysis and editing, to verify the analysis result and manuscript editing.

Compliance with ethical standards: This article is a completely original work of its authors; it has not been published before and will not be sent to other publications until the journal's editorial board decides not to accept it for publication.

References

[1] S. Karthick, "TDP: A Novel Secure and Energy Aware Routing Protocol for Wireless Sensor Networks," International Journal of Intelligent Engineering and Systems. vol. 11, no. 2, pp. 76-84, 2018.

[2] R. Preethi and R. Anbuselvi, "Energy Efficient Cluster Based Technique for Resource Allocation in MANET," 2017.

[3] R. Preethi and R. Anbuselvi, "A Game Theory Approach for Efficient Node Clustering Based on Unstable Ranging Transmission in MANET," International Journal of Engineering and Technology(UAE). 7. 21-26. 10.14419/ijet.v7i2.22.11802, 2018.

[4] I. G. A. Poornima and B. Paramasivan, "Anomaly detection in wireless sensor network using machine learning algorithm," Computer communications. vol. 151, pp. 331-337, 2020.

[5] L. Han, M. Zhou, W. Jia, Z. Dalil and X. Xu, "Intrusion detection model of wireless sensor networks based on game theory and an autoregressive model," Information sciences. vol. 476, pp. 491-504, 2019.

[6] W. Zhang, D. Han, K. C. Li and F. I. Massetto, "Wireless sensor network intrusion detection system based on MK-ELM," Soft Computing, vol. 24, pp. 12361-12374, 2020.

[7] A. Singh, V. Kotiyal, S. Sharma, J. Nagar and C. C. Lee, "A machine learning approach to predict the average localization error with applications to wireless sensor networks," IEEE Access. vol. 8, pp. 208253-208263, 2020.

[8] X. Tan, S. Su, Z. Huang, X. Guo, Z. Zuo, X. Sun and L. Li, "Wireless sensor networks intrusion detection based on SMOTE and the random forest algorithm," Sensors. vol. 19, no. 1, pp. 203, 2019.

[9] R. H. Dong, H. H. Yan and Q. Y. Zhang, "An Intrusion Detection Model for Wireless Sensor Network Based on Information Gain Ratio and Bagging Algorithm," Int. J. Netw. Secur., vol. 22, no. 2, pp. 218-230, 2020.

[10] P. Biswas and T. Samanta, "Anomaly detection using ensemble random forest in wireless sensor network," International Journal of Information Technology. vol. 13, no. 5, pp. 2043-2052, 2021.

[11] S. Ifzarne, H. Tabbaa, I. Hafidi and N. Lamghari, "Anomaly detection using machine learning techniques in wireless sensor networks," In Journal of Physics: Conference Series. vol. 1743, no. 1, 2021, pp. 012021. IOP Publishing.

[12] M. Alqahtani, A. Gumaei, H. Mathkour and M. Maher Ben Ismail, "A genetic-based extreme gradient boosting model for detecting intrusions in wireless sensor networks," Sensors. vol. 19, no. 20, pp. 4383, 2019.

[13] D. Hemanand, G. V. Reddy, S. S. Babu, K. R. Balmuri, T. Chitra and S. Gopalakrishnan, "An intelligent intrusion detection and classification system using CSGO-LSVM model for wireless sensor networks (WSNs)," International Journal of Intelligent Systems and Applications in Engineering. vol. 10, no. 3, pp. 285-293, 2022.

[14] Jiang, S., Zhao, J., & Xu, X. (2020). SLGBM: An intrusion detection mechanism for wireless sensor networks in smart environments. IEEE Access, 8, 169548-169558.

[15] Dao, T. K., Nguyen, T. T., Pan, J. S., Qiao, Y., & Lai, Q. A. (2020). Identification failure data for cluster heads aggregation in WSN based on improving classification of SVM. IEEE Access, 8, 61070-61084.
[16] Alqahtani, M., Gumaei, A., Mathkour, H., & Maher Ben Ismail, M. (2019). A genetic-based extreme gradient boosting model for detecting intrusions in wireless sensor networks. Sensors, 19(20), 4383.
[17] Subramani, N., Mohan, P., Alotaibi, Y., Alghamdi, S., & Khalaf, O. I. (2022). An efficient metaheuristic-based clustering with routing protocol for underwater wireless sensor networks. Sensors, 22(2), 415.
[18] Wang, J., Gao, Y., Liu, W., Sangaiah, A. K., & Kim, H. J. (2019). An improved routing schema with special clustering using PSO algorithm for heterogeneous wireless sensor network. Sensors, 19(3), 671.
[19] X. Zhu, S. Nie, C. Wang, X. Xi, J. Wang, D. Li and H. Zhou, "A noise removal algorithm based on OPTICS for photon-counting LiDAR data," IEEE Geoscience and Remote Sensing Letters. vol. 18, no. 8, pp. 1471-1475, 2020.
[20] H. Crisóstomo de Castro Filho, O. Abílio de Carvalho Júnior, O. L. Ferreira de Carvalho, P. Pozzobon de R. Bem, "dos Santos de Moura, A. Olino de Albuquerque and R. A. Trancoso Gomes, Rice crop detection using LSTM, Bi-LSTM, and machine learning models from Sentinel-1 time series," Remote Sensing. vol. 12, no. 16, pp. 2655, 2020.
[21] Z. Zhang, S. Ding and W. Jia, "A hybrid optimization algorithm based on cuckoo search and differential evolution for solving constrained engineering problems," Engineering Applications of Artificial Intelligence. vol. 85, pp. 254-268, 2019.
[22] Kiran, (2019) Kaggle: [https://www.kaggle.com/datasets/kiranmahesh/nslkdd?select=kdd].Acessed on 16-10-2023.

Authors' Profiles

Dr. R. Preethi, I'm working as Assistant Professor in the Department of Computer Science in Bishop Heber College, Affiliated to Bharathidasan University, Tiruchirappalli since 2020, and love to explore and research the processes and scientific mechanisms underlying the world we live in. Passion for Communication and Networking which paved the way to proceed with my research in Ad Hoc Networks. I have presented 15 papers and published 8 papers in reputed journals and attended 23 seminars, conferences, workshops, training programs, 1 book edited as chief editor and 1 book chapter, 1invited talk, guests for various occasions, and published 1 Patent. Organized 32 various Seminars, Workshops, Training Programs, 2 International Conferences.

How to cite this paper: R. Preethi, "Detection of Threats in Wireless Sensor Network Based on Optics Clustering With DE-BiLSTM Classifier", International Journal of Wireless and Microwave Technologies(IJWMT), Vol.14, No.3, pp. 14-30, 2024. DOI:10.5815/ijwmt.2024.03.02