DOI 10.15622/ia.22.6.2

K. KRISHNA, R. THIRUMURU

A BALANCED INTRUSION DETECTION SYSTEM FOR WIRELESS SENSOR NETWORKS IN A BIG DATA ENVIRONMENT USING CNN-SVM MODEL

Krishna K., Thirumuru R. **A Balanced Intrusion Detection System for Wireless Sensor Networks in a Big Data Environment Using CNN-SVM Model.**

Abstract. Wireless Sensor Networks (WSNs) were exposed to several distinct safety issues and attacks regarding gathering and sending data. In this scenario, one of the most prevalent WSN assaults that may target any tier of the protocol stack is the Denial of Service (DoS) attack. The current research suggested various strategies to find the attack in the network. However, it has classification challenges. An effective ensemble deep learning-based intrusion detection system to identify the assault in the WSN network was, therefore, suggested in this research to address this issue. The data pre-processing involves converting qualitative data into numeric data using the One-Hot Encoding technique. Following that, Normalization Process was carried out. Then Manta-Ray Foraging Optimization is suggested to choose the best subset of features. Then Synthetic Minority Oversampling Technique (SMOTE) oversampling creates a new minority sample to balance the processed dataset. Finally, CNN–SVM classifier is proposed to classify the attack kinds. The Accuracy, F-Measure, Precision, and Recall metrics were used to assess the outcomes of 99.75%, 99.21%, 100%, and 99.6%, respectively. Compared to existing approaches, the proposed method has shown to be extremely effective in detecting DoS attacks in WSNs.

Keywords: WSN, DoS attacks, artificial intelligence, deep learning, Convolutional Neural Networks (CNN), Support Vector Machine (SVM).

1. Introduction. Networking systems are used by people worldwide to solve problems, find fresh perspectives, and fulfil fundamental needs. One of the most recent and well-liked technical developments is the creation of sensors, which permit users to accept data from a long distance. Devices of the Internet of Things (IoT) are using sensors more frequently these days [1]. Wireless Sensor Networks are regarded as an important research topic. Many uses, including telecommunications, the military, healthcare, research, and agriculture, can benefit from the technology [2] and see natural catastrophes like earthquakes, floods, and volcanoes [3]. All of the nodes involved in the design of these applications are presumed to be reliable and cooperative. However, in actual deployments, nodes are subject to various intrusions and assaults that can substantially impact the system's performance and the network's capacity to function. Sadly, protecting this form of network from numerous harmful assaults is a challenging issue [4]. A DoS assault is the most frequent danger that WSNs encounter [5]. DoS attacks may harm a network service by blocking real traffic from connecting to the network by saturating it with many fraudulent requests. DoS assaults may occur at any

tier of the TCP/IP protocol stack [6, 7]. The WSN's security must be verified using the Intrusion Detection System (IDS).

IDSs keep an eye on system activity to spot and stop malicious traffic. Attacks may be immediately detected by choosing the usual network traffic design and size. IDS monitors and examines network-generated events to spot anything unusual and notify sensor nodes of an intrusion [8]. Data received from sensors may be used in real time by intrusion detection/prevention systems and cyber threat analysts to identify useful information. With the aid of this data, security solutions can be created to identify weaknesses and assaults.

One of the most popular ways to help IDSs better identify and distinguish intruders is to combine them with artificial intelligence (AI) approaches [9]. To detect DoS in WSN, the AI-based classification algorithm has been applied. This study used deep learning techniques to assess the defences against DoS attacks in WSN. To verify the efficiency of these defence methods in identifying and categorizing the many types of DoS assaults, they were assessed on a unique Wireless Sensor Network DataSet called WSN-DS, which included several normal and attack circumstances. The primary goal of attack detection, classified as a classification challenge, is determining if an assault is Flooding, Black Hole, Normal, Time Division Multiple Access (TDMA), or Gray Hole. Thus, this study suggested a unique deep-learning architecture to address the issues of attack detection. This study's primary contribution is as follows.

Pre-processing the data represents the initial stage of the suggested method. Here, the One-Hot Encoding approach changed the categorical data into numeral information. Following that, the normalization process is carried out to determine the normalized value using average and standard deviation values.

– Then, to select the best subset of aspects, this research proposed a Manta Ray Foraging Optimization, which makes every possible feature combination and returns the set of attributes that performs the best. Moreover, to balance the dataset, SMOTE oversampling approach is employed, which increases the classification performance.

– Finally, this research proposed a CNN-SVM-based deep learning approach to classify the attacks. Therefore, the suggested multi-class intrusion detection technique can properly categorize particular class assaults.

– This article is organized as follows: Section 2 reviews the existing artificial intelligence-based malicious threat detection approaches in WSN. Section 3 describes the proposed deep learning-based attack detection framework. Next, model implementation and their consecutive outcomes were discussed in Section 4. Finally, Section 5 concludes this research paper.

2. Literature survey. Articles on wireless sensor networks has been published more often in recent years WSN. Some efficient anomaly detection techniques, such as those based on the composition and properties of WSNs, classification methods, clustering algorithms, machine learning algorithms, and statistical learning models, are suggested in the literature. Despite the benefits of WSN, several security flaws make it vulnerable to DoS assaults. Many technological developments have made identifying and defending against DoS attacks simpler. However, the most efficient strategies for averting such safety problems and DoS attacks were supplied via deep learning.

Using edge and cloud data analysis, the authors in [10] presented a new method for automatic anomaly detection for heterogeneous sensor networks. The latter employs an artificial neural network method that is completely unsupervised, whilst the former uses a multi-parameterized edit distance algorithm created inside a residential structure and subsequently altered with various fictitious impairments. The findings demonstrate that the suggested technique can appropriately recognize abnormalities and self-adapt to environmental fluctuations.

In [11] the authors developed a highly scalable architecture for a hybrid intrusion detection warning system to create an adaptable and efficient IDS that can categorize unexpected and surprising cyberattacks. This system learns how to represent abstract and high-dimensional features by passing the IDS data through several hidden layers. However, the suggested system does not provide comprehensive details on the makeup and traits of the malware.

A new model, the genetic algorithm and an extreme gradient boosting (GXGBoost) model, was suggested by the authors in [12], using a XGBoost classifier to identify intrusion assaults. In the extremely inconsistent data flow of WSNs, this model seeks to develop the capability of conventional approaches to identifying certain threats. To lessen the number of aspects and increase the efficacy of incursion detection in WSN, this model does not concentrate on feature selection approaches.

To maintain the dataset's balance and develop the intrusion detection classifier, in study [13] developed a way using the SMOTE. However, this technique must further enhance its ability to recognize wireless sensor network intrusion data.

In paper [14] presented the backward sequence selection with Light Gradient Boosting Machine (SLGBM) approach to detect intrusions in wireless sensor networks. The Sequence Backward Selection (SBS) technique was initially used to minimize the feature space's data dimension to lower the real traffic data's computing cost. To identify various network

assaults, a LightGBM algorithm is then used. But this approach does not employ a distributed mechanism to further cut down on time.

For intrusion detection, in [15] SVM had the maximum accuracy compared to other algorithms, whereas Gradient Boosting Classifier had the minimum accuracy rating. It was discovered after examining the efficacy of several machine-learning techniques. However, this study is still not focused on multi-class classification and considering features to choose the most crucial aspects and improve accuracy.

In paper [16] the LightGBM method and autoencoders were suggested for constructing a network intrusion detection system, with the latter being used for feature selection and the former for training and detection. A foundation for intrusion detection is provided when data collection, including network intrusion behaviours, is fed into an autoencoder because there is a significant difference in restructural error between the input data and the output data produced by the autoencoder. An adequate threshold is determined based on the reconstruction error to discriminate between defensive and offensive actions symmetrically. However, this strategy was limited to binary class classification.

A unique feature selection technique called dynamic recursive feature selection algorithm (DRFSA), which determines how many characteristics are appropriate for categorization, was suggested by the authors in their paper published in Science [17]. Employing fuzzy temporal constraints as a smart addition to the decision tree algorithm is also suggested to categorize network users and traffic better accurately. Convolution neural networks are also used for data classification in huge volumes. No distributed environment was used to implement this work.

In paper [18] the authors suggested a network intrusion detection system based on LightGBM and adaptive synthetic (ADASYN) oversampling technology. The authors initially used information pre-processing to normalize and one-hot encode the original data to lessen the impact of the top or lowest ranking for all requirements. Next, the ADASYN oversampling method was used to deal with the poor rate of minority assault detection brought on by an imbalance in the training data. Finally, the system's temporal complexity is reduced while preserving detection accuracy by using the LightGBM ensemble learning model. A big data environment is not considered, though.

In study [19] the authors suggested a classification-based network attack detection method for massive amounts of data. The SMOTE + Tomek-Link – Hybrid Deep Learning (STL-HDL) method combines a hybrid deep learning network here with data balancing. The STL ensemble method was employed in the research for data balancing. The Convolutional

Neural Network-Long Short Term Memory (CNN-LSTM) is utilized for classification. However, a WSN-DS real-time dataset has not been utilized to evaluate how well the proposed strategy worked.

In study [20] the authors suggested a passive-aggressive online classifier and an information gain ratio-based model. First, the important parts of the sensor data are chosen using the information gain ratio. Second, several forms of Denial of Service assaults have been identified and categorized using the online Passive Aggressive algorithm. However, the data balance has not been taken into account.

Study [21] suggested a light-weight Intelligent Intrusion Detection Model for WSN. Utilizing the k-nearest neighbour algorithm (kNN) and the sine-cosine algorithm (SCA) allows for the intelligent identification of various threats, including unknown assaults, while significantly improving classification accuracy and reducing false alarm rates. The polymorphic mutation (PM) technique is employed to compensate for the decrease in optimization precision, and compact machine SCA (CSCA) is employed to reduce computation time and space. On the other hand, the big data environment is not considered.

In [22] the authors developed a memory-efficient, light-weight anomaly detection system that maintains high accuracy while reducing computing complexity. Concepts of dimension reduction and one-class learning were used in the study. The One-Class Support Vector Machine (OCSVM) was employed for one-class learning and size reduction, and the Candid Covariance-Free Incremental Principal Component Analysis (CCIPCA) approach was employed. The normal reference model must be adaptively learned over time to guarantee that sensor data is accurate because it gets rigid.

In a WSN, in study [23] the authors suggested a technique used for monitoring and thwarting DoS and distributed DoS assaults. The suggested approach uses an RNN as a classifier. However, the dispersed environment is not taken into account in this work.

Paper [24] provided a unique perspective on the malicious security threats in WSNs using an ensemble learning-based quick intrusion detection system that can handle continuous and dynamic data streaming. Although it generates superior results, the pre-processing with data reduction and parameter adjustment is not carried out, which might increase the classifier's effectiveness.

In [25] the authors developed a CNN-LSTM network to identify and categorize DoS intrusion threats. To enhance the accuracy of the classification findings, this strategy does not balance the dataset and yields worse performance outcomes.

To identify DoS assaults targeted specifically at WSNs, the authors in [26] suggested a special DoS Intrusion Detection System (DDS). STLGBM-DDS, the proposed system, integrates the LightGBM machine learning algorithm, data balance and feature selection techniques to create an ensemble intrusion detection system. It was created using the big data platform Apache Spark. The data imbalance's impact on system performance was reduced by treating the data imbalance utilizing the SMOTE and the STL sampling methods. Information Gain Ratio is also employed as a feature selection strategy during the data preparation. However, using incorrect neural network parameters has impacted classification accuracy.

Considering the literature review cited earlier, almost all studies that advocate for deploying intrusion detection systems acknowledge that the data imbalance problem is still unresolved. Furthermore, most research has overlooked feature selection. To illustrate the need for large data environments as the amount of WSNs data increases daily, not all required work is completed in one.

3. Proposed Method. WSN deployments have increased dramatically during the past several years. WSNs are being used in various applications across numerous sectors because of their tiny size and low cost. WSNs are exposed to several distinct security concerns and attacks regarding data collection and transmission. The DoS attack is among the most widespread WSN assaults that can target any protocol stack layer. Attack detection is a crucial responsibility in this case for protecting the network and the data. To provide security, this research proposed a deep learning framework. This framework consists of the following steps: data pre-processing, data normalization, feature selection, data balancing, and attack classification shown in Figure 1.

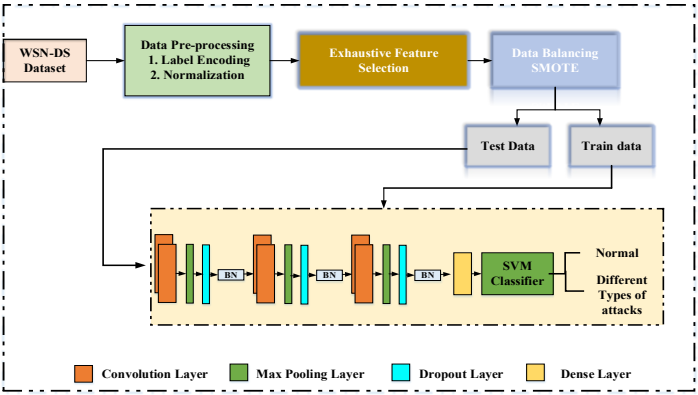

Fig. 1. Architecture of the proposed approach

The initial data pre-processing stage involves preparing the raw dataset for categorization algorithms – the One-Hot Encoding technique, which transforms the categorical data into numeric data. Then the Normalization Process was carried out to determine the normalized value using mean and standard deviation values. After normalization, removing unnecessary and redundant features and choosing the best subset of features is crucial using a feature selection approach that uses a process called Manta-Ray Foraging based Feature Selection that assesses each feature set using a fitness function. It implies that this method attempts to make every possible feature combination and returns the set of attributes that performs the best. After that, we must balance the imbalanced dataset to increase classification performance. The assault detection dataset contains imbalanced classes, which hurts classification performance. Oversampling with under-sampling approaches was used in our research. Noise and unwanted data size growth are caused when oversampling techniques are used alone. The imbalanced class problem is addressed in this research using the SMOTE approach. Following that, the dataset must be divided for testing and training purposes.

Attacks are finally identified using the CNN-SVM-based Deep Learning approach, in which the Convolution layer, Maxpooling layer, Batch Normalization, Dropout layer, and Dense layer are employed. Finally, the SVM classifier is utilized to classify the attack kinds. The ADAM optimizer will tune the deep learning model's hyper-parameter to detect intrusions with a high detection accuracy by minimizing the error rate. As a result, the proposed multi-class intrusion detection approach can accurately classify the specific classes such as Flooding, Blackhole, Normal, scheduling, or Grayhole. The following sections are arranged to explain the proposed method process in detail.

3.1. Dataset Description. A WSN-DS dataset was constructed in [27] to detect and classify the attacks on WSNs specifically. The Low-energy adaptive clustering hierarchy (LEACH) protocol was selected to create the dataset since it is one of the most popular and regularly utilized routing protocols in WSNs. The data were gathered in the NS-2 simulation environment. The dataset includes 23 characteristics that were obtained using the LEACH routing technique. The LEACH protocol is a routing protocol that employs 23 characteristics to indicate the status of each sensor node in the wireless network. There are 23 features, including Id, Time, Is_CH, Who_CH, RSSI, Dist_To_CH, M_D_CH, and A_D_CH. Energy that is now being used, consumed, ADV_S, ADV_R, JOIN_S, JOIN_R, ADV_SCH_S, ADV_SCH_R, rank, DATA_S, DATA_R, Data_Sent_BS, Dist_CH_BS, Send_code, and attack_type. However, as indicated in Table 1, there are only 19 features total in the dataset file, in addition to the class label.

Table 1. A detailed description of the attributes of the WSN-DS dataset

No	Attribute Name	Attribute Description
1	Id (Identification)	To recognize the sensor node in any round and at any stage, it is a special ID
2	Time	The sensor node's current simulation time is displayed
3	Is_CH	It acts as a flag to show whether the node is a CH (cluster head). Value 0 denotes a normal node, whereas value 1 denotes a CH
4	who_CH	It is a CH ID for the current round
5	Dist_To_CH	In the current round, it is the separation between the node and its CH
6	ADV_S (Advertise_Send)	The number was used to promote CH's broadcast messages delivered to nodes
7	ADV_R (Advertise_Receive)	The number was used to promote CH communications obtained from CHs
8	JOIN_S (Join_Send)	The amount of join request messages delivered to the CH by the nodes
9	JOIN_ R (Join_Receive)	It represents the number of join request messages the CH has received from the nodes
10	SCH_S (Schedule_Send)	It depends on how many advertised TDMA (Time Division Multiple Access) scheduled broadcast messages are sent to the nodes
11	SCH_ R (Schedule_Receive)	It is the quantity of advertised TDMA scheduling messages that the CH has received
12	Rank	It is this node's position in the TDMA schedule
13	DATA_S (Data packets_Send)	It refers to how many data packets a sensor delivers to its CH
14	DATA_R (Data packets_Receive)	It measures how many data packets were received from CH
15	Data_Sent_BS (Data packets _Send_ Base station)	It refers to how many data packets were transferred to the BS
16	Dist_CH_BS (Distance _Cluster Head_ Base station)	It measures the separation between the CH and the BS (base station).
17	Send_Code	It is the cluster-sending node
18	Consumed_Energy	It represents the energy that the sensor node used during the previous round
19	Attack_Type	Type of Attack (A black hole, Grayhole, Flooding, TDMA/Scheduling, Normal)

The WSN-DS dataset has 374.661 samples. The samples in the dataset are separated into five groups, four of which correspond to different forms of DoS attacks: Black hole, grey hole, flooding, TDMA/Scheduling, and normal. Table 2 displays the precise data distribution of each class.

Table 2. The number of samples in each class of WSN-DS Attack

Class	Number of Samples	Proportion (%)
Normal	340066	90.77
Gray hole	10049	2.68
Black hole	14596	3.90
TDMA/Scheduling	3312	0.88
Flooding	6638	1.77
Total	374661	100

This WSN-DS dataset served as the basis for all evaluations in the fundamental research. The dataset was selected as that includes DoS assaults specific to WSNs.

3.2. Data Normalization and Encoding. Constructing the raw dataset suitable for classification algorithms is the first phase in data pre-processing. Due to this, the One-Hot Encoding technique was employed in this work to first convert the dataset's category values to quantitative data, after which the normalization step was carried out by Equation (1).

$$x_z^{'} = \left| x_z - \mu \right| / \sigma \,, \tag{1}$$

where, x_z is the original value, x_z' are the normalized value, μ and σ are the mean and standard deviation values, respectively. As a result, each numerical value in the data set was changed to a number between 0 and 1.

For instance, the Attack Type attribute in the WSN-DS dataset contains textual descriptors like Normal, Grayhole, Blackhole, TDMA, and Flooding. AI-systems are prevented from using these textual descriptions to continue their calculations. Consequently, employing our suggested One-Hot Encoding in conjunction with the normalizing method is required to transform these descriptors into numeric values.

The process of One-Hot encoding can be applied in the WSN-DS dataset:

1. Identify the categorical variables: The attack type is described in textual format in this data set.

2. Create binary columns: Create a new binary column for each categorical variable for each unique category. In the case of "Attack Type,"

five binary columns will be created: "Normal," "Gray hole," "Black hole," "TDMA" and "Flooding".

3. Assign binary values: Assign a value of 1 to the appropriate binary column for each data instance corresponding to the respective category.

4. Remove the original categorical variable: Once the One-Hot encoding is applied, the original categorical variable.

The classification labels were changed into numbers via the One-Hot Encoding process, as shown in Table 3.

Table 3. Findings from Data Normalization and Encoding

Attribute Class	One-Hot Encoding Output
Normal	0
Gray hole	1
Black hole	2
TDMA	3
Flooding	4

Certain characteristics with large numerical values are prevented after normalization and encoding from impairing the algorithm's performance and having a negative impact. These data are then used in the feature selection method; this is covered in greater detail in the section below.

3.3. Feature Selection. Feature selection is an approach that determines the most effective feature subset by reducing noise and eliminating unnecessary and redundant features in the WSN-DS dataset. It can enhance computation performance by reducing the computational load. This research proposed a Manta Ray Foraging (MRF) optimization to choose the features, which examines each feature set that simulates manta ray foraging behaviours. Three foraging operators are used in this strategy: chain foraging, cyclone foraging, and somersault foraging.

3.3.1. Chain Foraging. The manta ray chain prefers to devour plankton as its main diet. Hence the MRF optimization algorithm indicates that a place with much plankton is the finest. According to the optimum location and the manta ray in front of it, the MRF algorithm modifies the manta ray's position by omitting the first one. The chain foraging update process is shown in Equation (2):

$$s_m^{iter+1} = \begin{cases} s_m^{iter} + rand_1\left(s_b^{iter} - s_m^{iter}\right) + \gamma\left(s_b^{iter} - s_m^{iter}\right), m = 1 \\ s_m^{iter} + rand_2\left(s_{m-1}^{iter} - s_m^{iter}\right) + \gamma\left(s_b^{iter} - s_m^{iter}\right), m = 2, \dots, M \end{cases}, \quad (2)$$

$$\gamma = 2 \times rand_3 \times \sqrt{\left|\log 2\left(rand_4\right)\right|}, \quad (3)$$

where, s_m^{iter} denotes where the m^{th} the manta ray is at iteration; $rand_{1,2,3,4}$ are randomly produced integers in the range [0, 1] that are distinct from one another; γ is the weight coefficient; $s_b^{iteration}$ is the maximum level of plankton concentration. The optimum plankton site, therefore, dictates the position update.

3.3.2. Cyclone Foraging. In addition to moving in the path of the plankton, each manta ray follows the manta ray in front of it as it swims. Equation (4) provides a mathematical representation of the cyclone foraging updating strategy.

$$s_m^{iter+1} = \begin{cases} s_b^{iter} + rand_5\left(s_b^{iter} - s_m^{iter}\right) + \propto \left(s_b^{iter} - s_m^{iter}\right), m = 1 \\ s_b^{iter} + rand_6\left(s_{m-1}^{iter} - s_m^{iter}\right) + \propto \left(s_b^{iter} - s_m^{iter}\right), m = 2, \ldots, M \end{cases}, \quad (4)$$

$$\propto = 2 \times e^{rand_7} \times \left(\frac{Max.iter - iter + 1}{iter}\right) \sin\left(2 \times \pi \times rand_8\right), \quad (5)$$

where, $\propto$ represent the weight coefficient; $Max.iter$ determines the most iterations possible; and $rand_{5.6}$ represent the different randomly produced numbers between [0, 1].

The MRF algorithm's exploration and exploitation are mostly driven by this step, which uses the best plankton as a benchmark (forcing the manta rays to move to a random place in the explore regions that is both distant from their current location and the best site for prey). Equations (6) and (7) show this hypothesized process.

$$s_{rp}^{iter} = Lw + rand_9\left(UP - Lw\right), \quad (6)$$

$$s_m^{iter+1} = \begin{cases} s_{rp}^{iter} + rand_{10}\left(s_{rp}^{iter} - s_m^{iter}\right) + \propto \left(s_{rp}^{iter} - s_m^{iter}\right), m = 1 \\ s_{rp}^{iter} + rand_{11}\left(s_{m-1}^{iter} - s_m^{iter}\right) + \propto \left(s_{rp}^{iter} - s_m^{iter}\right), m = 2, \ldots, M \end{cases}, \quad (7)$$

where, s_{rp}^{iter} is the fabricated random place within the permitted range; UP and Lw are the maximum and minimum values that can be found at a specific place; and $rand_{5,6,7,8,9,10,11}$ is the various randomly generated integers between 0 and 1.

3.3.3. Somersault Foraging. Equation 8 provides a mathematical description of this scenario where each manta ray circles this spot and drifts

to a different position. The reference point is the location with the greatest plankton concentration thus far.

$$s_m^{iter+1} = s_m^{iter} + smsf \times \left(rand_{12} \times s_b^{iter} - \left(rand_{13} \times s_m^{iter} \right) \right), \quad (8)$$

where $smsf$ is the somersault factor; and $rand_{12,13}$ are various randomly generated integers between 0 and 1.

Algorithm 1. MRF optimization algorithm

Input: Normalized data
Output: Extracted features
While criteria! Satisfied **do**
for m=1,…, M **do**
if $rand_1 < 0.5$ **then**
if ($iter/Max.iter$) < $rand_2$ **then // Cyclone Foraging**
Evaluate using equations (5 and 6)
else
Determine Equation (3)
end if
else // Chain Foraging
Calculate using Equation (1)
end if
// Calculate fitness for n^{th} manta ray f(s_m^{iter+1})
if f(s_m^{iter+1}) < f(s_b^{iter}) **then**
$s_b^{iter} = s_m^{iter+1}$
end if
end for
// Somersault Foraging
for m=1,…, M **do**
Evaluate using Equation (7)
// Calculate fitness for m^{th} manta ray f(s_m^{iter+1})
if f(s_m^{iter+1}) < f(s_b^{iter}) **then**
$s_b^{iter} = s_m^{iter+1}$
end if
end for
end while

The Id and ADV R features in the WSN-DS dataset show substantial correlations among themselves, even though many other characteristics in the dataset do not.

The correlation between these two features can be calculated using Pearson's correlation coefficient to measure the linear relationship between 2 variables. The corr ( ) function computes the pairwise correlations between the dataset's selected features (Id and ADV R). The Id property was thus dropped from the dataset due to this action. Id, ADV R, and SCH_R, respectively, have the attributes that have the least effect on the class, as shown in Table 4. As a result, the dataset did not include these attributes.

Table 4. Feature selection results of the proposed approach

Feature No	Attribute Name
2	Time
3	Is_CH
4	who_CH
5	Dist_To_CH
6	ADV_S
8	JOIN_S
9	JOIN_R
10	SCH_S
12	Rank
13	DATA_S
14	DATA_R
15	Data_Sent_BS
16	dist_CH_To_BS
17	Send_Code
18	Consumed_Energy

The number of features has decreased due to the feature selection procedure, leaving the WSN-DS dataset with only 16 features.

3.4. Data Balancing. The WSN-DS contains unevenly distributed classes, which has a detrimental effect on the precise classification. In study [28] the authors suggested the heuristic oversampling approach known as SMOTE to address the class imbalance in datasets. By oversampling the data from the minority class, this strategy creates fake data. By creating synthetic data, SMOTE also solves the overfitting issue brought on by random oversampling techniques. As a result, this study suggested using a SMOTE to modify the sparse spreading of the minority class samples and usually avoid overfitting.

SMOTE selects samples that are near the feature space. From the minority class, a random sample is taken, and the sample's nearest k neighbours are then identified. Following the random selection among the closest neighbours, the distinction between the two sample features is multiplied by a value between 0 and 1 and added to the selected sample value. A line then connects the two sample features, and synthetically samples are produced this way. The kNN algorithm are randomly picked the neighbours depending on the oversampling needed.

The following definition applies to SMOTE samples, which are linear combinations of the minority class's two comparable samples $\left(r^S - r\right)$:

$$n = r + v \times \left(r^S - r\right), \ 0 \le d \le 1, \tag{9}$$

where v is the variation between the two samples; and r^S is the randomly chosen model of r based on the closest neighbour digit. In order to address the class imbalance issue, SMOTE introduces new samples at random between neighbourhood samples from minority classes, greatly boosting the percentage of samples from minority classes. Figure 2 shows the raw dataset, and Figure 3 illustrates the after-data balancing.

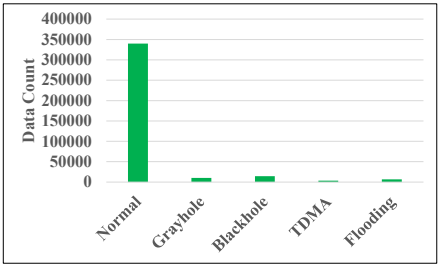

Fig. 2. Dataset before data balancing

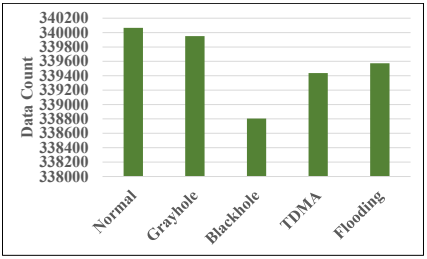

Fig. 3. Dataset after data balancing

Following that, the dataset is split into sections for testing and training. Finally, to classify the attacks, this research proposed a deep learning-based framework described in the following section.

3.5 CNN-SVM-based classification. Attacks are finally identified using the CNN-SVM based Deep Learning approach, in which the Convolution layer, max-pooling layer, batch normalization, Dropout layer, and dense layer are employed. Finally, the SVM classifier is utilized to provide the classification of the attack kinds.

CNN: The most analytical information can be gathered from raw data by CNN. In the proposed framework, the most recognizable features are taken from the raw data using a 5x5 kernel/filter. The input layer's n×n input neurons are convolved with an m×m filter in the convolutional layer, producing an output with the dimensions (n-m+1) × (n-m+1). Each layer's output serves as the following layer's input.

SVM: SVM attempts to denote a multi-dimensional dataset in a space where a hyperplane separates the data into different classes. The SVM classifier can minimize the generalization error.

CNN-SVM: The hybrid CNN-SVM model, SVM functions as a binary classifier and takes the place of CNN's softmax layer in this scenario, is introduced in the current work. Although CNN is a feature extractor, SVM is a binary classifier. Figure 4 describes a suggested hybrid CNN-SVM model's design.

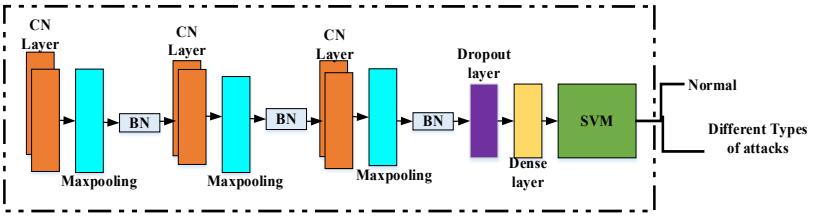

Fig. 4. Hybrid CNN-SVM approach

A 28×28 normalized data matrix is provided to the CNN layer as an input. The convolutional layers have a stride of size two and a 5×5 convolutional filtering. The feature map layer's convolution layers 1, 2, and 3 extract values that are also considered distinguishing properties of the input data. Here, dropout is a powerful regularization technique that can significantly reduce overfitting in Deep Neural Network (DNNs). By randomly dropping out neurons during training, the network learns to generalize better to new data. The CNN is trained after going through multiple epochs and until the training process converges. Here, the SVM classifier takes the place of CNN's last layer. The SVM classifier considers

the input features from the convolution layer 3 data as input. These new automatically generated training data features are initially used to train the SVM classifier. To identify the assaults utilized for testing, the trained SVM classifier is employed.

Table 5. Model Hyperparameter

Hyperparameter	Value
Iterations	100
Activation Function	ReLU
Loss Function	Categorical cross-entropy
Optimization algorithm	Adam
Learning rate	0.01

By reducing the error rate, as shown in Table 5, the hyper-parameter of the deep learning model will be tuned by the ADAM optimizer to identify intrusions with high detection accuracy. Consequently, the proposed CNN-SVM method can accurately categorize certain categories like Flooding, Blackhole, Normal, TDMA, or Grayhole.

4. Expriments and Evaluations. This segment details the performance and comparative findings of the suggested strategy and the implementation outcomes. The proposed method is implemented on the balanced WSN-DS dataset. For the evaluation purpose, we have split the dataset into testing and training in the ratio of 50:50. The simulation was carried out using the MATLAB 2021a tool operated in Windows 7 operating systems with 8GB RAM size of Intel premium processor.

4.1. Evaluation Parameters. This section describes the various performance indicators of the recommended deep learning system for classifying attacks in WSN.

Figure 5 demonstrates that after 32 iterations, the MRF optimization technique decreases the fitness function and converges on the value. To achieve the best characteristics, the optimization was done 100 times. As a result, the MRF technique has a good performance in terms of feature selection.

Figure 6 shows the multi-intrusion detection system outcomes of the suggested CNN-SVM-based method. Figure 6 classifies the assaults in the WSN-DS dataset. A confusion matrix displays the assaults properly detected on the main diagonal (top left to bottom right). The erroneous labels are visible in the remaining cells, referred to as true negatives or false negatives. The suggested hybrid model, therefore, produces superior outcomes. From that confusion matrix, we obtain the following performance values.

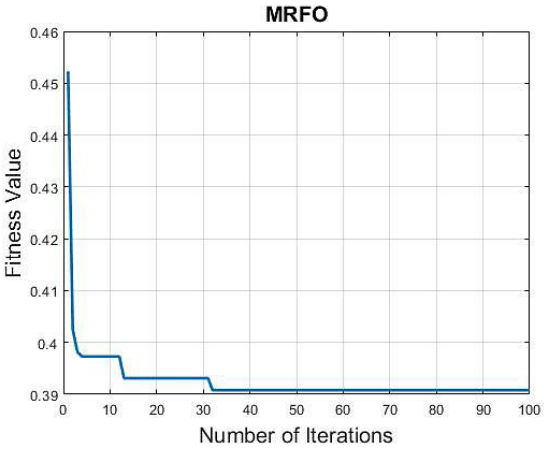

Fig. 5. Results of Manta Ray Foraging Optimization

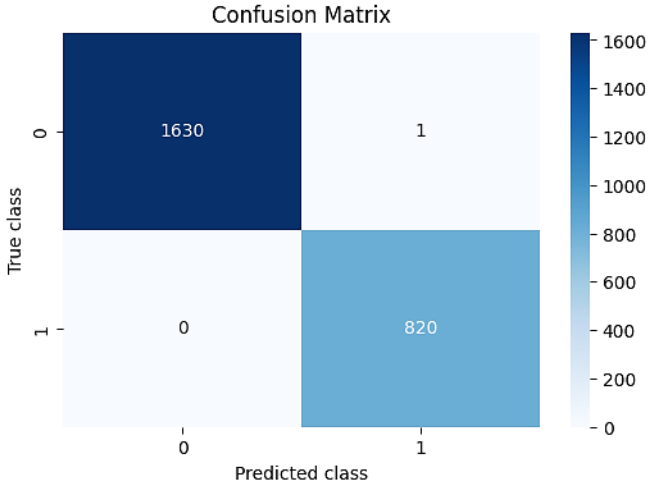

Fig. 6. Confusion Matrix

The efficiency of the unique strategy for attack categorization is assessed using the performance of our suggested method and several measures, including accuracy, F1 Score, precision, and recall. The following formulas were used to assess the Accuracy, Precision, Recall, and F1 Score performance metrics.

$$Accuracy = \frac{\text{True Positive} + \text{True Negative}}{\text{True Positive} + \text{True Negative} + \text{False Positive} + \text{False Negative}}, \quad (10)$$

$$\text{Precision} = \frac{\text{True Positive}}{\text{True Positive} + \text{False Positive}}, \quad (11)$$

$$\text{Recall} = \frac{\text{True Positive}}{\text{True Positive} + \text{False Negative}}, \quad (12)$$

$$F1\ Score = \frac{2 * Precision * Recall}{Precision + Recall}. \quad (13)$$

The suggested method's performance evaluation measures are shown in Figure 7.

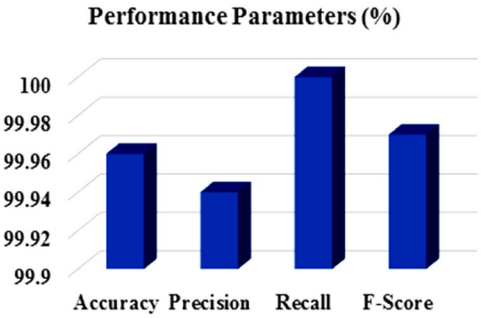

Fig. 7. Performance parameters of the proposed approach

The resulting values of accuracy, F1 Score, recall and F-score are 99.75%, 99.60%, 99.21% and 100.0%, tabulated in Table 6.

Table 6. Performance Evaluation

Performance Parameters	Value (%)
Accuracy	99.96
Precision	99.94
Recall	100
F-Score	99.97

By applying data normalization and encoding, the performance of our suggested method achieves greater accuracy, F1 score, precision, recall,

feature selection using the MRF approach, and data balancing and classification of attacks using the CNN-SVM approach.

4.2. Comparison Analysis. This segment contrasts the suggested method with accepted practices to evaluate it against current practices. The following algorithms have been reported: Decision Tree (DT) [26], Random Forest (RF) [26], Naive Bayes (NB) [26], Logistic Regression (LR) [26], Multi-Layer Perceptron (MLP) [26], CNN [26], LSTM [26], and CNN-LSTM [26].

Figure 8 displays a comparison of all accuracy levels. Using CNN-SVM enhances the accuracy of the suggested method. As related to the baseline, our suggested technique achieves greater accuracy than DT, RF, NB, LR, MLP, CNN, LSTM, and CNN-LSTM such as 97.53%, 97.67%, 92.53%, 97.16%, 96.24%, 98.03%, 99.31%, and 98.04%.

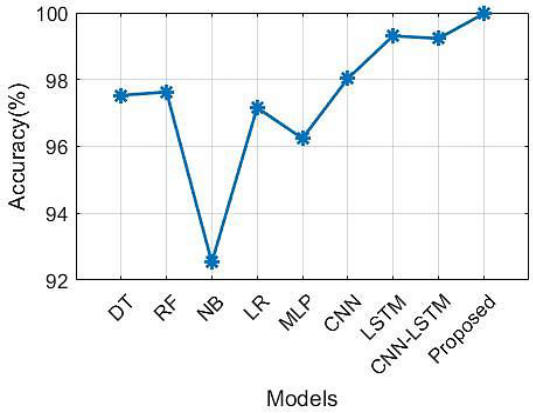

Fig. 8. Comparison of Accuracy

Consequently, our original distinguishing technique has a 99.96% accuracy rate, outperforming conventional techniques.

Figure 9 shows the comparison of complete precision. Using CNN-SVM enhances the suggested method's accuracy. As compared to the baseline, our suggested strategy achieves greater precision than DT, RF, NB, LR, MLP, CNN, LSTM, and CNN-LSTM such as 0.97%, 0.98%, 0.88%, 0.97%, 0.96%, 0.98%, 0.99%, and 0.98%. Our innovative distinguishing method has a 0.9994% precision, outperforming conventional techniques.

Figure 10 shows the entire recall comparison. Using CNN-SVM increases the recall of the suggested method. We are comparing the recall of

our suggested method to the baseline DT, RF, NB, LR, MLP, CNN, LSTM, and CNN-LSTM such as 0.98%, 0.98%, 0.93%, 0.97%, 0.96%, 0.98%, 0.99%, and 0.98%. Consequently, our original, distinctive technique has a 100.0% recall compared to standard methods.

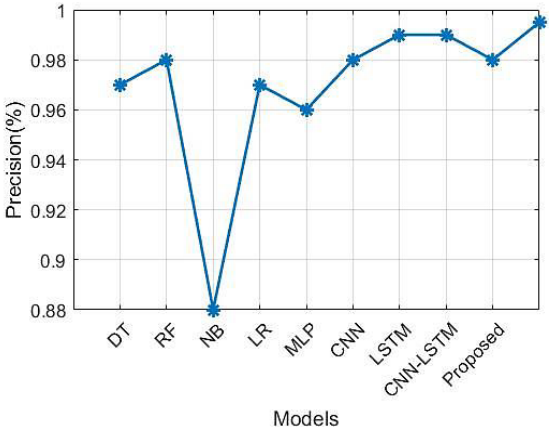

Fig. 9. Comparison of Precision

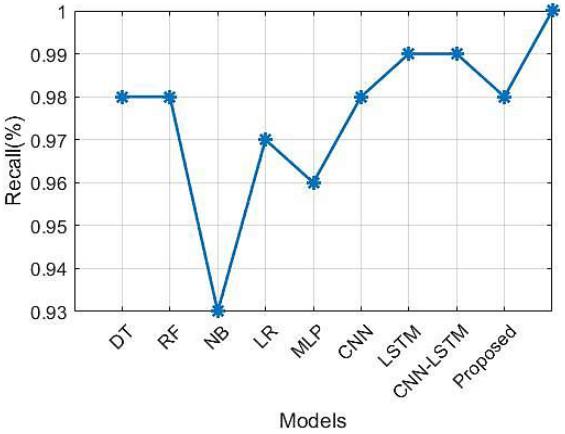

Fig. 10. Comparison of Recall

Figure 11 displays the comparison of the total F-Score. Using CNN-SVM raises the F-Score of the suggested method. As compared to the baseline, our suggested method outperforms DT, RF, NB, LR, MLP, CNN, LSTM, and CNN-LSTM such as 0.97%, 0.97%, 0.90%, 0.97%, 0.96%,

0.98%, 0.98%, and 0.98%. Because of this, our innovative, distinctive method has a 0.9997% F-Score, which is better than baseline methods. The overall comparison of the performance parameters is tabulated in Table 7.

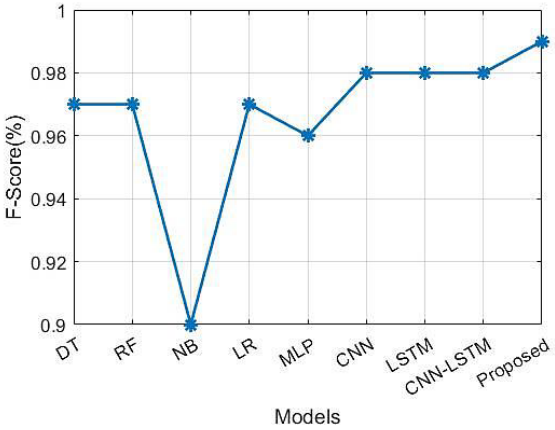

Fig. 11. Comparison of F-Score

Table 7. Comparison of Performance Parameters

Methods	Accuracy (%)	Precision (%)	Recall (%)	F-Score (%)
DT	97.53	97	98	97
Random Forest	97.67	98	98	97
Naïve Bayes	92.53	88	93	90
Logstic Regression	97.16	97	97	97
Multi-layer Perceptron	96.24	96	96	96
CNN	98.03	98	98	98
LSTM	99.31	99	99	98
CNN-LSTM	98.04	98	98	98
Proposed	99.75	99.21	100	99.6

4.3. Discussion. The striking aspect of the results is that, on the WSN-DS dataset, the CNN and LSTM deep learning algorithms outperform the CNN-SVM hybrid strategy separately. Data balancing enhances the performance of the DoS intrusion detection system greatly. In research published in the literature, hybrid approaches frequently outperformed individual deep learning techniques, whereas individual methods outperformed hybrid methods on the WSN-DS dataset with the feature

selection algorithm. The peculiar feature structure of the WSN-DS dataset is thought to be the root cause of this circumstance. The NB algorithm had the worst results. Since it could not recognize data from the TDMA and Flooding classes with a high rate, the Naive Bayes method performed the worst. Figures 8–11 and Table 7 compare classification algorithms based on accuracy, precision, recall, and F1-Score characteristics. When the outcomes of these parameters are analyzed, the suggested technique yields the best results for each parameter.

5. Conclusion. This work proposes a novel classification-based DoS intrusion detection method to identify DoS assaults targeted at WSNs. The suggested CNN-SVM technique combines data balance and feature selection operations in the deep learning-based approach. The SMOTE ensemble technique was utilized in this work to balance the data. For the attack categorization procedure, CNN-SVM was used. Also, the impact of each feature was evaluated, and using the Manta Ray Optimization feature selection approach, the number of features was lessened from 19 to 16. It was observed that the feature selection process led to an improvement in the algorithm's performance. The WSN-DS dataset was used for experimental experiments. The metrics Accuracy, Precision, Recall, and F-Score were used to assess the performance of the suggested technique. According to the test findings, the proposed approach performed better than the alternatives, with an accuracy, F-Measure, Precision, and Recall of 99.75%, 99.21%, 100%, and 99.6%. In the future, it is intended to integrate the machine learning algorithm with various deep learning techniques, such as Gated Recurrent Unit (GRU) and Auto Encoder (AE), to increase performance and analyze the outcomes. Furthermore, several oversampling and undersampling approaches will be explored for data balance. Furthermore, the suggested method's performance on various datasets will be evaluated. Finally, research is being conducted to improve the dependability and transparency of intrusion detection systems using Explainable Artificial Intelligence approaches.

References

1. Kopetz H. Internet of things. Real-time systems. 2011. pp. 307–323. DOI: 10.1007/978-1-4419-8237-7_13.
2. Alsulaiman L., Al-Ahmadi S. Performance evaluation of machine learning techniques for DOS detection in wireless sensor network. arXiv preprint arXiv:2104.01963. 2021.
3. Aziz N.A.A., Aziz K.A. Managing disaster with wireless sensor networks. 13th international conference on advanced communication technology (ICACT2011). 2011. pp. 202–207.
4. Butun I., Morgera S.D., Sankar R. A survey of intrusion detection systems in wireless sensor networks. IEEE Communications surveys and tutorials. 2013. vol. 16. no. 1. pp. 266–282.

5. Pelechrinis K., Iliofotou M., Krishnamurthy S.V. Denial of service attacks in wireless networks: The case of jammers. IEEE Communications surveys and tutorials. 2010. vol. 13. no. 2. pp. 245–257.
6. López J., Zhou J. (Eds.). Wireless sensor network security. Ios Press, 2008. 320 p.
7. Das S.K., Kant K., Zhang N. Handbook on securing cyber-physical critical infrastructure. Elsevier Inc., 2012. 848 p.
8. Rassam M.A., Maarof M.A., Zainal A. A survey of intrusion detection schemes in wireless sensor networks. American Journal of Applied Sciences. 2012. vol. 9. no. 10. pp. 1636–1652.
9. Mahbooba B., Sahal R., Alosaimi W., Serrano M., Alosaimi W. Trust in intrusion detection systems: an investigation of performance analysis for machine learning and deep learning models. Complexity. 2021. vol. 2021. 23 p. DOI: 10.1155/2021/5538896.
10. Cauteruccio F., Fortino G., Guerrieri A., Liotta A., Mocanu D.C., Perra C., Terracina G., Vega M.T. Short-long term anomaly detection in wireless sensor networks based on machine learning and multi-parameterized edit distance. Information Fusion. 2019. vol. 52. pp. 13–30.
11. Vinayakumar R., Alazab M., Soman K.P., Poornachandran P., Al-Nemrat A., Venkatraman S. Deep learning approach for intelligent intrusion 936 detection system. IEEE Access. 2019. vol. 7. pp. 41525–41550.
12. Alqahtani M., Gumaei A., Mathkour H., Maher Ben Ismail M. A genetic-based extreme gradient boosting model for detecting intrusions in wireless sensor networks. Sensors. 2019. vol. 19(20). no. 4383. DOI: 10.3390/s19204383.
13. Tan X., Su S., Huang Z., Guo X., Zuo Z., Sun X., Li L. Wireless sensor networks intrusion detection based on SMOTE and the random forest algorithm. Sensors. 2019. vol. 19(1). DOI: 10.3390/s19010203.
14. Jiang S., Zhao J., Xu X. SLGBM: An intrusion detection mechanism for wireless sensor networks in smart environments. IEEE Access. 2020. vol. 8. pp. 169548–169558.
15. Abhale A.B., Manivannan S.S. Supervised machine learning classification algorithmic approach for finding anomaly type of intrusion detection in wireless sensor network. Optical Memory and Neural Networks. 2020. vol. 29. pp. 244–256.
16. Tang C., Luktarhan N., Zhao Y. An efficient intrusion detection method based on LightGBM and autoencoder. Symmetry. 2020. vol. 12(9). no. 1458. DOI: 10.3390/sym12091458.
17. Nancy P., Muthurajkumar S., Ganapathy S., Santhosh Kumar S.V.N., Selvi M., Arputharaj K. Intrusion detection using dynamic feature selection and fuzzy temporal decision tree classification for wireless sensor networks. IET Communications. 2020. vol. 14. no. 5. pp. 888–895.
18. Liu J., Gao Y., Hu F. A fast network intrusion detection system using adaptive synthetic oversampling and LightGBM. Computers and Security. 2021. vol. 106. no. 102289.
19. Al S., Dener M. STL-HDL: A new hybrid network intrusion detection system for imbalanced dataset on big data environment. Computers and Security. 2021. vol. 110. no. 102435.
20. Ifzarne S., Tabbaa H., Hafidi I., Lamghari N. Anomaly detection using machine learning techniques in wireless sensor networks. Journal of Physics: Conference Series. 2021. vol. 1743(1). no. 012021. DOI: 10.1088/1742-6596/1743/1/012021.
21. Pan J.S., Fan F., Chu S.C., Zhao H.Q., Liu G.Y. A Light-weight Intelligent Intrusion Detection Model for Wireless Sensor Networks. Security and Communication Networks. 2021. vol. 2021(2). 15 p. DOI: 10.1155/2021/5540895.

22. Zamry N.M., Zainal A., Rassam M.A., Alkhammash E.H., Ghaleb F.A., Saeed F. Light-weight Anomaly Detection Scheme Using Incremental Principal Component Analysis and Support Vector Machine. Sensors. 2021. vol. 21(23). no. 8017. DOI: 10.3390/s21238017.
23. Yadav A., Kumar A. Intrusion Detection and Prevention Using RNN in WSN. Proceedings of Inventive Computation and Information Technologies (ICICIT). 2022. pp. 531–539.
24. Tabbaa H., Ifzarne S., Hafidi I. An Online Ensemble Learning Model for Detecting Attacks in Wireless Sensor Networks. arXiv preprint arXiv:2204.13814. 2022. 15 p.
25. Salmi S., Oughdir L. CNN-LSTM Based Approach for Dos Attacks Detection in Wireless Sensor Networks. International Journal of Advanced Computer Science and Applications. 2022. vol. 13. no. 4. pp. 835–842.
26. Dener M., Al S., Orman A. STLGBM-DDS: An Efficient Data Balanced DoS Detection System for Wireless Sensor Networks on Big Data Environment. IEEE Access. 2022. vol. 10. pp. 92931–92945.
27. Almomani I., Al-Kasasbeh B., Al-Akhras M. WSN-DS: A dataset for intrusion detection systems in wireless sensor networks. Journal of Sensors. 2016. vol. 2016. 16 p. DOI: 10.1155/2016/4731953.
28. Chawla N.V., Bowyer K.W., Hall L.O., Kegelmeyer W.P. SMOTE: Synthetic minority over-sampling technique. Journal of artificial intelligence research. 2002. vol. 16. pp. 321–357.

Krishna Kuraganty Phani Rama — Assistant professor, Electronics and communication department, Prasad V. Potluri Siddhartha Institute of Technology; Researcher, Koneru Lakshmaiah Education Foundation (Deemed to be University), Vaddeshwaram. Research interests: wireless sensor networks, wireless communication. The number of publications — 17. kprkrishna007@gmail.com; Teacher's Colony, Bhavanipuram, 520012, Vijayawada, India; office phone: +91(866)258-1699.

Thirumuru Ramakrishna — Ph.D., Professor, Electronics and communication department, Koneru Lakshmaiah Education Foundation (Deemed to be University), Vaddeshwaram. Research interests: speech signal processing, wireless sensor networks, wireless communication. The number of publications — 25. ramakrishnaece@kluniversity.in; Green Fields, Vaddeswaram, 522302, Guntur, Russia; office phone: +91(8645)350-0200.

DOI 10.15622/ia.22.6.2

К. КРИШНА, Р. ТИРУМУРУ

СБАЛАНСИРОВАННАЯ СИСТЕМА ОБНАРУЖЕНИЯ ВТОРЖЕНИЙ ДЛЯ БЕСПРОВОДНЫХ СЕНСОРНЫХ СЕТЕЙ В СРЕДЕ БОЛЬШИХ ДАННЫХ С ИСПОЛЬЗОВАНИЕМ МОДЕЛИ CNN-SVM

Кришна К., Тирумуру Р. **Сбалансированная система обнаружения вторжений для беспроводных сенсорных сетей в среде больших данных с использованием модели CNN-SVM.**

Аннотация. Беспроводные сенсорные сети (WSN) подвергались нескольким различным проблемам безопасности и атакам, связанным со сбором и отправкой данных. В этом сценарии одной из наиболее распространенных атак WSN, которая может быть нацелена на любой уровень стека протоколов, является атака типа «отказ в обслуживании» (DoS). Текущее исследование предлагает различные стратегии обнаружения атаки в сети. Однако у него есть проблемы с классификацией. Поэтому в этом исследовании для решения этой проблемы была предложена эффективная система обнаружения вторжений на основе ансамблевого глубокого обучения для выявления атак в сети WSN. Предварительная обработка данных включает преобразование качественных данных в числовые с использованием метода One-Hot Encoding. После этого был проведен процесс нормализации. Затем предлагается выбрать лучшее подмножество функций с помощью Manta-Ray Foraging Optimization. Затем метод передискретизации синтетического меньшинства (SMOTE) создает новую выборку меньшинства для балансировки обработанного набора данных. Наконец, предлагается классификатор CNN-SVM для классификации видов атак. Метрики Точность, F-мера, Прецизионность и Отзыв использовались для оценки результатов 99,75%, 99,21%, 100% и 99,6% соответственно. По сравнению с существующими подходами предложенный метод оказался чрезвычайно эффективным при обнаружении DoS-атак в WSN.

Ключевые слова: беспроводная сенсорная сеть, DoS-атаки, искусственный интеллект, глубокое обучение, сверточные нейронные сети, метод опорных векторов.

Литература

1. Kopetz H. Internet of things. Real-time systems. 2011. pp. 307–323. DOI: 10.1007/978-1-4419-8237-7_13.
2. Alsulaiman L., Al-Ahmadi S. Performance evaluation of machine learning techniques for DOS detection in wireless sensor network. arXiv preprint arXiv:2104.01963. 2021.
3. Aziz N.A.A., Aziz K.A. Managing disaster with wireless sensor networks. 13th international conference on advanced communication technology (ICACT2011). 2011. pp. 202–207.
4. Butun I., Morgera S.D., Sankar R. A survey of intrusion detection systems in wireless sensor networks. IEEE Communications surveys and tutorials. 2013. vol. 16. no. 1. pp. 266–282.
5. Pelechrinis K., Iliofotou M., Krishnamurthy S.V. Denial of service attacks in wireless networks: The case of jammers. IEEE Communications surveys and tutorials. 2010. vol. 13. no. 2. pp. 245–257.
6. López J., Zhou J. (Eds.). Wireless sensor network security. Ios Press, 2008. 320 p.
7. Das S.K., Kant K., Zhang N. Handbook on securing cyber-physical critical infrastructure. Elsevier Inc., 2012. 848 p.

8. Rassam M.A., Maarof M.A., Zainal A. A survey of intrusion detection schemes in wireless sensor networks. American Journal of Applied Sciences. 2012. vol. 9. no. 10. pp. 1636–1652.
9. Mahbooba B., Sahal R., Alosaimi W., Serrano M., Alosaimi W. Trust in intrusion detection systems: an investigation of performance analysis for machine learning and deep learning models. Complexity. 2021. vol. 2021. 23 p. DOI: 10.1155/2021/5538896.
10. Cauteruccio F., Fortino G., Guerrieri A., Liotta A., Mocanu D.C., Perra C., Terracina G., Vega M.T. Short-long term anomaly detection in wireless sensor networks based on machine learning and multi-parameterized edit distance. Information Fusion. 2019. vol. 52. pp. 13–30.
11. Vinayakumar R., Alazab M., Soman K.P., Poornachandran P., Al-Nemrat A., Venkatraman S. Deep learning approach for intelligent intrusion 936 detection system. IEEE Access. 2019. vol. 7. pp. 41525–41550.
12. Alqahtani M., Gumaei A., Mathkour H., Maher Ben Ismail M. A genetic-based extreme gradient boosting model for detecting intrusions in wireless sensor networks. Sensors. 2019. vol. 19(20). no. 4383. DOI: 10.3390/s19204383.
13. Tan X., Su S., Huang Z., Guo X., Zuo Z., Sun X., Li L. Wireless sensor networks intrusion detection based on SMOTE and the random forest algorithm. Sensors. 2019. vol. 19(1). DOI: 10.3390/s19010203.
14. Jiang S., Zhao J., Xu X. SLGBM: An intrusion detection mechanism for wireless sensor networks in smart environments. IEEE Access. 2020. vol. 8. pp. 169548–169558.
15. Abhale A.B., Manivannan S.S. Supervised machine learning classification algorithmic approach for finding anomaly type of intrusion detection in wireless sensor network. Optical Memory and Neural Networks. 2020. vol. 29. pp. 244–256.
16. Tang C., Luktarhan N., Zhao Y. An efficient intrusion detection method based on LightGBM and autoencoder. Symmetry. 2020. vol. 12(9). no. 1458. DOI: 10.3390/sym12091458.
17. Nancy P., Muthurajkumar S., Ganapathy S., Santhosh Kumar S.V.N., Selvi M., Arputharaj K. Intrusion detection using dynamic feature selection and fuzzy temporal decision tree classification for wireless sensor networks. IET Communications. 2020. vol. 14. no. 5. pp. 888–895.
18. Liu J., Gao Y., Hu F. A fast network intrusion detection system using adaptive synthetic oversampling and LightGBM. Computers and Security. 2021. vol. 106. no. 102289.
19. Al S., Dener M. STL-HDL: A new hybrid network intrusion detection system for imbalanced dataset on big data environment. Computers and Security. 2021. vol. 110. no. 102435.
20. Ifzarne S., Tabbaa H., Hafidi I., Lamghari N. Anomaly detection using machine learning techniques in wireless sensor networks. Journal of Physics: Conference Series. 2021. vol. 1743(1). no. 012021. DOI: 10.1088/1742-6596/1743/1/012021.
21. Pan J.S., Fan F., Chu S.C., Zhao H.Q., Liu G.Y. A Light-weight Intelligent Intrusion Detection Model for Wireless Sensor Networks. Security and Communication Networks. 2021. vol. 2021(2). 15 p. DOI: 10.1155/2021/5540895.
22. Zamry N.M., Zainal A., Rassam M.A., Alkhammash E.H., Ghaleb F.A., Saeed F. Light-weight Anomaly Detection Scheme Using Incremental Principal Component Analysis and Support Vector Machine. Sensors. 2021. vol. 21(23). no. 8017. DOI: 10.3390/s21238017.
23. Yadav A., Kumar A. Intrusion Detection and Prevention Using RNN in WSN. Proceedings of Inventive Computation and Information Technologies (ICICIT). 2022. pp. 531–539.

24. Tabbaa H., Ifzarne S., Hafidi I. An Online Ensemble Learning Model for Detecting Attacks in Wireless Sensor Networks. arXiv preprint arXiv:2204.13814. 2022. 15 p.
25. Salmi S., Oughdir L. CNN-LSTM Based Approach for Dos Attacks Detection in Wireless Sensor Networks. International Journal of Advanced Computer Science and Applications. 2022. vol. 13. no. 4. pp. 835–842.
26. Dener M., Al S., Orman A. STLGBM-DDS: An Efficient Data Balanced DoS Detection System for Wireless Sensor Networks on Big Data Environment. IEEE Access. 2022. vol. 10. pp. 92931–92945.
27. Almomani I., Al-Kasasbeh B., Al-Akhras M. WSN-DS: A dataset for intrusion detection systems in wireless sensor networks. Journal of Sensors. 2016. vol. 2016. 16 p. DOI: 10.1155/2016/4731953.
28. Chawla N.V., Bowyer K.W., Hall L.O., Kegelmeyer W.P. SMOTE: Synthetic minority over-sampling technique. Journal of artificial intelligence research. 2002. vol. 16. pp. 321–357.

Кришна Кураганти Пхани Рама — доцент, отдел электроники и связи, Прасад В. Потлури Технологический институт Сиддхартхи; научный сотрудник, Образовательный фонд Конеру Лакшмайи (Считается университетом). Область научных интересов: беспроводные сенсорные сети, беспроводная связь. Число научных публикаций — 17. kprkrishna007@gmail.com; Учительская колония, Бхаванипурам, 520012, Виджаявада, Индия; р.т.: +91(866)258-1699.

Тирумуру Рамакришна — Ph.D., профессор, отдел электроники и связи, Образовательный фонд Конеру Лакшмайи (Считается университетом). Область научных интересов: обработка речевых сигналов, беспроводные сенсорные сети, беспроводная связь. Число научных публикаций — 25. ramakrishnaece@kluniversity.in; Зеленые поля, Ваддесварам, 522302, Гунтур, Россия; р.т.: +91(8645)350-0200.