International Journal of Data Science and Artificial Intelligence (IJDSAI)
Volume 02, Issue 04, July – August (2024)

RESEARCH ARTICLE

CHICKEN SWARM OPTIMIZATION BASED ENSEMBLED LEARNING CLASSIFIER FOR BLACK HOLE ATTACK IN WIRELESS SENSOR NETWORK

K. Vijayan[1, *], S.V. Harish[2] and R.A. Mabel Rose[3]

[1]Department of Electronics and Communication Engineering, Sapthagiri NPS university, Banglore, Karnataka 560057, India.
[2]Department of Electronics and Communication Engineering, NIE(South), Mysuru, Karnataka 570008, VTU Affiliated College, India.
[3]Department of Computer science and Engineering, Panimalar Engineering College, Varadharajapuram, Poonamallee, Chennai, 600 123, India.

[*]*Corresponding e-mail: vijayankvijayan@gmail.com*

Abstract – Wireless Sensor Networks (WSNs) are an inevitable technology prevalently used in various critical and remote monitoring applications. The security of WSNs is compromised by various attacks in wireless medium. Even though, various attacks are present, the Black hole attack degrades the network performance and resource utilization resulting in poor network lifetime. Therefore, the proposed research suggests an effective Intrusion Detection System for WSN to detect and classify black hole attacks based on ensemble ML classifiers. The BDD dataset is used for the analysis which is subjected to Chicken Swarm Optimization based feature selection. The selected features are balanced through SMOTE and TOMEK based STL data balancing module. An ensemble of five baseline ML classifiers such as SMO, NB, J48, KNN and RF utilizing voting ensemble approach is suggested to classify the attacks in the dataset. The performance of the algorithm is analyzed through evaluation metrics such as accuracy, precision, recall and F1-score. The comparison of proposed model with six ML and DL classifiers exposes the superiority of the proposed model's classification performance.

Keywords *– WSN, Black hole attack, SMOTE, BDD dataset, Chicken Swarm Optimization, Ensemble classifier.*

1. INTRODUCTION

Since WSNs function in a limited resources environment, they differ from the typical OSI paradigm. WSN sensor nodes are grouped or clustered closely around a certain region which is defined as sensor area. These nodes are maintained remotely and have limited computational power and bandwidth. Because nodes in WSNs are frequently left unattended, an attacker can simply seize a node. Furthermore, the nodes such as sensors in WSN are susceptible for a variety of malfunctions, and the communication link is unstable. As a result, WSN security is both a tough and critical task [1]. In general, assaults are classified as active or passive. Active assaults include black holes, wormholes, flooding, and overlay network wormholes [2]. Among them, black hole attacks can significantly affect network performance and resources [3, 24]. Several approaches for detecting and preventing such assaults have been developed. A targeted feature selection technique for determining the most significant traits may be advantageous [23]. The selected characteristics can then be implemented for designing an effective learning model for classification. Moreover, the adoption of attack-specific dataset can aid in enhanced speed and accuracy of detection [25].

In the proposed study, a metaheuristic optimization-based Chicken Swarm Optimization features selection, dataset balancing and voting ensemble ML classifier are used to design a Black hole attack classification-based IDS in WSN. The suggested algorithm's classification success has been evaluated with metrics such as accuracy, recall, confusion matrix, precision and F1-score on the BDD dataset. The following are the primary contributions of the proposed research:

- A Black hole attack detection system specialized to WSNs based on classification was built in the study. In identifying WSN-specific intrusions, a voting ensemble ML strategy including Random Forest, KNN, SMO, Naive Bayes, and J48 as base classifiers.
- On the BDD dataset, feature selection was conducted using Chicken Swarm Optimization (CSO) to reduce computational complexity while increasing classification accuracy.
- For data balance, the STL link that contains oversampling with SMOTE and under sampling with Tomek-Links methods are coupled. The drawbacks of both oversampling and under sampling approaches are avoided, resulting in improved classification performance.

Section 2 provides the review of literature relevant to black hole detection, Section 3 explains the behavior of BH attack, Section 4 explains the proposed methodology and Section 5 provides the simulation and experimentation results of the study. Section 6 describes the conclusions of the study.

2. RELATED WORK

A DL based WSN black hole and wormhole attack detection framework has been designed by Pawar, M. V. in 2023. The attack classification has been performed by LSTM and Whale Optimization Algorithm based on Fitness Rate (FR-WOA) can calculate the shortest path along with Round Trip Time (RTT) validation process and Bait process [4]. To decrease BH assaults, Dhanaraj, R. K., et al. proposed an Enhanced Gravitational Search Algorithm (EGSA) module for Simulated Annealing Black-hole assault Detection (SABD) in 2021. EGSA-SABD is used to detect and isolate attacking nodes in WSN [5].

H. Kalkha suggested a Hidden Markov Model approach to recognize fraudulent nodes in WSNs by preventing black hole attacks in 2019. It proposed a novel routing method that assesses the shortest way to prevent malicious node paths [6]. Suma, S., and B. Harsoor (2022) used On-demand Link and Energy Aware Dynamic Multipath (O-LEADM) routing strategy for MANETs to identify black-hole node by incorporating bait approach to discriminate packet loss due to congestion or malicious node. While accessing the channel, the activity of the node is analyzed employing control messages reply-sequence (rep-Seq) and destination-sequence (des-Seq) [7]. Gite, P., et al. 2023 proposes a lightweight model for identifying black hole, wormhole, grey hole, and DDoS attacker nodes in a WSN with no sensor node burden and uses the C4.5 and CART classifiers (decision tree algorithms) [8]. J. Kolangiappan and A. S. Kumar (2022) proposed a blackhole attack avoidance strategy based on a Deep Belief Network (DBN) with a larger number of hidden layers. [9].

Umamaheswari, S., et al. develop an IDS to categorize WSN assaults on WSN-DS dataset using ML classifier in 2021 to assess system performance. For feature extraction, a decision tree classifier is employed, and feature selection is accomplished using Fisher Score, Correlation Score, and Kruskal-Wallis (KW) based Statistical Analysis, Relief algorithm and Minimum Redundancy Maximum Relevance (MRMR) method [10]. Pawar, M. V., and A. Jagadeesan (2021) proposed IDS for WSN that detects blackhole and wormhole threats. A Self Adaptive-Multi-Verse Optimization (SA-MVO) approach is used to extract the optimum unique characteristics. Subsequently, the best characteristics were exposed to Deep Belief Network (DBN) analysis [11].

In 2022, Tabbaa, H., et al. investigate the use of different homogeneous ensemble HAT and an Adaptive Random Forest (ARF) based heterogeneous ensemble paired to the Hoeffding Adaptive Tree (HAT) algorithm in WSN-DS dataset to recognize attack types: Grayhole, Blackhole attack, Scheduling and Flooding across WSN traffic [12]. Rezvi, M. A., and colleagues will present in 2021 a data mining approach for different kinds of classification algorithms to identify Grayhole, Blackhole, Flooding, and TDMA. Several data mining approaches, including Support Vector Machine (SVM), KNN, Nave Bayes, Logistic Regression, and ANN algorithms, are used to the dataset and their performance in identifying assaults is evaluated [13].

3. BLACK HOLE ATTACKS IN WSN

If a suspicious node enters a network and captures the transferred data traffic along the network and drops the data packets without further transmission is termed as black hole attack and it produces Denial of Service (DoS) [14]. The attacker selectively drops packets or all control and data packets routed through him in this attack. This is done in two phases by malicious nodes. First, the attacker node displays a fictitious low-rank value in order to entice neighbors to choose it as their parent (Sink hole attack) [15]. Second, it may drop selected packets depending on predetermined criteria (Selective Forwarding attack), or it might drop all packets from other nodes [16]. As a result, every packet passing via this intermediary malicious node is susceptible to partial or complete data loss. Black hole attacks have an impact on network performance. Black hole attacks have a large impact on throughput, packet delivery rate, and latency, but only a moderate impact on battery drain and control packet overhead [17].

In general, there are two ways for the malicious node to obtain the data packet in this type of attack. In first method, a Route Reply control message (RREP) message is send using the routing protocol to the source node by the malicious node as soon as a receiving a Route Request control message (RREQ) is received in order to enter the network as neighboring node having shortest path to reach destination. Such bogus route can be used by the source node to transfer data packets. Second, whether the malicious node is capable of intercepting data transfers without transmitting control message (RREP) to source node. Under both cases, the malicious node dumps the transferred data on receiving it. Consequently, the data transfer from source to destination is interrupted affecting connection and performance of network.

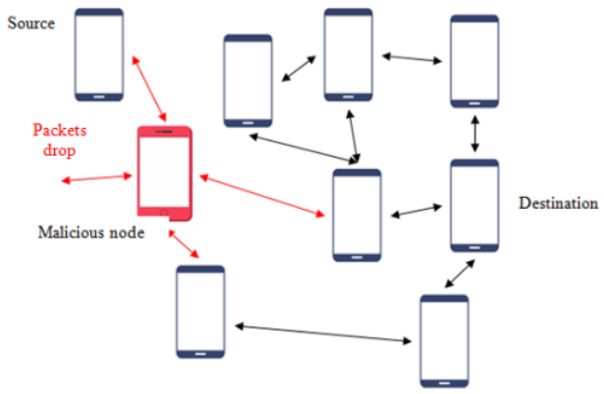

Figure 1. Illustration of black hole attack

Figure 1 illustrates the black hole attack scenario. "Source" node intends to communicate with node "Destination," thus it broadcasts "RReq" to the neighboring nodes. Node "M" injects itself and responds quickly,

claiming to have the best path. When communication begins, node "M" discards any data supplied through it.

Algorithm for Black hole Attack

Necessity: Node-ID of Attacker

If ID of Node is similar to ID of Attacker, then

Reduce the rank value

Retain higher rank parents

Discard data packets generated by nodes other than parents

else

Maintain clear calculations of rank

End if

4. PROPOSED METHODOLOGY

The proposed blackhole attack classification using ensemble ML classifier, CSO feature selection and STL data balancing approach consists of four phases. Figure 2 illustrates the representation of proposed method.

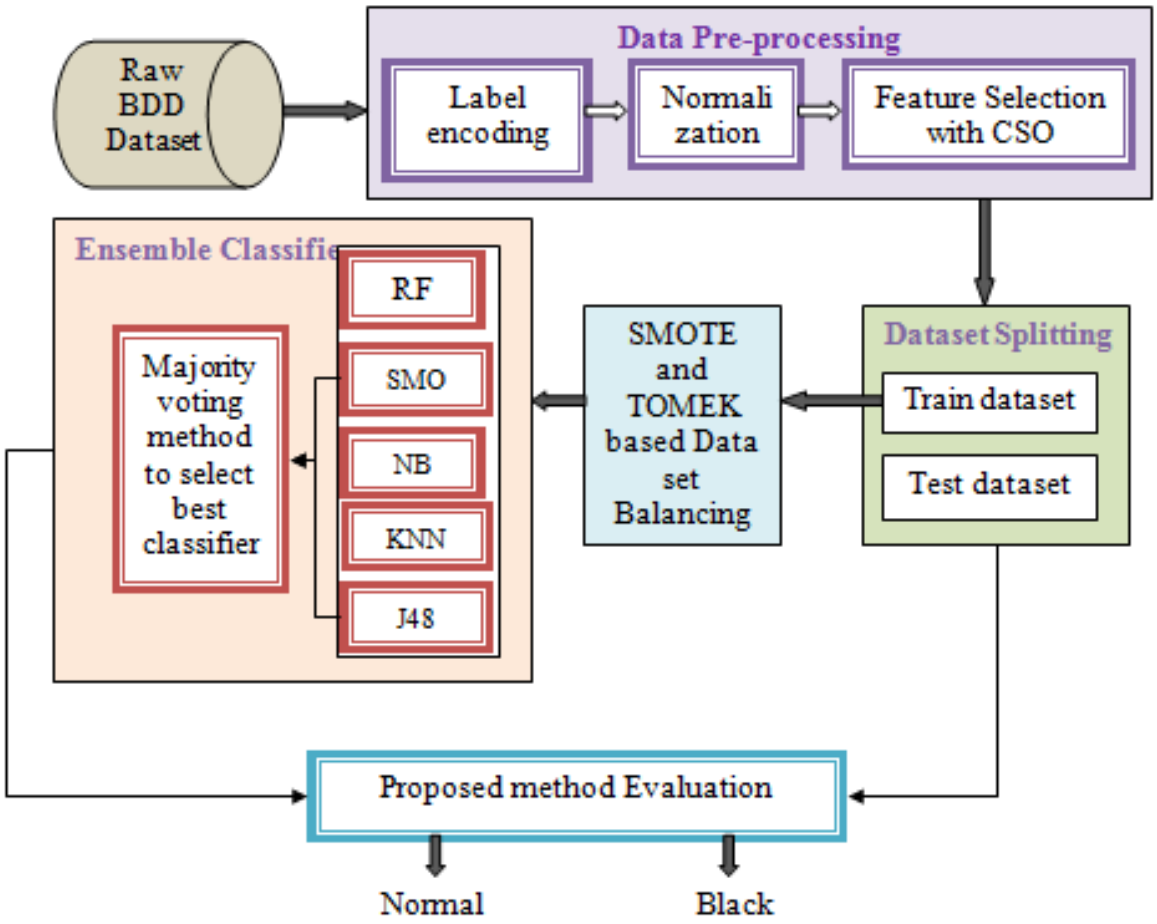

Figure 2. Overview of Proposed method

The initial phase is data preprocessing of BDD dataset. The data preprocessing consists of label encoding with one-hot encoding, normalization and feature selection. In this work, the Chicken Swarm Optimization (CSO) method was utilized for selecting the ideal feature combination to maximize classification performance while minimizing the number of chosen features. The next phase is data splitting into training and testing dataset. The third phase is data balancing to mitigate the imbalance problems in dataset. The final phase is classification based on ensemble ML classification with voting ensemble approach.

4.1. Dataset

The behavior of MANETs in the presence of black hole assaults is investigated in order to build the Behavior-Driven Development (BDD) dataset. This is designed especially for black hole attacks with labels. [18]. GloMoSim 2.03 simulator was utilized to collect audit data by simulating regular and black hole attack situations. There are 29 characteristics in the BDD dataset. Moreover, there are 1289 instances/nodes in dataset having 1189 normal nodes and 100 black nodes. Figure 3 depicts the characteristics observed in the BDD dataset.

4.2. Normalization and Label Encoding

The raw BDD dataset is subjected to One-hot encoding process to convert the categorical values as numerical values before proceeding to classification algorithm. The mathematical representation of normalization is explained in equation 1. It converts the numerical entries in the dataset into a value among 0 and 1.

$$x^n = \frac{x-\overline{x}}{\emptyset} \quad (1)$$

where x represents the actual entry, x^n represents the normalised value, $\overline{x}$ represents the mean and $\emptyset$ represent the standard deviation, accordingly. This normalization process reduces entries having high numerical value without affecting the classification output and network performance. This one hot encoding procedure alters labels/classes as value.

Total RREQ transmitted	Total RREP transmitted	Total RREP forwarded	RREP transmitted with route	Average Nodal speed
Total RERR transmitted	Total RERR retransmitted	Total selected routes	Total transmitted data	Total generated data
Total received data	Total discarded data	Total hop counts	Total control packets transmitted	Total retries of broken link
Total number of broken links	Total received RREQ	Total received RREP	Total data packets in node	Total low hop counts to destination
Total high hop counts to destination	Intermediate hops count	No. of Nodes transmitting maximum data	Highest number of reply from node	Total number of bytes transmitted
Total number of bytes received	Nodes functioning as source	Nodes functioning as destination	Fast replying AVG	

Figure 3. Features in BDD Dataset

4.3. Feature Selection

The Features Selection Algorithm may be used to locate the most important and trustworthy characteristics in finding and preventing black hole attacks before incorporating the entire BDD dataset into the classifier. Using the feature selection procedure resulted in a reduction in model complexity and computation time, which helped in designing the optimal learning model.

CSO Algorithm

CSO algorithm has three classifications of roles: roosters, hens, and chicks, each with its own set of behavior criteria. The following are the fundamental preliminaries of CSO algorithm.

i. A chicken swarm is divided into groups each having a rooster, few chickens and more hens.

ii. Roosters, chickens and hens have unique characteristics mentioned by corresponding fitness score. Based on which, the chicks are worst, roosters are best and hens are inter mediate. Each hen picks one rooster at random as her mate and joins his group, and each chick chooses one hen at random as its mother.

iii. For iterative cycle of G generations, the unique identities, maternal and spouse relationships remain unaltered. After G generations all these values are updated.

iv. Hens follow their partner rooster to locate food in each group of the entire population and thrive to get food at random among other members in a group. Members with high fitness can get food.

The position of each chicken describes its location. Let $R_n,\ C_n,\ H_n, and\ M_n$ denote the number of rooster, chicken, hen and mother hens correspondingly. Let the position of i^{th} chicken in the t^{th} on the j^{th} dimensional space be $x_{i,j}^t$. The value of t lies between$\{1,2, \dots, T\}$, j lies between $\{1,2,3, \dots J\}$ and i lies between $\{1,2, \dots I\}$ in which the number of chickens is represented by I, the dimension number is represented by J and the maximum iterations are denoted by T. Individual location update formula exists for a chicken, a hen and a rooster. The rooster's recurrent location can be mathematically represented as,

$$x_{i,j}^{t+1} = (Randn(0, \emptyset^2) + 1) \quad (2)$$

$$\emptyset^2 = \begin{cases} 1, & if\ f_i \le f_k \\ \exp\left(\frac{f_k - f_i}{\varepsilon + |f_i|}\right) & otherwise, k \in [1, NR], where\ k \neq i \end{cases} \quad (3)$$

In the above equation, $(Randn(0, \emptyset^2)$ is the random number having $\emptyset^2$ variance and zero expectation following Gaussian distribution. k denotes the number of randomly selected rooster, ε denotes a constant of small value, the fitness value of k^{th} and i^{th} rooster is f_k and f_i accordingly. The hen's recurrent location can be denoted mathematically as,

$$x_{i,j}^{t+1} = Rand * C_1 * (x_{r1,j}^t - x_{i,j}^t) + Rand * C_2 * (x_{r2,j}^t - x_{2,j}^t) x_{i,j}^t \quad (4)$$

The random number $Rand$ lies between [0,1] in the uniform distribution, the learning factors denoted by C_1 and C_2. The hen's recurrent location is mathematically represented as,

$$x_{i,j}^{t+1} = RF * \left(x_{m,j}^t - x_{i,j}^t\right) + x_{i,j}^t \quad (5)$$

The term RF denotes the random factor in the range [0,2], $x_{m,j}^{t}$ denotes the mother hen.

Input: parameters N, T, NR, NH, NC, NM, G

Output: Selection of best features

Steps:

(1) Randomly assign locations to chickens.

(2) Each chickens' fitness score is calculated; the local and global optimal location of each chicken is selected. Iteration time is set to $t = 1$.

(3) If $t\%G = 0$ (% is the remainder function), then based on descending order of fitness score, chickens are arranged. Roosters, NR have best values, NC chicks have worst value and NH hens are others. The swarm is classified as groups having a rooster, various hens and chicks in which moms and spouses are randomly selected.

(4) The positions of roosters, chicks and hens are updated using formulae (2), (4), and (5), and fitness values are evaluated.

(5) The population's global best location and each individual's local best position are updated.

(6) Iterate $t = t + 1$ times; if t equals M or the solution meets the accuracy criteria, CSO outputs the final result; otherwise, go to Step 3.

CSO based Feature Selection

The following four characteristics are identified as the most significant features based on the analysis results: (1) Total RREQ features transmitted. (2) Total RREPs with forward feature. (3) High destination sequence number features. (4) A low number of hops to the destination feature. Furthermore, despite the fact that they do not have a very significant fitness value, the findings of the evaluation of these two features indicate that they're capable of an important impact in prevention and detection of BH attack, and they are: (5) The total acts that serve as the source feature. (6) Act as a destination feature. The BDD dataset after CSO based feature selection has six features alone that are most significant for classification.

4.4. Dataset Splitting

During the dataset splitting step, the BDD dataset is separated into a training dataset and a testing dataset. The training step employs a labeled training set to train a particular classifier, which is then utilized in the testing phase to categorize test instances as black or normal.

4.5. Data Balancing

The unequal distribution of classes has a detrimental impact on categorization performance. Minority groups, in particular, have a detrimental impact on the detection rate. IDS designs cannot adequately identify the class imbalance issue in dataset. The individual use of undersampling approach for class imbalance leads to removal of important data transmitted and reduces data quality significantly. The use of oversampling approach results in unwanted noise and data volume increase. To address the unbalanced class problem, Synthetic Minority Oversampling Technique (SMOTE) oversampling and Tomek-Links undersampling approaches named as STL approach is presented in this work.

SMOTE

SMOTE was suggested by Chawla et al. [19], a heuristic oversampling approach, to overcome the class imbalance issue in datasets. Minority class data are oversampled to generate synthetic data in this approach. The overfitting issue is also eliminated with the generated synthetic data. The overfitting problem of class imbalance can be reduced with random non-metaheuristic sampling approach which is widely used recently [20]. The class imbalance issue is eliminated with SMOTE through increasing the number of minorities labeled instances along with its neighbors. Samples in proximity to the feature space are selected by SMOTE. An arbitrary instance is selected from the minority label and its proximal neighbor's k is identified. A neighbor is selected at random and the variation among both samples is multiplied with a value from 0 to 1 which is combined with randomly selected sample value. The line created across the two sample attributes is then used to create synthetic samples. Randomly selected neighbors from k proximal neighbors determine the required oversampling quantity.

The linear groupings of two minority labeled identical samples $(X, X^r)$are mathematically represented as,

$$m = x + g.(X^r - X), \quad 0 \le g \le 1 \tag{6}$$

The sample X was selected randomly as X^rin correspondence to the nearest proximal number having difference g among two instances.

TOMEK-LINKS

Tomek-Links is a Tomek-developed approach for undersampling unbalanced datasets. It may be thought of as an enhanced variant of the Nearest Neighbour Rule. The specimens on the Tomek link can be deleted from the provided dataset using this method [21]. It generates sample data pairs within the same dataset and from separate labels. These paired data are referred to as Tomek linkages [22]. Its primary aim was to segregate the majrotiy and minority labels. Let u, v be the proximal neighbors and u belong to a class and v belongs to another class. The distance among instances u and v is represented as,

$$Tomex\ link\ be\ T(u,v), for\ any\ value\ of\ i, * d(u,v) < d(u,i) ord(u,v) < d(v,i) \tag{7}$$

T-links connect the two classes. This link's data samples are deemed noise. The removal of majority class noises improves class separation and stabilizes the data distribution. Thus, the noise instances are eliminated from the majority labels.

4.6. Ensemble classifier to detect Black hole attack

This module compares the test data to the network's typical profile using a predetermined classifier to determine if the data is normal or malicious. If there is any divergence from the network's typical behavior, the incident is classified as an attack. Otherwise, it is seen as normal. Classification is

the process of learning a model (classifier) from a training module with labeled data to categorize test data as labels. Ensemble learning is a popular ML approach employed for the categorization and detection of WSN-attacks. These strategies are often built by solving the identical problems with different ML classifiers and combining the results with one of the voting procedures. It combines many basic models to create one optimal prediction model. Bagging, boosting, stacking, and majority voting are examples of ensemble procedures. We describe an ensemble-based ML technique for detecting WSN-attacks by combining a number of different base models. To distinguish the most effective classifier for detecting assaults, many classifiers have been trained to test the evaluation measures. This suggested method was developed utilizing a collection of heterogonous supervised ML approaches, including SMO, NB, J48, KNN, and RF. Ensemble methods are commonly used in machine learning approaches.

Sequential Minimal Optimization (SMO) is a novel training approach for Support Vector Machines (SVM). SMO was created in order to address the Quadratic Programming (QP) SVM training issue. A Naive Bays (NB) algorithm is a basic probabilistic algorithm for classification which uses the Bayes' theorem to categorize a fresh occurrence based on robust independent assumptions about the characteristics. J48 classifier was a basic incorporation of the C4.5 decision tree technique utilizing the training set's attribute values to generate a binary tree. A non-parametric K-Nearest Neighbor (K-NN) supervised technique for regression and classification issues that may quickly determine the category or class of a given dataset. Random Forest algorithm is a classification process on dataset using various decision trees on various data sub-groups. The prediction accuracy of provided dataset can be enhanced through averaging them. In this study, the majority voting approach is used to combine predictions from many other models. When employing majority voting, performance can be improved over when using a single model. When all models perform equally well, the voting ensemble approach is utilised. As basic models, SMO, NB, J48, KNN, and RF are employed. Different classifiers combine using the majority vote approach in majority voting. The greatest likelihood of the chosen class determined the final forecast. Figure 4 depicts the ensemble voting mechanism.

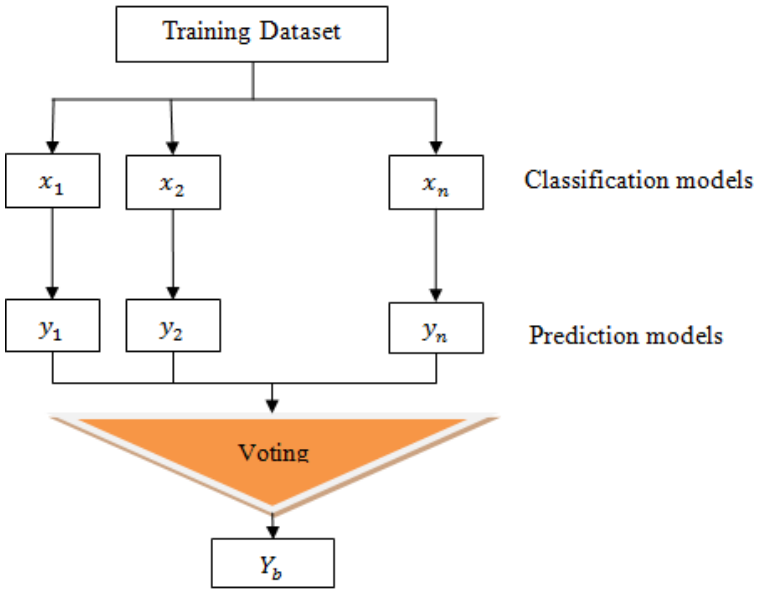

Figure 4. Voting based Ensemble method

Individual learners SMO, NB, J48, KNN, and RF, as well as majority voting, are implemented on the BDD dataset utilizing feature selection and data balance in the suggested technique. The results of both methods are then compared using the criteria accuracy, precision, and recall.

5. SIMULATION RESULTS

The suggested approach has been evaluated on the unbalanced BDD dataset in this part. The research is carried out in Python programming language with Pyspark tool, using Google Colab framework in Apache Spark environment. Keras and Scikit-learn libraries of PySpark MLib tool has been employed for ensemble, ML and DL techniques. The suggested technique was tested against six various ML and DL algorithms, and the results were analyzed.

5.1. Evaluation Metrics

The study uses the famous metrics such as accuracy, precision, recall and F1-sore. Many classification problems make use of these assessment factors. The confusion matrix data is used to calculate these parameters. The confusion matrix's primary constituents are true-positive (t_p), false-positive (f_p), true-negative (t_n), and false-negative (f_n). Accuracy defines the percentage of rightly identified samples. Precision can be termed as the ratio of the number of retrieved relevant instances to the number of retrieved samples (relevancy and non-relevancy). Recall is the ratio of rightly categorized nodes as black to the total black nodes in the data set. A harmonic mean of Recall and Precision is F-score. The mathematical formula for accuracy, precision, recall and F-score is explained below in equations.

$$Accuracy = \frac{t_p+t_n}{t_p+f_p+t_n+f_n} \quad (8)$$

$$Recall = \frac{t_p}{f_n+t_p} \quad (9)$$

$$Precision = \frac{t_p}{t_p+f_p} \quad (10)$$

$$f1-score = \frac{2t_p}{\frac{t_p}{2t_p+f_n+f_p}} \quad (11)$$

5.2. Performance Analysis

The performance of the proposed ensemble ML classifier has been examined based on the above evaluation metrics. The values of accuracy, precision, recall and F1-score for the baseline ML models such as SMO, NB, J48, RF and Voting Ensemble classifier are calculated and its numerical values are illustrated in Table 1.

Table 1. Performance Evaluation of proposed model

Models	Accuracy	Recall	Precision	F1-score
SMO	0.9876	0.9958	0.9907	0.9932
NB	0.9930	0.9983	0.9941	0.9962
J48	0.9961	0.9992	0.9966	0.9979
KNN	0.9899	0.9924	0.9966	0.9945
RF	0.9977	0.9992	0.9983	0.9987
Ensemble Classifier with Majority Voting	**0.9953**	**0.9983**	**0.9966**	**0.9975**

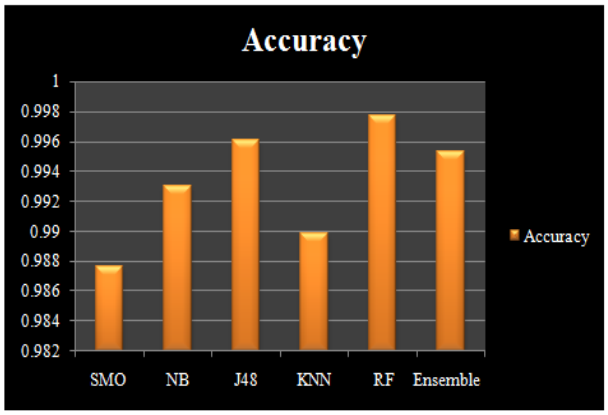

Figure 5. Accuracy Comparison of baseline models and ensemble classifier

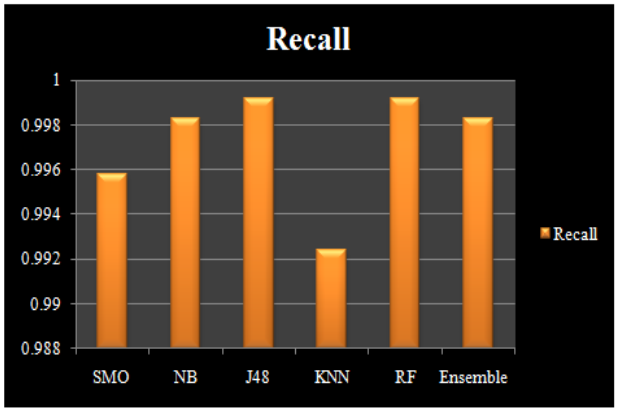

Figure 6. Recall Comparison of baseline models and ensemble classifier

As depicted in Table 1, the RF classifier has the highest accuracy of 0.9977, recall of 0.9992, precision of 0.9983 and F1-value of 0.9987 among other baseline classifiers. The ensemble classifier produces values of 0.9953 accuracy, 0.9983 recall, 0.9966 precision and 0.9975 F1-score respectively. The graphical representation of comparison of accuracy of five base models and proposed model is illustrated in Figure 5. Similarly Figure 6, 7 and 8 represent the graphical comparison of recall, precision and F1-score values respectively.

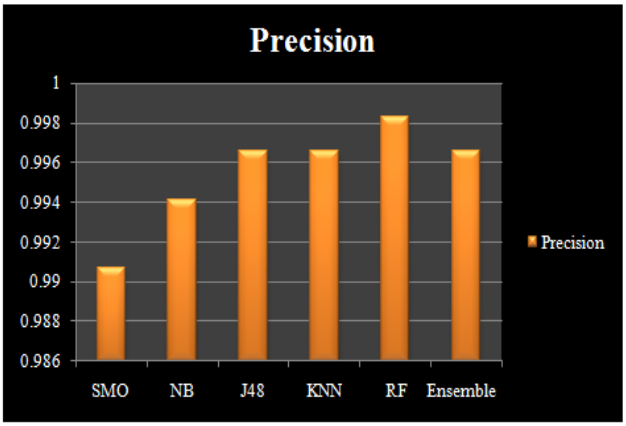

Figure 7. Precision Comparison of baseline models and ensemble classifier

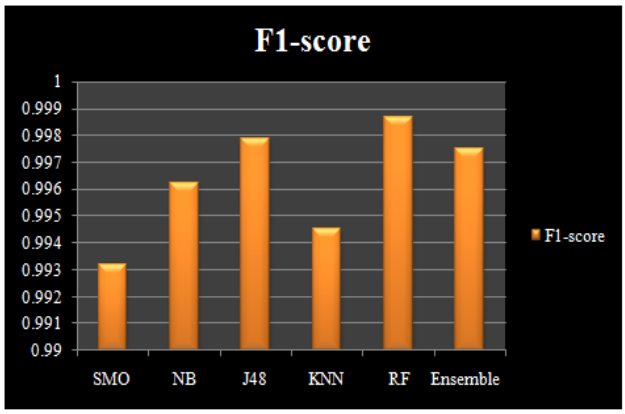

Figure 8. F1-score Comparison of baseline models and ensemble classifier

Table 2. Confusion matrix for classification baseline models and ensemble classifier

	RF		KNN		NB		J48		SMO		Ensemble	
Actual vs. Predicted	**Normal**	**Black**	**Normal**	**Black**	**Normal**	**Black**	**Normal**	**Black**	**Normal**	**Black**	**Normal**	**Black**
Normal	1184	2	1178	4	1181	7	1185	4	1177	11	1183	4
Black	1	100	9	98	2	99	1	99	5	96	2	100

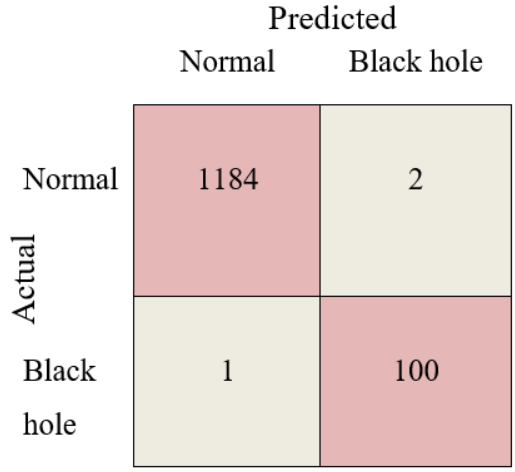

Figure 9. Confusion matrix of RF classifier

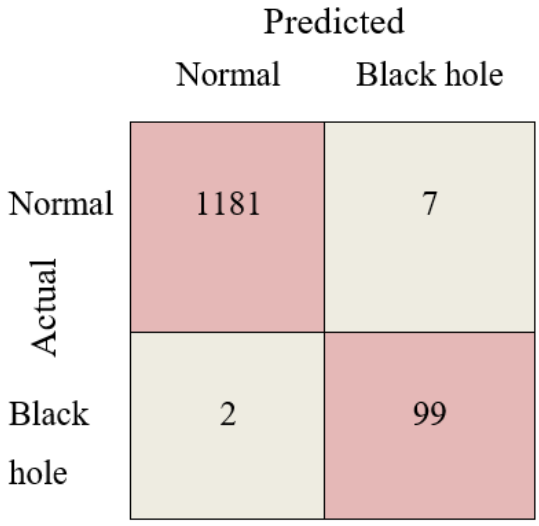

Figure 10. Confusion matrix of NB classifier

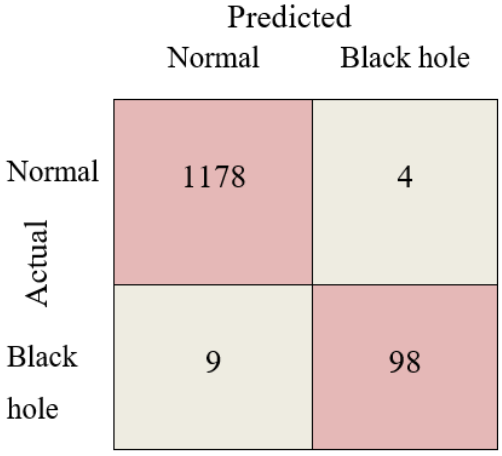

Figure 11. Confusion matrix of KNN classifier

The value of actual classes to the predicted classes are explained in Table 2 and from its interpretation, the RF classifier has the best classification performance among the comparable baseline models.

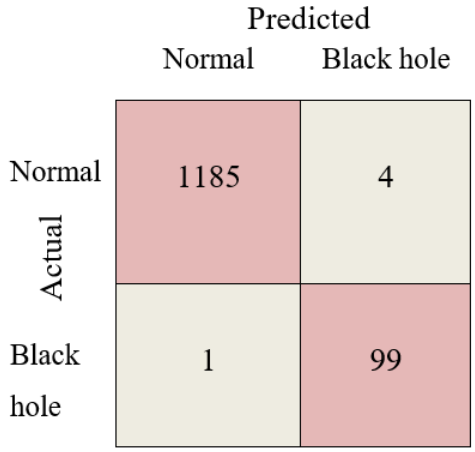

Figure 12. Confusion matrix of J48 classifier

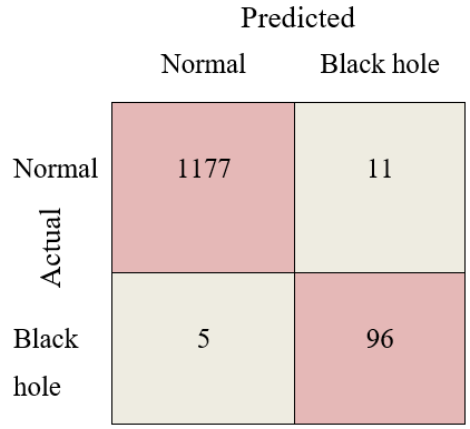

Figure 13. Confusion matrix for SMO classifier

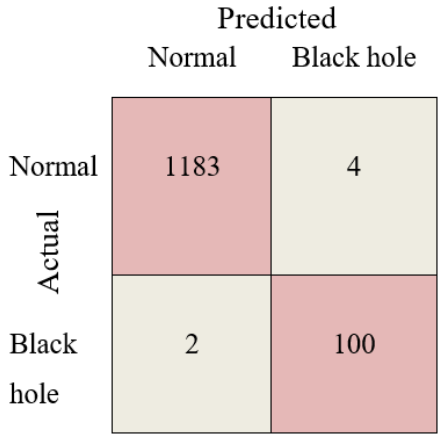

Figure 15. Confusion matrix for Ensemble classifier

5.3. Comparative Analysis

The effective performance of the proposed model can be analyzed through evaluation with other similar existing models. In the study, the accuracy, precision, recall and F1-score values are compared with three ML methods and three DL methods in order to examine the proposed model's classification performance.

Table 3. Comparison of Performance of various models with proposed ensemble classifier

Models	Accuracy	Recall	Precision	F1-score
SVM	0.9818	0.9965	0.9845	0.9905
Logistic Regression	0.9672	0.7181	0.9099	0.8028
ANN	0.9856	0.9124	0.9066	0.90954
CNN	0.9879	0.9297	0.9486	0.9372
DNN	0.9704	0.8201	0.8280	0.8208
RNN	0.9648	0.6911	0.8562	0.7537
Proposed	0.9953	0.9983	0.9966	0.9975

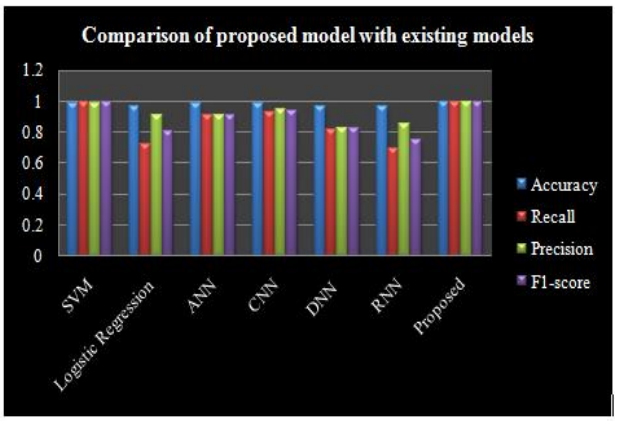

Figure 16. Comparison of proposed model with existing approaches

Table 3 provides the values of performance metrics of ML models such as Support Vector Machine (SVM) [13], Logistic Regression and Artificial Neural Network (ANN) [14] along with three DL models such as Deep Neural Network (DNN), Convolutional Neural Network (CNN) and Recurrent Neural Network (RNN) [15]. The graphical illustration of the performance comparison of the above models with proposed ensemble classifier is explained in Figure 16. From the results, the proposed model shows the highest accuracy, precision, recall and F1-score values of 0.9953, 0.9983, 0.9966 and 0.9975 respectively. Apart from this, the CNN model has the highest classification accuracy of 0.9879 among DL models and ANN model exhibit 0.9856 accuracy among ML models. In terms of recall, precision and F1-score, the SVM has highest value of 0.9965, 0.9845 and 0.9905 respectively followed by CNN with0.9297 recall, 0.9486 precision and 0.9372 F1-score respectively.

6. CONCLUSIONS AND RECOMMENDATIONS

The study proposes a classification-based IDS in WSN to detect black hole attack. The BDD dataset is used for the analysis. The dataset is preprocessed with label encoding and normalization followed by feature selection. The optimal features suitable for classification can be viewed as an optimization problem and can be selected with the swarm-based Chicken Swarm Optimization (CSO) approach. The CSO algorithm reduces the BDD dataset with 29 features into 6 features most significant for attack classification. The dataset with selected features are spitted into training and testing dataset. The training dataset is fed into data balancing module which eliminates the class imbalance problem in dataset using SMOTE and TOMEK Link (STL) approach for upsampling and downsampling. The ensemble classifier with five base ML classifiers such as RF, KNN, J48, NB and SMO are utilized and combined with voting ensemble method. The performance of the proposed model has been analyzed with accuracy, precision, recall and F1-score. The comparative analysis of the ensemble classifier with three ML and three DL models illustrates the superiority of the proposed classification approach. In future, the proposed study can be enhanced with adapting fuzzy based feature selection techniques. Moreover, the attack classification approach can be extended to detect other attack types such as wormhole attack, grayhole attack and flooding.

CONFLICTS OF INTEREST

The authors declare that there is no conflict of interest.

FUNDING STATEMENT

Not applicable.

ACKNOWLEDGEMENTS

The author would like to express his heartfelt gratitude to the supervisor for his guidance and unwavering support during this research for his guidance and support.

REFERENCES

[1] U. D. Maiwada, A. A. Muazu and N. Noor, "The Security Paradigm That Strikes a Balance Between a Holistic Security Mechanism and The WSN's Resource Constraints", *East Asian Journal of Multidisciplinary Research*, vol. 1, no. 3, pp. 343-352, 2022. [CrossRef] [Google Scholar] [Publisher Link]

[2] M. Keerthika and D. Shanmugapriya, "Wireless sensor networks: Active and passive attacks-vulnerabilities and countermeasures", *Global Transitions Proceedings*, vol. 2, no. 2, pp. 362-367, 2021. [CrossRef] [Google Scholar] [Publisher Link]

[3] G. Farahani, "Black hole attack detection using K-nearest neighbor algorithm and reputation calculation in mobile ad hoc networks", *Security and Communication Networks*, vol. 2021, pp. 1-15, 2021. [CrossRef] [Google Scholar] [Publisher Link]

[4] M. V. Pawar, "Detection and prevention of black-hole and wormhole attacks in wireless sensor network using optimized LSTM", *International Journal of Pervasive Computing and*

Communications, vol. 19, no. 1, pp. 124-153, 2023. [CrossRef] [Google Scholar] [Publisher Link]

[5] R. K. Dhanaraj and R. H. Jhaveri, L. Krishnasamy, G. Srivastava, & P. K. R. Maddikunta, "Black-hole attack mitigation in medical sensor networks using the enhanced gravitational search algorithm", *International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems*, vol. 29, no. Suppl 2, pp. 297-315, 2021. [CrossRef] [Google Scholar] [Publisher Link]

[6] H. Kalkha, H. Satori and K. Satori, "Preventing black hole attack in wireless sensor network using HMM", *Procedia computer science*, vol. 148, pp. 552-561, 2019. [CrossRef] [Google Scholar] [Publisher Link]

[7] S. Suma and B. Harsoor, "An approach to detect black hole attack for congestion control utilizing mobile nodes in wireless sensor network", *Materials Today: Proceedings*, vol. 56, pp. 2256-2260, 2022. [CrossRef] [Google Scholar] [Publisher Link]

[8] P. Gite, K. Chouhan, K. M. Krishna, C. K. Nayak, M. Soni and A. Shrivastava, "ML Based Intrusion Detection Scheme for various types of attacks in a WSN using C4. 5 and CART classifiers", *Materials Today: Proceedings*, vol. 80, pp. 3769-3776, 2023. [CrossRef] [Google Scholar] [Publisher Link]

[9] J. Kolangiappan and A. S. Kumar, "A novel framework for the prevention of black-hole in wireless sensors using hybrid convolution network", 2022. [CrossRef] [Google Scholar] [Publisher Link]

[10] S. Umamaheswari, K. H. Priya and D. Allinjoe, "Towards Building Robust Song and Location Detection System Using LBP Features", In 2021 *International Conference on Advancements in Electrical, Electronics, Communication, Computing and Automation (ICAECA)*, pp. 1-6, 2021. [CrossRef] [Google Scholar] [Publisher Link]

[11] M. V. Pawar and A. Jagadeesan, "Detection of blackhole and wormhole attacks in WSN enabled by optimal feature selection using self-adaptive multi-verse optimiser with deep learning", *International Journal of Communication Networks and Distributed Systems*, vol. 26, no. 4, pp. 409-445, 2021. [CrossRef] [Google Scholar] [Publisher Link]

[12] H. Tabbaa, S. Ifzarne and I. Hafidi, "An online ensemble learning model for detecting attacks in wireless sensor networks", arXiv preprint arXiv:2204.13814, 2022. [CrossRef] [Google Scholar] [Publisher Link]

[13] M. A., Rezvi, S. Moontaha, K. A. Trisha, S. T. Cynthia and S. Ripon, "Data mining approach to analyzing intrusion detection of wireless sensor network", Indonesian J. Electric. Eng. Comput. Sci, vol. 21, no. 1, pp. 516-523, 2021. [CrossRef] [Google Scholar] [Publisher Link]

[14] A. Kumar, V. Varadarajan, A. Kumar, P. Dadheech, S. S. Choudhary, V. A. Kumar and K. C. Veluvolu, "Black hole attack detection in vehicular ad-hoc network using secure AODV routing algorithm", *Microprocessors and Microsystems*, vol. 80, pp. 103352, 2021. [CrossRef] [Google Scholar] [Publisher Link]

[15] P. P. Ioulianou, V. G. Vassilakis and S. F. Shahandashti, "A trust-based intrusion detection system for RPL networks: Detecting a combination of rank and blackhole attacks", *Journal of Cybersecurity and Privacy*, vol. 2, no. 1, pp. 124-153, 2022. [CrossRef] [Google Scholar] [Publisher Link]

[16] S. Singh and H. S. Saini, "Learning-based security technique for selective forwarding attack in clustered WSN", *Wireless Personal Communications*, vol. 118, no. 1, pp. 789-814, 2021. [CrossRef] [Google Scholar] [Publisher Link]

[17] G. Farahani, "Black hole attack detection using K-nearest neighbor algorithm and reputation calculation in mobile ad hoc networks", *Security and Communication Networks*, vol. 2021, pp. 1-15, 2021. [CrossRef] [Google Scholar] [Publisher Link]

[18] Y. Khamayseh, M. B. Yassein and M. Abu-Jazoh, "Intelligent black hole detection in mobile AdHoc networks", *International Journal of Electrical and Computer Engineering*, vol. 9, no. 3, pp. 1968, 2019. [CrossRef] [Google Scholar] [Publisher Link]

[19] N. V. Chawla, K. W. Bowyer, L. O. Hall and W. P. Kegelmeyer, "SMOTE: synthetic minority over-sampling technique", *Journal of artificial intelligence research*, vol. 16, pp. 321-357, 2002. [CrossRef] [Google Scholar] [Publisher Link]

[20] 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. [CrossRef] [Google Scholar] [Publisher Link]

[21] R. M. Pereira, Y. M. Costa and C. N. Silla Jr, "MLTL: A multi-label approach for the Tomek Link undersampling algorithm", *Neurocomputing*, vol. 383, pp. 95-105, 2020. [CrossRef] [Google Scholar] [Publisher Link]

[22] M. Kamaladevi, V. Venkataraman and K. R. Sekar, "Tomek link Undersampling with Stacked Ensemble classifier for Imbalanced data classification", *Annals of the Romanian Society for Cell Biology*, pp. 2182-2190, 2021. [CrossRef] [Google Scholar] [Publisher Link]

[23] A. I. Al-issa, M. Al- Akhras, M. S. ALsahli and M. Alawairdhi, "Using machine learning to detect DoS attacks in wireless sensor networks", In 2019 *IEEE Jordan International Joint Conference on Electrical Engineering and Information Technology (JEEIT)*, pp. 107-112, 2019. [CrossRef] [Google Scholar] [Publisher Link]

[24] M. A. Rezvi, S. Moontaha, K. A. Trisha, S. T. Cynthia and S. Ripon, "Data mining approach to analyzing intrusion detection of wireless sensor network", *Indonesian J. Electric. Eng. Comput. Sci*, vol. 21, no. 1, pp. 516-523, 2021. [CrossRef] [Google Scholar] [Publisher Link]

[25] S. Salmi and L. Oughdir, "Performance evaluation of deep learning techniques for DoS attacks detection in wireless sensor network", *Journal of Big Data*, vol. 10, no. 1, pp. 1-25, 2023. [CrossRef] [Google Scholar] [Publisher Link]

AUTHORS

K. Vijayan completed his PhD from Vels Institute of Science, Technology and Advanced Studies (VISTAS) formerly known as VELS University. His research interest includes Wireless sensor networks, Internet of things, VLSI circuits and systems, Machine learning, routing and networking in wireless sensor networks. He has completed his Masters in engineering in VLSI design from college of engineering Guindy - Anna University in 2003. He has 21 years of teaching experience and currently working as Professor in Department of Electronics and communication Engineering Sapthagiri NPS University. He has supervised 15 post graduate students and 45 undergraduate students for their academic projects. He is having two international patents and two national patents. He has published more than 25 papers in international journals and 7 papers in international conferences. He is also a life member of ISTE, IETE, SESI, IAENG and IACSIT.

S.V. Harish received his Post Graduate in 2010 from VTU, Belagavi. He is a assistant professor in the dept of Electronics and communication engineering NIE(South), Mysuru. He has 26 years of experience in academics 11 years in research field. His area of research interest are Wireless sensor networks, communication protocols and power electronics.

R.A. Mabel Rose received her B.E. degree and M.E. degree in Computer Science and Engineering from Anna University, Chennai, India. She started her career as Lecturer and has 13 years of experience. Currently she is working as Assistant Professor in Panimalar Engineering College, Chennai. Her research interests include Cyber Security and Cloud Computing. She is a lifetime member of ISTE.

Arrived: 16.07.2024
Accepted: 21.08.2024