Mitigating DDoS Attacks in Wireless Sensor Networks using Heuristic Feature Selection with Deep Learning Model

Abdul Rahaman Wahab Sait[1], Irina Pustokhina[2], M. Ilayaraja[3]

[1] King Faisal University, Kingdom of Saudi Arabia

[2] Plekhanov Russian University of Economics, Moscow, Russia

[3] Department of Computer Science and Information Technology, Kalasalingam Academy of Research and Education, Krishnankoil, India

Emails: asait@kfu.edu.sa ; ivpustokhina@yandex.ru, ilayaraja.m@klu.ac.in

Abstract

A wireless sensor network (WSN) encompasses a massive set of sensors with limited abilities for gathering sensitive data. Since security is a significant issue in WSN, there is a possibility of different types of attacks. In Distributed Denial of Service (DDOS) attack, the malicious node can adapt to several attacks, namely flooding, black hole, warm hole, etc., to interrupt the working of the WSN. The recently developed deep learning (DL) models can effectively detect DDoS attacks in the network. Therefore, this article proposes a heuristic feature selection with a Deep Learning-based DDoS (HFSDL-DDoS) attack detection model in WSN. The proposed HFSDL-DDoS technique intends to identify and categorize the occurrence of DDoS attacks in WSN. In addition, the HFSDL-DDoS technique involves the immune clonal genetic algorithm (ICGA) based feature selection (FS) approach to improve the detection performance. Moreover, a fruit fly algorithm (FFA) with bidirectional long, short-term memory (BiLSTM) based classification model is employed. The experimental analysis of the HFSDL-DDoS technique is performed, and the results are examined interms of several performance measures. The resultant experimental results pointed out the betterment of the HFSDL-DDoS technique over the other techniques.

Keywords: WSN, Security, DDoS attacks, Deep learning, Feature selection, Metaheuristics.

1. Introduction

Wireless sensor network (WSN) represents specific classes of ad-hoc networks [1]. They are composed of multiple smart sensors of smaller size, constrained energy, at lower cost, and multi-tasking (termed as nanocomputer). Theoretically, this network node has spontaneous modes of the organization since they are designed to be placed arbitrarily and quickly in a domain region. They are controlled by a power unit (battery) of constrained capability. They could acquire (or collect) a physical quantity from the environments like wind speed, relative humidity, temperature, etc. Also, they can detect realtime events, process data, and interact with one another to bring the data gathered to a set point named Base Station (BS)/sink node [2]. Then, the data is transferred through transport networks to a processing center, wherein feasible decision-making analyzes and interpretations were performed with an end-user. WSN is becoming a hot research topic because of its broad spectrum of real-time applications such as battlefields, crucial military surveillance, forest fire monitoring, healthcare, and building security monitoring [3]. The structure of this application assumes which each node included is trustworthy and cooperative. But this isn't the case in realtime deployment. At the same time, the node is revealed with

DOI: https://doi.org/10.54216/JCIM.000106

distinct kinds of attack and intrusion, which could harm the appropriated working of the networks and degrade method performances.

Inappropriately, guaranteeing the safety of this kind of network against several malicious attack events is a tedious process, mainly while the node is composed of low-cost electronic devices with constrained hardware capacities [4]. The cryptographic algorithm requires necessary processing, memory, and energy consumption. Generally, cryptographic and authentication algorithm provides the service of authentication, confidentiality, and integrity, even though, using a security level management and encryption algorithm, it is complex for guarantying that the information is genuine and didn't suffer another kind of attacks which extract sensitive information, and utilize them for a malicious purpose [5]. While the cryptographic technique solution was made for reducing the cyberattacks, however, they haven't removed them entirely. Then, detection-based methods are presented for protecting WSN from popular and novel cyberattacks as second-line protection.

Protecting and providing the privacy of this information is a significant problem since there are many frightening the occurrence of this network. This attack is categorized into denial-of-service (DoS), communication, protocol-specific attack, node compromise, and impersonation. Amongst other, DoS attack tries to prevent the transmission of the sensor by avoiding more than one network which implements routing function [6]. While beneficial data distribution among the nodes is hindered, the networks could not serve their purposes. Thus, it is essential for the WSN models to analyze and detect DoS attacks appropriately, also for taking privacy measures beforehand facing the attacks. DoS attack is performed at every five layers of the TCP/IP protocol stack [7]. Even though a different type of DoS attack exists at every layer, the network layer DoS attack has the maximum diversity and number in this study.

This article proposes a heuristic feature selection with Deep Learning-based DDoS (HFSDL-DDoS) attack detection model in WSN. The proposed HFSDL-DDoS technique intends to identify and categorize the occurrence of DDoS attacks in WSN. In addition, the HFSDL-DDoS technique involves the immune clonal genetic algorithm (ICGA) based feature selection (FS) approach to improve the detection performance. Moreover, a fruit fly algorithm (FFA) with bidirectional long short term memory (BiLSTM) based classification model is employed. The experimental analysis of the HFSDL-DDoS technique is performed, and the results are examined interms of several performance measures.

2. Existing Works on DDoS Attack Detection

Deepa et al. [8] presented the hybrid ML technique for protecting the controller in a DDoS attack. Experimental outcomes manifest which the hybrid ML technique gives further accuracy, detection rate, and less false alarm rate related to easy ML techniques. Yang and Zhao [9] introduce an SDN structure for identifying and defending DDoS attacks dependent upon ML. This structure has three parts: traffic collection process, DDoS attack identification component, and flow table delivery. The traffic collection part removes traffic features for preparation for traffic identification. Employing the flexible and multi-dimension features of the SDN network framework in utilizing the DDoS attacks detection method, the controller removes the network traffic features with statistical flow table data and uses the SVM technique to identify the attack traffics.

Li et al. [10] progress a novel structure named PCA-RNN (Principal Component Analysis-Recurrent Neural Network) for identifying DDoS attacks. For widely understanding network traffic, one can choose one of the network features to explain the traffic. A more utilize the PCA technique to reduce the dimensional of the feature and reduce the time complexity of detections. With implementing PCA, the forecast time has considerably decreased, but one of the original data is still be contained. The data after dimensional decrease has been fed as to RNN for training and getting detection technique.

Yuan et al. [11] presented a DL-based DDoS attacks detection manner (DeepDefense). In DL manner is automatically remove high-level feature in low-level ones and obtain powerful representation and inference. It plans the recurrent DNN for learning designs in order of network traffic and trace network attack performances. Wankhede and Kshirsagar [12] purpose for detecting DoS attacks efficiently utilizing ML and NN techniques. The detection has been concentrated on application layer DoS attacks

DOI: https://doi.org/10.54216/JCIM.000106

detection before transport as well as network DoS attacks detection. The modern DoS attack dataset CIC IDS2017 dataset has been utilized experimentally.

He et al. [13] projected a DOS attack detection model from the cloud-dependent upon ML approaches on the source side. This technique leverages statistical data combined with the cloud server hypervisor and virtual machine (VM) to prevent network packages from being sent out to the outside networks. It can be estimated 9 ML techniques and carefully relate its efficiency. Zekri et al. [14] intended a DDoS detection system that depends on C.4.5 approach for mitigating the DDoS threats. This technique, coupled with signature detection methods, creates the DT for performing automatic, effectual detection of signature attacks to DDoS flooding attacks. For validating our model, it is chosen other ML algorithms and related to the attained outcomes.

3. The DDoS Attack Detection Model

A novel HFSDL-DDoS attack detection model is designed to identify and categorize the occurrence of DDoS attacks in WSN. The HFSDL-DDoS technique involves preprocessing, ICGA based FS, BiLSTM based classification, and FFA-based hyperparameter optimization. The working of every module is offered in the following subsections.

3.1 Overview of ICGA based FS Technique.

To attain the higher efficacy and high classification performance of typical genetic models, few hybrid GA for FS method has been proposed by integrating GA's adequate global search capacity with few practical local search heuristic algorithms. In these works, a new immune clonal genetic model that depends on the clonal immune method, known as ICGA, is developed to solve the FS problems. The immune clone method is an inspiration of the immune system that can find the bacteria and variety, and the search target contains specific independence and dispersion. ICA could efficiently retain the diversity among the population of antibodies and speed up the global convergence speed. The ICGA method has additional exploitation and exploration capacities because of the clonal selection concept that antibodies can be cloning a few related antibodies in the solution space. All antibodies in the search space specify a subset of the potential feature. In the ICGA method, all the antibodies in the population represent a candidate solution to the FS problems. The process employs a binary coding model that"0" implies "unselected" and "1" means "selected". Thus, the chromosome represents a string of binary digits of 0's and 1's, and every gene in the chromosome corresponds to a feature. They develop an affinity function that incorporates classification performance with F-score, i.e., the calculation standard for the FS model. The affinity function is determined by:

$$\underset{s_i \in S}{\text{affinity}}(i) = \lambda_1 \times ass(s_i) + \lambda_2 \times \frac{1}{|S|} \times \frac{\sum_{j=1}^{|S|} F\left(FS(s_j)\right)}{\sum_{j=1}^{|S|} F(s_j)}. \quad (1)$$

Where $FS(s_j)$ has been equivalent to the sample of feature i if feature i has been chosen, else $FS(s_j)$ has been matching to 0, $\lambda_1 + \lambda_2 = 1$.

The three major operations of ICGA, involving selection, clonal, and mutation. Mutation operations would take the binary mutation operations in a typical genetic model. Clonal is the large antibodies affinity for a specific scale replication. Clone size is estimated by:

$$size(i) = \left\lceil \frac{|D|}{|F|} \times \frac{affinity(i)}{\sum_{j=1}^{N} affinity(i)} \right\rceil. \quad (2)$$

Where, $|D|$ & $|F|$ represents the number of components in the set D & F, correspondingly. N represent the amounts of antibodies in the population. The main concept of selection operation is given below. First of all, choose the n maximum affinity antibody and produce the number of clones for them. Then, the antibody which is directly elected were maintained to the upcoming generation.

3.2 Working of FFA-BiLSTM based Detection Technique

DOI: https://doi.org/10.54216/JCIM.000106

During the detection process, the BiLSTM model is executed to classify the DDoS attacks. The typical RNN process the input in a specific direction and processes the data which have in the future. Such problems are resolved by executing the bi-directional topology of LSTM. Bi-directional LSTM extracts complete time data at time t by taking into account the future and past data. This technique divides the hidden neuron of the typical RNN into backward and forward states. The neuron in the forward state isn't linked to the neuron in the backward state, and the neuron in the backward state isn't linked to the neuron in the on-state—the fundamental framework of the three-time steps of the two-way LSTM extension. Without a reverse state, this framework is equivalent to a typical one-way RNN. Using this framework is not necessary to add other delays as in typical RNN [15]. Fig. 1 shows the framework of Bi-LSTM.

The hyperparameter tuning of the BiLSTM model is performed using the FFA. FFA is a novel heuristic technique that inspires fruit fly's foraging actions in nature for seeking an optimum solution to objective functions. The fundamental phases are as follows [16]:

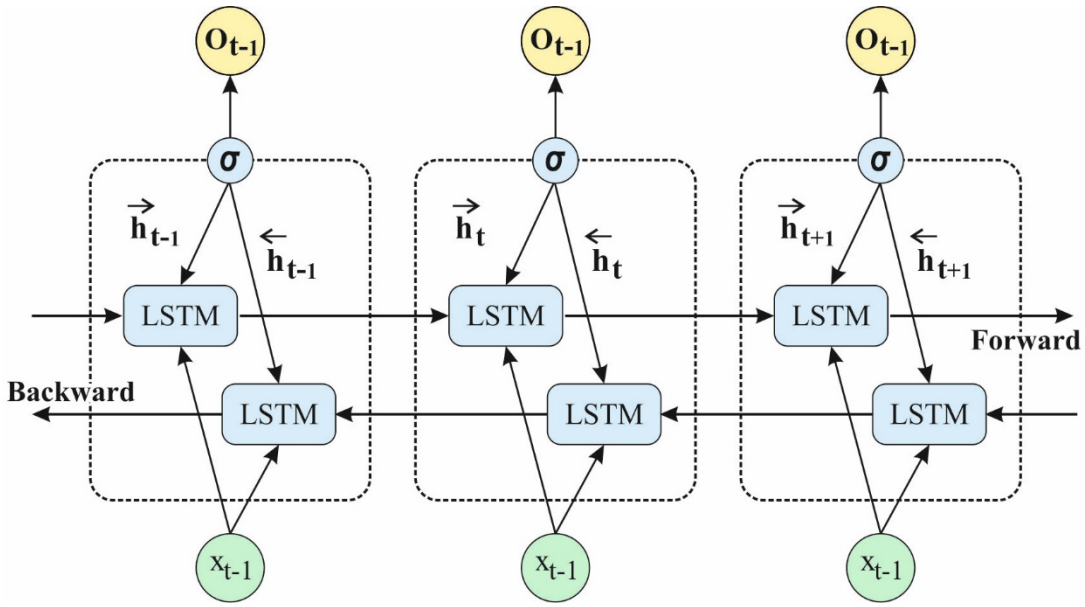


Fig. 1. Structure of Bi-LSTM

Step 1: Initialization parameters. Fixed *Sizepop* and *Maxgen* of the population sizes and initialization the population place:

$$(\mathrm{X}_{-\mathrm{axis}}, \mathrm{Y}_{-\mathrm{axis}}). \tag{3}$$

Step 2: The fruit flies search from the olfactory model that creates the search way and the search step arbitrarily. The random value (RV) remains the search distance, and the place of the population has been upgraded concurrently:

$$\begin{cases} \mathrm{X_i} = \mathrm{X}_{-\mathrm{axis}} + \mathrm{RV} \\ \mathrm{Y_i} = \mathrm{Y}_{-\mathrm{axis}} + \mathrm{RV} \end{cases}. \tag{4}$$

Step 3: As the right place of the food has been unidentified, it can be required for calculating the distance ($Dist_i$) amongst the fruit fly and the origin of coordinates, later compute the taste concentration parameters (S_i):

$$Dist_i = \sqrt{\mathrm{X_i^2} + \mathrm{Y_i^2}} \tag{5}$$

$$S_i = \frac{1}{\mathrm{Dist_i}}. \tag{6}$$

DOI: https://doi.org/10.54216/JCIM.000106

Step 4: Additional, the fruit flies favor concentration resolve value (S_i) as to the taste concentration decision functions, the fitness function, afterward, it can attain the individuals taste concentration of fruit fly $Smell_i$.

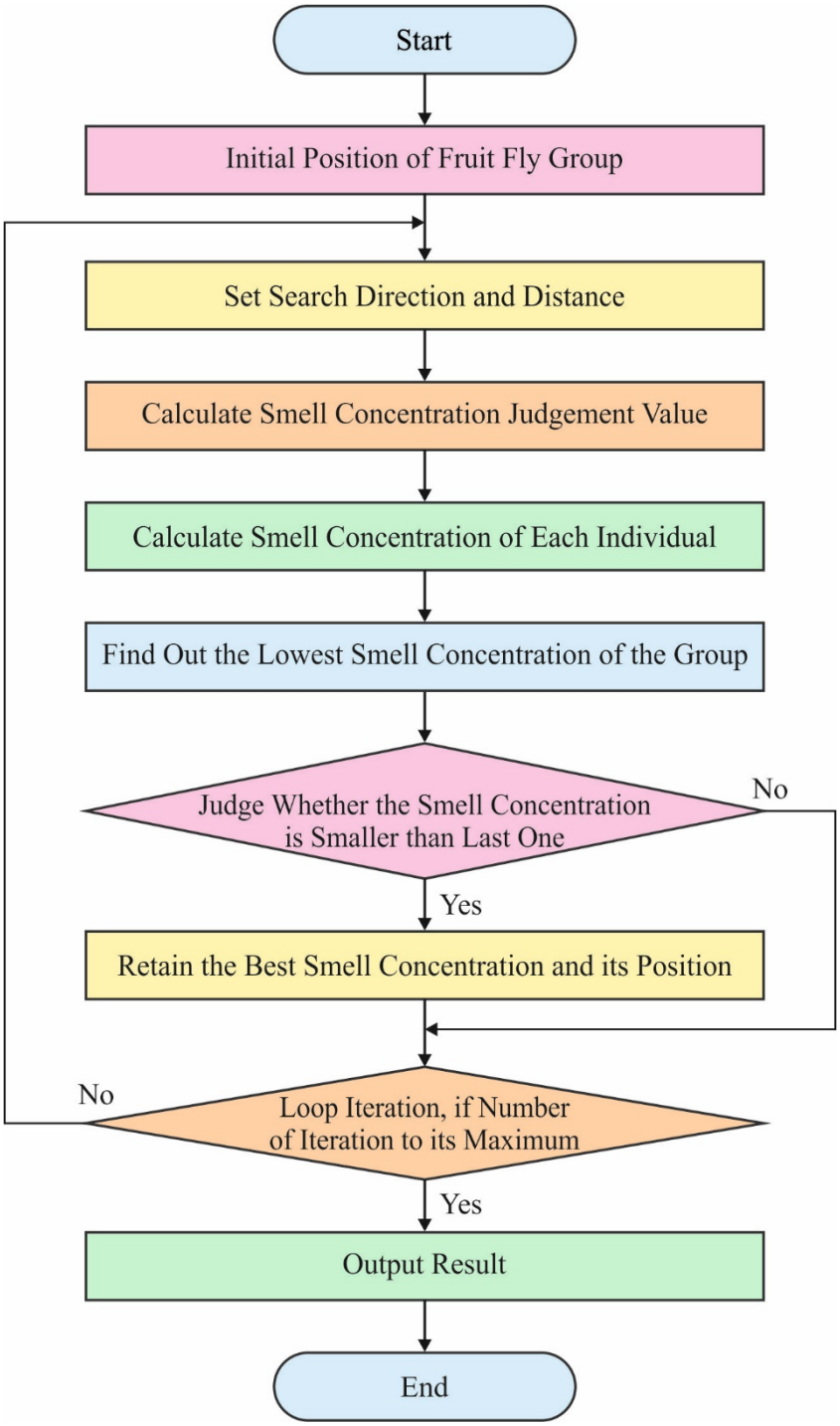


Fig. 2. Flowchart of FFA

$$Smell_i = Fitness(S_i). \quad (7)$$

Step 5: Recognize the individuals with maximum favor concentration from the drosophila populations.

$$[bestSmell, bestIndex] = min\ (Smell). \quad (8)$$

Step 6: Maintain the optimum favor concentration values and coordinates, and other individuals from the population fly to this place:

DOI: https://doi.org/10.54216/JCIM.000106

$$SmellBest = bestSmell. \tag{9}$$

$$\begin{cases} X_{-axis} = X\,(bestIndex) \\ Y_{-axis} = Y\,(bestIndex) \end{cases} \tag{10}$$

Step 7: End state, judge if the concentration of the optimum place is superior to that preceding generation, and obtain the maximal amount of iterations; else, skip step 2 to enter the iterative optimize. Fig. 2 illustrates the flowchart of FFA [17].

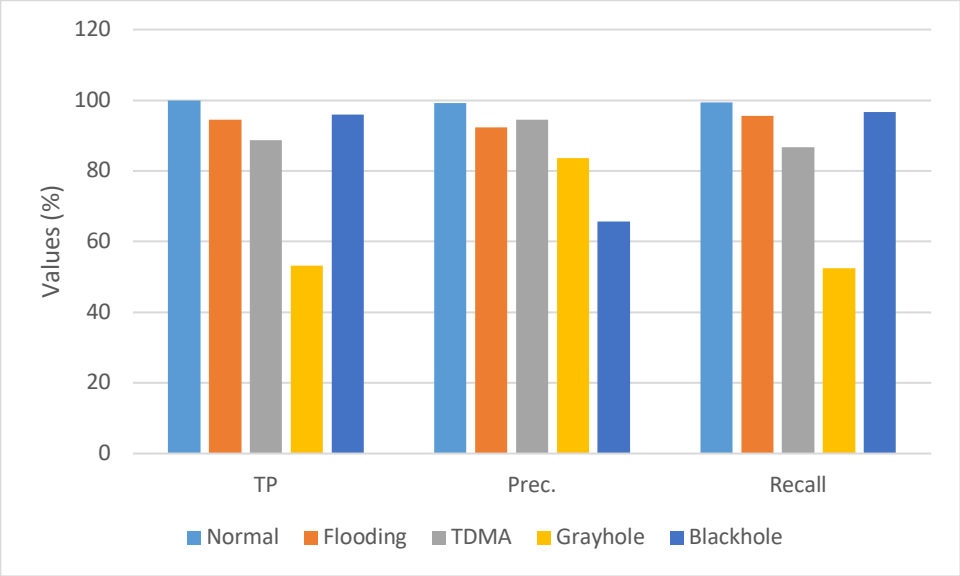
4. Results and Discussion

The performance validation of the HFSDL-DDoS technique is investigated. The results demonstrated that the HFSDL-DDoS technique is under training and testing sets. The results are investigated interms of TP, FP, precision, and recall.

Table 1 and Fig. 3 offer the analysis of the results of the HFSDL-DDoS technique is performed interms of different measures on the training dataset. The results have shown that the HFSDL-DDoS technique has gained effective outcomes on the classification of distinct classes. The HFSDL-DDoS technique has classified the 'Normal' class with the TP, FP, precision, and recall of 99.9%, 9.2%, 99.2%, and 99.4%. Also, the HFSDL-DDoS manner has classified the 'Flooding' class with the TP, FP, precision, and recall of 94.5%, 0.2%, 92.3%, and 95.5%. In addition, the HFSDL-DDoS approach has classified the 'TDMA' class with the TP, FP, precision, and recall of 88.7%, 0.3%, 94.4%, and 86.7%. Besides, the HFSDL-DDoS method has classified the 'Grayhole' class with the TP, FP, precision, and recall of 53.2%, 0.7%, 83.6%, and 52.5%. In line with this, the HFSDL-DDoS system has classified the 'Blackhole' class with the TP, FP, precision, and recall of 95.9%, 1.9%, 65.6%, and 96.7%.

Table 1 Result analysis of HFSDL-DDoS model on the training dataset

Class	TP	FP	Prec.	Recall
Normal	99.9	9.2	99.2	99.4
Flooding	94.5	0.2	92.3	95.5
TDMA	88.7	0.3	94.4	86.7
Grayhole	53.2	0.7	83.6	52.5
Blackhole	95.9	1.9	65.6	96.7
Average	86.44	2.46	87.02	86.16



DOI: https://doi.org/10.54216/JCIM.000106

Fig. 3. Result analysis of HFSDL-DDoS model on the training dataset

Table 2 and Fig. 4 provide the outcomes analysis of the HFSDL-DDoS algorithm concerning distinct measures on the testing dataset. The outcomes demonstrated that the HFSDL-DDoS algorithm had obtained effective outcomes on the classification of varying classes. The HFSDL-DDoS system has classified the 'Normal' class with the TP, FP, precision, and recall of 99.7%, 8.9%, 98.6%, and 99.6%. Similarly, the HFSDL-DDoS technique has classified the 'Flooding' class with the TP, FP, precision, and recall of 95.9%, 0.2%, 92.3%, and 94.3%. Moreover, the HFSDL-DDoS approach has classified the 'TDMA' class with the TP, FP, precision, and recall of 88.2%, 0.4%, 94.7%, and 88%. Furthermore, the HFSDL-DDoS manner has classified the 'Grayhole' class with the TP, FP, precision, and recall of 50.9%, 1.1%, 81.9%, and 52.3%. Along with that, the HFSDL-DDoS methodology has classified the 'Blackhole' class with the TP, FP, precision, and recall of 96.5%, 1.2%, 66.6%, and 96.8%.

Table 2 Result analysis of HFSDL-DDoS model on the testing dataset

Class	TP	FP	Prec.	Recall
Normal	99.7	8.9	98.6	99.6
Flooding	95.9	0.2	92.3	94.3
TDMA	88.2	0.4	94.7	88
Grayhole	50.9	1.1	81.9	52.3
Blackhole	96.5	1.2	66.6	96.8
Average	86.24	2.36	86.82	86.2

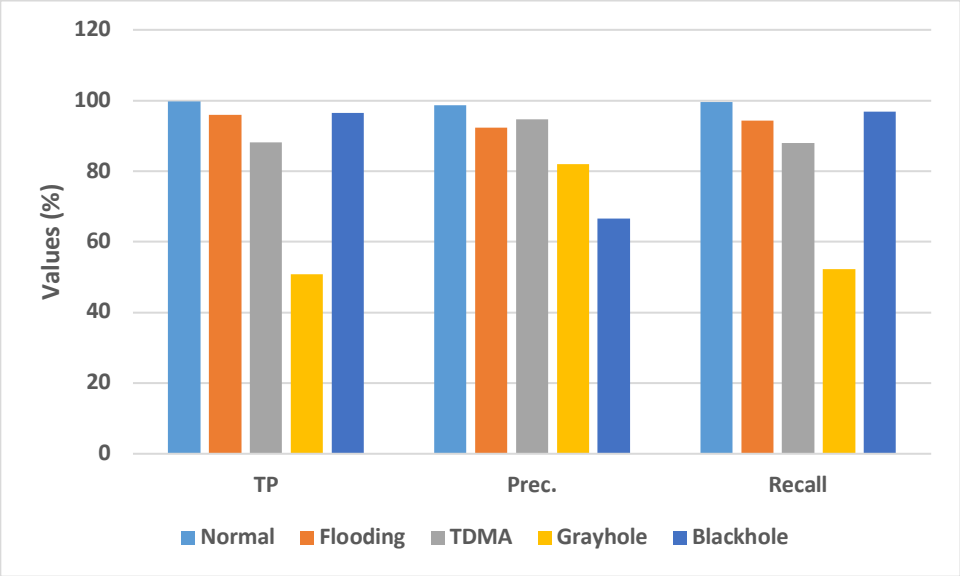


Fig. 4. Result analysis of HFSDL-DDoS model on the testing dataset

An average results analysis of the HFSDL-DDoS technique on the applied training and the testing dataset is given in Table 3 and Fig. 5. The results demonstrated that the HFSDL-DDoS technique had accomplished effective outcomes with the TP, FP, precision, and recall of 99.70%, 8.90%, 98.60%, and 99.60% on the training dataset. Moreover, the results portrayed that the HFSDL-DDoS scheme has accomplished effectual outcomes with the TP, FP, precision, and recall of 86.24%, 2.36%, 86.82%, and 86.20% on the testing dataset.

DOI: https://doi.org/10.54216/JCIM.000106

Table 3 Average values of HFSDL-DDoS model on training and testing datasets

Average Values	TP	FP	Prec.	Recall
Average Training	99.70	8.90	98.60	99.60
Average Testing	86.24	2.36	86.82	86.20

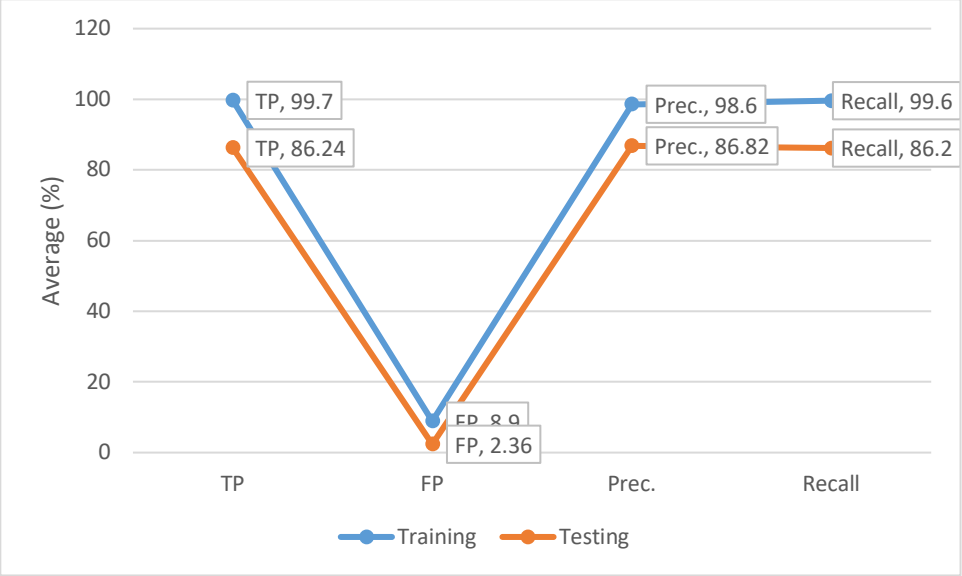


Fig. 5. Average analysis of HFSDL-DDoS model on training and testing datasets

To portray the improved performance of the HFSDL-DDoS technique, a brief comparative result analysis is made interms of performance measures in Table 4 and Fig. 6. The results have shown that the SVM technique has reduced performance with the TP, FP, precision, and recall of 85.06%, 2.18%, 85.58%, and 85.06% on the training dataset. In addition, the HFSDL-DDoS technique has resulted in a maximum performance with the TP, FP, precision, and recall of 99.70%, 8.90%, 98.60%, and 99.60% on the training dataset.

Table 4 Comparative analysis of HFSDL-DDoS model with different measures

Methods	TP	FP	Prec.	Recall
SVM	85.06	2.18	85.58	85.06
HFSDL-DDoS	99.70	8.90	98.60	99.60

DOI: https://doi.org/10.54216/JCIM.000106

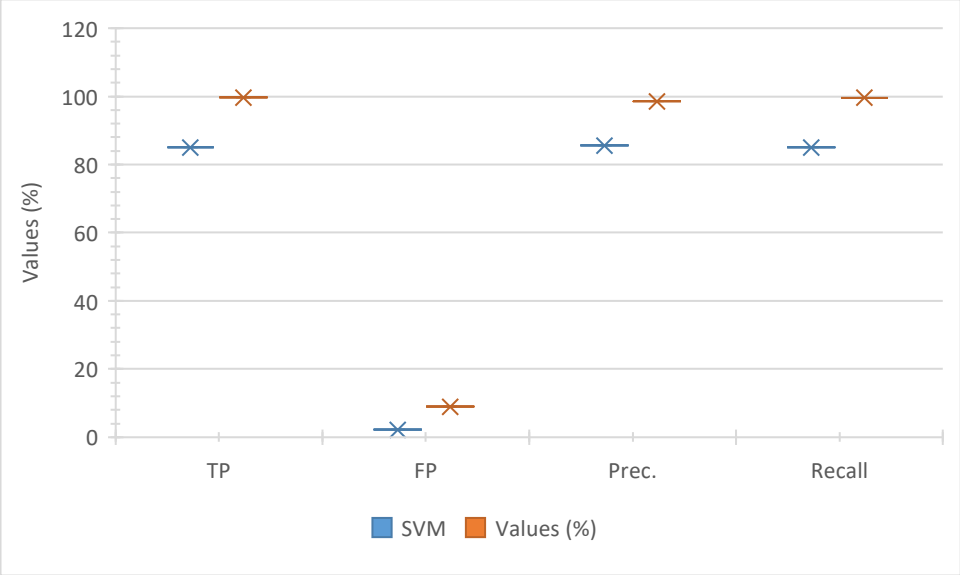


Fig. 6. Comparative analysis of HFSDL-DDoS model with distinct measures

5. Conclusion

In this article, a novel HFSDL-DDoS attack detection model is designed to identify and categorize the occurrence of DDoS attacks in WSN. The HFSDL-DDoS technique involves preprocessing, ICGA based FS, BiLSTM based classification, and FFA-based hyperparameter optimization. Besides, the HFSDL-DDoS technique involves the ICGA based FS approach to improve the detection performance. Furthermore, the FFA with BiLSTM based classification model is employed. The experimental analysis of the HFSDL-DDoS technique is performed, and the results are examined interms of several performance measures. The resultant experimental results pointed out the betterment of the HFSDL-DDoS technique over the other techniques. In the future, the HFSDL-DDoS technique can be deployed in the 5G-enabled Internet of Things (IoT) systems.

References

[1] Almon, L., Riecker, M. and Hollick, M., 2017, October. Lightweight Detection of Denial-of-Service Attacks on Wireless Sensor Networks Revisited. In 2017 IEEE 42nd Conference on Local Computer Networks (LCN) (pp. 444-452). IEEE.

[2] Gavric, Z. and Simic, D., 2018. Overview of DOS attacks on wireless sensor networks and experimental results for simulation of interference attacks. Ingeniería e Investigación, 38(1), pp.130-138.

[3] Osanaiye, O., Alfa, A.S. and Hancke, G.P., 2018. A statistical approach to detect jamming attacks in wireless sensor networks. Sensors, 18(6), p.1691.

[4] Almomani, I.M. and Alenezi, M., 2018. Efficient Denial of Service Attacks Detection in Wireless Sensor Networks. J. Inf. Sci. Eng., 34(4), pp.977-1000.

[5] Mazur, K., Ksiezopolski, B. and Nielek, R., 2016. Multilevel modeling of distributed denial of service attacks in wireless sensor networks. Journal of Sensors, 2016.

[6] Mansouri, D., Mokddad, L., Ben-Othman, J. and Ioualalen, M., 2015, June. Preventing denial of service attacks in wireless sensor networks. In 2015 IEEE International Conference on Communications (ICC) (pp. 3014-3019). IEEE.

[7] Dhar, M. and Singh, R., 2015. A review of security issues and denial of service attacks in wireless sensor networks. International Journal of Computer Science and Information Technology Research, 3(1), pp.27-33.

[8] Deepa, V., Sudar, K.M. and Deepalakshmi, P., 2018, December. Detection of DDoS attack on SDN control plane using hybrid machine learning techniques. In 2018 International Conference on Smart Systems and Inventive Technology (ICSSIT) (pp. 299-303). IEEE.

DOI: https://doi.org/10.54216/JCIM.000106

[9] Yang, L. and Zhao, H., 2018, October. DDoS attack identification and defense using SDN based on machine learning method. In 2018 15th International Symposium on Pervasive Systems, Algorithms and Networks (I-SPAN) (pp. 174-178). IEEE.

[10] Li, Q., Meng, L., Zhang, Y. and Yan, J., 2018, September. DDoS attacks detection using machine learning algorithms. In International Forum on Digital TV and Wireless Multimedia Communications (pp. 205-216). Springer, Singapore.

[11] Yuan, X., Li, C. and Li, X., 2017, May. DeepDefense: identifying DDoS attack via deep learning. In 2017 IEEE International Conference on Smart Computing (SMARTCOMP) (pp. 1-8). IEEE.

[12] Wankhede, S. and Kshirsagar, D., 2018, August. DoS attack detection using machine learning and neural network. In 2018 Fourth International Conference on Computing Communication Control and Automation (ICCUBEA) (pp. 1-5). IEEE.

[13] He, Z., Zhang, T. and Lee, R.B., 2017, June. Machine learning based DDoS attack detection from source side in cloud. In 2017 IEEE 4th International Conference on Cyber Security and Cloud Computing (CSCloud) (pp. 114-120). IEEE.

[14] Zekri, M., El Kafhali, S., Aboutabit, N. and Saadi, Y., 2017, October. DDoS attack detection using machine learning techniques in cloud computing environments. In 2017 3rd international conference of cloud computing technologies and applications (CloudTech) (pp. 1-7). IEEE.

[15] Huang, Z., Xu, W. and Yu, K., 2015. Bidirectional LSTM-CRF models for sequence tagging. arXiv preprint arXiv:1508.01991.

[16] Pan, Q.K., Sang, H.Y., Duan, J.H. and Gao, L., 2014. An improved fruit fly optimization algorithm for continuous function optimization problems. Knowledge-Based Systems, 62, pp.69-83.

[17] Dong, C.C., Shen, X., Zhou, J.X., Wang, T. and Yin, Y.J., 2016. Optimal design of feeding system in steel casting by constrained optimization algorithms based on InteCAST. China Foundry, 13(6), pp.375-382.

DOI: https://doi.org/10.54216/JCIM.000106