SMO-ANN: A HYBRID CLASSIFIER FOR NETWORK INTRUSION DETECTION SYSTEM USING SPIDER MONKEY OPTIMIZATION ALGORITHM

[1]Deepshikha Kumari, [1]Abhinav Sinha, [1]Sandip Dutta and [1]Prashant Pranav
[1]Department of Computer Science and Engineering, Birla Institute of Technology, Mesra, Jharkhand
deepshikha3377@gmail.com, sinhaabhinav337@gmail.com, prashantpranav19@gmail.com, sandipdutta@bitmesra.ac.in

Abstract: Threat from an eavesdropper has always been a concern for government and other private organizations. Be it in a form of traffic analysis, modification of messages, disruption of normal network services or identity mismatch, attackers have always come up with ways to disrupt normal communication between two parties. They try to intrude into a server or a system with malign motives which eventually harms the organizations and government alike. Intrusion Detection System is a very naïve field which focusses on the detection of abnormal network traffic and report in real time if any malicious activity is detected. In our present research work Spider Monkey Optimization (SMO) algorithm is used with Artificial Neural Network (ANN) to detect attacks or intrusion in the system. The developed model was rigorously examined on a publicly available dataset LuFlow20 to classify any incoming traffic as attack or normal. The ANN–SMO model gives an accuracy of 99.82 % on LuFlow20, which is a labelled dataset.

Keywords: Intrusion Detection System; Spider Monkey Optimization Algorithm (SMO); Machine Learning; LUflow dataset; Attacks; Artificial Neural Network (ANN).

1. INTRODUCTION

Network security has become essential to preserving the availability, confidentiality, and integrity of information in the modern world due to our growing reliance on digital technology and the internet. The area covers a broad spectrum of tools, procedures, and methods to protect digital assets from different types of cyberattacks. It's a continuous process that needs to be updated, watched over, and adjusted to new threats. A key component of total cyber security is network security. Network security is an extremely technical and specialised field of study and practice. It addresses risk management, malware, honeypots, packets and flows, intrusion detection and prevention systems, defencelessness, and misconducts [1].

Attacks on cyber security can take many different forms and target various information system components. The several cyber security attacks are listed below:

Malware is software that is intentionally created to do harm or take advantage of systems. Malware includes programmes such as viruses, worms, ransomware, trojans, spyware, and more. **Phishing** is the term used to describe deceptive attempts to obtain personal information by deceiving people, like credit card numbers or passwords. The terms **Denial of Service (DoS)** and **Distributed Denial of Service (DDoS)** attacks refer to overloading a system, network, or services to prevent regular operation or render it unavailable to users. A **man-in-the-middle attack** involves potentially influencing and stopping communication between two parties without their knowledge. SQL Injection is the process of gaining unauthorised access to a website's database by taking advantage of flaws in the system. **Ransomware** is a software that encrypts a user's data or systems and demands payment to be unlocked (often in cryptocurrency). **Social engineering** is the practice of tricking someone into disclosing private information or taking acts against their better judgement. Credential stuffing is the practice of breaking into other accounts by utilising a password and username that you have already got.

Insider threats are those that come from people who work for a company, such as contractors or employees, and who utilise their access for nefarious ends. **Zero-day exploits** are cyberattacking that take use of IOT device vulnerabilities to obtain unauthorised access or control.

The purpose of an Intrusion Detection System (IDS) is to identify and guard against unauthorised access to the system. An essential component of any cybersecurity strategy is an intrusion detection system, which offers proactive monitoring and detection capabilities to protect networks and systems against evolving threats. Real-time detection and response to any security problems is the main function of an intrusion detection system. There are two forms of intrusion detection systems: host-based and network-based. The architecture of IDS is displayed below in *fig 1.*

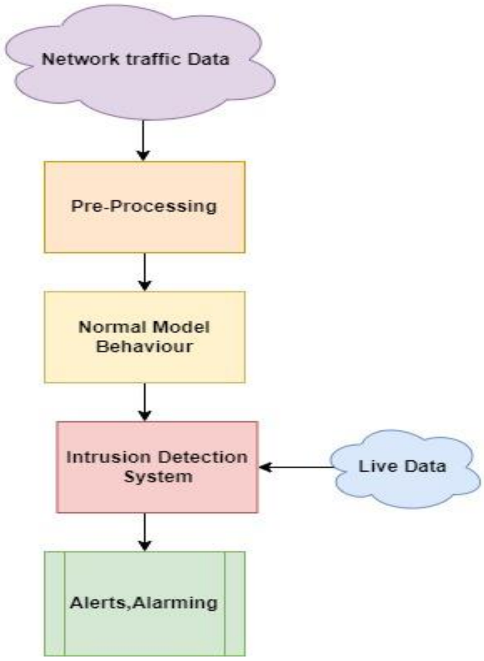

Fig 1: Architecture of IDS

IDS employ a specific set of analytics approaches to find attack targets, identify their resources, notify network administrators, and maybe lessen the impact of an attack. IDS concludes data from a computer network to distinguish system misuse and assaults. Only a small percentage of IDS attempt to thwart attacks as soon as they are launched; the majority only assess attacks. IDS uses three sorts of data: system status files, system level test data, and network traffic data [3].

2. RELATED WORK

For detecting cyber security threats on the internet of things, P. Vijayalaxmi [4] suggested a hybrid dual channel convolutional neural network (DCCNN) with spider monkey optimisation (SMO). Deep learning algorithms have demonstrated as a highly beneficial and productive in this research project when it comes to IOT security as compared to earlier methods.

A Spider Monkey-based Elman Spike Neural Network (SM-ESNN) was proposed by Panem Charanarur [5] to detect intrusion risks in software-defined networks (SDNs). ISCXIDS 2012 datasets are used in this paper. A detection module and a mitigation module are two submodules that make up the produced programme. The accuracy of the suggested SM-ESNN method was found to be 98.24%. The hybrid spider monkey Battle Royale algorithm (SMBR), which uses a range-based localization method to locate sensor nodes, was proposed by R. Shalilla [6]. Several metrics, including localization error, delivery ratio, localization coverage, energy consumption, latency, packet drop, and throughput, are to be determined using this model. To detect threats in an IOT environment, Ethala Sandhya [7] presented Spider Monkey Optimisation using Random Forest (SMO-RF). The NSL-KDD dataset was pre-processed. When compared to the current SVM parameter optimisation, SMO-RF yielded the best accuracy. Balaji R. M. [8] suggested using the CSE-CIC-IDS2018 dataset in a hybrid Deep Learning algorithm (PCA+SMO-FCM+AE) for intrusion detection on the AWS cloud. This model has a 95% accuracy rate when it was compared to 11 other models.

Sandhaya Ethala [9] presented hybrid optimisation techniques that handle massive volumes of intrusion data classification problems and enhance detection accuracy by shrinking false alarm rate. These techniques integrate Spider Monkey Optimisation (SMO) and hierarchical Particle Swarm Optimisation (HPSO). Several publicly accessible benchmark malware datasets for DNN and other traditional machine learning technique experiments are covered by R. Vinayakumar [10]. The author suggests Scale-hybrid-IDS-AlertNet, an abundantly climbable and hybrid DNNs Framework that may be used to proactively warn against cyberattacks by supervising network traffic flow in real-time and host-level incidents. To identify and thwart a malicious assault, Nevrus Kaja [11] presented a revolutionary two-stage intelligent Intrusion Detection System (IDS). The suggested techniques are divided into two phases: the first phase uses K-means to identify attacks, while the second phase uses supervised machine learning algorithms to categorise attacks and reduce the false alarm rate. The method result classifies assaults with 99.97% accuracy based on J48 algorithms. Sharma Anshika [12] Through the use of performance metrics, different machine learning techniques such as SVM, KNN, and RF have been employed to precisely detect threats and malformations in the IOT environment. This research makes use of the latest Luflow dataset.

RF exhibited the highest accuracy of 97.7%, while KNN and SVM had respective accuracy of 94.7% and 92.8%. For intrusion detection, Suad Mohammed Othman [13] introduced the Spark-chi-SVM model. This model uses the Cup 99 dataset from the Apache Spark Big Data platform and ChiSqSelector for feature selection to develop an intrusion detection model using SVM classifier. The author then compares the Chi-Logistic and Chi-SVM classifiers. Consequently, Chi-SVM classifier outperformed Chi-logistic classifier in terms of shrinking the training time and is very systematic for big data.

3. METHODOLOGY

There are numerous labelled and unlabelled intrusion detection system datasets available, including UNSW-NB15, KYOTO 2006+, CICIDS, LuFlow, NSL-KDD, DARPA98, and KDD-CUP99. In this research project, we have used the tagged dataset LuFLow 20. Initially, the gathered LuFlow20 dataset underwent pre-processing to eliminate any redundant data. The following steps make up the current work's methodology:

3.1. Data Collection

LuFlow20 dataset is used in this study. A robust ground truth is provided by LuFlow, a flow-based network intrusion detection dataset that correlates harmful conduct. In terms of characteristics, the dataset is not that vast. The LuFlow dataset has sixteen features.

3.2. Data Pre-Processing

After collection of data, data cleaning is necessary. Removing noise from the datasets would aid in enhancing the models' performance. Data cleansing, normalisation, integration, and transformation are some of the pre-processing procedures that were used. The data has nan values, duplicate values, and outlier values eliminated. Using the label encoder, we have assigned a 1 to malicious and a 0 to benign labels.

4. PROPOSED ALGORITHM

The proposed work demonstrates a methodology that involves optimizing the layer of neural networks by **Spider Monkey Optimisation (SMO)** algorithm, a population-based optimisation technique inspired by nature. The social interactions and communication styles of foraging spider monkeys in the jungle served as inspiration for the algorithm. SMO is an element of the swarm intelligence algorithm that is depleted to determine the best answer to optimisation problems. The fission-fusion social structure is the foundation of spider monkeys' foraging behaviour. Improving the IDS's performance is connected to the use of spider monkey optimisation. Reducing false positives and false negatives while progressing the system's detection skills is the major objective. Six groups of spider monkeys are used for food foraging. SMO involves following steps to select the best solution:

Step 1: Initializing the Population: In the proposed work, SMO is allocated with the population P=30, where SM_p (*p=1,2,3……. p*) and SM_p is known as the population for the p^{th} monkey. Each SM_p is initialized as follow:

$$SM_{pq} = SM_{minq} + UR\ (0,1) * \left(SM_{maxq} - SM_{minq}\right)$$

Where SM_{maxq} and SM_{minq} are the upper and lower bound of the search space in the q^{th} dimension and UR is the random number in the range with uniform distribution of (0,1) [14].

Step 2: Local Leader Phase (LLP): In this dynamic phase of SMO algorithm, all the spider monkey gets a change to update their positions. Any modifications made on the location of the monkeys is dependent on the local leader and local group members proficiencies. Each spider monkey's fitness value is computed at its new location. If it is more than that of its previous location, it will renew otherwise not [14]. Below is the equation of position update:

$$SM_{newpq} + UR(0{,}1) * \left(LL_{kq} - SM_{pq}\right) + \mathrm{UR}(-1{,}1) * (SM_{rq} - SM_{pq})$$

where LL_{kq} is the q^{th} dimention from the randomly chosen k^{th} SM of a k^{th} local group having their leader position.

Step 3: Global Leader Phase (GLP): Once LLP is processed, the GLP is modified. The location of SM is calculated using the equation (3) and the location of SM is updated is calculated as follows.

$$SM_{newpq} = SM_{pq} + UR(0{,}1) * \left(GL_{kq} + SM_{pq}\right) + UR(-1{,}1) * (SM_{rq} - SM_{pq})$$

Where GL_{kq} is referred as the global leader location having q^{th}dimension (q = 1,2,3….M) having the haphazardly chosen index.

Step 4: Global Leader Learning Phase (GLL): The SMO algorithm establishes optimal solution for the whole swarm in this phase. The established SM is considered as the swarm's global leader. Furthermore, the global leader's position is proven and if it is not updated then the counter correlated with the global leader called as global limit count (GLC) is upgraded to 1, else 0[15]. GLC is verified by Global leader and then cross-referenced with the Global Leader limit.

Step 5: Local Leader Learning Phase (LLL): Greedy selection model is used for a local group to update the LLL with the SM location based on the fitness value of that precise local group [16]. The optimal location has a value allocated to the local leader. Increase it to 1 if there is no update present. This value will be inserted to the LLC.

Step 6: Local Leader Decision Phase (LLD): If the local leader is not updated for their location amid the fixed LLL then the local groups existing are having the candidates altered with the random location from the step1 that utilizes the evidence from the global leader and local leader occupying on pursuing equation (4).

$$SM_{newpq} + SM_{pq} + UR(0,1) * \left(GL_{kq} - SM_{pq}\right) + UR(0,1) * (SM_{rq} - LL_{pq}) \qquad (4)$$

Step 7: Global Leader Decision Phase (GLD): It is comparable to the LLD phase if the global leader is not reorganized to meet the GLL condition, then the global leader splits the swarm into small groups.

$$fitnesss_{features\ importance} = \frac{Number\ of\ samples\ that\ reaches\ the\ nodes}{Total\ number\ of\ samples} \qquad (5)$$

The best value of features is selected based on the peak value of the fitness function using equation (5).

The proposed SMO algorithm is used in **ANN (Artificial Neural Network)** which is a machine learning algorithm stimulate by the performance of biological neurons found in the brain and central nervous system [17]. Using SMO and ANN, we propose a hybrid algorithm to detect and optimize intrusions in a network. We call the algorithm as **Hybrid Classification (SMO-ANN)** and demonstrates its efficiency using the Luflow20 dataset.

Hybrid Classification (SMO-ANN): In this portion the developed hybrid algorithm is exhibited. We propose a method to build a hybrid IDS that conglomerates with one of the machine learning algorithms which is ANN algorithm. In SMO, Spider monkeys update their positions by food foraging behaviour. The workflow diagram of SMO-ANN is shown in below in *Fig 2.*

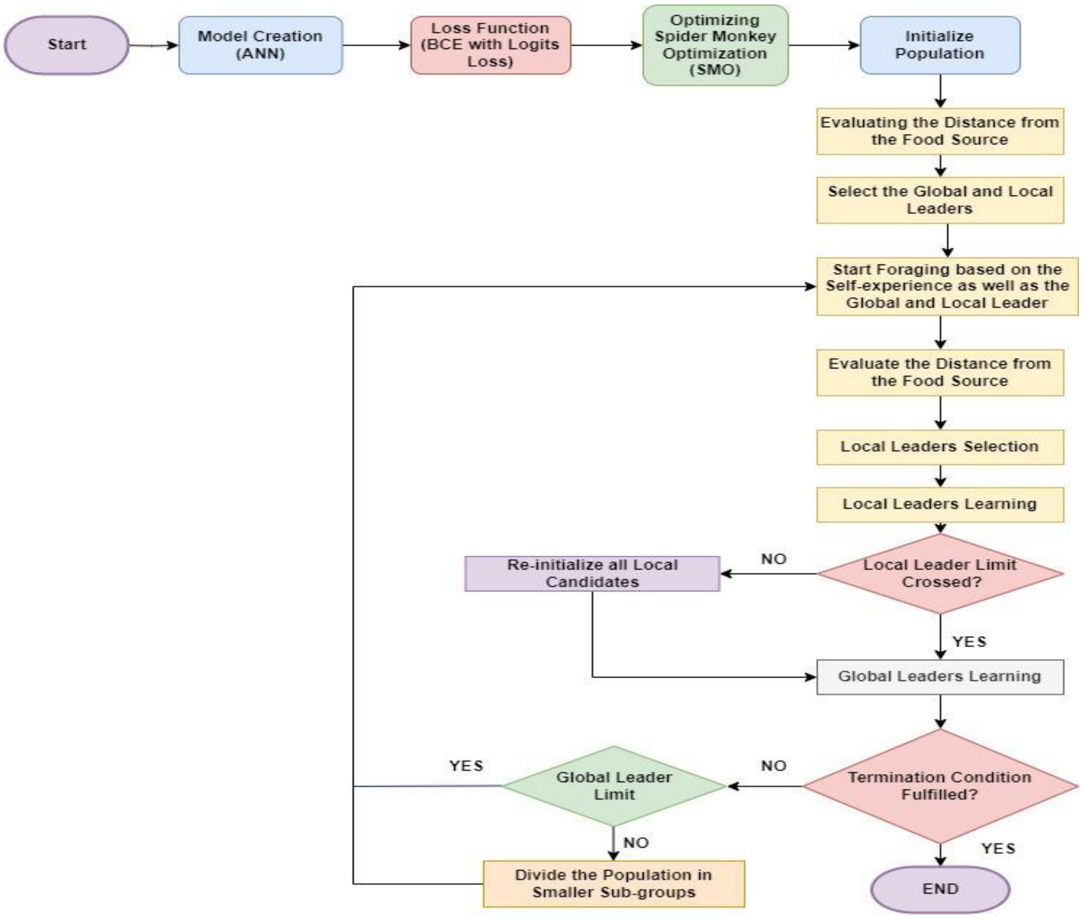
Start
Model Creation (ANN)
Loss Function (BCE with Logits Loss)
Optimizing Spider Monkey Optimization (SMO)
Initialize Population
Evaluating the Distance from the Food Source
Select the Global and Local Leaders
Start Foraging based on the Self-experience as well as the Global and Local Leader
Evaluate the Distance from the Food Source
Local Leaders Selection
Local Leaders Learning
Local Leader Limit Crossed?
NO
Re-initialize all Local Candidates
YES
Global Leaders Learning
Termination Condition Fulfilled?
NO
YES
Global Leader Limit
YES
NO
Divide the Population in Smaller Sub-groups
END

Fig 2: SMO-ANN Algorithm Workflow Diagram

5. RESULT AND DISCUSSION

System Specification: The proposed algorithm was run on a system with the following specifications:

- Operating system: Windows 10
- RAM:83 GB
- GPU: 40 GB
- Processor:12th Gen Intel(R)core (TM) i5-1235U

Performance Evaluation: Luflow20 dataset is used in the current research work for performing the testing. The proposed SMO-ANN method performance is evaluated using accuracy, precision, recall and f1 score.

Comparative Analysis: As mentioned, populations size is 30 and 50 iterations are done to get the best performance. The proposed model SMO-ANN has achieved better accuracy of 99.82 % than ANN. The following *table 1* and *figure 3* shows the precision, recall, f1 score and accuracy of the hybrid SMO-ANN models and simple ANN for network intrusion detection for normal traffic of Luflow20 dataset.

TABLE 2: COMPARISON OF ANN AND HYBRID SMO-ANN (BENIGN)

Metrices	ANN	SMO-ANN
Accuracy	71.65	99.82
Precision	73.11	99.78
Recall	61.11	99.91
F1-score	83.22	99.68

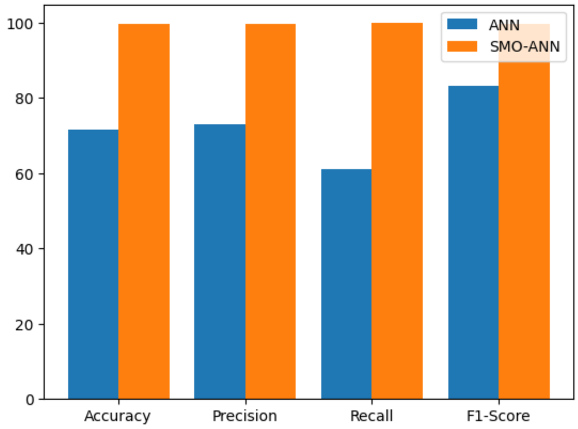

Fig 2: Comparison of SMO-ANN and ANN for Normal Traffic

Table 2 and *figure 4* below shows precision, recall, f1 score and accuracy of the hybrid SMO-ANN models and simple ANN for network intrusion detection for attack scenario of Luflow20 dataset.

TABLE 2: COMPARISON OF ANN AND HYBRID SMO-ANN (MALICIOUS)

Metrices	ANN	SMO-ANN
Accuracy	71.65	99.82
Precision	65.71	99.84
Recall	57.43	99.59
F1-score	63.12	99.88

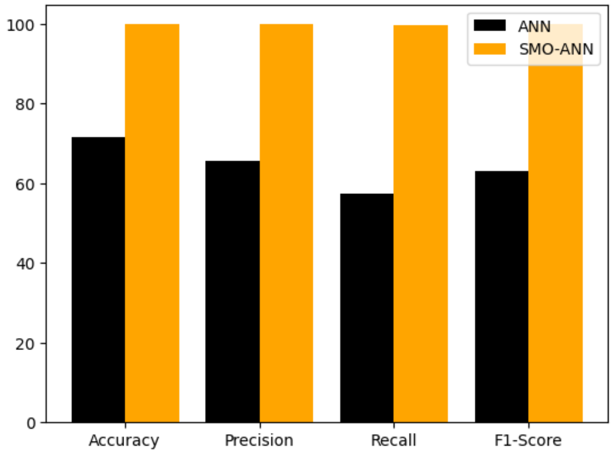

Fig 3: Comparison of SMO-ANN and ANN for Attack Scenario

As can be seen from the result SMO-ANN outperform ANN in all the performance metrices for both normal traffic and attack scenario of Luflow20 dataset.

6. CONCLUSION

Intrusion Detection System is like a watchful guardian for network. Every time it monitors all the network traffic and examines for any signs of suspicious or unauthorized activity. Using machine learning algorithm and optimization techniques, a hybrid classifier SMO-ANN (Spider Monkey Optimization -Artificial Neural Network) model for intrusion is introduced in this paper. By comparing ANN and our proposed model viz. SMO-ANN, we conclude that SMO-ANN has the best Accuracy of 99.82 for both benign and malicious activities in Luflow 20 dataset. As the model has currently been tested for binary classification, in the future the work can be extended to check other available cyber-attack datasets containing both binary and multiclass data.

REFERENCES

1. Shiravi, Hadi, Ali Shiravi, and Ali A. Ghorbani. "A survey of visualization systems for network security." *IEEE Transactions on visualization and computer graphics* 18, no. 8 (2011): 1313-1329.
2. Vijayalakshmi, P., and D. Karthika. "Hybrid dual-channel convolution neural network (DCCNN) with spider monkey optimization (SMO) for cyber security threats detection in internet of things." *Measurement: Sensors* 27 (2023): 100783.
3. Aydın, M.A., Zaim, A.H. and Ceylan, K.G., 2009. A hybrid intrusion detection system design for computer network security. *Computers & Electrical Engineering*, *35*(3), pp.517-526.
4. KS, D. and Ramakrishna, B., 2013. An artificial neural network-based intrusion detection system and classification of attacks. *International Journal of Engineering Research and Applications*, *3*, pp.1959-1964.
5. Charanarur, Panem, Bui Thanh Hung, Prasun Chakrabarti, and S. Siva Shankar. "Design optimization-based software-defined networking scheme for detecting and preventing attacks." *Multimedia Tools and Applications* (2024): 1-19.
6. Shakila, R. and Paramasivan, B., 2024. An improvised optimization algorithm for submarine detection in underwater wireless sensor networks. *Microsystem Technologies*, pp.1-12.
7. Sandhya, Ethala, and Annapurani Kumarappan. "Enhancing the Performance of an Intrusion Detection System Using Spider Monkey Optimization in IoT." *International Journal of Intelligent Engineering & Systems* 14, no. 6 (2021).
8. RM, Balajee, and Jayanthi Kannan MK. "Intrusion detection on AWS cloud through hybrid deep learning algorithm." *Electronics* 12, no. 6 (2023): 1423.
9. Ethala, Sandhya, and Annapurani Kumarappan. "A Hybrid Spider Monkey and Hierarchical Particle Swarm Optimization Approach for Intrusion Detection on Internet of Things." *Sensors* 22, no. 21 (2022): 8566.
10. Vinayakumar, R., Alazab, M., Soman, K.P., Poornachandran, P., Al-Nemrat, A. and Venkatraman, S., 2019. Deep learning approach for intelligent intrusion detection system. *Ieee Access*, *7*, pp.41525-41550.
11. Kaja, Nevrus, Adnan Shaout, and Di Ma. "An intelligent intrusion detection system." *Applied Intelligence* 49 (2019): 3235-3247

12. Sharma, Anshika, and Himanshi Babbar. "LUFlow: Attack Detection in the Internet of Things Using Machine Learning Approaches." In *2023 International Conference on Distributed Computing and Electrical Circuits and Electronics (ICDCECE)*, pp. 1-5. IEEE, 2023
13. Othman, Suad Mohammed, Fadl Mutaher Ba-Alwi, Nabeel T. Alsohybe, and Amal Y. Al-Hashida. "Intrusion detection model using machine learning algorithm on Big Data environment." *Journal of big data* 5, no. 1 (2018): 1-12.
14. Bansal, Jagdish Chand, Pramod Kumar Singh, and Nikhil R. Pal, eds. *Evolutionary and swarm intelligence algorithms*. Vol. 779. Cham: Springer, 2019.
15. Sharma, Harish, Garima Hazrati, and Jagdish Chand Bansal. "Spider monkey optimization algorithm." *Evolutionary and swarm intelligence algorithms* (2019): 43-59.
16. Chua, Tuan-Hong, and Iftekhar Salam. "Evaluation of machine learning algorithms in network-based intrusion detection system." *arXiv preprint arXiv:2203.05232* (2022).
17. Shenfield, A., Day, D. and Ayesh, A., 2018. Intelligent intrusion detection systems using artificial neural networks. *Ict Express*, *4*(2), pp.95-99.