Network Intrusion Detection in MANET Using Improved Whale Optimization Algorithm-SVM

Majid Hamid Ali

Computer science department, Collage of computer and mathematics science, Tikrit University, Iraq
Majid.hamid@tu.edu.iq

ABSTRACT

Presently, the MANET is indispensable for network administration because of the collaborative efforts of its individual nodes. Due to the dynamic nature of the network's nodes, this type of node might emerge for many different causes. In a MANET, you could be vulnerable to DDoS attacks, probing assaults, R2L attacks, and U2R attacks, to name a few. These kinds of threats are detrimental to the MANET. Consequently, a large toolkit of techniques is used to detect attacks and wipe them out of MANET. It is important to limit the possibility of attacks, and optimization is a major factor in doing so. This paper presents a new approach for intrusion detection using the Improved Whale Optimization Algorithm (IWOA) to select the most relevant features from the NSL-KDD dataset. The selected features are then used to train a Support Vector Machine (SVM) classifier for improved intrusion detection performance. Experimental results on the NSL-KDD dataset show that the proposed approach outperforms existing methods in terms of accuracy and efficiency.

Keywords: malware detection, classification, SVM, WOA, MANET, GA etc .

I.Introduction

Network Intrusion Detection (NID) is an important area of research in the field of cyber security. In recent years, the rise of Mobile Ad hoc Network (MANET) has posed new challenges for efficient and reliable NID systems. MANETs are self-configuring and self-organizing networks which are vulnerable to malicious attacks due to their open nature. Therefore, an efficient NID system is essential to protect these networks from intrusions.

In order to address these challenges, researchers have started exploring into intrusion detection algorithms that employ ML approaches. Analytical model building can be mechanised using the data-processing technique known as machine learning. Since ML has a proven track record of high detection rates [3], it has become increasingly attractive for use in cyber security, especially for the aim of detecting network anomalies. Naive Bayes (NB) [4, Random Forests (RF) [5, Decision Trees (DT) [6, K Nearest Neighbor (KNN) [7, and Logistics Regression (LR) [8] are just a few of the ML algorithms that have found use in IDS. Most of our focus in this research has been on the evolution and use of support vector machines (SVMs) [9], a subset of the larger family of machine learning algorithms. Support vector machines (SVMs) are a type of supervised machine learning model that can perform both classification and regression analysis. SVMs are exceptional in that they can learn and perform rather well even with very small sample sizes. SVMs excel in this particular area. As a result of their consistency, malleability, and efficiency on datasets of any size, they have gained widespread adoption, especially in the field of intrusion detection [10].

Previous works in MANETs have faced several problems such as the high false-positive rate and low detection rate which have limited the performance of NID systems. To overcome these problems, various optimization algorithms have been proposed to improve the detection rate.

Among them, the Whale Optimization Algorithm (WOA) has shown promising results in terms of accuracy and performance.

1.1 Contribution & Objective

This research paper provides an overview of the use of the improved Whale Optimization Algorithm-SVM to detect network intrusions in mobile ad-hoc networks (MANETs). MANETs are vulnerable to malicious attacks due to their lack of a centralized infrastructure and lack of authentication between nodes. Intrusion detection systems (IDS) are used to detect these malicious attacks and protect the network. Traditional IDSs are not effective in MANETs due to their dynamic topology and resource constraints. The Improved Whale Optimization Algorithm-SVM (IWOA-SVM) is a novel evolutionary algorithm that has been proposed to detect network intrusions in MANETs. The IWOA-SVM uses a whale optimization algorithm to optimize a support vector machine (SVM) model, which is trained with labeled data from the MANET. The IWOA-SVM improves the detection accuracy and reduces the false positives of the conventional SVM model. This paper presents a detailed study of the IWOA-SVM approach and provides an evaluation of its performance in terms of detection accuracy and false positive rate. The results of this study demonstrate the effectiveness of the IWOA-SVM approach in detecting network intrusions in MANETs.

1.2 Manuscript Organization

In the flow of this paper, we are discussing about dataset and whale optimization algorithm (WOA) in section 2 which is followed by the proposed work in section 3. The results are discussed in section 4 and work is concluded in section 5.

II. Preliminaries

2.1 NSL-KDD DATA SET

The NSL-KDD dataset is a dataset of network-based intrusion detection data. It was created as a data mining task to develop a model to detect and classify attacks on computer networks. The dataset consists of a total of 41 features and a total of 48984 instances. The features are a combination of quantitative and qualitative features and include characteristics such as duration, protocol type, service, flag, and source and destination IP addresses. The dataset includes four main types of attack: denial of service (DoS), unauthorized access from a remote machine (R2L), unauthorized access to local superuser (U2R) privileges, and probing. The DoS attack is a malicious attempt to make a machine or network resource unavailable. The R2L attack is an unauthorized access from a remote machine to another machine. The U2R attack is an unauthorized access to local superuser privileges. The probing attack is an attempt to gain information about a system. The NSL-KDD dataset is organized into two sets: a training set and a testing set. The training set consists of a total of 125973 instances and the testing set consists of a total of 22544 instances. Each instance consists of 41 features and one of the four attack types mentioned above. The dataset is available in the UCI Machine Learning Repository.

2.2 Whale Optimization Algorithm

The Whale Optimization Algorithm (WOA) is an optimization algorithm inspired by the movement patterns of humpback whales [2]. The algorithm is based on a competition and cooperation mechanism between the whales that mimics the foraging behavior of humpback whales. The algorithm is designed to solve optimization problems with a large number of decision variables and parameters.

The WOA algorithm consists of three main steps: initialization, exploration, and exploitation. In the initialization step, the initial population of solutions is generated using a random search. In the exploration step, the solutions are evaluated and compared based on a fitness function, and the best solutions are selected for further exploration. In the exploitation step, the selected solutions are combined and modified in order to improve the fitness of the resulting solutions. The algorithm iterates through these steps until a satisfactory solution is found.

In the encircling prey, the behavior of whale is mathematically presented as:

$$\vec{D} = \left|\vec{C}.\overrightarrow{X^*}(t) - \vec{X}(t)\right| \quad (2.1)$$

$$\vec{X}(t+1) = \overrightarrow{X^*}(t) - \vec{D}.\vec{A} \quad (2.2)$$

$\vec{C}$ & $\vec{A}$ are coefficients, $\overrightarrow{X^*}$ is the position vector of best position obtained so far and $\vec{X}$ is the position vector. The A & C are calculated as:

$$\vec{A} = 2.\vec{a}\vec{r} - \vec{a} \quad (2.3)$$

$$\vec{C} = 2.\vec{r} \quad (2.4)$$

where $\vec{a}$ decreased from 2 to 0. This keep the balance between the exploration and exploitation phase of the WOA.

The exploitation phase is conditionally dependent and mimics the behavior of shrinking the search space circle and updating the spiral position. Mathematically, it can be represented as:

$$\vec{X}(t+1) = \begin{cases} \overrightarrow{X^*}(t) - \vec{D}.\vec{A} & \text{if p} < 0.5 \\ \overrightarrow{X^*}(t) + \overrightarrow{D'}.e^{bl}.cos(2\pi l) & if\ p \geq 0.5 \end{cases} \quad (2.5)$$

Where p is a random value as the selection criteria to select the encircling or spiral position update. $\vec{A}$ selects the random value in between $[-a, a]$.

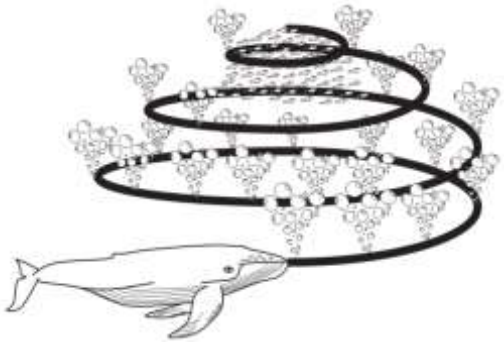

Fig. 1 Exploitation phase behaviour of humpback whales [2]

In the exploration phase, the position of the whale is updated on the basis of randomly selected search agent instead of best solution as in the exploitation phase. Mathematic depiction of this phase is as:

$$\vec{D} = \left|\vec{C}.\overrightarrow{X_{rand}}(t) - \vec{X}(t)\right| \quad (2.6)$$

$$\vec{X}(t+1) = \overrightarrow{X_{rand}}(t) - \vec{D}.\vec{A} \quad (2.7)$$

Here $\overrightarrow{X_{rand}}(t)$ is the random selected solution. The algorithmic steps of the WOA are:

1. Generate a random population of whales $X = \{X_1, X_2, X_3, \ldots, X_n\}$, where X_i represents the solution vector for the ith whale.
2. Evaluate the fitness of the whales in the population by calculating the objective function $F(X_i)$.
3. Calculate the distances between whales in the population by calculating the Euclidean distance $D(X_i, X_j)$ between each pair of whales X_i and X_j.
4. For each whale X_i, calculate its local search direction $A(X_i)$ using the equations 2.3 to 2.5 in the exploitation phase.
5. Move each whale X_i in the local search direction $A(X_i)$ to a new position $X_i(t+1)$.
6. Evaluate the new position $X_i(t+1)$ of the whale and calculate its new fitness $F'(X_i(t+1))$.
7. Replace the old solution with the new one if the new position has better fitness than the old one, i.e., if $F'(X_i) > F(X_i(t+1))$.
8. If none of the whales in the population has better fitness than the global best solution, then update the global best solution with the whale having the best fitness in the population.
9. Repeat steps 2 to 8 until a termination condition is met.

III. Proposed Work

We've broken this process down into five stages:

1. Consider the master one data set's training and test datasets in light of the NSL-KDD dataset.
2. When that was done, we used improved WOA (IWOA) to further refine our features. The primary goal is to better identify the attack.
3. Use the SVM classifier's reduce function to incorporate both training and test data.
4. Validate the trained SVM model's prediction accuracy on test data.
5. Contrast the GA-reduced feature findings with those of the suggested technique.

3.1 Feature Reduction using IWOA

In both the training and testing phases of data analysis, NSL-KDD makes use of a huge feature set consisting of 41 different components. It is necessary to preprocess the data in order to incorporate it into a machine learning (ML) model, as the data is currently only available in its raw version. All ML models can only communicate in numbers. Since data features

contain strings, we must first transform them into statistics format. Some variables have an infinite number of zeros, all of which are irrelevant to classification and so influence the training of the network. We then eliminate them from the set of features using an algorithmic selection process. Some characteristics have very large numerical values, while others have very small ones. The machine learning model is also skewed by this massive discrepancy. So features have to be normalized as:

$$normalised\ attribute = \sum_{i=1}^{n} \frac{f_i - \min(f)}{\max(f) - \min(f)} \quad (2.8)$$

Where n is number of attributes, f_i is the statistics value of i[th] attribute.

The data was reduced to 16 attributes after we eliminated the columns where 50% or more of the samples had zero values. We are at a loss as to which set of traits is most important, given that they have all thus far failed to improve accuracy. We chose an innovative optimization approach inspired by whale foraging techniques to achieve this end. Training a predictive model might take a long time if there is a lot of data involved and if a big number of features is used. This time drops once features have been chosen.

The IWOA is a recursive method that can maximise or minimise any objective function. The fundamentals of the Whale optimization algorithm were covered in the prior section. The framework of WOA is where the novel IWOA takes shape. The exploration step of the WOA is modifed to preent it to as IWOA. It can track the location of the whale's food in relation to its motion. The whale's location shifts between the largest and smallest possible search regions. The accuracy of the attack detection procedure will be enhanced by using IWOA. In the event of IWOA, the assault can be detected in less time.

Even though the Improved Whale optimization method and the fetaure selection algorithms are separate, they both function in a closed loop system. Figure 2 is a blog diagram depicting their exchanges.

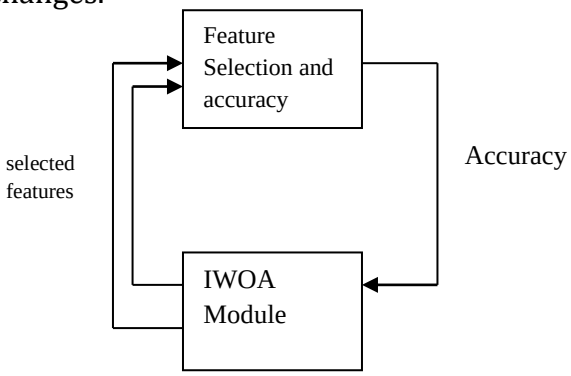

Fig. 2: Connection between feature-choice/machine-learning-selection and WOA-optimization

Both modules work in equilibrium; IWOA provides the input as a binary matrix to the ML module, while the ML module provides IWOA with the precision it needs in its own input. This binary matrix represents the set of characteristics that have to be included. In the matrix, the value "1" reflects the property that is selected, and the value "0" indicates that this feature is not selected. Through the process of training and testing the SVM model, the ML module determines the accuracy for this particular collection of chosen features. This accuracy is passed back to the IWOA module, which uses it to update the feature set based on the information. The accuracy is calculated as:

$$Accuracy\ (acc) = \frac{True\ Positives + True\ Negatives}{True\ Positives + False\ Positives + True\ Negatives + False\ Negatives} \quad (2.9)$$

This accuracy is the objective function used for the features selection b y IWOA.

$$f\big(x_i(t)\big) = acc(x_i(t), x_i'(t)) \quad (2.10)$$

Here $x_i'(t)$ is the ideal labelled data. The next section discusses the suggested improvement in WOA.

3.1.1 Improved WOA

The weak exploitation capabilities of the WOA was shown in the previous section. This is due to the fact that the algorithm constantly swaps out the current whale for a different one within the population, which may slow down the rate at which solutions converge to the optimal one. Because of this, the algorithm may need the more time to arrive to a better-optimized solution. In addition, local exploration inside the regions where whales are situated is limited, therefore the search for whales within regions may need numerous iterations. For this reason, it is important to do study within the parts of the whale that have not been explored by any other whales. This motivates the proposed approach, which seeks to improve the algorithm's exploitative capabilities in the vicinity of the

best-so-far answer. Mathematically, the exploitation phase in WOA can be updated as:

$$\vec{X}(t+1) = \begin{cases} \overrightarrow{X^*}(t) - \vec{D}.\vec{A} & \text{if p} < 0.5 \\ \overrightarrow{X^*}(t) + \overrightarrow{D'}.e^{bl}.cos(2\pi l) + \text{r} * (\vec{X}_{r1}(t) - \vec{X}_{r2}(t)) + (1-\text{r}) * (\overrightarrow{X^*}(t) - \vec{X}_{r3}(t)) & if\ p \geq 0.5 \end{cases} \quad (2.11)$$

Here $r1, r2, r3$ are the radomly selected three solutions in every iteration. r is a control parameter which controls the weightage of the both new additions in equation 2.11 for the spiral position update phase

IV. Results And Discussuion

For the purpose of developing a more effective intrusion detection system, we have recommended a thorough investigation into the use of the IWOA (Whale Optimization Technique) optimization algorithm for feature reduction (IDS). The algorithm is being tested in MATLAB, and it has been evaluated on the NSL KDD dataset. The results were generated through simulations run on a workstation equipped with 12 GB of RAM, an Intel i5 processor running at 2.4 GHz, and a 2 GB Nvidia graphics processing unit. The evaluation has concentrated on two types of attacks: denial-of-service attacks and intrusion-probe assaults. The evaluation criteria of accuracy, precision, recall, and specificity were applied to the results. For the state-of-the art comparison, we have compared the results with the conventional WOA and genetic algorthm in the same simulation enviornment.

To validate the statement that reduced number of features vs accuracy, we have calculated the accuracy with varying number of features for both DOS and probe attack. Figure 3 demonstrate the variation of the accuracy with respect to number of features. In figure 3(a), for the DOS attack, the accuracy is decreasing with the increae in number of features. There is sag in the accuracy at the 10 features' usage. A similar behavior can also be observed in figure 3(b) for probe attack. So there can be a tradeof here in accurcy and number of features. So, the optimization algorithm works well at this point.

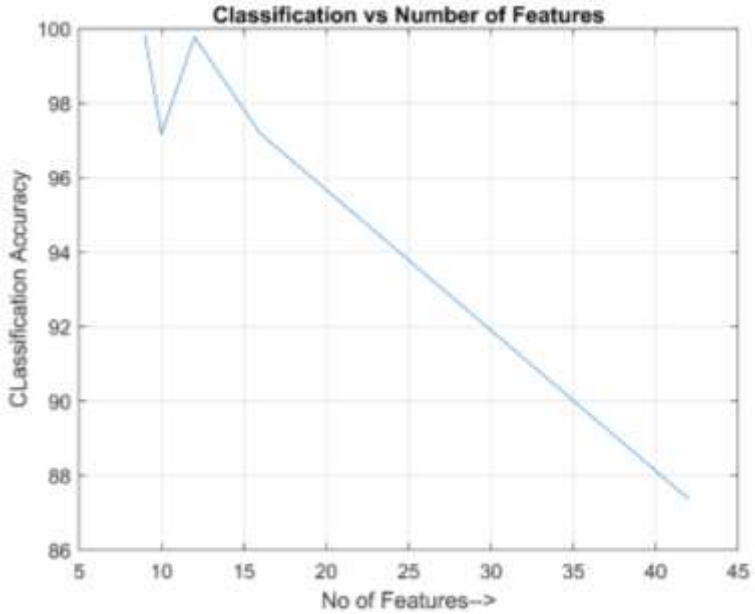

Figure 3(a): Accuracy vs number of features for DOS attack in NSL-KDD dataset

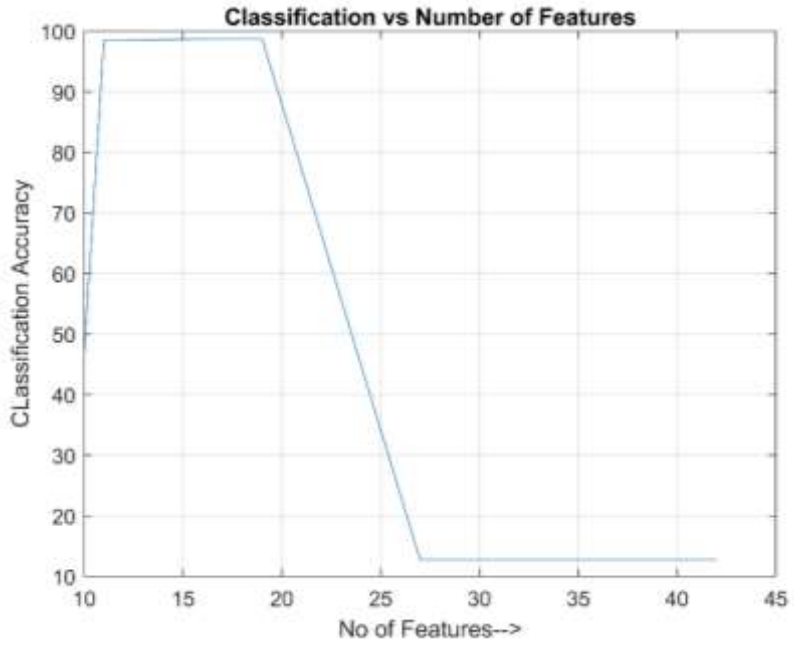

Figure 3(b): Accuracy vs number of features for Probe attack in NSL-KDD dataset

To evaluate the performance of the otimization algorithm, the convergence curve is to plotted. Since the objective function is the accurcy for both proposed IWOA and WOA, the convergece curve must be inreasing with iterations and must be converging after few iterations. Figure 4 shows the convergence curve for both probe and DOS attack. The convergenc plot is higher by the propsed feature seletion algorithm for both cases.

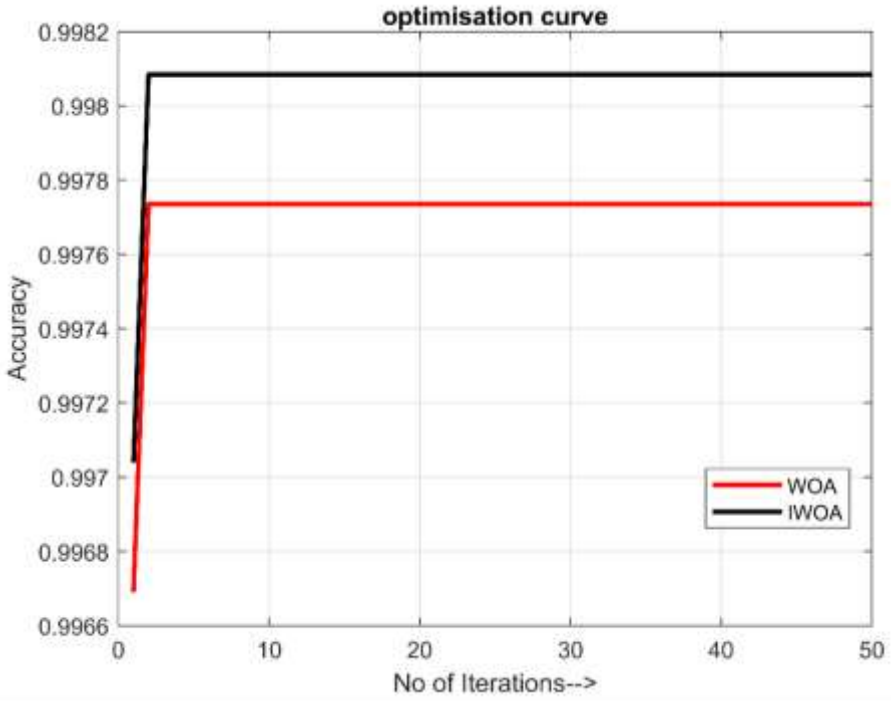

Figure 4(a): Convergence curve plot for the DOS attack by the proposed IWOA and WOA

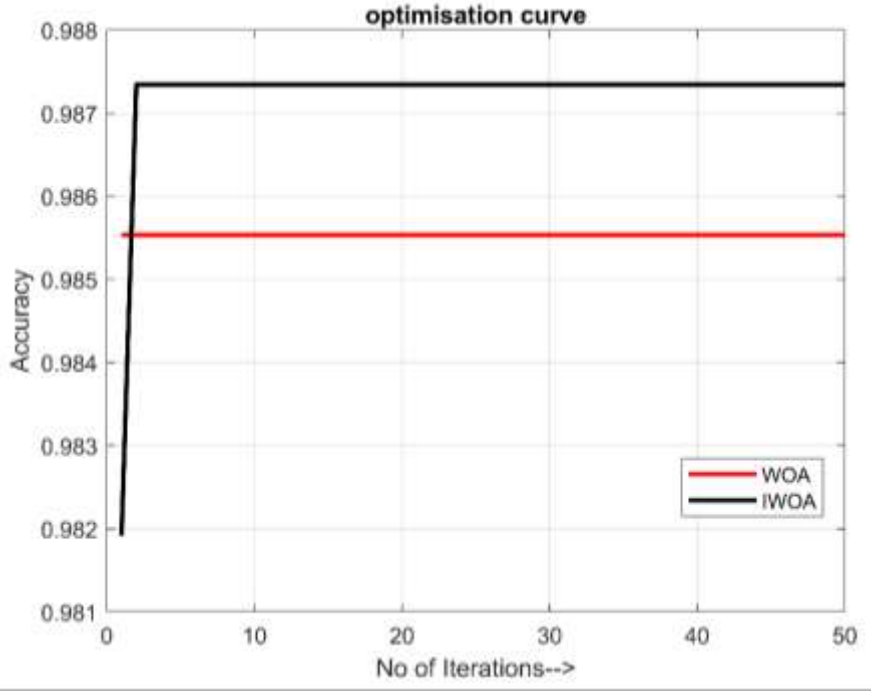

Figure 4(b): Convergence curve plot for the probe attack by the proposed IWOA and WOA

In the next part of this section, we will discuss the cases separately.

Case-1 Denial of Service (DOS) Attack

DOS attacks are further broken down into six kinds in the NSL-KDD dataset: back, land, Neptune, smurf, pod, and teardrop. In each attack type, we are comparing the accuracy for the complete set of features trained on the SVM, preprocessed features to remove the number of zeros and undefined values in the dataset, reduced features by IWOA, WOA and GA. A pie chart in figure 5 is plotted to correlate the number of features and accuracy for DOS attack. It can be observed from the chart that for the 9 selected features by the proposed method, the accuracy is highest amongst all cases. Nonetheless, the comparable accuracy is also observed by the WOA selected features, however, the number of features are 16 which is much resource consuming than IWOA. All 42 features are the least performing set of features.

The alignment of the feature with true labelled data is shown in figure 6 to present a visualization of the attack detection. The detected attacks are perfectly aligning with the original labels.

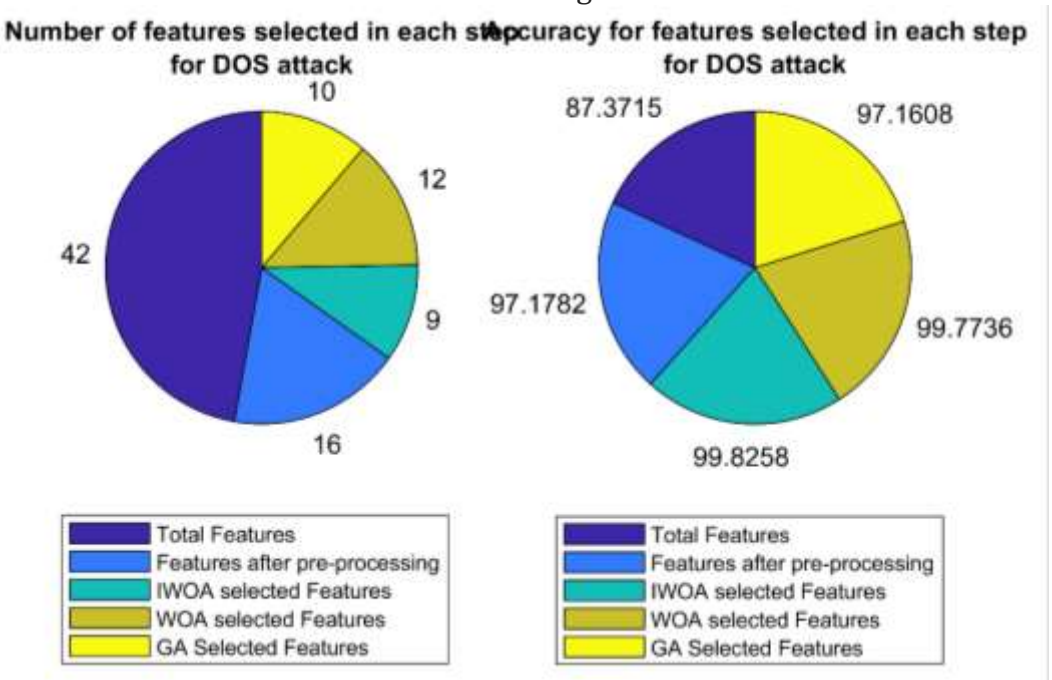

Figure 5: Pie chart comparison for the accuracy vs features for the DOS attack

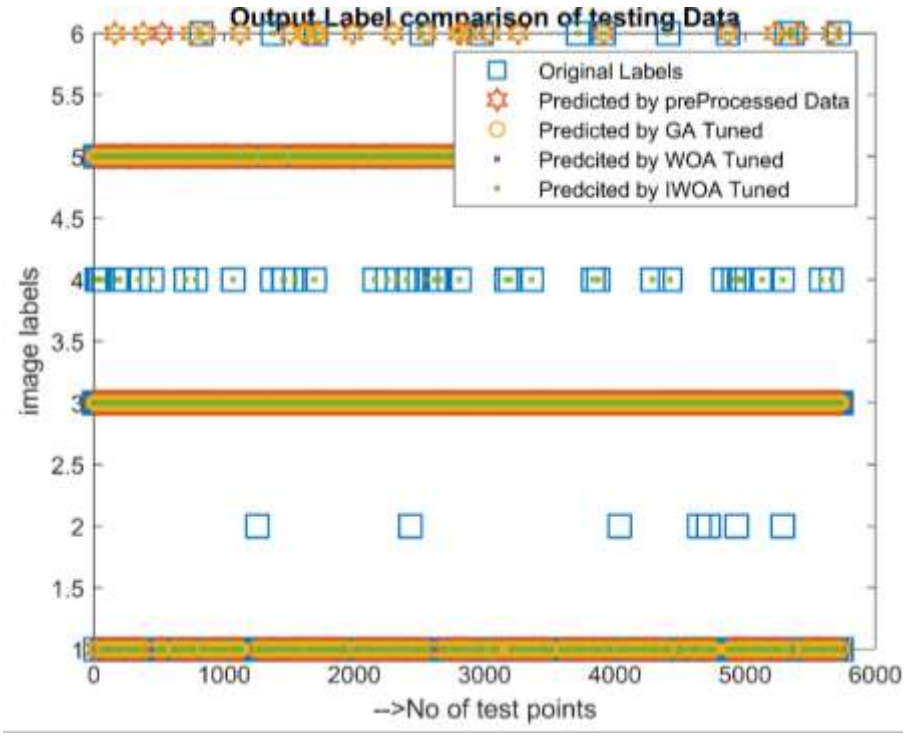

Figure 6: Output label comparison for the IWOA, WOA and GA

The evaluation on the four parameters as discussed above is presented as comparative bar chart with the state-of-the-art algorithm in figure 7. The proposed scheme is performing best amongst all algorithms.

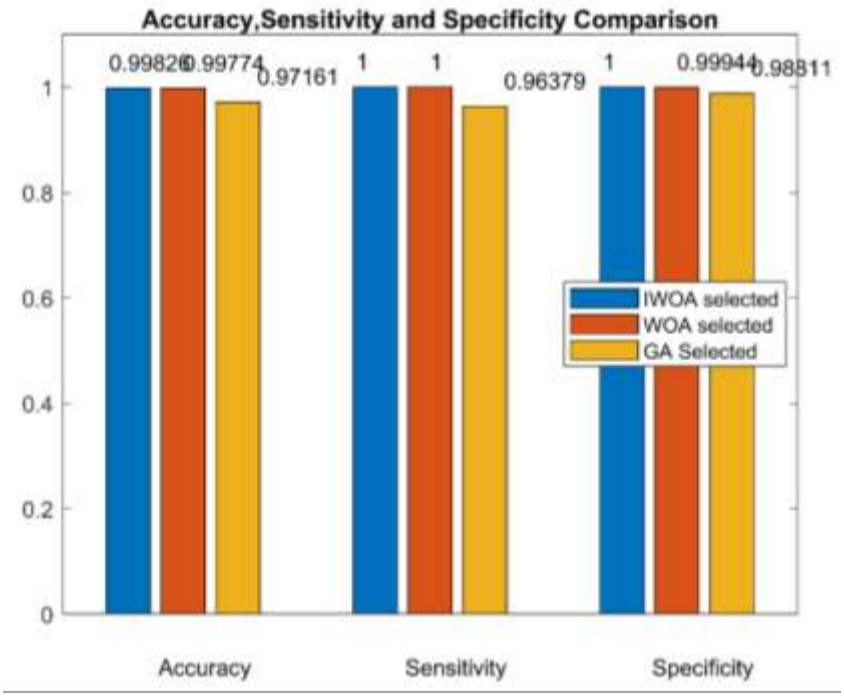

Figure 7: Accuracy, sensitivity and specificity comparison of GA and WOA based method

Case-2 Probe Attack

In the case of probe attack, the selected features are lesser than state-of-the-art algorithms and plotted as pie chart in figure 8.

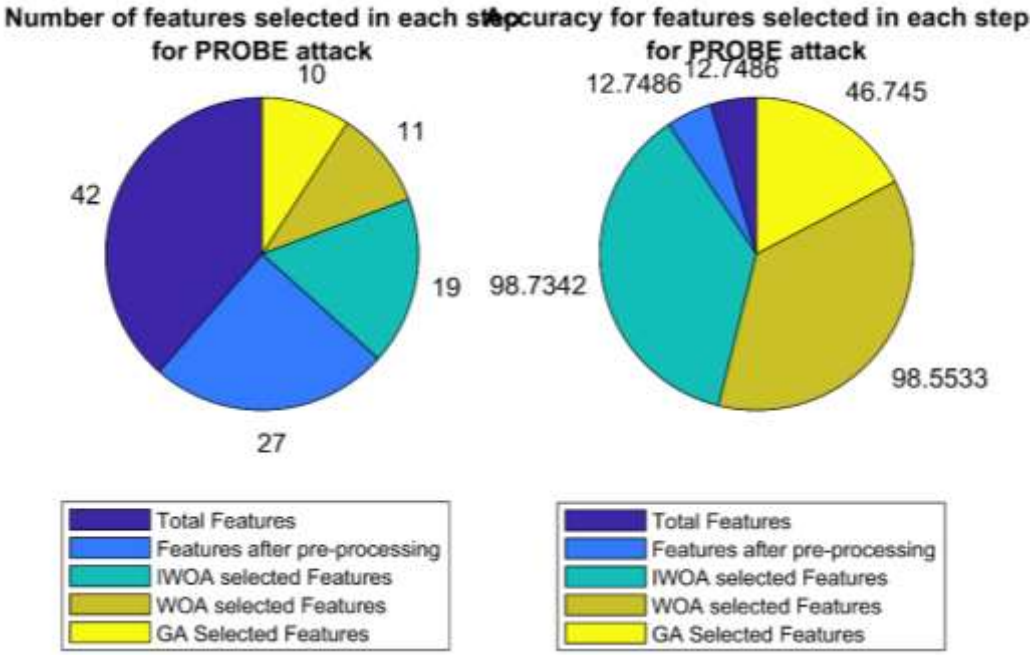

Figure 8: Pie chart comparison for the accuracy vs features for the Probe attack

V. Conclusion

The research paper has presented an improved Whale Optimization Algorithm-SVM (IWOA-SVM) for network intrusion detection in Mobile Ad-hoc Networks (MANETs). The proposed IWOA-SVM has shown better performance than existing Whale Optimization Algorithm (WOA) by selecting 4.75% less features and achieving an accuracy of 99.7736%. This improved accuracy is primarily due to the proposed IWOA-SVM's ability to select optimal features. This eliminates the need for feature selection techniques such as Wrapper, Filtering, and Embedded methods. The proposed method has shown promising results and can be used as a reliable algorithm for network intrusion detection in MANETs. Moreover, the IWOA-SVM can be used in other domains such as image processing, text classification, and face recognition. It can also be used to detect various types of cyber-attacks such as Distributed Denial of Service (DDoS) attacks, Phishing attacks, and Password Guessing attacks. Overall, the proposed IWOA-SVM can be used as an efficient and reliable algorithm for network intrusion detection in MANETs. The proposed algorithm has shown better accuracy than existing.

References

1. Z. Ullah, M. S. Khan, I. Ahmed, N. Javaid and M. I. Khan, "Fuzzy-Based Trust Model for Detection of Selfish Nodes in MANETs," *2016 IEEE 30th International Conference on Advanced Information Networking and Applications (AINA)*, Crans-Montana, 2016, pp. 965-972.
2. M. A. Abdelshafy and P. J. B. King, "Dynamic source routing under attacks," *2015 7th International Workshop on Reliable Networks Design and Modeling (RNDM)*, Munich, 2015, pp. 174-180.
3. C. Alocious, H. Xiao and B. Christianson, "Analysis of DoS attacks at MAC Layer in mobile adhoc networks," *2015 International Wireless Communications and Mobile Computing Conference (IWCMC)*, Dubrovnik, 2015, pp. 811-816.
4. A. Quyoom, R. Ali, D. N. Gouttam and H. Sharma, "A novel mechanism of detection of denial of service attack (DoS) in VANET using Malicious and Irrelevant Packet Detection Algorithm (MIPDA)," *International Conference on Computing, Communication & Automation*, Noida, 2015, pp. 414-419.
5. A. M. Shabut, K. P. Dahal, S. K. Bista and I. U. Awan, "Recommendation Based Trust Model with an Effective Defence Scheme for MANETs," in *IEEE Transactions on Mobile Computing*, vol. 14, no. 10, pp. 2101-2115, Oct. 1 2015.
6. A. Menaka Pushpa and K. Kathiravan, "Resilient PUMA (Protocol for Unified Multicasting through Announcement)

against internal attacks in Mobile Ad hoc Networks," *2013 International Conference on Advances in Computing, Communications and Informatics (ICACCI)*, Mysore, 2013, pp. 1906-1912.

7. M. A. Abdelshafy and P. J. B. King, "Analysis of security attacks on AODV routing," *8th International Conference for Internet Technology and Secured Transactions (ICITST-2013)*, London, 2013, pp. 290-295.
8. A. M. Kurkure and B. Chaudhari, "Analysing credit based ARAN to detect selfish nodes in MANET," *2014 International Conference on Advances in Engineering & Technology Research (ICAETR - 2014)*, Unnao, 2014, pp. 1-5.
9. S. Biswas, P. Dey and S. Neogy, "Trusted checkpointing based on ant colony optimization in MANET," *2012 Third International Conference on Emerging Applications of Information Technology*, Kolkata, 2012, pp. 433-438.
10. D. Das, K. Majumder and A. Dasgupta, "A game-theory based secure routing mechanism in mobile ad hoc network," *2016 International Conference on Computing, Communication and Automation (ICCCA)*, Noida, 2016, pp. 437-442.
11. T. Poongothai and K. Duraiswamy, "Intrusion detection in mobile AdHoc networks using machine learning approach," *International Conference on Information Communication and Embedded Systems (ICICES2014)*, Chennai, 2014, pp. 1-5.
12. D. A. Varma and M. Narayanan, "Identifying malicious nodes in Mobile Ad-Hoc Networks using polynomial reduction algorithm," *2016 International Conference on Electrical, Electronics, and Optimization Techniques (ICEEOT)*, Chennai, 2016, pp. 1179-1184.
13. Bandana Mahapatraa and Prof.(Dr) Srikanta Patnaik, "Self Adaptive Intrusion Detection Technique Using Data Mining concept in an Ad-Hoc Network," 2nd International Conference on Intelligent Computing, Communication & Convergence(ICCC-2016)
14. Manjula C. Belavagi and BalachandraMuniyal, "Performance Evaluation of Supervised Machine Learning Algorithms for Intrusion Detection," 12th International Multi-Conference on Information Processing-2016 (IMCIP-2016).
15. PreetiAggarwala and Sudhir Kumar Sharmab, "Analysis of KDD Dataset Attributes - Class wise For Intrusion Detection," 3rd International Conference on Recent Trends in Computing 2015 (ICRTC-2015).
16. Ciza Thomas, Vishwas Sharma and N. Balakrishnan, "Usefulness of DARPA Dataset for Intrusion Detection System Evaluation
17. P.Natesan and P.Balasubramanie, "Multi Stage Filter Using Enhanced Adaboost for Network Intrusion Detection," International Journal of Network Security & Its Applications (IJNSA), Vol.4, No.3, May 2012
18. M. Tavallaee, E. Bagheri, W. Lu and A. A. Ghorbani, "A detailed analysis of the KDD CUP 99 data set," *2009 IEEE Symposium on Computational Intelligence for Security and Defense Applications*, Ottawa, ON, 2009, pp. 1-6.
19. Seyedali Mirjalili, Andrew Lewis, "The Whale Optimization Algorithm," Advances in Engineering Software, Volume 95, 2016, Pages 51-67, ISSN 0965-9978