International Journal of Computer Information Systems and Industrial Management Applications
ISSN 2150-7988 Volume 18(17s) 2026 pp. 611-619
https://doi.org/10.70917/ijcisim-2026-4786
Article

ENHANCING INTRUSION DETECTION SYSTEMS' ROBUSTNESS FROM ADVERSARIAL ATTACKS USING METAHEURISTIC ALGORITHMS

Kanak Giri[1], Pankaj Dadheech[2], Mukesh Kumar Gupta[3]

[1]Department of Computer Science and Engineering, Swami Keshvanand Institute of Technology, Management & Gramothan (SKIT), Jaipur, Rajasthan-302017, Email: kanakgiri88@gmail.com, ORCID 000000200580776
[2]Department of Computer Science and Engineering, Swami Keshvanand Institute of Technology, Management & Gramothan (SKIT), Jaipur, Rajasthan-302017 pankajdadheech777@gmail.com ORCID 0000-0001-5783-1989
[3]Scientist-F & HOD, Data Governance & Strategy Division, National Informatics Section (NIC),New Delhi-110003 (India) drmukesh5766@gmail.com ORCID 0000-0002-4907-0259

Abstract: Intrusion Detection Systems (IDS) play a crucial role in safeguarding computer networks from malicious activities. The increasing prevalence of malicious attacks on IDS bases systems has posed a significant threat to their robustness and effectiveness. It has become essential to build strong defense procedure to confirm the transparency and dependability of these systems. This research introduces an efficient methodology that incorporates ensemble learning technique like GBM, genetic operators, DNN-Deep Neural Network and Particle Swarm Optimization (PSO) with defense system to enhance the robustness of IDS against non-friendly system attacks. In order to maximize the selection of the most appropriate features from the given dataset, the suggested methodology starts with a feature engineering stage that makes use of PSO and GBM. In order to create a robust and discriminative feature subset, genetic operators are then used to further recheck the process of selection using different features.
Keywords: Intrusion detection system, GBO, PSO, GO (Genetic Operators), DNN (Deep Neural Network), NSL-KDD Dataset

1. Introduction

In IDS, learning algorithms play a significant role. Using predetermined criteria, machine learning techniques classify system events or network traffic as either malicious or lawful. Two deep learning-based approaches, convolutional neural networks (CNN) and recurrent neural networks (RNN), make use of neural networks' capacity to recognize complex patterns and identify anomalies in network traffic or system behavior. Despite these advancements, IDS still requires assistance in accurate detection of intricate and dynamic threats, such adversarial attacks. Adversarial assaults are created expressly to trick IDS and avoid detection. They can evade conventional detection methods by altering or manipulating network traffic and behavior of the system.

The intentional alteration of system behavior or network traffic with the goal of tricking IDS and avoiding detection is known as an adversarial attack. Adversarial attacks are made to take advantage of IDS flaws and get around their detection systems. The two main types of adversarial attacks are poisoning attacks and evasion attacks [5]. employed in order to neutralize the impact of adversarial attacks on IDS are comprehensively outlined in this section.

The below table (Table-1) summarizes the study which includes the important conclusions of various techniques based on adversarial assaults from the perspectives of deep learning and machine learning. Injecting the

malicious data in order to spoof the system is a part of poisoning attack which may degrade and intrude the systems' resilience.

In order to minimize the effects of adversarial attacks, researchers have suggested a number of defense strategies. These include assemble techniques that integrate several models to increase detection accuracy and selection of effic9ient feature, where features are carefully selected to be resistant to manipulation of various adversarial threats [12].

1.1 Adressing adversarial attacks based on various existing techniques

As adversarial assaults on intrusion Detection Systems (IDS)get more complex, researchers have developed a number of strategies to address these challenges. These techniques aim to improve the accuracy of attack detection as well as the IDS's robustness and resilience to adversarial assaults [8, 11]. The techniques currently.

Table-1: Comparative study of various methodologies for IDS

Work by	Approach	Important conclusions
Akhtar et al. (2021) [6]	Investigated adversarial threats	There are serious security dangers associated with adversarial attacks.
Wang et al. (2021) [7]	Focus area Attacks: GAN-based	Highly effective GAN based examples can be obtained
Liu et al. (2021) [8]	Examined attacks- Physical	Physical domains are also attack prone by adversial attacks
Luo et al. (2020) [9]	Examined transferability	Universally transferrable of these attacks is complex
Sitawarin et al. (2022) [10]	Examined poisoning attacks	Deep learning models can easily be avoided by black box threats
Tramèr et al. (2017) [11]	Examined- adversarial training	Model robustness against attacks can be improved by adversarial training
Xu et al. (2021) [12]	Examined methods based on defence mechanism	Robustness is increased by combining several defense strategies

2. PROPOSED APPROACH

2.1 An outline of the suggested methodology

In this study, we provide a unique approach to addressing the issues that adversarial attacks cause in various IDS based systems. By incorporating cutting-edge methods from the machine learning and cyber security fields, the suggested method seeks to enhance the robustness and effectiveness of IDS models. I hope to develop a robust and efficient defense methodology that can identify and mitigate adversary attacks more accurately and effectively by utilizing the advantages of these strategies. The proposed approach flow chart and suggested algorithm consist of a number of essential elements and procedures are described below:

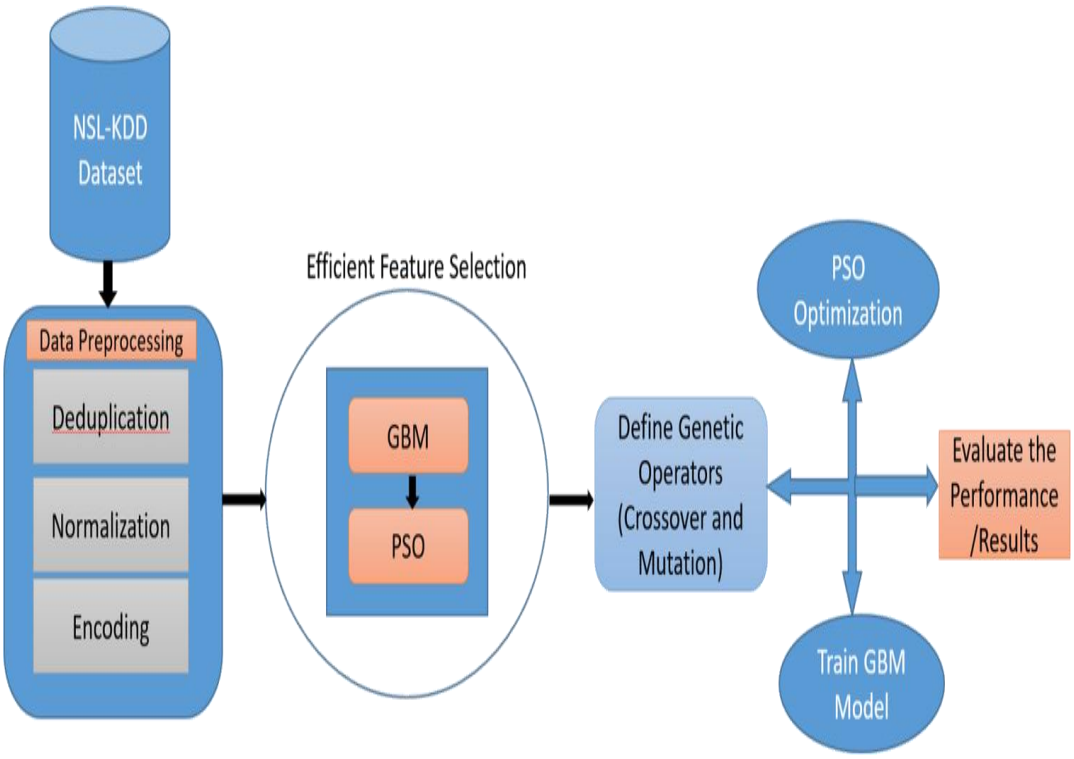

Fig 1: Flow of Proposed Approach

Algorithm 1: Pseudo code of Proposed

Approach Start

1. Insert/Upload: Dataset (NSL-KDD)
2. Dataset preprocessing using techniques like normalization, feature encoding, data cleaning
3. Divide the dataset into labels (y) and features (X)
4. Divide the data: test & training sets

5. Describe function (fitness) used to assess the particle's efficacy:
   a. Based on the particle's location, set the chosen features
   b. Set up the GBM classifier and train it using the chosen features
   c. Analyze the GBM model's accuracy using the validation set
   d. Fitness value: returned as accuracy value

6. Describe: function particle swarm optimization:
   a. Establish the PSO's parameters, such as the number of particles, number of iterations, IW (inertia weight), and acceleration parameters
   b. Optimizer generation such as Global-Best-PSO

c. To optimize PSO is used to select features, Use the fitness function as the objective function and the combined iterations which comprises maximum number of iterations when calling the optimizer's optimize() method.
d. Obtain the optimizer's best features and accuracy
e. Apply genetic operations (mutation and crossover) to the finest traits

7. Describe genetic operations- crossover & mutation:
a. Apply crossover:
i. Choose 2 main particles at random
ii. Create a child particle by performing a crossover operation, such as a one-point crossover
b. Apply mutation:
i. Repeat for every particle feature
ii. Execute a mutation operation (such as flipping a bit) on the feature if the mutation probability is satisfied

8. Establish the parameters of the genetic algorithm such as the crossover and mutation probabilities

9. Execute the particle swarm optimisation to choose the efficient parameters:
a. Use the optimisation (PSO) function to find the best accr and efficient parameters

10. Use the chosen features to train the GBM model:
a. Take the marked elements out of the training set.
b. Set up the GBM classifier and train it using the chosen features

11. Analyze the GBM model's performance using test set:
a. Take the marked features out of the test set
b. Utilize the supervised GBM model to make predictions
c. Determine accr and additional asses metrics {where accr variable stands for accuracy}

12. Produce the outcomes:
a. Produce: marked features
b. Produce: the highest accuracy attained
c. Produce the test set's final accuracy

Stop

3. EXPERIMENTAL SETUP

3.1 Proposed approach- Implementation

To achieve reliable and accurate intrusion identification on NSL-KDD dataset, the suggested approach's implementation requires a number of essential elements and configurations. The implementation takes into account the following factors:

- **Machine Learning Framework:** The suggested method is implemented using an appropriate ML based framework like TensorFlow or PyTorch [2, 5, 10].
- **Configuration of Gradient Boosting Machine:** To maximize its effectiveness on the dataset, the GBM method is set up with the appropriate hyperparameters. **Architecture of DNN:** The architecture based on DNN employed in the model based on IDS is meant to capture the complex patterns and features of the NSL-KDD dataset.

- **Integration- Defence Mechanisms:** Several defense mechanisms are included in the recommended approach to strengthen the IDS model's resilience to attacks.
- **Training Approach:** The NSL-KDD dataset which has been preprocessed [9, 12] is fed into the model as part of the training phase, and an optimization algorithm is used to update the model's parameters iteratively.

The Code is given the sample values listed below:

PSO parameters:
There are twenty (20) particles (n_particles)
50: Total iterations (n_iterations)
Weight of Inertia (w): 0.8
Parameters- Cg (c1): 1.4
Parameters- Sc (c2): 1.4
{Where Cg stands for cognitive and Sc stands for social.}

$$v_i^{t+1} = \underbrace{v_i^t}_{\text{Inertia}} + \underbrace{c_1 r_1 (pbest_i^t - p_i^t)}_{\text{Personal influence}} + \underbrace{c_2 r_2 (gbest^t - p_i^t)}_{\text{Social influence}}$$

Parameters- Gradient Boosting Machine algo:
Learning Rate: 0.2
Estimators number: 100
Maximum depth: 4
Performance measures: Acc

Population of pop_size = 52
Maximum Iterations (max_iter) = 100
Total available parameters (num_features) = 41
Total features to be selected (num_selected_features) = 11

Iteration Termination: When reached Maximum iterations

3.2 Performance measures and evaluation metrices

A range of evaluation metrics and measurements of performance are used to evaluate the efficacy and effectiveness of the suggested approach in relation to the NSL-KDD dataset [7, 11]. These metrics give information about the capacity of model in order to identify cyber breaches, malicious attacks on network, and uphold a high degree of robustness.

- **Accuracy:** When it divides the combined number of occurrences by the true negative & true positive predictions, accuracy is calculated.

$$\text{Accuracy} = \frac{(TN+TP)}{FP+FN+(TN+TP)}$$

- **Precision:** Precision represents the proportion of correct positive forecasts among all instances where a favorable result was anticipated.

$$\text{Precision} = \frac{(TP+FP)}{(TP)}$$

- **Recall (Sensitivity):** Recall, The percentage of genuine positives that the model accurately recognized is sometimes referred to as sensitivity or the true positive rate.

$$\text{Recall} = \frac{(FN+TP)}{(TP)}$$

- **F-score:** Harmonic mean of recall and precision generates the value of F1-score.

$$\text{F1 Score} = \frac{2^* (\text{Recall} * \text{Precision})}{(\text{Recall} + \text{Precision})}$$

4. ANALYSIS OF RESULTS

Here in the below section, the suggested methodology is compared and contrasted with current intrusion detection and adversarial attack defense strategies. This analysis shows the benefits of the suggested method over alternative approaches and assesses its superiority. The findings shed light on how well the model performs in identifying intrusions, thwarting hostile attacks, and preserving a high degree of robustness.

The proposed technique's performance is compared with that of current methodologies based on various previously mentioned

assessment metrics. The presentation focuses on the significant improvements that the proposed technique achieved in terms of

accuracy, intrusion detection rates, FP reduction, and resilience opposed to attackers. The findings demonstrate where the

recommended methodology is better than contemporary approaches. The below table-2 shows the values of performance matrices against various methodologies:

Table 2: Comparative Performances: Various Existing & Proposed Method

Methodologies	**Accuracy**	**Precision**	**Recall**	**F-Score**
ANN: Artificial Neural N/W	91.17	92.84	91.14	92.08
SVM	94.76	92.44	93.64	94.73
NB- Naïve Bayes	86.84	84.96	87.54	84.38
Genetic operators and fuzzy classifier	88.56	86.86	86.32	86.77
Support vector and particle swarm	91.35	92.90	88.72	91.79
Ensemble based IDS using two level classifier	92.68	92.54	91.64	91.46
Using semi supervised extreme ML	86.46	86.83	85.55	87.05
Using KNN method	83.42	86.88	87.89	87.68

Proposed Method	**98.34**	**95.86**	**95.56**	**96.56**

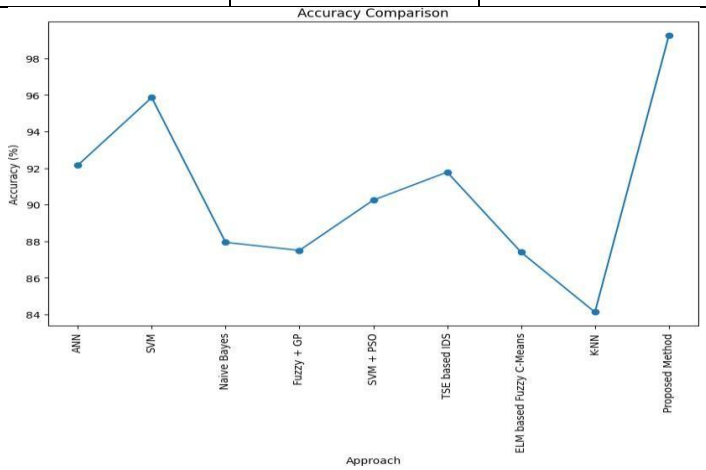

Fig 2: Line Graph (Accuracy Comparison)

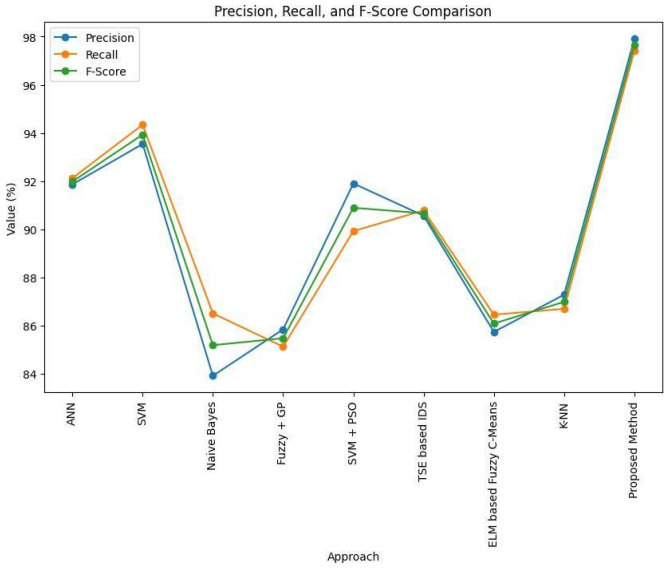

Fig 3: Line Graph (Precision, Recall, F-Score Comparison)

The above graphs shows the value of accuracy (Fig-2) which is very high in the proposed methodology and that makes the defence system more resilient towards the adversial attacks and in the next graph other values are showing the best outcomes as well for precision, f-score and recall (Fig-3). Below figure (Fig-4) shoes the bar graph comparison between various evaluation matrices of other and proposed methodology.

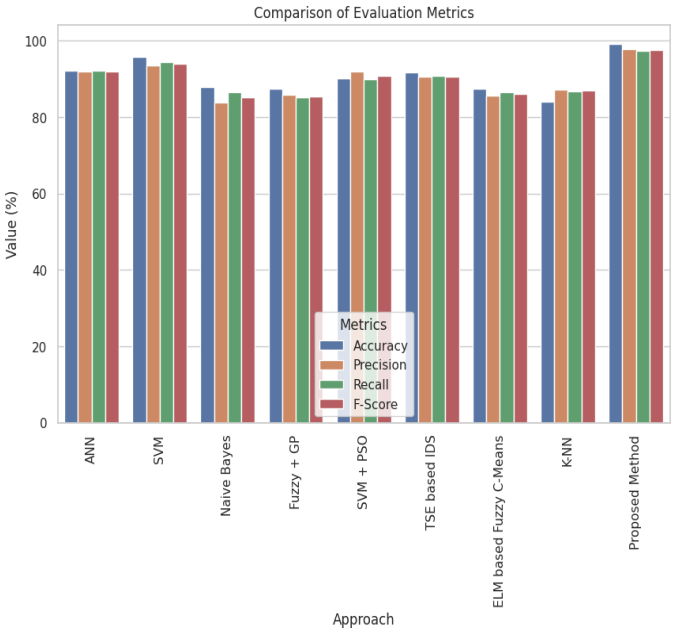

Fig 4: Grouped Bar Graph (Comparison of Evaluation Metrics)

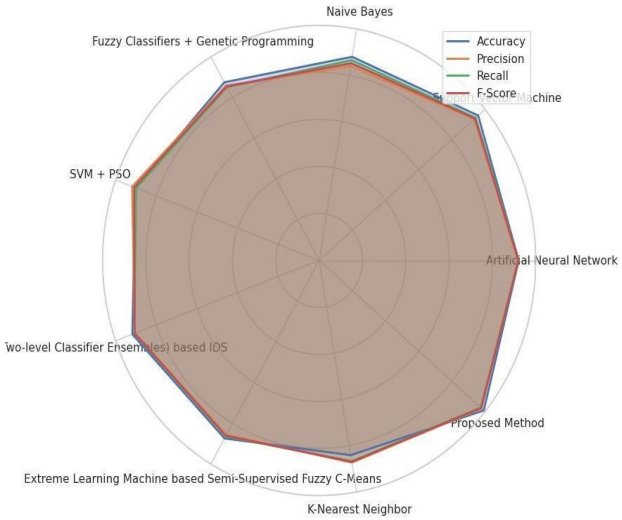

Fig 5: Radar chart (Comparison of various ML classifiers with proposed method)

The above radar chart (Fig-5) visually compares the performance of several machine learning classifiers and ensemble methods on four metrics: Precision**,** Accuracy, F-Score and Recall. All methods perform very similarly, with

their performance lines nearly perfectly overlapping, suggesting consistent evaluation across the metrics for each model.

"Proposed Method", along with TSE (Two-level Classifier Ensembles) based IDS, Machine learning based Semi-Supervised Fuzzy C-Means, and Artificial Neural Network, shows a slightly more regular, almost circular or octagonal, shape in the plot, implying balanced scores for all four metrics.

5.CONCLUSION

An effective mechanism for intrusion detection and defense against harmful attacks in cyber systems has been proposed in this research project. The goal of the study was to identify the shortcomings of current approaches and enhance the robustness & precision of IDS based systems during any malicious assault activity. The field of intrusion detection and network system defense counter malicious attacks has benefited greatly from this study. The accomplishments and contributions of this work significantly advance the fields of intrusion detection and adversarial attack defense. The suggested method shows promise for improving network system security and dependability. The results lay the groundwork for future investigations and developments in the field of intrusion detection and network security. The study has contributed significantly to the corpus of information already in existence and acts as a springboard for further advancements in the discipline.

- Comparative analysis: A comprehensive comparison with cutting-edge methods in the field is part of the study project. The comparative study highlights the benefits and advancements made possible by the suggested strategy. It emphasizes how combining PSO, GBM, genetic operators, and defense mechanisms results in better performance, robustness, and detection rates.

Research insights: The findings and observations of the study provide important new insights on how to handle IDS attackers. The proposed approach offers new perspectives on mitigating the negative impacts of adversarial attacks on N/W safety and also simultaneously creating ID

References:

1. T. Das, R. M. Shukla and S. Sengupta, "Poisoning the Well: Adversarial Poisoning on ML-Based Software-Defined Network Intrusion Detection Systems," in IEEE Transactions on Network Science and Engineering, vol. 12, no. 1, pp. 252-262, Jan.-Feb. 2025, doi: 10.1109/TNSE.2024.3492032
2. Dhumal, Chaitrali T. and Pingale, Dr. S. V., Analysis of Intrusion Detection Systems: Techniques, Datasets and Research Opportunity (March 6, 2024). Proceedings of the International Conference on Innovative Computing & Communication (ICICC 2024), Available at SSRN: https://ssrn.com/abstract=4749820 or http://dx.doi.org/10.2139/ssrn.4749820
3. Z. Azam, M. M. Islam and M. N. Huda, "Comparative Analysis of Intrusion Detection Systems and Machine Learning-Based Model Analysis Through Decision Tree," in IEEE Access, vol. 11, pp. 80348-80391, 2023, doi: 10.1109/ACCESS.2023.3296444.
4. Pranpaveen Laykaviriyakul, Ekachai Phaisangittisagul, Collaborative Defense-GAN for protecting adversarial attacks on classification system, Expert Systems with Applications,Volume 214,2023,118957,ISSN 0957-4174, https://doi.org/10.1016/j.eswa.2022.118957.
5. Alotaibi, A.; Rassam, M.A. Adversarial Machine Learning Attacks against Intrusion Detection Systems: A Survey on Strategies and Defense. Future Internet 2023, 15, 62. https://doi.org/10.3390/fi15020062
6. Akhtar, N., Hussain, M.Z., & Sohail, A. (2021). Adversarial attacks and defenses in deep learning: A comprehensive survey. Journal of Ambient Intelligence and Humanized Computing, 12(5), 6349-6380.
7. Wang, Y., Zhang, Y., & Qi, G.J. (2021). Generative adversarial attacks: An overview. ACM Computing Surveys, 54(5), 1-34.
8. Liu, B., Xiao, J., & Chen, C.L.P. (2021). Adversarial attacks and defenses in deep learning for computer vision: A survey. ACM Transactions on Multimedia Computing, Communications, and Applications, 17(4), 1-23S.
9. Kumar, S. Gupta and S. Arora, "Research Trends in Network-Based Intrusion Detection Systems: A Review," in IEEE Access, vol. 9, pp. 157761-157779, 2021, doi: 10.1109/ACCESS.2021.3129775.
10. Sitawarin, C., Chaisopon, N., & Damrongrat, C. (2022). A survey on black-box adversarial attacks and defenses for deep learning. Information Fusion, 75, 244-264.
11. Tramèr, F., Kurakin, A., Papernot, N., Boneh, D., & McDaniel, P. (2017). Ensemble adversarial training: Attacks and defenses. arXiv preprint arXiv:1705.07204.
12. Xu, Y., Yin, J., Cao, Z., Lu, Z., & Zhang, J. (2021). A comprehensive survey of GAN-based adversarial attacks. Information Fusion, 74, 37-59.