International Journal of
DATA SCIENCE AND IOT MANAGEMENT SYSTEM
ISSN: 3068-272X www.ijdim.com Original Research Paper

SUPERVISED MACHINE LEARNING-BASED NETWORK INTRUSION DETECTION WITH OPTIMIZED FEATURE SELECTION

[1]E.AJITH KUMAR, [2]SAGI AKANKSHA

[1]Assistant Professor & TPO, CSE, Tallapadmavathi College of Engineering, Somidi, Kazipet, Hanumakonda – 506003.Email-id: ajithkumartpce@gmail.com.

[2]Research Scholar, H.no: 24UC1D5811, CSE, Tallapadmavathi College of Engineering, Somidi, Kazipet, Hanumakonda – 506003,Email-id: akanksharaosagi14@gmail.com.

ABSTRACT

This study proposes a supervised machine learning system to classify network traffic as malicious or benign. By combining feature selection with learning algorithms, the model optimizes detection accuracy. Experiments on the NSL-KDD dataset show that an Artificial Neural Network (ANN) with wrapper-based feature selection outperforms Support Vector Machine (SVM) techniques. Comparative results demonstrate that the proposed approach improves intrusion detection success rates, offering a more efficient and reliable solution for network security.

Index Terms:-Intrusion Detection, Machine Learning, Deep Learning, Neural Networks, Support Vector Machine, Feature Selection.

Received: 23-09-2025 Accepted: 27-10-2025 Published: 03-11-2025

1. INTRODUCTION

The rapid expansion of internet usage and online content has led to a corresponding rise in cybercrime [1–2]. Intrusion detection is a critical first line of defense against such attacks, and security solutions including Firewalls, Intrusion Detection Systems (IDS), Unified Threat Management (UTM), and Intrusion Prevention Systems (IPS) have become essential research and operational priorities. IDSs detect potential security breaches by collecting and analyzing data from various systems and network sources [3].

Network-based IDS operate using two main approaches: signature-based and anomaly-based detection. While signature-based systems have seen widespread commercial adoption, anomaly-based detection remains an active area of research due to its ability to identify novel attacks [4–5]. Anomaly detection, however, presents significant challenges, as it must differentiate between normal and malicious traffic without prior knowledge of new attack patterns. To overcome these challenges, machine learning techniques have increasingly been applied in recent years [6].

Despite their utility, IDSs cannot address all security vulnerabilities, such as weak authentication or protocol flaws. The study of intrusion detection began in the 1980s, with the first model introduced in 1987 [7]. Although significant research and investment have occurred since then, intrusion detection technology is still maturing and remains partially effective [7].

Various machine learning algorithms have been applied to anomaly-based IDS, including Linear Regression, Support Vector Machines (SVM), Genetic Algorithms, Gaussian Mixture Models, k-Nearest Neighbor, Naive Bayes, and Decision Trees [3,5]. Among these, SVM is widely adopted due to its consistent performance across different problem domains [10]. A major limitation of anomaly-based approaches is the high false alarm rate, often caused by the difficulty of accurately modeling normal network behavior from training datasets [11].

Artificial Neural Networks (ANNs), commonly trained using the backpropagation algorithm

International Journal of
DATA SCIENCE AND IOT MANAGEMENT SYSTEM
ISSN: 3068-272X www.ijdim.com Original Research Paper

developed in the 1970s [12], provide an alternative approach. Evaluating IDS performance is often hindered by the scarcity of comprehensive real-world datasets [13]. While earlier studies primarily relied on the KDD CUP 99 dataset [14], this work utilizes the NSL-KDD dataset [15], a widely recognized benchmark for network intrusion detection.

In this study, we propose a supervised machine learning model to classify previously unseen network traffic based on patterns learned from historical data. Both SVM and ANN algorithms are implemented to determine the most accurate and reliable classifier for effective intrusion detection.

2.LITERATURE SURVEY

H. Song, M. J. Lynch, and J. K. Cochran examined whether macro-level opportunity indicators influence cyber-theft victimization. Using criminal opportunity theory, they analyzed state-level internet access patterns and structural characteristics such as unemployment and non-urban population. Their findings indicate that higher proportions of users accessing the internet only at home are positively associated with cyber-theft victimization. The study highlights how structural conditions shape exposure to cybercrime.

P. Alaei and F. Noorbehbahani addressed challenges in Intrusion Detection Systems (IDS) for streaming data. They proposed an online classification method using an incremental Naive Bayesian classifier combined with active learning to reduce labeling costs. The approach, which includes offline preprocessing and an NADAL online method, demonstrated improved accuracy and Kappa metrics compared to conventional methods, making it suitable for IDS applications.

M. Saber, S. Chadli, M. Emharraf, and I. El Farissi presented an evaluation framework for IDS performance based on component-level assessment. Implementing IDS components using embedded hardware platforms and testing with traffic generators like Linux KALI and Metasploit, they showed that IDS effectiveness depends heavily on the characteristics of its individual components.

Manjula C. Belavagi and Balachandra Muniyal focused on predictive IDS models using machine learning algorithms, including Logistic Regression, Gaussian Naive Bayes, Support Vector Machine, and Random Forest. Testing on the NSL-KDD dataset, they found that Random Forest outperforms other classifiers in detecting normal versus malicious network traffic.

N. Chakraborty reviewed the evolution of intrusions in computing environments and the corresponding security measures. The paper discusses the functionalities, performance, and effectiveness of Intrusion Detection Systems (IDS) and Intrusion Prevention Systems (IPS) in securing computing resources and networks against malicious activities.

3. EXISTING SYSTEM

Signature-based network IDS have achieved widespread commercial adoption and are effectively deployed across technology-driven organizations worldwide. In contrast, anomaly-based network IDS have not experienced the same level of success. As a result, anomaly detection remains a major focus of research and development in the IDS domain. However, before these systems can be deployed on a large scale, several key challenges must be addressed. Moreover, the existing literature provides limited comparative analysis of intrusion detection performance using supervised machine learning techniques.

4. PROPOSED SYSTEM

Machine learning has demonstrated significant potential across various real-world applications, and its contribution to cybersecurity continues to grow. In this work, we hypothesize that supervised machine learning techniques can

International Journal of
DATA SCIENCE AND IOT MANAGEMENT SYSTEM
ISSN: 3068-272X www.ijdim.com Original Research Paper

effectively address the challenge of detecting novel or zero-day attacks in network traffic. To test this, we developed a supervised learning model capable of classifying unseen network traffic based on patterns learned from historical data. Both Support Vector Machine (SVM) and Artificial Neural Network (ANN) algorithms were implemented to determine the classifier with the highest accuracy and reliability for intrusion detection.

5.SYSTEM MODEL:

The system proposed is composed of feature selection and Machine Learning algorithm, as shown in the figure. The feature selection component is responsible for extracting the most relevant features or attributes to identify each instance within a particular group or class. The Machine Learning algorithm component builds the necessary intelligence or knowledge using the results from the feature selection component. Using the training dataset, the model is trained to develop its intelligence. The learned knowledge is then applied to the testing dataset to measure the accuracy of how well the model correctly classifies unseen data.

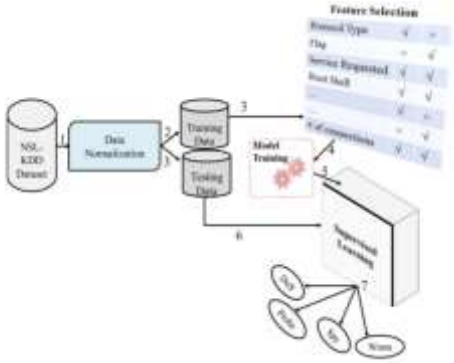


Fig. System Model

6.MODULES:

Upload NSL-KDD Dataset:

Users upload the NSL-KDD dataset, which serves as the benchmark dataset for network intrusion detection.

Preprocess Dataset:

The dataset is cleaned, normalized, and encoded to make it suitable for training machine learning models.

Generate Training Model:

Training models are created using the preprocessed data, building intelligence for detecting attacks.

Run SVM Algorithm:

The Support Vector Machine (SVM) algorithm is applied to classify network traffic and learn patterns of normal and malicious activities.

Run ANN Algorithm:

The Artificial Neural Network (ANN) algorithm is applied for classification, enabling the system to detect complex attack patterns.

Upload Test Data & Detect Attack:

Users upload test datasets, and the trained models classify incoming traffic as normal or malicious.

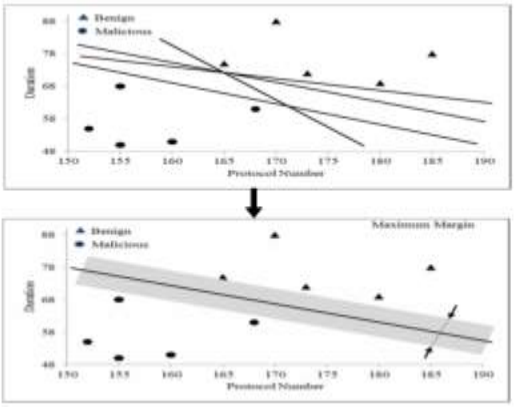
Accuracy Graph:

Performance metrics such as accuracy, detection rate, and false positive rate are visualized using graphs for comparison and evaluation.

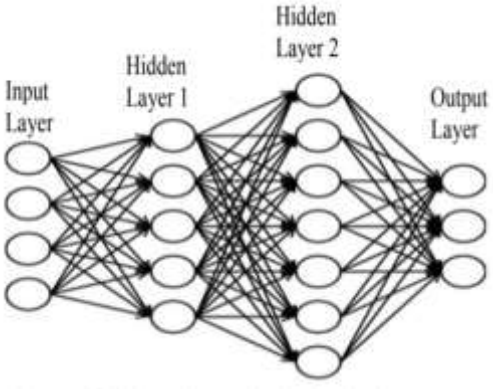
7.ALGORITHMS:

Support Vector Machine (SVM): In SVM, a separating hyperplane defines the decision boundary for classification based on the dataset and problem type. For a one-dimensional dataset, the hyperplane is represented as a point, whereas for a two-dimensional dataset, it is represented as a line, as illustrated in the figure below.

International Journal of
DATA SCIENCE AND IOT MANAGEMENT SYSTEM
ISSN: 3068-272X www.ijdim.com Original Research Paper



Artificial Neural Network (ANN): Artificial Neural Network (ANN) is a machine learning model inspired by the human brain, designed to replicate its learning process. It typically consists of an input layer, one or more hidden layers, and an output layer, as shown in the figure. ANN employs a technique called backpropagation to minimize the difference between the predicted output and the expected result, thereby improving classification accuracy.



8.RESULTS

Double click on 'run.bat' file to get below screen

In above screen click on 'Upload NSL KDD Dataset' button and upload dataset

In above screen I am uploading 'intrusion_dataset.txt' file, after uploading dataset will get below screen

Now click on 'Pre-process Dataset' button to clean dataset to remove string values from dataset and to convert attack names to numeric values

After pre-processing all string values removed and convert string attack names to numeric values such as normal signature contains id 0 and anomaly attack contains signature id 1.

International Journal of
DATA SCIENCE AND IOT MANAGEMENT SYSTEM
ISSN: 3068-272X www.ijdim.com Original Research Paper

Now click on 'Generate Training Model' to split train and test data to generate model for prediction using SVM and ANN

In above screen we can see dataset contains total 1244 records and 995 used for training and 249 used for testing. Now click on 'Run SVM Algorithm' to generate SVM model and calculate its model accuracy

In above screen we can see with SVM we got 84.73% accuracy, now click on 'Run ANN Algorithm' to calculate ANN accuracy

In above screen we got 96.88% accuracy, now we will click on 'Upload Test Data & Detect Attack' button to upload test data and to predict whether test data is normal or contains attack. All test data has no class either 0 or 1 and application will predict and give us result. See below some records from test data

In above test data we don't have either '0' or '1' and application will detect and give us result

In above screen I am uploading 'test_data' file which contains test record, after prediction will get below results

In above screen for each test data we got predicted results as 'Normal Signatures' or 'infected' record for each test record. Now click on 'Accuracy Graph' button to see SVM and ANN accuracy comparison in graph format

From above graph we can see ANN got better accuracy compare to SVM, in above graph x-axis contains algorithm name and y-axis represents accuracy of that algorithms

9.CONCLUSION

In this paper, we presented various machine learning models using different algorithms and feature selection techniques to identify the best-

performing model. The results indicate that the model built using Artificial Neural Networks (ANN) combined with wrapper-based feature selection outperformed other models, achieving a detection rate of 94.02% in classifying network traffic. These findings highlight the potential of machine learning in developing intrusion detection systems capable of identifying both known and novel attacks. Current IDS solutions are largely limited to detecting known threats, while zero-day attacks remain a significant research challenge due to high false positive rates in existing systems.

10FUTURE SCOPE

To enhance security, future IDS systems can be designed to process each incoming request in real time. Requests from genuine users are forwarded to the server, while requests containing attack signatures are blocked and logged. This logged data can be incorporated into the system's dataset to improve detection of future attacks, enabling continuous learning and more effective prevention of both known and emerging threats.

REFERENCES

[1] H. Song, M. J. Lynch, and J. K. Cochran, "A macro-social exploratory analysis of the rate of interstate cyber-victimization," American Journal of Criminal Justice, vol. 41, no. 3, pp. 583–601, 2016.

[2] P. Alaei and F. Noorbehbahani, "Incremental anomaly-based intrusion detection system using limited labeled data," in Web Research (ICWR), 2017 3th International Conference on, 2017, pp. 178–184.

[3] M. Saber, S. Chadli, M. Emharraf, and I. El Farissi, "Modeling and implementation approach to evaluate the intrusion detection system," in International Conference on Networked Systems, 2015, pp. 513–517.

[4] M. Tavallaee, N. Stakhanova, and A. A. Ghorbani, "Toward credible evaluation of anomaly-based intrusion-detection methods," IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews), vol. 40, no. 5, pp. 516–524, 2010.

[5] A. S. Ashoor and S. Gore, "Importance of intrusion detection system (IDS)," International Journal of Scientific and Engineering Research, vol. 2, no. 1, pp. 1–4, 2011.

[6] M. Zamani and M. Movahedi, "Machine learning techniques for intrusion detection," arXiv preprint arXiv:1312.2177, 2013.

[7] N. Chakraborty, "Intrusion detection system and intrusion prevention system: A comparative study," International Journal of Computing and Business Research (IJCBR) ISSN (Online), pp. 2229–6166, 2013.

[8] P. Garcia-Teodoro, J. Diaz-Verdejo, G. Maciá-Fernández, and E. Vázquez, "Anomaly-based network intrusion detection: Techniques, systems and challenges," computers & security, vol. 28, no. 1–2, pp. 18–28, 2009.

[9] M. C. Belavagi and B. Muniyal, "Performance evaluation of supervised machine learning algorithms for intrusion detection," Procedia Computer Science, vol. 89, pp. 117–123, 2016.

[10] J. Zheng, F. Shen, H. Fan, and J. Zhao, "An online incremental learning support vector machine for large-scale data," Neural Computing and Applications, vol. 22, no. 5, pp. 1023–1035, 2013.

[11] F. Gharibian and A. A. Ghorbani, "Comparative study of supervised machine learning techniques for intrusion detection," in Communication Networks and Services Research, 2007. CNSR'07. Fifth Annual Conference on, 2007, pp. 350–358.

[12] J. Schmidhuber, "Deep learning in neural networks: An overview," Neural networks, vol. 61, pp. 85–117, 2015.

[13] N. Moustafa and J. Slay, "UNSW-NB15: a comprehensive data set for network intrusion detection systems (UNSW-NB15 network data set)," in Military Communications and

Information Systems Conference (MilCIS), 2015, 2015, pp. 1–6.

[14] T. Janarthanan and S. Zargari, "Feature selection in UNSW-NB15 and KDDCUP'99 datasets," in Industrial Electronics (ISIE), 2017 IEEE 26th International Symposium on, 2017, pp. 1881–1886.

[15] L. Dhanabal and S. P. Shantharajah, "A study on NSL-KDD dataset for intrusion detection system based on classification algorithms," International Journal of Advanced Research in Computer and Communication Engineering, vol. 4, no. 6, pp. 446–452, 2015.

[16] A. Javaid, Q. Niyaz, W. Sun, and M. Alam, "A deep learning approach for network intrusion detection system," in Proceedings of the 9th EAI International Conference on Bio-inspired Information and Communications Technologies (formerly BIONETICS), 2016, pp. 21–26.

[17] M. Panda, A. Abraham, and M. R. Patra, "Discriminative multinomial naive bayes for network intrusion detection," in Information Assurance and Security (IAS), 2010 Sixth International Conference on, 2010, pp. 5–10.

[18] B. Ingre and A. Yadav, "Performance analysis of NSL-KDD dataset using ANN," in Signal Processing And Communication Engineering Systems (SPACES), 2015 International Conference on, 2015, pp. 92–96.

[19] L. M. Ibrahim, D. T. Basheer, and M. S. Mahmod, "A comparison study for intrusion database (Kdd99, Nsl-Kdd) based on self organization map (SOM) artificial neural network," Journal of Engineering Science and Technology, vol. 8, no. 1, pp. 107–119, 2013.