Received 3-07-2024; Accepted 23-09-2024; Published 30-09-2024

Toward Reducing IDS Misclassification Using Hybrid DL and ML Approach

Mohammed Alyahya malyahya0016@stu.kau.edu.sa
King Abdulaziz University,
Faculty of Computing and Information Technology, Department of Information Technology,
Medical Statistics and Medical Informatics
Prince Majed Road, Jeddah, Saudi Arabia

Husam Lahza hlahza@kau.edu.sa
King Abdulaziz University, Faculty of Computing and Information Technology, Department of Information Technology,
Medical Statistics and Medical Informatics
Prince Majed Road, Jeddah, Saudi Arabia

Rayan Mosli rmosli@kau.edu.sa
King Abdulaziz University, Faculty of Computing and Information Technology, Department of Information Technology,
Medical Statistics and Medical Informatics
Prince Majed Road, Jeddah, Saudi Arabia

Corresponding Author: Mohammed Alyahya

Abstract

Operation centers often face challenges due to the high rate of misclassifications caused by the lower precision in Intrusion Detection System (IDS) models. Despite several research contributions ranging from machine learning and deep learning techniques aiming to reduce false positives and negatives, researchers and security experts consistently encounter a trade-off between these two types of errors. This indicates a significant opportunity for further contributions in this field. We propose a hybrid model that combines Recurrent Neural Networks (RNN) feature extraction capabilities with Support Vector Machines (SVM) classification abilities. Our model achieves an impressive accuracy rate of 98.2% and significantly reduces misclassification errors compared to contemporary state-of-the-art models. This work shows the potential of hybrid approaches in improving accuracy and reducing false positive and negative errors.

Keywords: Intrusion Detection System, Enhancing Accuracy of Detecting Attack, Cybersecurity, Behavior-based Detection, Reduce False Classification, Deep Learning, Machine Learning.

Citation: Mohammed Alyahya, et al. Toward Reducing IDS Misclassification Using Hybrid DL and ML Approach. Advances in Artificial Intelligence and Machine Learning. 2024;4(3):161.

1. INTRODUCTION

In today's interconnected world, integrating data with the Internet is essential across various domains such as infrastructure, healthcare, e-government, and personal communication devices. This trend is driven by the evolution of cloud computing and the Internet of Things (IoT). However, digital progress increases the vulnerability of online data to cyber threats. For instance, the United States faced an estimated $6.9 billion in losses due to cyberattacks, with 5.7 million breaches reported in 2021. According to John P. Mello's analysis, global spending on cybersecurity measures is expected to reach $10 billion by 2027 [1, 2].

The core principles of cybersecurity are confidentiality, integrity, and availability, collectively called the CIA triad. These principles are complemented by non-repudiation, access control, and authentication mechanisms. FIGURE 1 shows an illustration of the Intrusion Detection Systems (IDS) are critical in maintaining the CIA triad by monitoring network traffic for suspicious activities and alerting the Security Operation Center (SOC) on threat detection [3, 4].

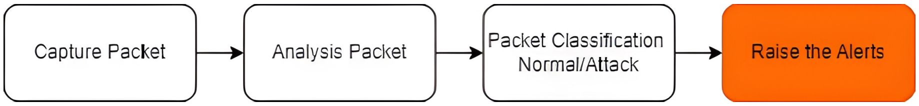


Figure 1: Workflow of IDS

Among the two main IDS types, HIDS (High Interaction Detection Systems) is the most reliable last line of defense but has challenges and flaws. One major problem in intrusion detection systems is the frequent generation of false positive alerts. False positive alarms in IDS impact the system performance by reducing accuracy, increasing the False positive alarms by lowering accuracy, and reducing the true positive ratio. False positives don't directly impact or impair threat detection. They can induce alarm fatigue and result in potential threats being overlooked, indirectly impacting detection. Investigating each false alarm demands considerable time and effort, resulting in inefficient resource utilization. This burdens analysts, diverts computational and network resources, and hinders the performance of other systems. Thus, false positives undermine the reliability of an IDS, waste resources, and increase the risk of overlooking legitimate threats. Effective IDS operation balances accuracy and detection ability while minimizing false positives [5, 6].

While HIDS are effective, they encounter challenges, particularly in generating false positive alerts, which can compromise vigilance security teams and mistakenly restrict legitimate system users. IDS methods are categorized into Signature-based and Anomaly-based detection. Signature-based detection accurately identifies known threats by comparing data packet signatures with patterns in the IDS database. Conversely, Anomaly-based detection struggles with identifying new or zero-day threats that lack pre-existing signatures.

Anomaly-based IDS relies on deep learning (DL) and machine learning (ML) techniques to overcome these challenges and detect obscure threats. The efficiency of these approaches varies among the specific algorithms and models employed. This study presents a hybrid model that integrates ML and DL approaches, which are detailed in the methodology section. Our model demonstrates superior attack detection accuracy and reducing misclassification errors, as illustrated in the confusion matrix in the results section.

The contributions of this paper are as follows:

- Creation and development of a hybrid model integrating ML and DL techniques to enhance IDS model efficiency and reduce attack classification errors.
- Detailed performance analysis of various ML and DL algorithms and comparing their accuracy on the same dataset of the proposed model.

2. RELATED WORK

Adeyemo Victor Elijah et al. compare the artificial intelligence applications in IDS. They evaluated the DL method LSTM and two hybrid methods (homogeneous and heterogeneous) using the UNSW_NB15 dataset. Results indicated that a homogeneous hybrid model is the most effective application of DL, with a detection accuracy of 97.96%, and the DL LSTM model, with the least accurate detection of 80% [7].

Albara Awajan's study emphasized the criticality of real-time intruder detection for enhancing the reliability of Internet of Things (IoT) technology. The research provided a Deep Neural Network (DNN) model to detect intruders with a detection accuracy of 93.74% [8].

Modern AI-based IDS systems operate more efficiently than their predecessors. However, addressing and minimizing false alarms remains crucial to reducing dependence on human intervention [9–11].

Reliance on cybersecurity operators to investigate IDS-generated alerts poses some serious negative consequences associated with significant drawbacks, including a high incidence of false positives. Studies have discussed DL methods and their usage in IDS systems to automate attack detection more efficiently [11, 12].

Guru et al. proposed neural network NN methods (FCNN and LSTM) to distinguish between the alarms as an automated process and differentiate between true positive (real attack) and false positive (fake alarm). The proposed method was performed on the NSL-KDD dataset. The results demonstrated that the FCNN algorithm has an accuracy of 95.8%. In comparison, traditional ML methods like Decision Trees achieved a lower accuracy of 91.9% [12].

Prachiti et al. surveyed ML techniques, revealing different supervised and unsupervised ML techniques. The results showed that the supervised techniques accurately identify known attacks, while the unsupervised techniques better detect anomaly attacks (behavior-based). This advantage stems from unsupervised techniques autonomously extracting features from the data [13].

Ankit Thakkar et al. emphasized the necessity of continually updating datasets to enhance research efficacy in response to evolving attack patterns [14].

Ensemble learning methods have gained attention in network intrusion detection for enhancing prediction accuracy by integrating multiple classifiers. Studies confirm the effectiveness of ensemble techniques in detecting complex intrusion patterns. For example, Thokchom et al present a study that presented a novel ensemble learning-based model that uses diverse classifiers to improve detection rates [15]. Similarly, researcher Saheed proposed a voting-based ensemble model optimized using the Gray Wolf Optimizer, demonstrating superior performance in identifying various intrusions [16]. These findings highlight the robustness and effectiveness of ensemble approaches in intrusion detection systems (IDS).

Artificial intelligence (AI) enhances intrusion detection mechanisms, particularly in wireless communication environments. Jeyanthi et al provision a study proposed an effective scheme using AI-enabled modified learning strategies, highlighting the adaptability and accuracy of AI-driven models in dynamic network environments [17]. Additionally, a subset of AI has been explored for Anomaly-based network intrusion detection. Another researcher Idrissi et al introduced a federated learning framework that facilitates decentralized data processing, ensuring robust detection and addressing privacy concerns linked to centralized data storage [18].

Data imbalance presents a critical challenge in training effective IDS models, potentially biasing detection outcomes. Researcher Awotunde et al have addressed this challenge through oversampling and advanced feature engineering, enhancing model performance on imbalanced datasets [19]. The impact of dataset imbalance on IDS performance in SCADA systems was extensively studied by Balla et al, underscoring the importance of balanced datasets in achieving reliable detection results [20]. Researcher Talukder et al employed oversampling techniques and advanced feature engineering methods to improve model performance on imbalanced datasets [21]. Alternative innovative approaches, like employing Particle Swarm Optimization in neural networks for IDS [22]. These diverse methodologies aimed to improve intrusion detection in various network environments. Additionally, transitioning from traditional electric grids to smart microgrids requires incorporating intrusion detection systems to ensure system stability. Research by Turukmane introduced an Enhanced Deep Belief Network (EDBN) for detecting intrusions in smart microgrids, demonstrating significant enhancements in detection accuracy and reduction in false alarms compared to current methods [23].

Table 1: Comparing Related Work Models.

Ref.	**Year**	**Model**	**Dataset**	**Accuracy**	**Limitation**
[22]	2024	PSO + ANN (Hybrid)	WSN	97.5%	WSN dataset quality issues, such as noise and unreliability, hinder anomaly detection, while unbalanced datasets challenge machine learning models, affecting their generalizability [24, 25].
[23]	2024	M-MultiSVM	UNSW-NB15	97.5%	The M-MultiSVM model's complex calculations, particularly with the Mud Ring optimization, result in extended training durations.
[8]	2023	Deep Neural Network - DNN	Private Dataset	93.74%	A private dataset was used that cannot be measured and has poor reliability
[12]	2022	NN model	NSL-KDD	95.8%	The accuracy achieved is low.
[11]	2022	Decision Tree	NSL-KDD	91.9%	The accuracy achieved is lower than acceptable.

TABLE 1 compares the models proposed by the 3 researchers mentioned in the related work.

3. BACKGROUND

3.1 Intrusion Detection System (IDS)

Intrusion detection systems (IDS) detect attacks and raise alerts with the security team to verify the alert and take appropriate action. There are two main types: Host-based IDS (HIDS) and Network-based IDS (NIDS). HIDS monitors computers and endpoints by collecting logs, analyzing them, and raising suspicious alerts that arise from the endpoint.

Host-based IDS is better suited for detecting long-term and internal attacks but is harder to install because it needs to be installed on all endpoints.

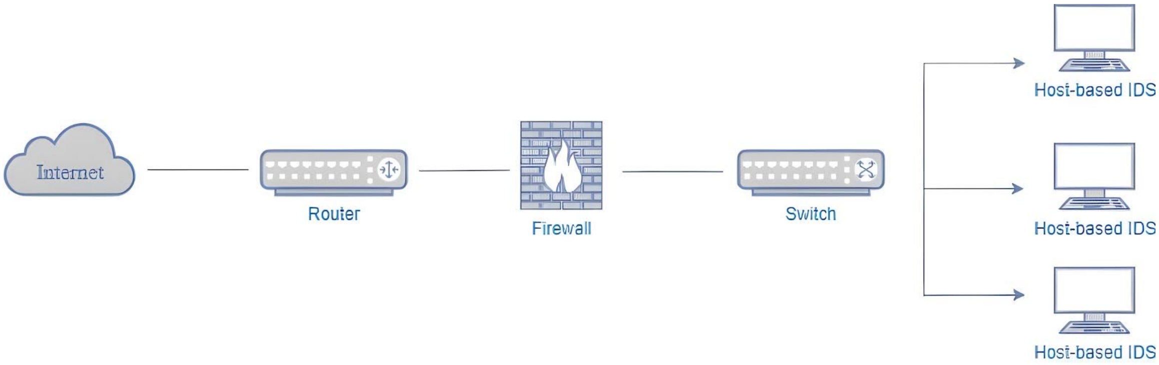


Figure 2: Host-Based IDS

FIGURE 2 illustrates the HIDS architecture. As shown in the FIGURE 3, NIDS is a network-based intruder detection system that monitors intruders in real-time. It is highly efficient at detecting attacks and is easier to install than HIDS.

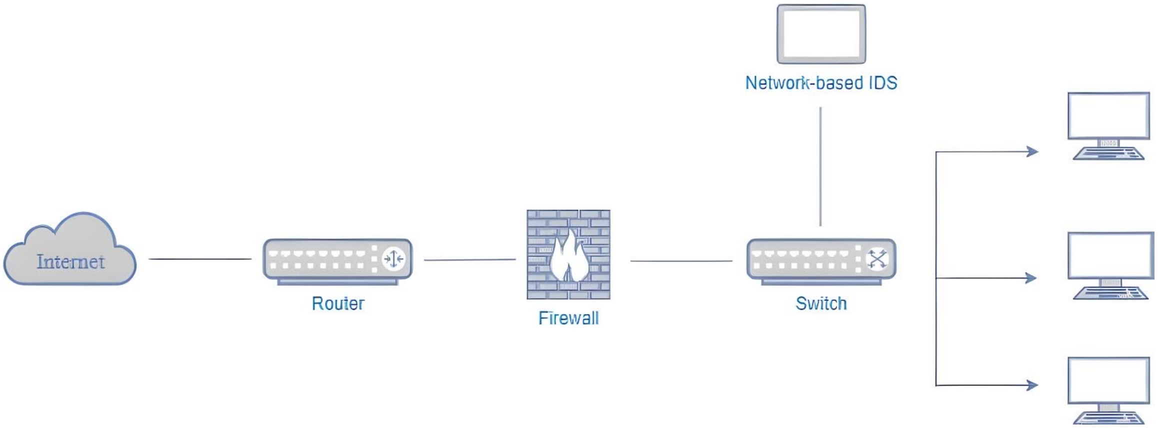


Figure 3: Network-Based IDS

Intrusion detection systems (IDS) are crucial for security operations centers, systems, and applications. For instance, antivirus software integrates IDS to identify and thwart threats at the device level proactively, safeguarding individual users from potential cyber-attacks.

3.2 Intrusion Prevention System (IPS)

An Intrusion Prevention System (IPS) enhances an Intrusion Detection System (IDS) by detecting and preventing intruders, unlike an IDS, which only detects intrusions, an IPS actively takes measures to block them.

3.3 Security Operation Center (SOC)

The Security Operations Center (SOC) has solutions and tools to safeguard systems against cyber-attacks. It comprises a team of experts organized into three layers, with a specialized employee in each layer. The team is divided into layers based on experience level and investigation escalation.

3.4 Security Information and Event Management (SIEM)

SIEM is a main tool and solution in the SOC, gathering logs from different systems, displaying them in an organized and uniform manner, linking the relevant logs, and conducting analysis with ML and DL methods.

3.5 Attack Detection

Attacks are identified using two methods: Signature-based and Anomaly-Based.

Signature-based detection matches file signatures (hashe) with known attacks recorded in attack detection systems' databases. Therefore, databases of known attacks must be constantly updated. Detecting this type of attack is highly accurate.

Anomaly-based IDS detects attacks by identifying abnormal behaviors. This approach is more challenging than a Signature-based technique but offers the advantage of detecting unknown attacks.

Anomaly-based attack detection uses protocols in packets that pass through the network to analyze them and raise alerts.

3.6 Accuracy of IDS

IDS utilize DL and ML techniques employing a Confusion Matrix to evaluate detection accuracy, precision, and recall. The current research utilizes the Confusion Matrix to evaluate the accuracy of the proposed IDS model.

3.7 IDS-based ML and DL

ML and DL techniques increase improvement, increase the accuracy of detecting attacks, and minimize positive errors. The research presents different models and machine learning applications for improving the performance of intruder detection systems.

4. METHODOLOGY

4.1 Proposed Model Methodology

Our novel hybrid model combines two distinct artificial intelligence algorithms: Support Vector Machine (SVM) from Machine Learning (ML) and Recurrent Neural Network (RNN) from Deep Learning (DL). This model combines both algorithms into a unified unit, diverging from heterogeneous models that rely on a voting mechanism for ensemble learning.

The rationale for this integration is to utilize the advantages of both ML and DL. DL algorithms excel in identifying novel, unfamiliar cyber threats with high precision, whereas ML algorithms effectively identify and respond to known threats accurately [26].

FIGURE 4 illustrates the concept of the proposed model and how it works.

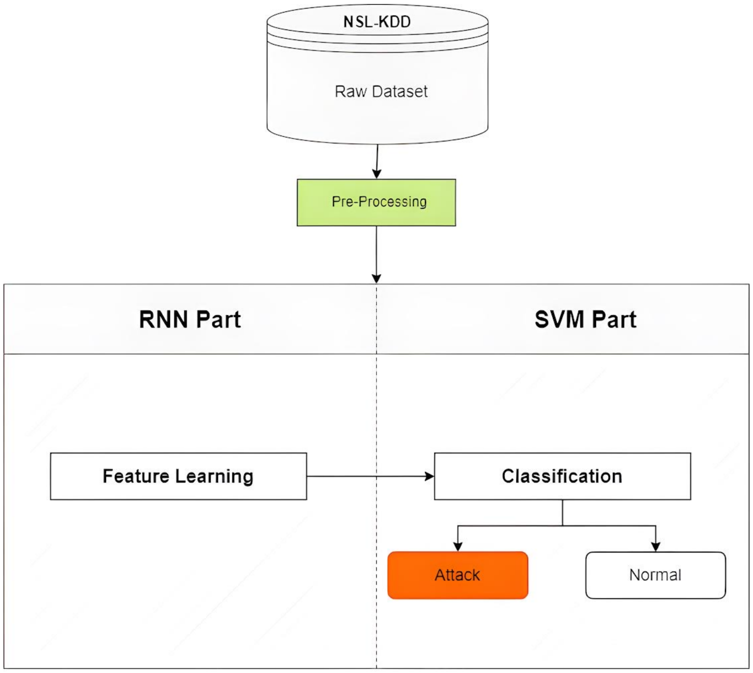


Figure 4: Hybrid Homogenies Model (RNN+SVM)

FIGURE 4 depicts our model's architecture schematically. Initially, the dataset undergoes a preprocessing phase to optimize its compatibility with our model. The dataset section elaborates on the chosen dataset, outlining the preprocessing techniques and their characteristics.

Subsequently, we developed the hybrid model, integrating RNN and SVM into a unified operational framework, as illustrated in FIGURE 4. The model consists of two connected components. The first component, RNN, handles feature extraction and includes all layers except the final one. This segment is crucial in identifying the dataset's features and distinguishing between normal behavior and potential threats. The second component is the classification layer, which utilizes the traits of the SVM algorithm to classify the data as benign or suspicious. The primary obstacle to effectively combining diverse models stems from the complexities of programming. A thorough comprehension of these complexities, coupled with the identification of suitable integration techniques, is essential for successful model integration.

In our comparative analysis, we subjected the hybrid model and four other models (RNN, SVM, Logistic Regression, and Naive Bayes) to the NSL-KDD dataset to assess and validate our model's effectiveness.

4.2 Dataset

We utilized the NSL-KDD dataset, an improved version of the original KDD dataset specifically curated for IDS research. This dataset has been carefully pruned to eliminate duplicate records and address bias, reflecting the evolving nature of cyber threats.

The balance of the dataset is crucial for accuracy; FIGURE 5 shows that our dataset comprises 53% normal and 47% suspicious records, representing a distribution that is considered equitable given the total count of 125,972 records.As Ankit noted, not all features in the NSL-KDD dataset are essential for threat detection [14]. Hence, we refined the feature set to 40 through Principal Component Analysis (PCA), a technique acclaimed for its efficacy in feature reduction by excluding extraneous attributes that hold no significant value in classification.

Table 2: Dataset Description

No. of Attribute	43
No. of records	125,972
No. of Attack Record	59,207
No. of Normal Record	66,765
Rate of Attack Record	47%
Rate of Normal Record	53%

TABLE 2 shows the number of attributes and records 4 in the NSL-KDD dataset used in the research.

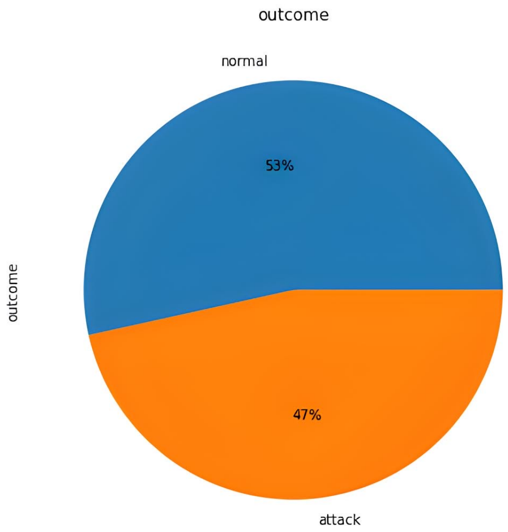


Figure 5: Balance of Dataset

4.3 Model Evaluation

4.3.1 Evaluation metrics

To evaluate the proposed hybrid model, we used a set of equations for evaluating the performance of DL and ML models, depending on these metrics: False Positive (FP), Fale Negative (FN), True Positive (TP), and True Negative (TN).

We used the following equations: accuracy, precision, recall, and F1-score. The methods for calculating the equations are shown below:

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN}$$
$$Precision = \frac{TP}{TP + FP}$$
$$Recall = \frac{TP}{TP + FN}$$
$$F1 - score = 2X\frac{Precision \times Recall}{Precision + recall}$$

4.3.2 Tuning parameter

During the proposed hybrid model experiment, an experiment was conducted to modify the PCA from 25 to 45 and record the result of each modification. The optimal result was achieved with a PCA of 40. We found that Dropout layers, which add extra burden when PCA reduces feature numbers, needed to be minimized. Thus, the proposed model added a Dropout layer only after the third layer, yielding the best results.

4.4 Functional Specifications of the Proposed Model RNN + SVM

Architecture: We implemented a robust RNN architecture with four layers dedicated to feature extraction, effectively capturing the temporal dependencies within the data.

New Layer Addition: At the culmination of the RNN model, we integrated a novel layer designed to seamlessly transform the RNN's output into a format optimized for the SVM.

1. **Layer Type**: A 'dense' layer with 'sigmoid' activation.
2. **Output Transformation**: This layer generates a feature vector that encapsulates the temporal patterns identified by the RNN, ensuring compatibility with the SVM.

Input: The hybrid model processes attributes from the NSL-KDD dataset, including duration, network traffic data, packet size, and more, as detailed in TABLE 3.

Table 3: Attribute of NSL-KDD

No.	Features	No.	Features
1	duration	**23**	count
2	protocol_type	**24**	srv_count
3	service	**25**	serror_rate
4	flag	**26**	srv_serror_rate
5	src_bytes	**27**	rerror_rate
6	dst_bytes	**28**	srv_rerror_rate
7	land	**29**	same_srv_rate
8	wrong_fragment	**30**	diff_srv_rate
9	urgent	**31**	srv_diff_host_rate
10	hot	**32**	dst_host_count
11	num_failed_logins	**33**	dst_host_srv_count
12	logged_in	**34**	dst_host_same_srv_rate
13	num_compromised	**35**	dst_host_diff_srv_rate
14	root_shell	**36**	dst_host_same_src_port_rate
15	su_attempted	**37**	dst_host_srv_diff_host_rate
16	num_root	**38**	dst_host_serror_rate
17	num_file_creations	**39**	dst_host_srv_serror_rate
18	num_shells	**40**	dst_host_rerror_rate
19	num_access_files	**41**	dst_host_srv_rerror_rate
20	num_outbound_cmds	**42**	outcome
21	is_host_login	**43**	level
22	is_guest_login		

TABLE 3 shows all the attributes in the NSL-KDD 9 dataset

Training Parameters: The hybrid RNN + SVM model was meticulously trained using the 'adam optimizer' over five epochs, ensuring optimal performance.

Output: The final output of the hybrid model is a classification result, accurately determining whether the network traffic is benign or malicious.

5. RESULTS

5.1 Proposed Hybrid Model (RNN+SVM) Performance

The data obtained from the confusion matrix, shown in FIGURE 6, support the efficiency of our hybrid model. Our hybrid model demonstrates exceptional precision and recall rates of 98.7% and 97.5%, respectively. These metrics result in an overall detection accuracy of 98.2%, outperforming individual RNN and SVM models and all other models assessed in this study. It correctly identified 13,243 as negative (non-threats) and 11,517 as positive (threats), with only minimal misclassification of 143 negatives as positives and 292 positives as negatives. This results in just 435 classification errors, the lowest among all models tested. Because the time criterion is important in measuring the performance of models, the time taken for this model is 92s

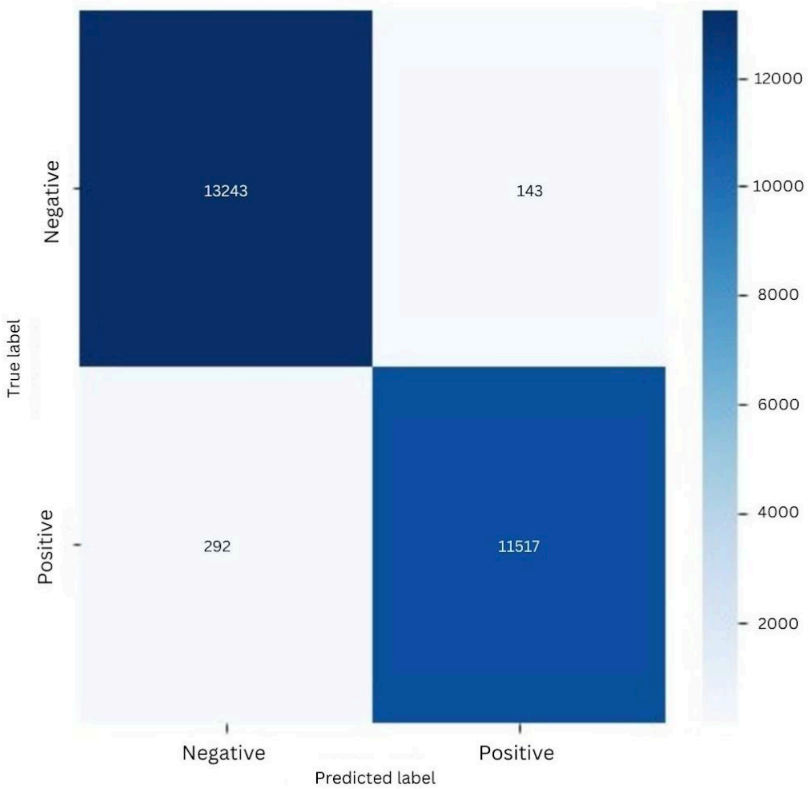
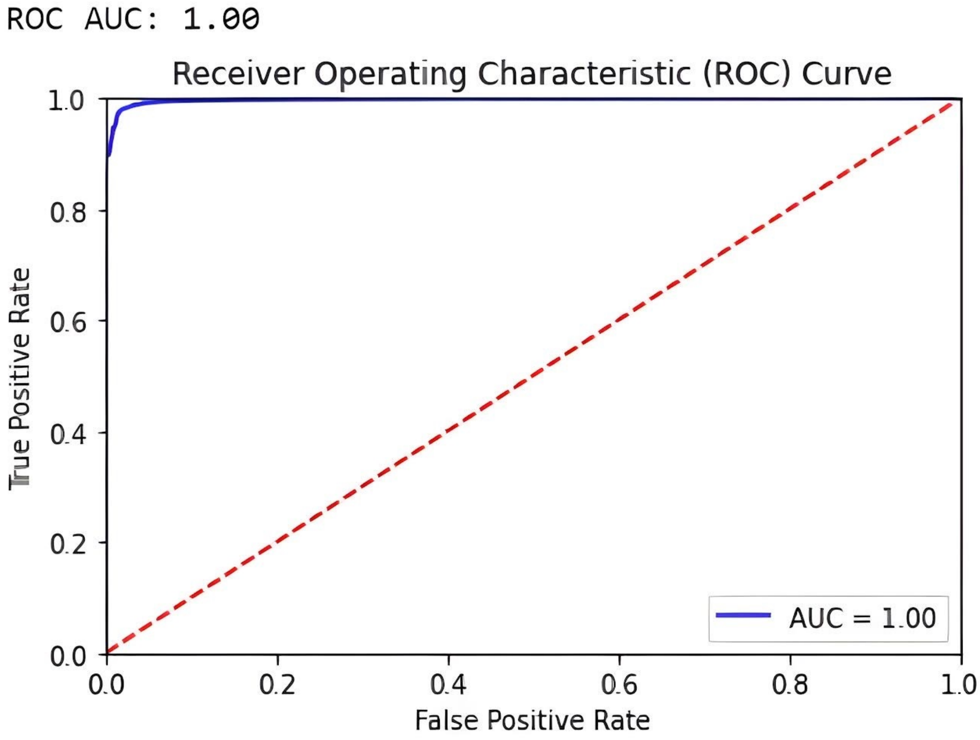


Figure 6: Confusion Matrix Analysis on Proposed Hybrid Model

FIGURE 7 shows that the AUC is equal to 1, which 49 indicates that the proposed model can classify the 50 attack or not with high accuracy.



```
ROC AUC: 1.00
```

Figure 7: Receiver Operating Characteristic (ROC).

5.2 RNN Performance

FIGURE 8 presents the confusion matrix for the RNN deep learning model, highlighting its impressive metrics: 99.6%, a recall of 95.9%, and 97.9% overall accuracy. The RNN model has a remarkably low count of false positives of 35, significantly contributing to its high precision rate. However, the recall is slightly impacted by a higher count of false negatives, 474. The RNN model has 509 classification errors, 74 more than our proposed hybrid model.

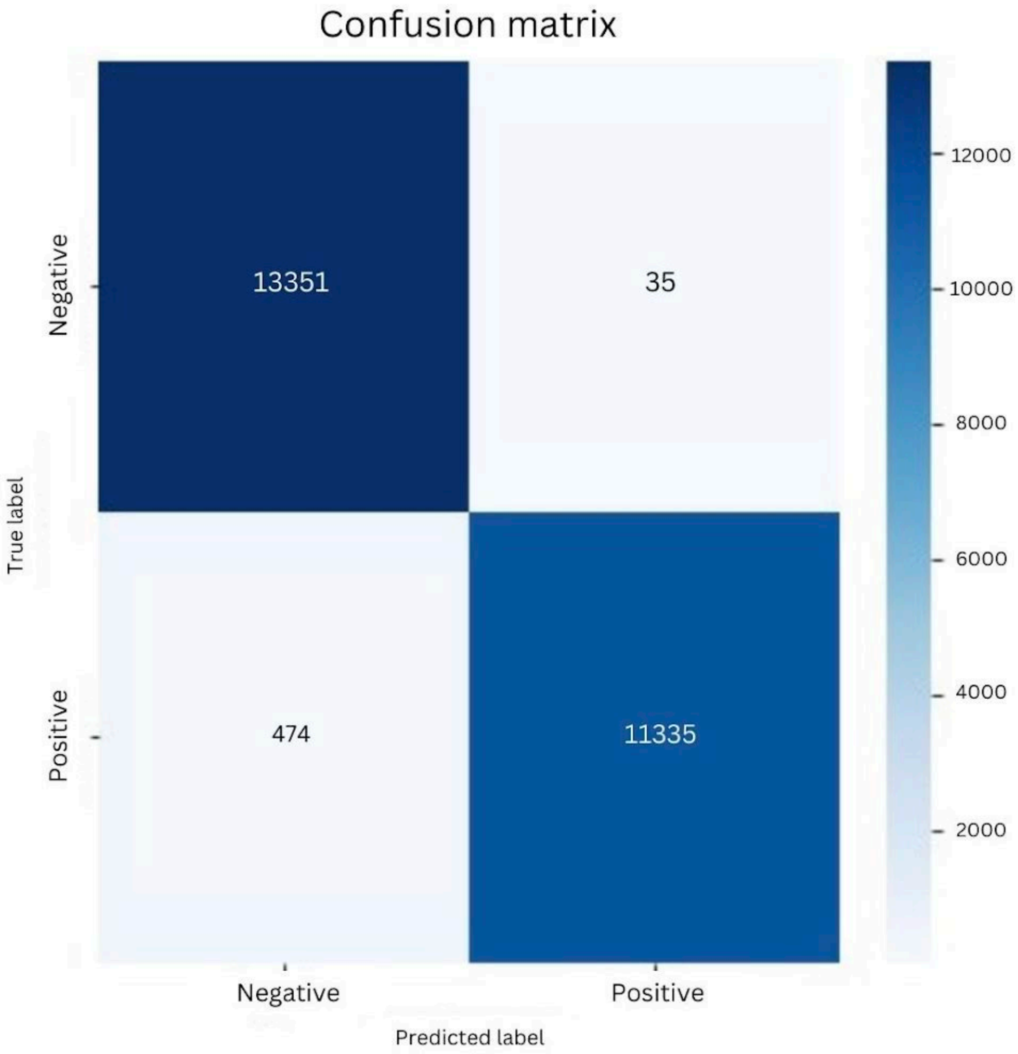


Figure 8: Confusion Matrix Analysis on the RNN Model

5.3 SVM Performance

The SVM model, a fundamental part of supervised ML, was thoroughly tested using our research dataset. It was compared to the proposed hybrid model for performance evaluation. The experiment yielded the following metrics: a precision of 97.5% and a recall of 96.6%, resulting in a detection accuracy of 97.2%. FIGURE 9, the confusion matrix, reveals 287 false positives and 396 false negatives, totaling 683 misclassifications. This indicates a higher error rate than the proposed hybrid model.

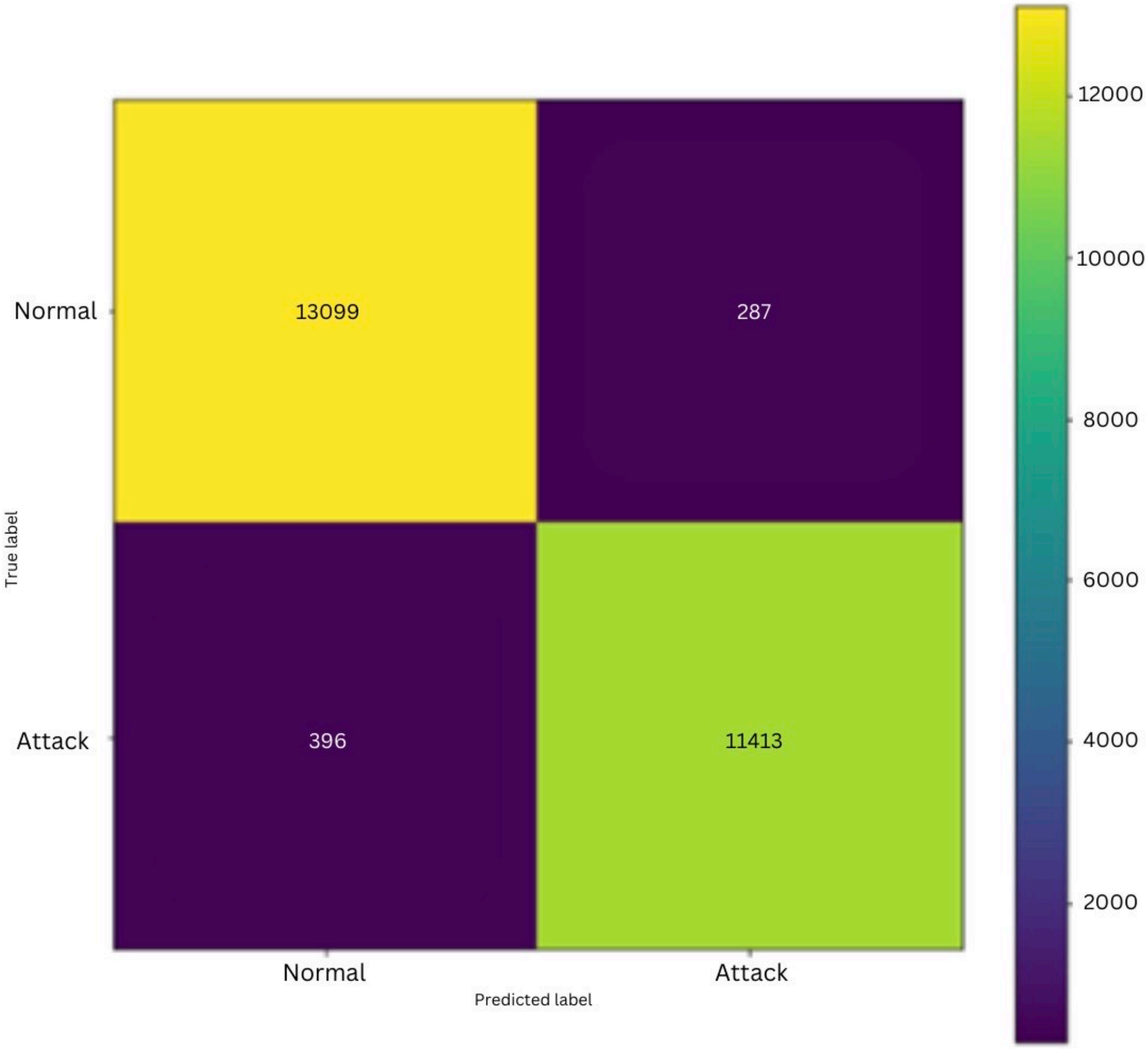


Figure 9: Confusion Matrix Analysis on SVM Model

5.4 Logistic Regression Performance

Logistic Regression, a predictive analysis technique in ML, is particularly effective in estimating binary outcomes - determining whether an event occurs (1) or not (0). Applied to our research dataset, the model demonstrated a precision of 83.5% and a recall of 91.6%, with an overall detection accuracy of 87.6%. As shown in FIGURE 10, the confusion matrix reveals that the model correctly identified 10,818 instances as attacks and 11,259 instances as normal. However, it also mislabeled 2,127 normal instances as attacks (false positives) and 991 attacks as normal (false negatives), resulting in 3,118 misclassifications. This figure represents the highest error rate among all the models evaluated in this study.

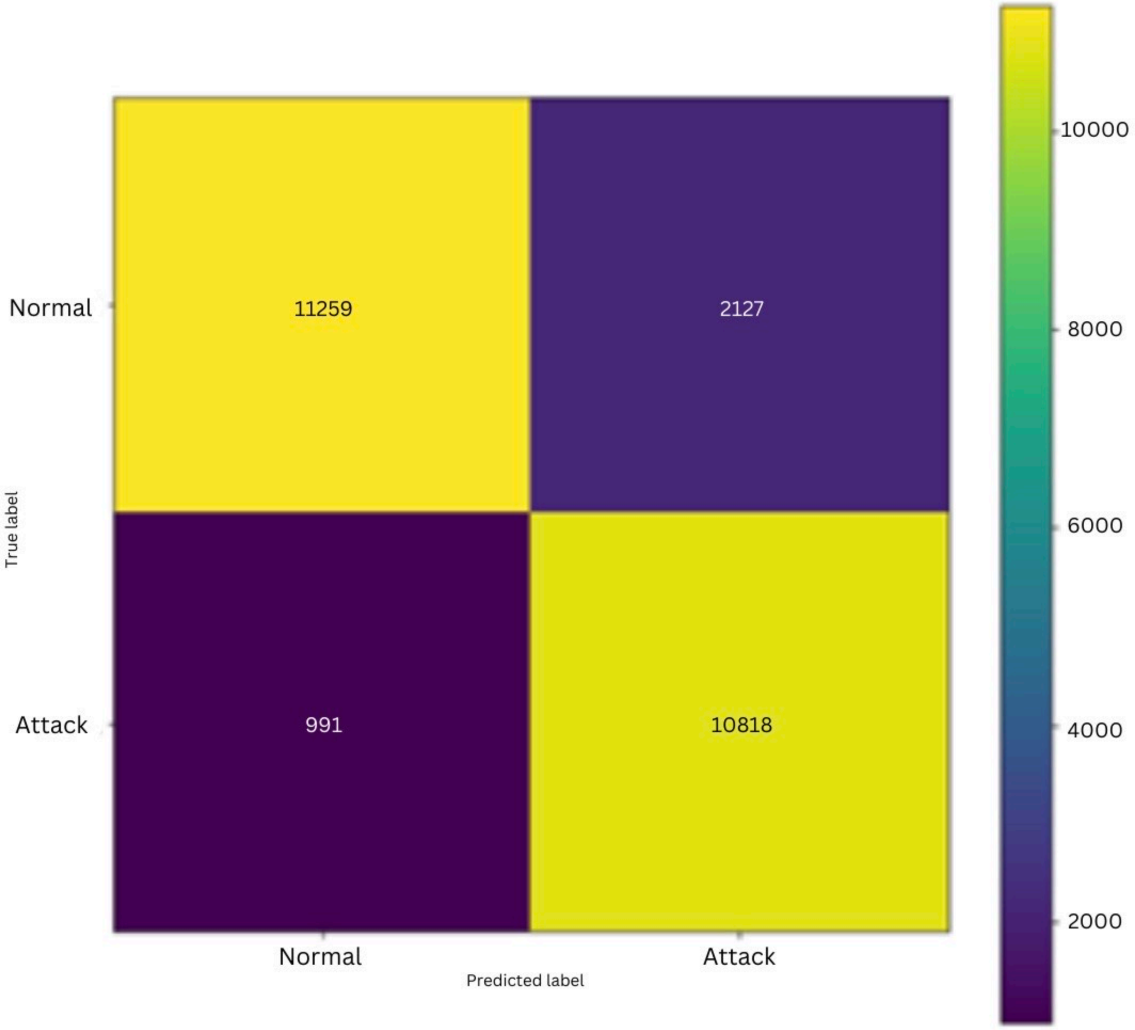


Figure 10: Confusion Matrix Analysis on Logistic Regression Model

5.5 Naïve Bayes Performance

The Naive Bayes classifier, a fundamental supervised ML algorithm, was thoroughly evaluated using our research dataset. Its performance was compared to the proposed hybrid and other models in the study. The confusion matrix reveals that the Naive Bayes classifier had the highest number of false negatives, totaling 1264 (FIGURE 11). The model achieved a precision of 92.5% and recall of 89.2%, resulting in an overall detection accuracy of 91.6%. The relatively lower recall rate is due to the significant number of false negatives, indicating an improvement in the model's ability to detect true positives.

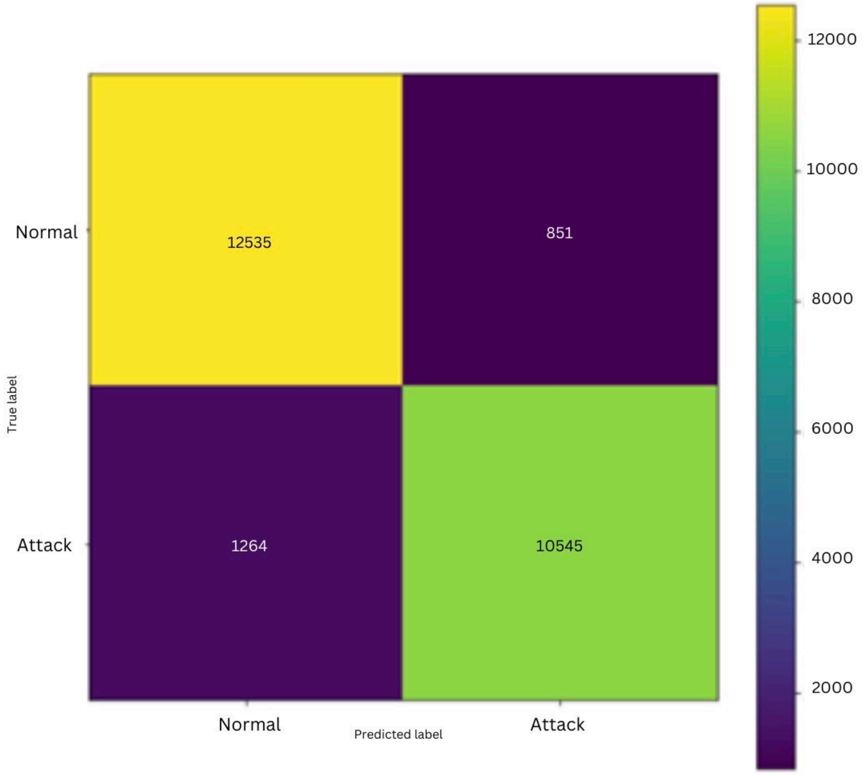


Figure 11: Confusion Matrix Analysis on Naïve Bayes Model

5.6 Overall Performance

TABLE 4 compares the DL and ML models, especially the proposed hybrid model. The research results show that the hybrid model's detection accuracy is superior to other models.

Table 4: Overall performance.

Approach	Accuracy	Precision	Recall	F1-score
Hybrid (RNN + SVM)	98.2%	98.7%	97.5%	98%
RNN	97.9%	99.6%	95.9%	98%
SVM	97.2%	97.5%	96.6%	96.6%
Logistic Regression	87.6%	83.5%	91.6%	88.3%
Naïve Bayes	91.6%	92.5%	89.2%	91.1%

The hybrid model, RNN, and SVM demonstrate exceptional performance with an accuracy of 98.2%, precision of 98.7%, recall of 97.5%, and an F1-score of 98%. It surpasses other models, proving a strong justification for its adoption. The hybrid model achieves versatility and accuracy by integrating RNN's sequential data processing capability with SVM's robust classification capabilities. Its minimal error rate ensures reliability in precision-dependent applications. It optimally balances precision and recall, effectively identifying threats with minimal false positives. The model's high accuracy and precision have significant implications for detection, reducing false positives.

6. LIMITATION & FUTURE WORK

Our research into IDS revealed a significant disparity in reported model accuracy and actual performance. The dataset and the preprocessing method influence the accuracy of the model outcomes. In the future, we advocate developing and publishing an updated dataset for IDS and processing in alignment with the facts of cyber incidents. This approach allows researchers to concentrate on enhancing DL and ML models rather than splitting their attention between model creation and dataset preprocessing.

7. CONCLUSION

Based on the findings, the hybrid model outperforms the RNN and SVM models in threat detection accuracy. Analysis of the confusion matrix (FIGURE 6) supports this, showing exceptional precision and recall rates of 98.7% and 97.5% for the hybrid model. This translates to an unparalleled detection accuracy of 98.2%. The results represent a significant advancement compared to the individual RNN and SVM models.

Practically, the hybrid model demonstrated its effectiveness by accurately identifying 13,243 negative instances and 11,517 positive instances, with only 435 misclassifications. This is the lowest error count among all the evaluated models, highlighting the strength and reliability of the hybrid model.

Further analysis reveals that although RNN and SVM models demonstrate high precision and recall, they fall short of the performance of the hybrid model, with 509 and 683 classification errors, respectively. The Logistic Regression and Naive Bayes models exhibit a higher tendency for classification errors, underscoring the hybrid model's superiority.

Our study confirms that combining RNN and SVM models can significantly reduce classification errors and enhance threat detection system accuracy. The hybrid model excels in theoretical metrics and offers practical advantages, making it a powerful tool for combating security threats. These results validate the potential of machine learning and deep learning models to transform predictive analytics.

References

[1] https://cybersecurityventures.com/security-62awareness-training-report/63

[2] Haris Uddin Sharif Md, Ali M. Mohammed, A Literature Review Of Financial Losses Statistics For Cyber Security And Future Trend . World J. Adv. Respir Res. 2022;15:138-156.

[3] Varanasi VR, Razia S. Intrusion detection using machine learning and deep learning. Int J Recent Technol Eng. 2019;778:9704-9719.

[4] Stallings W. Cryptography And Network Security: Principles And Practice. sixth edition. Pearson. 2013.

[5] Zhu G. Automated False Positive Filtering For Esnetwork Alerts. 2022. Arxiv Preprint: https://arxiv.org/pdf/2208.12729

[6] Scaife N, Carter H, Traynor P, Butler KRB . CryptoLock (and Drop It): Stopping Ransomware Attacks on User Data. In: 36th International Conference on Distributed Computing Systems (ICDCS). IEEE PUBLICATIONS. 2016:303-312.

[7] Elijah AV, Abdullah A, JhanJhi N, Supramaniam M, Abdullateef B. Ensemble And Deep-Learning Methods For Two-Class And Multi-Attack Anomaly Intrusion Detection: An Empirical Study. Int J Adv Comput Sci Appl. 2019;10:520-528.

[8] Awajan A. A Novel Deep Learning-Based Intrusion Detection System For Iot Networks. Computers. 2023;12:34.

[9] Rozendaal K, Dissanayake-Mohottalalage T, Mailewa A. Neural Network Assisted IDS/IPS: An Overview Of Implementations Benefits And Drawbacks. Int J Comput Appl. 2022;184:21-28.

[10] Nisar A. Intrusion detection systems: categories, attack detection and response.2023. Avilable at: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4478816

[11] Alahmadi BA, Axon L, Martinovic I. 99% False Positives: A Qualitative Study of SOC Analysts' Perspectives on Security Alarms. In 31st USENIX Security Symposium (USENIX Security 22). 2022:2783-2800.

[12] Reddy PGM, Suleman S, Sainath G, Goud EVK, Muzamil S. Cyber Threat Detection Based On Deep Learning Through AI. Int J Res Publ Rev. 2022;3:235-238.

[13] Parkar P, Bilimoria A. A Survey On Cybersecurity IDS Using ML Methods. In: 5th International Conference On Intelligent Computing And Control Systems (ICICCS). IEEE. 2021:352-360.

[14] Thakkar A, Lohiya R. A Review Of The Advancement In Intrusion Detection Datasets. Procedia Comput Sci. 2020;167:636-45.

[15] Thockchom N, Singh MM, Nandi U. A Novel Ensemble Learning-Based Model For Network Intrusion Detection. Complex Intell Syst. 2023;9:5693-5714.

[16] Saheed YK, Misra S. A Voting Gray Wolf Optimizer-Based Ensemble Learning Models For Intrusion Detection In The Internet Of Things. Int J Inf Secur. 2024;23:1557-1581.

[17] Jeyanthi DV, Indrani B. An Effective Intrusion Detection Scheme Over Wireless Communication Environment Based On Artificial Intelligence Enabled Modified Learning Strategy. ACS J Sci Eng. 2024;3:1-11.

[18] Idrissi MJ, Alami H, El Mahdaouy A, El Mekki A, Oualil S, et al. Fed-ANIDS: Federated Learning For Anomaly-Based Network Intrusion Detection Systems. Expert Syst Appl. 2023;234:121000.

[19] Awotunde JB, Folorunso SO, Imoize AL, Odunuga JO, Lee CC, et al. An Ensemble Tree-Based Model For Intrusion Detection In Industrial Internet Of Things Networks. Appl Sci. 2023;13:2479.

[20] Balla A, Habaebi MH, Elsheikh EA. The Effect of Dataset Imbalance on the Performance of SCADA Intrusion Detection Systems. Sensors. 2023;23:758.

[21] Talukder MA, Islam MM, Uddin MA, Hasan KF, Sharmin S, et. al. Machine Learning-Based Network Intrusion Detection For Big And Imbalanced Data Using Oversampling, Stacking Feature Embedding And Feature Extraction. J Big Data, 2024;11:33.

[22] Srivastava A, Addimulam SC, Basu MT, Sindhuri BP, Maurya RK. Network Intrusion Detection System (NIDS) For WSN Using Particle Swarm Optimization Based Artificial Neural Network. Int. J. Intell. Syst. Appl. Eng .2024:12:143-150 .

[23] Turukmane AV, Devendiran R. M-Multi SVM: An Efficient Feature Selection Assisted Network Intrusion Detection System Using Machine Learning. Comput Secur. 2024;137:103587.

[24] Haque A, Chowdhury NUR, Soliman H, Hossen MS, Fatima T, et al. Wireless Sensor Networks Anomaly Detection Using Machine Learning: A Survey. In: Arai K, (eds). Intelligent Systems and Applications. Springer Nature. 2024;824:491-506.

[25] Kumar R, Tripathi S, Agrawal R. Handling Dynamic Network Behavior and Unbalanced Datasets for Wsn Anomaly Detection. J Ambient Intell Humaniz Comput. 2023;14:10039-10052.

[26] Ahmad R, Alsmadi I, Alhamdani W, Tawalbeh LA. A Deep Learning Ensemble Approach to Detecting Unknown Network Attacks. J Inf Secur Appl. 2022;67:103196.