Decision-Making Framework for Intrusion Detection in Networks Using XGBoost-Based Feature Selection and Deep Neural Networks

Vijay Kumar Sharma[1*], Dr. Navin Kumar Agrawal[2]

[1]Research Scholar, Department of Electronics & Communication Engineering, Bhabha University, Bhopal

[2]Professor, Department of Electronics & Communication Engineering, Bhabha University, Bhopal

ORCID ID: [0009-0004-8922-666X][1], [0009-0006-9631-6238][2]

Email ID: vijaykumarsharma24@gmail.com[1], nka765@gmail.com[2]

Abstract

The growing complexity and frequency of cyberattacks have rendered intrusion detection a vital aspect of network security. The conventional Intrusion Detection Systems (IDS) tend to be challenged with high-dimensional data and poor capability to detect variegated patterns of attacks. To address such issues, this study introduces a decision-making framework for intrusion detection through the integration of XGBoost-based feature selection and Deep Neural Network (DNN) classification. The suggested methodology was tested on two popular benchmark datasets: NSL-KDD and CIC-IDS 2017. Feature selection was first carried out by using the XGBoost algorithm to eliminate redundant and less informative features while preserving the most relevant attributes. This step enhanced the efficiency and accuracy of the model. The features were then classified using a DNN, which took advantage of its robust representation learning ability in identifying different types of attacks. Experimental results confirm the efficacy of the presented XGBoost–DNN approach. On the NSL-KDD dataset, the model performed with an accuracy of 99.69%, successfully identifying major attack categories such as. Likewise, on the CIC-IDS 2017 dataset, the model performed at 99.25% accuracy, showing excellent accuracy and recall for several contemporary attack types. These comparison of results with previous work demonstrate the strengths of the proposed methodology in dealing with heterogeneous and complex network traffic data.

Keywords: Decision-Makin, IDS, XGBoost, Deep Neural Network, NSL-KDD, CIC-IDS.

1. Introduction

Network security is now a cornerstone of information technology infrastructure in this modern age of digital communication. Computer networks have become immensely dependent with the explosive proliferation of connected devices, cloud computing, and web applications. It is now an attractive target for countless types of malicious cyber attacks. Malicious behavior like unauthorized access, sensitive information disclosure, identity theft, malware injection, and mass denial-of-service attacks now threaten organizations and individuals equally. These threats compromise the three most important principles of cybersecurity: confidentiality, integrity, and availability (CIA) of network resources. Therefore, providing robust security controls within networks has become a necessary prerequisite for enterprises and governments alike.

In response to these threats, Intrusion Detection Systems (IDSs) have been created as smart monitoring and defense systems [1]. An IDS constantly monitors incoming and outgoing network traffic, examines pattern behavior, and identifies any abnormal or unauthorized activity. By offering near real-time responses to anomalies, IDSs are a critical component in enhancing general security [2]. They can detect a wide range of cyberattacks, such as denial-of-service (DoS) and distributed denial-of-service (DDoS) attacks, port scan

attempts, malware spreading, and insider threats. Older IDS technologies are mostly rule-based detection and signature-matching based, and they are effective in identifying known attack vectors. These legacy systems tend to not work when faced with zero-day exploits, dynamic attack tactics, and highly sophisticated intrusion attempts [3].

To overcome these issues, researchers have suggested sophisticated methodologies that utilize contemporary computational intelligence. One such promising approach lies in the prediction and modeling of cyberattacks, which works much like forecasting the weather—via the study of patterns to predict and reduce damage prior to escalation [4]. In these situations, deep learning methodologies have exhibited superior performance over traditional IDS methodologies. Specifically, Convolutional Neural Networks (CNNs) and Deep Convolutional Neural Networks (DCNNs) have the capability to learn implicit patterns from high-volume network traffic data without the need for handcrafted features [5]. These networks can automatically extract intricate spatial and temporal relationships from raw inputs, thus enhancing detection accuracy and responsiveness towards new threats [6].

Deploying strong IDS frameworks along with efficient machine learning models is essential to secure valuable assets and ensure user, customer, and stakeholder confidence [7]. For example, the horrific effects of DoS and DDoS attacks are hugely discussed since they target disabling network services either through bandwidth resource burnout or computation burnout. In these cases, attackers usually overwhelm the victim's system with too many requests, making it unable to serve actual users.

This paper presents an all-around benchmark for detecting cyberattacks on the basis of two extensively used intrusion detection datasets, NSL-KDD and CICIDS2017. For effective and efficient detection, the discussed framework uses a two-stage approach: the XGBoost-based attribute reduction is used initially to eliminate irrelevant and redundant attributes; next, an Optimized Deep Neural Network (DNN) is utilized to classify network intrusions effectively.

- Creation of a feature selection technique based on XGBoost to filter and rank important features from unprocessed network data.
- Building of a training and testing pipeline that uses the outlined optimized DNN architecture to enhance detection accuracy.
- Systematic performance assessment of the framework on benchmark datasets, with comparisons to state-of-the-art IDS methods.

The remainder of this paper is organized as: Section 2 gives a comprehensive review of existing works in the field of intrusion detection. Section 3 elaborates the proposed IDS model's system architecture and methodology. Section 4 summarizes experimental configuration, i.e., datasets and parameter settings. Section 5 illustrates results and performance comparison with current models. Lastly, the Conclusion section summarizes main findings and states potential future research directions.

2. Related Work

The rapid growth of information produced by contemporary communication networks, combined with the growth in evasiveness of cyberattacks, has strongly impacted the direction of research on intrusion detection systems (IDSs). Conventional manual security measures are not effective anymore, so new automated, adaptive, and intelligent intrusion detection solutions have been developed. With the emergence of machine

learning (ML) and deep learning (DL), researchers have suggested many models to enhance the efficiency and scalability of IDSs [8].

A number of research works have proven the efficiency of machine learning algorithms in identifying a large variety of intrusions [9]. As an example, Çavuşoğlu et al. presented a layered IDS system in which various ML algorithms were dynamically chosen according to the particular type of identified attack [10]. Support Vector Machines (SVM), Naïve Bayes, Random Forest (RF), and some clustering algorithms are widely used ML methods. Zhao suggested a hybrid kernel function optimized Least Squares Support Vector Machine (LSSVM), whose parameters were tuned through Particle Swarm Optimization (PSO) to enhance the classification accuracy [11]. Analogously, Thaseen et al. introduced an ensemble method that integrated Chi-square feature selection and SVM, Modified Naïve Bayes (MNB), and LPBoost using majority voting to provide more accurate predictions [12]. In yet another study, Sumaiya et al. constructed a model that combined Chi-square feature selection with multi-class SVM and achieved better detection rates and much fewer false alarms [13].

Researchers have also tried optimizing model-driven approaches. Tao et al., for instance, introduced the FWP-SVM-GA model that incorporated feature selection, weight optimization, and parameter tuning through Genetic Algorithms (GA), thus performing better than typical SVM-based IDSs on both detection rate and error minimization [14]. Peng et al. combated the problem of high-dimensional data by uniting Mini-Batch K-means clustering with Principal Component Analysis (PCA), thus allowing their IDS to handle big data environments in an efficient manner [15].

Deep learning has also progressed the intrusion detection field. It has achieved significant success in identifying intricate hierarchical patterns within network traffic data [16]. Yin et al., for example, used Recurrent Neural Networks (RNNs) for IDS and explored its effectiveness in both binary and multi-class scenarios through the examination of the effect of hyperparameters like neuron numbers and learning rate [17]. Kim further followed up on this work by utilizing the Long Short-Term Memory (LSTM) in the RNN architecture to show better performance in coping with sequential dependencies of network traffic [18]. These works underscore the significance of deep models in acquiring both spatial and temporal correlations in network flows.

Class imbalance remains an issue in intrusion datasets. In real-world networks, specific attacks seem much less common but can be more devastating than frequent attacks. Various detection approaches are aimed only at reducing total false positives at the expense of disproportionately greater minority attack misclassification. To remedy this issue, Bamakan et al. proposed a robust approach based on Ramp Loss K-Support Vector Classification-Regression (K-SVCR) specifically designed for highly imbalanced intrusion datasets [20]. Analogously, Yan et al. also suggested a modified Synthetic Minority Oversampling Technique (SMOTE) named Mean SMOTE (M-SMOTE) that enhanced detection precision by balancing minority and majority classes more effectively [21].

Another important issue influencing IDS performance is the judicious choice of hyperparameters [22] like the number of layers, learning rate, activation functions, and kernel sizes in deep learning algorithms. Inadequate tuning of hyperparameters can significantly restrict accuracy as well as generalization capability. To address this problem, researchers have presented randomized search methods [23] and even bio-inspired optimisation algorithms such as lion swarm optimisation, which were applied in fine-tuning an Optimized CNN-Hierarchical Multiscale LSTM (OCNN-HMLSTM) model [24]. Even with these developments, it should be noted that addressing one challenge tends to bring about others, thus creating loopholes hindering the attainment of a universal IDS solution.

3. Proposed Work

The suggested work presents a decision-making platform for network intrusion detection that combines XGBoost-based feature selection with Deep Neural Network (DNN) classification to overcome the drawbacks of traditional intrusion detection systems. Traditional IDSs are known to struggle with high-dimensional data, redundant features, and inconsistent detection accuracy against diverse categories of attacks. To that end, the experiments were chosen with two benchmark datasets: NSL-KDD and CIC-IDS 2017. The first one is refined KDD'99 with attack categories such as DoS, Probe, R2L, and U2R and the second contains realistic modern traffic including DDoS, Brute Force, Botnet, and Infiltration attacks.
Figure Proposed XGBoost classifier model for network intrusion detection.

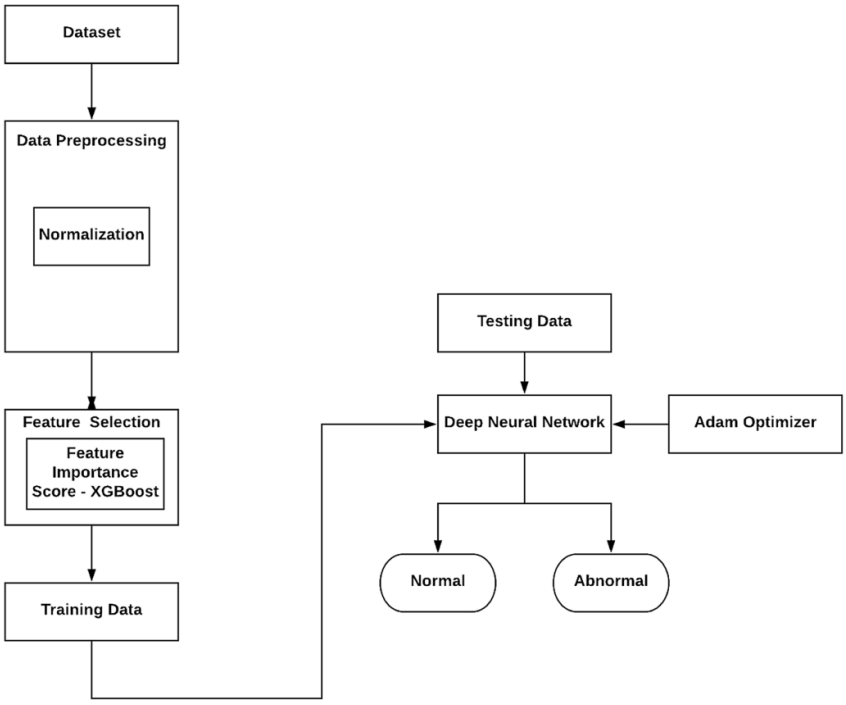

Figure 1. Proposed XGBoost classifier model for network intrusion detection

The proposed approach adheres to a structured methodology. The first step is to utilize preprocessing techniques like data cleaning, normalization, and categorical encoding to get the datasets ready. Then, the XGBoost algorithm is used to carry out feature selection in terms of computing feature importance scores and removing irrelevant and redundant attributes. This decreases the dimension of the dataset and enhances computational efficiency. The chosen characteristics are then grouped based on a multi-layer Deep Neural Network that uses ReLU activation in hidden layers and Softmax activation in the output layer to manage multi-class categorization. The network is trained using the Adam optimizer and cross-entropy loss for reliable learning.

Algorithm of Proposed Model

Algorithm: XGBoost Classifier-Based Intrusion Detection using NSL-KDD Dataset

Input:

- NSL-KDD OR CIC-IDS-2017 Dataset with labeled instances (features + class labels)
- Hyperparameters: max_depth, learning_rate, n_estimators, gamma, subsample, colsample_bytree

Output:

- Trained XGBoost classification model
- Predicted labels for test data

- Performance evaluation metrics

Step 1: Data Preprocessing

1.1 Load the NSL-KDD dataset DDD
1.2 Remove any missing or duplicate records
1.3 Encode categorical features (e.g., protocol_type, service, flag) using label encoding or one-hot encoding
1.4 Normalize or standardize numerical features to a common scale
1.5 Split the dataset into:
  1.5.1 Training set DtrainD
  1.5.2 Testing set Dtest

Step 2: Feature Selection (Optional but Recommended)

2.1 Compute feature importance using mutual information or an initial XGBoost run
2.2. Select top-kkk features to reduce dimensionality and improve performance

Step 3: Model Initialization

3.1. Initialize the XGBoost classifier with chosen hyperparameters:

Step 4: Model Training

4.1. Train the XGBoost classifier on the training set:
model.fit(Xtrain,Ytrain)(X_{train}, Y_{train})model.fit(Xtrain,Ytrain)

Step 5: Prediction

5.1. Predict class labels on test data:
Y^test=model.predict(Xtest)\hat{Y}_{test} = \text{model.predict}(X_{test})Y^test =model.predict(Xtest)

Step 6: Evaluation

6.1. Evaluate model performance using classification metrics:

- Accuracy
- Precision
- Recall
- F1-score
- Confusion Matrix

6.2. Optionally, compute ROC Curve and AUC score if binary classification is used

Step 7: Result Interpretation

7.1. Analyze which features contributed most to the prediction (feature importance)
7.2. Identify classes with high false positives or false negatives
7.3. Adjust hyperparameters and repeat training if needed (tuning loop)
End of Algorithm

4. Results

These findings suggest that the model possesses a high potential to discriminate between malicious and benign traffic patterns, with extremely high precision and very low prediction error. Figure 2 and figure 3 shows results accuracy and loss graph capture by proposed method. Specifically, the high precision (99.43%) suggests that the model produces very few false positive predictions (i.e., classifying normal traffic as attacks), which is essential for real-world IDS applications where false alarms may overwhelm security teams and erode trust in the system.

Figure 2. Training and Testing Accuracy Curve of proposed model

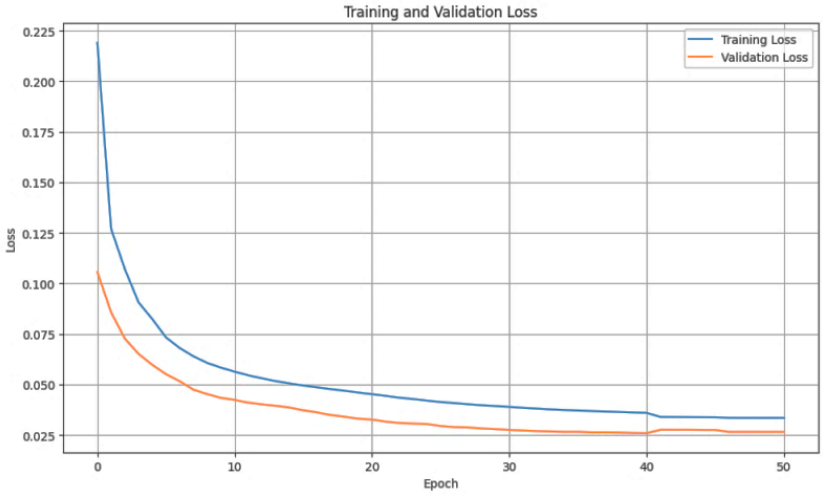

Figure 3. Training and Testing Loss Curve of proposed model

The classification report given in figure 4 verifies that the model has very high performance in identifying normal traffic with almost perfect recall (1.00) and precision (0.99). Although the attack traffic detection (label 1.0) has slightly lower recall (0.93), the precision is still very high (0.98), meaning that when the model classifies an attack, it will be accurate most of the time.

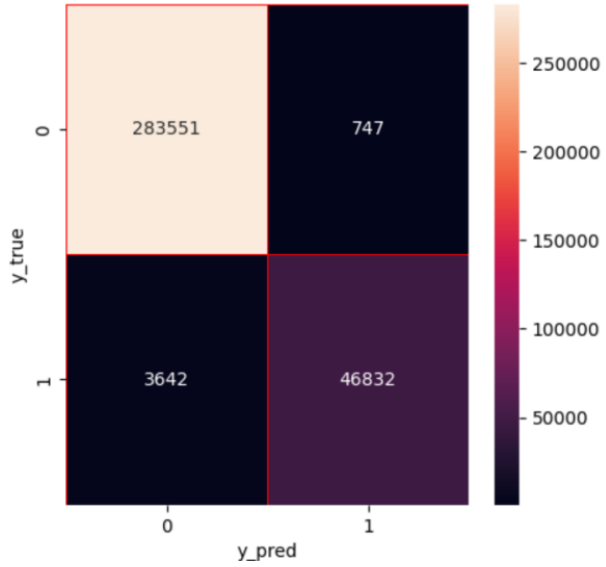

Figure Confusion Matrix for CIC-IDS 2017 dataset prediction

F1 score of 0.96 for attacks reflects a high degree of balance between precision and recall. Macro average F1-score of 0.97 and weighted average F1-score of 0.99 also reaffirm that the model performs consistently on both classes in the presence of a high degree of class imbalance.

4.1 Comparison with existing methods

In order to assess the efficacy of the designed framework, its performance was compared against some state-of-the-art intrusion detection methods presented in recent work with NSL-KDD dataset in Table 1. As evident from CIC-IDS in Table 2, 90.25% accuracy was obtained by Nidaa et al. with a DNN-based solution, while W. Elmasry et al. obtained 92.41% accuracy using an LSTM model. Farhan et al. enhanced detection to 95% employing a hybrid DNN coupled with Binary Particle Swarm Optimization (BPSO). Lin et al. utilized an LSTM with Attention Mechanism (AM) with 96.20% accuracy, and Ashik Elahi et al. had 96.23% accuracy using a GRU-based approach. Comparatively, the proposed XGBoost–CNN approach far exceeds these current approaches by obtaining an accuracy of 99.25%, clearly outshining in identifying diversified network intrusions.

Table 1. The accuracy comparison of proposed model with existing models using NSL-KDD dataset

Author	**Year**	**Technique**	**Accuracy**
Mulyanto et al [27]	2020	FL-CNN	0.7833
Xiangyu et al [25]	2020	AESMOTE	0.8209
Lan Liu et al [26]	2020	DSSTE+AlexNet	0.8284
Rahbar et al [28]	2022	SMOTE+VC	0.8300
Zhengfa et al [29]	2024	VWLM	0.8345
Proposed Model		**XGBoost+DNN**	**0.9969**

Table 2. Accuracy comparison of proposed model with current models utilizing CIC-IDS 2017 dataset

Author	**Year**	**Technique**	**Accuracy**
Nidaa et al [32]	2020	DNN	90.25%
W. Elmasry et al [30]	2019	LSTM	92.41%
Farhan et al [31]	2022	DNN + BPSO	95%
Lin et al [33]	2019	LSTM + AM	96.20%
Ashik Elahi et al [34]	2024	GRU	96.23%
Proposed Method		**XGBoost+DNN**	**99.25%**

The better performance of the proposed XGBoost–CNN model can be explained by its excellent feature selection ability integrated with the robust learning capability of deep neural networks. XGBoost removes redundant and irrelevant features, thus lowering data dimensionality and increasing computational efficiency, while the CNN part learns intricate hierarchical patterns in the chosen features for better classification. This collaboration helps the model attain greater accuracy and reliability than single deep learning models like DNN, LSTM, or GRU. The findings affirm that the combination of feature selection and deep learning not only improves detection accuracy but also provides a more solid decision-making paradigm for intrusion detection systems in contemporary network environments.

5. Conclusion

This work proves that combining XGBoost-based feature selection and a Deep Neural Network offers a productive and dependable framework for intrusion detection in network environments. Through minimizing feature redundancy and prioritizing the most important attributes, the XGBoost approach improved the effectiveness of the classification process, while the DNN showed better ability in precise detection of both conventional and new cyberattacks. The experimental tests on two benchmark sets verify the stability of the proposed method with 99.69% accuracy on NSL-KDD and 99.25% accuracy on CIC-IDS 2017, with stable performance across various attack types. These findings confirm that the XGBoost–DNN model not just enhances detection accuracy but also enhances decision-making in intrusion detection systems. Therefore, this research contributes to developing intelligent, scalable, and highly accurate IDS models that can successfully protect contemporary networks against changing cybersecurity threats.

References

[1] H.S. Sánchez, D. Rotondo, T. Escobet, V. Puig, J. Quevedo, Bibliographical review on cyber attacks from a control oriented perspective, Annu. Rev. Control 48 (2019) 103–128,

[2] A. Al-Abassi, H. Karimipour, A. Dehghantanha, R.M. Parizi, An ensemble deep learning-based cyber-attack detection in industrial control system, IEEE Access 8 (2020) 83965–83973,

[3] A.J. Gallo, M.S. Turan, F. Boem, T. Parisini, G. Ferrari-Trecate, A distributed cyberattack detection scheme with application to dc microgrids, IEEE Trans. Automat Contr. 65 (9) (2020) 3800–3815,.

[4] R.V. fect Deshmukh, K.K. Devadkar, Understanding ddos attack (2015) & its effect in cloud environment, Procedia Comput. Sci. 49 202–210, doi:10.1016/j.procs.2015.04.245.

[5] P. Kamboj, M.C. Trivedi, V.K. Yadav, V.K. Singh, Detection techniques of ddos attacks: a survey, in: 2017 4th IEEE Uttar Pradesh Section International Conference on Electrical, Computer and Electronics (UPCON), 2017, pp. 675–679. 10.1109/UPCON.2017.8251130

[6] D. Gautam, V. Tokekar, A novel approach for detecting ddos attack in manet, Mater. Today:. Proc. 29 (2020) 674–677, doi:10.1016/j.matpr.2020.07.332.

[7] A. Bhardwaj, V. Mangat, R. Vig, S. Halder, M. Conti, Distributed denial of service attacks in cloud: state-of-the-art of scientific and commercial solutions, Comput. Sci. Rev. 39 (2021) 100332, doi:10.1016/j.cosrev.2020.100332.

[8] Hnamte, V., Hussain, J. (2023). DCNNBiLSTM: An efficient hybrid deep learning-based intrusion detection system. Telematics and Informatics Reports, 10, 100053.

[9] Wang, X. (2018). Design of temporal sequence association rule based intrusion detection behavior detection system for distributed network. Modern Electron. Techn , 41 (3), 108-114.

[10] Çavuşoğlu, Ü. (2019). A new hybrid approach for intrusion detection using machine learning methods. Applied Intelligence, 49, 2735-2761.

[11] Liu, Z., Wang, J., Liu, G., & Zhang, L. (2019). Discriminative low rank preserving projection for dimensionality reduction. Applied soft computing, 85, 105768.

[12] Thaseen, I. S., Kumar, C. A., & Ahmad, A. (2019). Integrated intrusion detection model using chisquare feature selection and ensemble of classifiers. Arabian Journal for Science and Engineering, 44, 3357-3368.

[13] Thaseen, I. S., & Kumar, C. A. (2017). Intrusion detection model using fusion of chi-square feature selection and multi class SVM. Journal of King Saud University-Computer and Information Sciences, 29(4), 462-472.

[14] Tao, P., Sun, Z., & Sun, Z. (2018). An improved intrusion detection algorithm based on GA and SVM. IEEE Access , 6 , 13624-13631.

[15] Sarker, M. N. I., Peng, Y., Yiran, C., & Shouse, R. C. (2020). Disaster resilience through big data: Way to environmental sustainability. International Journal of Disaster Risk Reduction, 51, 101769.

[16] Ferrag, M. A., Maglaras, L., Janicke, H., & Smith, R. (2019, September). Deep learning techniques for cyber security intrusion detection: A detailed analysis. In 6th International Symposium for ICS & SCADA Cyber Security Research 2019 6 (pp. 126-136).

[17] Yin, C., Zhu, Y., Fei, J., & He, X. (2017). A deep learning approach for intrusion detection using recurrent neural networks. IEEE Access , 5 , 21954-21961.

[18] He, K., Kim, D. D., & Asghar, M. R. (2023). Adversarial machine learning for network intrusion detection systems: a comprehensive survey. IEEE Communications Surveys & Tutorials.

[19] Talukder, M. A., Hasan, K. F., Islam, M. M., Uddin, M. A., Akhter, A., Yousuf, M. A., ... & Moni, M. A. (2023). A dependable hybrid machine learning model for network intrusion detection. Journal of Information Security and Applications, 72, 103405.

[20] Bamakan, S. M. H., Wang, H., & Shi, Y. (2017). Ramp loss K-Support Vector ClassificationRegression; a robust and sparse multi-class approach to the intrusion detection problem. Knowledge-Based Systems, 126, 113-126.

[21] Yan, B., Han, G., Huang, Y., & WANG, X. (2018). New traffic classification method for imbalanced network data. Journal of Computer Applications, 38(1), 20.

[22] Xu, H., Sun, Z., Cao, Y., & Bilal, H. (2023). A data-driven approach for intrusion and anomaly detection using automated machine learning for the Internet of Things. Soft Computing, 1-13.

[23] Naseer, S., & Saleem, Y. (2018). Enhanced network intrusion detection using deep convolutional neural networks. KSII Transactions on Internet and Information Systems (TIIS) , 12 (10), 5159- 5178.

[24] Kanna, P. R., & Santhi, P. (2021). Unified deep learning approach for efficient intrusion detection system using integrated spatial–temporal features. Knowledge-Based Systems, 226, 107132.

[25] Xiangyu Ma, Wei Shi, Aesmote: Adversarial reinforcement learning with smote for anomaly detection, IEEE Trans. Netw. Sci. Eng. 8 (2) (2020) 943–956.

[26] Lan Liu, Pengcheng Wang, Jun Lin, Langzhou Liu, Intrusion detection of imbalanced network traffic based on machine learning and deep learning, IEEE Access 9 (2020) 7550–7563.

[27] Mulyanto Mulyanto, Muhamad Faisal, Setya Widyawan Prakosa, Jenq-Shiou Leu, Effectiveness of focal loss for minority classification in network intrusion detection systems, Symmetry 13 (1) (2020) 4.

[28] Rahbar Ahsan, Wei Shi, Jean-Pierre Corriveau, Network intrusion detection using machine learning approaches: Addressing data imbalance, IET Cyber-Phys. Syst.: Theory Appl. 7 (1) (2022) 30–39.

[29] Li, Zhengfa, Chuanhe Huang, and Wanyu Qiu. "An intrusion detection method combining variational auto-encoder and generative adversarial networks." Computer Networks 253 (2024): 110724.

[30] Wisam Elmasry, Akhan Akbulut, and A. Zaim. 2019. Evolving deep learning architectures for network intrusion detection using a double PSO metaheuristic. Computer Networks 168 (12 2019), 107042.

[31] Rawaa Farhan, Abeer Maolood, and Nidaa Flaih. 2022. Optimized Deep Learning with Binary PSO for Intrusion Detection on CSE-CIC-IDS2018 Dataset. Journal of Al-Qadisiyah for Computer Science and Mathematics 12 (08 2022), 16–27.

[32] Rawaa Farhan, Abeer Maolood, and Nidaa Hassan. 2020. Performance analysis of flow-based attacks detection on CSE-CIC-IDS2018 dataset using deep learning Deep learning Flow-based intrusion detection Internet of thing (IOT). Indonesian Journal of Electrical Engineering and Computer Science 20 (12 2020), 1413–1418. https://doi.org/10.11591/ijeecs.v20.i3.pp1413-1418

[33] Peng Lin, Kejiang Ye, and Cheng-Zhong Xu. 2019. Dynamic Network Anomaly Detection System by Using Deep Learning Techniques. In Cloud Computing– CLOUD 2019, Dilma Da Silva, Qingyang Wang, and Liang-Jie Zhang (Eds.). Springer International Publishing, Cham, 161–176.

[34] Elahi, Md Ashik, Rafi Ahammed Songram, and Md Shahid Uz Zaman. "Network-Shield: Exploring the Efficacy of GRU Model in Intrusion Detection Using CIC-IDS 2018 Dataset." Proceedings of the 3rd International Conference on Computing Advancements. 2024.