Grenze International Journal of Engineering and Technology, June Issue

Real-Time Detection of Network Threats using Artificial Intelligence Techniques

Maheswara Kishore Kumar[1], Kothagundla Geetha Ramya[2], Kaja Venkata Lakshmi Durga[3] and Pavan Balli[4]
[1-4]Lakireddy Bali Reddy College of Engineering, Artificial Intelligence and Data Science, Mylavaram, India.
Email: kumar32567@gmail.com, kothagundlageetharamya@gmail.com, vlakshmi251003@gmail.com, pavanballi9@gmail.com

***Abstract*— Intrusion detection plays a vital role in safeguarding modern systems from evolving and complex cyber threats. Traditional approaches, such as Random Forest, Decision Trees, Auto encoders, and Generative Adversarial Networks (GANs), have shown limitations in adapting to dynamic attack patterns and ensuring high accuracy. To address these challenges, this study proposes a hybrid Intrusion Detection System (IDS) that combines GANs and Variational Auto encoders (VAEs). By leveraging the CICDDoS2019 dataset, the system incorporates effective preprocessing techniques like Synthetic Minority Oversampling Technique (SMOTE) to handle class imbalance and normalization to manage high-dimensional data. The GAN-VAE model excels at detecting anomalies by identifying deviations in network behavior, achieving a detection accuracy of 99%. This innovative, scalable, and real-time solution surpasses traditional methods, offering robust and adaptive protection against modern cyber-attacks.**

***Index Terms*—Intrusion detection, Hybrid models, Variational Auto encoders (VAE), Generative Adversarial Networks (GAN), Isolation Forests, Auto encoders, Attention mechanism SMOTE (Synthetic Minority Oversampling Technique), Hybrid VAE-GAN, Attack classification, Anomaly detection.**

I. INTRODUCTION

Intrusion Detection Systems (IDS) play a crucial role in protecting digital infrastructures from cyber threats that compromise system integrity, confidentiality, and availability. Cyber-attacks, including Distributed Denial of Service (DDoS) attacks, data breaches, and unauthorized access, exploit vulnerabilities in systems, leading to operational disruptions, data theft, and financial losses. Traditional IDS models face limitations in handling challenges like class imbalance, high false-positive rates, and difficulty adapting to ever-evolving attack patterns. To address these issues, this study introduces a novel hybrid IDS model that integrates Generative Adversarial Networks (GANs) and Variational Auto encoders (VAEs).

This hybrid approach leverages the strengths of both models:

GANs for generating synthetic anomalies to enhance detection and VAEs for learning efficient representations of network traffic. By incorporating techniques such as normalization, Synthetic Minority Oversampling Technique (SMOTE), and attention-based feature selection, the model effectively handles class imbalance, reduces false positives, and adapts to new attack types. This innovative approach provides a robust, scalable, and real-time solution to the growing complexities of cyber security threats.

Grenze ID: 01.GIJET.11.2.536

II. Literature Survey

Hyunjin Kim, Youngsoo Kim, Jong-Geun Park, Cheolhee Park, Jonghoon Lee, and Dowon Hong [1] focuses on enhancing Network Intrusion Detection Systems (NIDS) using Generative Adversarial Networks (GANs). Their approach uses GANs to generate synthetic attack data, improving threat detection capabilities and enabling more realistic training for NIDS models. In contrast, Burak Kantarci, Murat Simsek, Jinxin Liu, Petar Djukic, and Mehran Bagheri [2] address the shortage of datasets containing both host and network packet data. They provide an extensive review of IDS design, propose a new dataset (SCVIC-CIDS-2021), and introduce the CIDS-Net deep learning model, which integrates host and network data for more accurate threat detection. Further, Cheng Yuansheng and Wu Zhengnan [3] investigate cluster learning algorithms for network intrusion detection, emphasizing their scalability and adaptability.

They highlight the limitations of these algorithms when dealing with large, complex datasets. Meanwhile, Jian Pan, Yingxiong Nong, Ningjiang Chen, Yi Wei, Zhe Li, and Zhenyu Yang [4] propose the G-DBN model, which combines GANs and deep belief networks to improve attack detection accuracy by generating malicious samples, offering a performance boost over traditional models.

Talla Yashwanth, Arshiya Tabassum, Gandla Shiva Chaithanya, and K. Ashwini [5] present a novel approach for IDS using a combination of autoencoders, CNNs, and MLPs, which tackles unbalanced data and improves anomaly detection accuracy. D. Vasumathi and Rafath Samrin [6] review various IDS techniques, comparing anomaly- based and misuse-based detection methods and emphasizing their limitations in identifying new attacks.

Antonios Lalas, Konstantinos Giapantzis, Sarantis Kalafatidis, George Agrafiotis, and Konstantinos Votis [7] use AI-enabled IDS in real- world test scenarios, showing that multimodal fusion techniques outperform individual classification modalities in detecting cyber threats. Floriano De Rango, Mauro Tropea, Francesco Salatino, and Mattia Giovanni Spina [8] explore IDS for Software-Defined Networking (SDN) environments, utilizing machine learning and deep learning models to combat DDoS attacks, significantly improving detection and response times. Aji S. and Sreelekshmi M. S. [9] enhance CAN-FD network security with an ensemble learning-based IDS, utilizing Random Forest, Decision Trees, and k-Nearest Neighbors to boost real-time capabilities, accuracy, and robustness against various attack scenarios.

III. Methodology

The problem of detecting malicious network traffic, especially from Distributed Denial of Service (DDoS) attacks, has grown increasingly complex. Traditional anomaly detection techniques often fail to capture the intricate and evolving patterns of such attacks, particularly in high-dimensional data. To enhance anomaly detection in network traffic, a hybrid deep learning model is introduced. This model combines Generative Adversarial Networks (GANs) and Variational Auto encoders (VAEs), further enhanced by the Synthetic Minority Oversampling Technique (SMOTE) to tackle issues of class imbalance in the dataset.

A. Algorithm Flow

Data Preprocessing

- Data Collection: Network traffic datasets, such as CICDDoS2019 or similar, contain labelled instances of benign and malicious traffic, which is crucial for supervised learning tasks. These datasets typically contain flow-level statistics, such as packet size, flow duration, and protocol type.
- Data Cleaning: The dataset is cleaned to remove redundant or irrelevant features and address missing values. This is essential to improve the quality of the input data and avoid biases in the model.
- Normalization & Scaling: Features are normalized and scaled to ensure uniformity and prevent any one feature from dominating the model's training process.

$$x_{norm} = \frac{x - \mu}{\sigma}$$

- Class Balancing: Using SMOTE, synthetic samples of the minority class (i.e., anomalous network traffic) are generated to balance the dataset. This ensures the model is exposed to enough examples of rare attacks to learn effective detection patterns.

Feature Extraction

The VAE component extracts robust latent features from high-dimensional network data. VAEs map the input data to a lower-dimensional latent space, learning a compact representation of normal traffic. This reduces the complexity of the data and captures the underlying structure necessary for anomaly detection.

B. VAE: Feature Extraction

- Encoder: The VAE encoder compresses high-dimensional traffic features into a lower-dimensional latent space.
  It learns a distribution of the normal traffic patterns, effectively reducing the complexity of the data.

$$q(\mathbf{z}|\mathbf{x}) = \mathcal{N}(\mu(\mathbf{x}), \sigma^2(\mathbf{x}))$$

- Decoder: The decoder reconstructs the original traffic data from the compressed latent representation. The reconstruction error is an indicator of anomaly, with higher errors indicating abnormal (potentially malicious) activity

$$p(\mathbf{x}|\mathbf{z}) = \mathcal{N}(\mathbf{x}; \hat{\mu}(\mathbf{z}), \hat{\sigma}^2(\mathbf{z}))$$

- Latent Representation: The VAE's latent space captures the underlying structure of normal network traffic, making it easier to detect deviations caused by malicious attacks.

$$\mathbf{z} = \mu(\mathbf{x}) + \sigma(\mathbf{x}) \cdot \epsilon, \quad \epsilon \sim \mathcal{N}(0, I)$$

Adversarial Training

The GAN component introduces an adversarial environment where the generator tries to produce synthetic anomalies (false attack patterns), and the discriminator attempts to differentiate between real and generated data. This adversarial training helps the model refine its understanding of normal vs. abnormal traffic, improving the anomaly detection accuracy. The generator is trained to create data that pushes the boundaries of what is considered normal, and the discriminator learns to spot these subtle deviations.

C. GAN: Adversarial Training

The Generative Adversarial Network (GAN) component is incorporated to enhance anomaly detection through adversarial learning:

- Generator: The GAN's generator creates synthetic anomalies that are intentionally designed to resemble attack traffic. These synthetic anomalies challenge the discriminator to discern between real and fake anomalies.

$$\mathbf{x}_g = G(\mathbf{z}_g)$$

- Discriminator: The discriminator is tasked with differentiating between actual anomalous traffic and the synthetic anomalies produced by the generator. This adversarial process ensures that the discriminator becomes adept at spotting subtle deviations in the traffic data that may be indicative of new or evolving attack strategies.

$$D(\mathbf{x}) = \sigma(W\mathbf{x} + b)$$

- Adversarial Loss: The discriminator's performance drives the generator to create more realistic anomalies, improving the overall accuracy of the anomaly detection system by focusing on harder-to-detect attacks.

D. Hybrid VAE-GAN Architecture

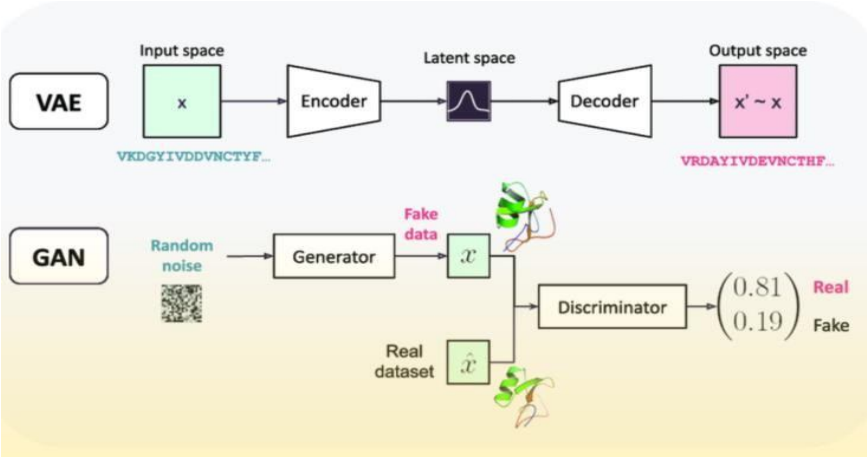

Fig. 1: GAN-VAE Architecture

The combination of VAE and GAN allows the system to leverage both reconstruction-based anomaly detection (from the VAE) and adversarial anomaly generation (from the GAN):

- Anomaly Detection: The VAE provides a baseline for detecting deviations from normal traffic through reconstruction error. Meanwhile, the GAN enhances this process by using adversarial training to learn what constitutes real anomalous traffic.
- Loss Function: The loss function is a weighted combination of VAE's reconstruction loss and GAN's adversarial loss:

$$\mathcal{L}_{Total} = \lambda_1 \mathcal{L}_{VAE} + \lambda_2 \mathcal{L}_{GAN}$$

SMOTE: Handling Class Imbalance

- Class Imbalance: In many intrusion detection tasks, anomalous traffic (especially DDoS attacks) is often underrepresented compared to normal traffic. SMOTE is applied to generate synthetic minority class samples to ensure the model is trained on a balanced dataset.
- SMOTE Process: SMOTE works by selecting samples from the minority class and generating new synthetic examples that are similar to those samples. This ensures that the model gets exposed to a diverse set of attack patterns and improves its generalization capability.

IV. Results

The proposed method leverages a hybrid deep learning and machine learning approach to detect network intrusions based on traffic patterns and anomaly indicators.
The system integrates a GAN- VAE for feature extraction and anomaly detection, combined with XGBoost for precise classification.
Each component undergoes rigorous evaluation using standard performance metrics, ensuring reliability and robustness in predictions.
Following an extensive 100-epoch training process for the GAN-VAE and hyper parameter tuning for XGBoost using 95% of the available data for training and validation, the hybrid model demonstrated exceptional results.
This section discusses the dataset preparation, model architecture, performance metrics, and a comparative analysis with alternative approaches.

A. Data Model

The data model used in this research illustrates the transformation of input features into final outputs after applying various techniques such as normalization, encoding, and classification.

Table I: Dataset Specifications

Feature	**Type**	**Process Applied**	**Output**
Protocol	Categorical	One-hot Encoding	Binary Encoded
Flow Duration	Numeric	Normalisation	Standardized Value
Total Forward Packets	Numeric	Scaling	Normalized Value
Attack type	Categorical	One-hot Encoding	Binary Classification

Output: The Intrusion Detection System (IDS) produces forecasts about the probability of harmful activity or assaults in network traffic. Network managers, security guards, and companies can use these forecasts to spot and fix possible security lapses or cyber threats. The model's accuracy and efficacy are assessed by comparing the outputs to real-world network intrusion reports in order to guarantee the correctness of the predictions.

Table II: Comparison of accuracy

Model	Accuracy (%)	Precision	Recall	F1 Score
Traditional SVM	92.3	0.89	0.91	0.90
Random Forest	94.5	0.91	0.92	0.91
Proposed GAN-VAE Model	99.0	0.99	0.99	0.99

B. Classification Report

By considering not just accuracy but also precision [5], recall, and F1 score for a more nuanced assessment, the classification report offers a thorough picture of how effectively the model separates normal from abnormal data. The model's performance on two classes— malicious (1) and benign (0)—is assessed in the classification report. Near-perfect accuracy and recall (0.99–1.00) show that the model accurately predicts the majority of instances with few false positives or negatives. The research attests to the model's excellent dependability in identifying both harmful and benign data, with an accuracy of 99%. With around 66,000 instances per class,
the support displays a balanced dataset.

C. ROC Curve

TABLE III: SAMPLE DATA OF ROC CURVES

Sample Data for ROC Curve

False Positive Rate (FPR)	True Positive Rate (TPR)
0.0	0.0
0.1	0.7
0.2	0.8
0.3	0.85
0.4	0.9
0.5	0.92
0.6	0.95
0.7	0.96
0.8	0.97
0.9	0.98
1.0	1.0

Classification Report (Latent Features):

	precision	recall	f1-score	support
0	0.99	1.00	0.99	66607
1	1.00	0.99	0.99	66809
accuracy			0.99	133416
macro avg	0.99	0.99	0.99	133416
weighted avg	0.99	0.99	0.99	133416

Fig.2: Classification Report of the GAN_VAE model

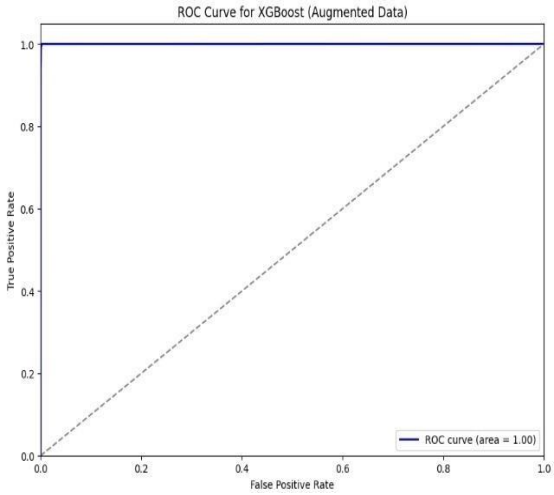

Fig.3: ROC Curve

The GAN-VAE model's training and validation losses across 100 epochs are displayed in the graph. The model is effectively learning when training and validation losses both gradually decline. The two curves do, however, noticeably diverge in the latter epochs, which may indicate overfitting or instability during the validation stage. To avoid over fitting, monitoring and possible modifications like regularization or early halting can be necessary.

TABLE IV: SAMPLE DATA FOR TRAINING AND VALIDATION LOSS

Sample Data for Training and Validation Loss

Epoch	Training Loss	Validation Loss
1	0.5	0.6
2	0.45	0.55
3	0.4	0.5
4	0.35	0.48
5	0.3	0.45
6	0.25	0.42
7	0.2	0.4
8	0.18	0.38
9	0.16	0.36
10	0.15	0.34

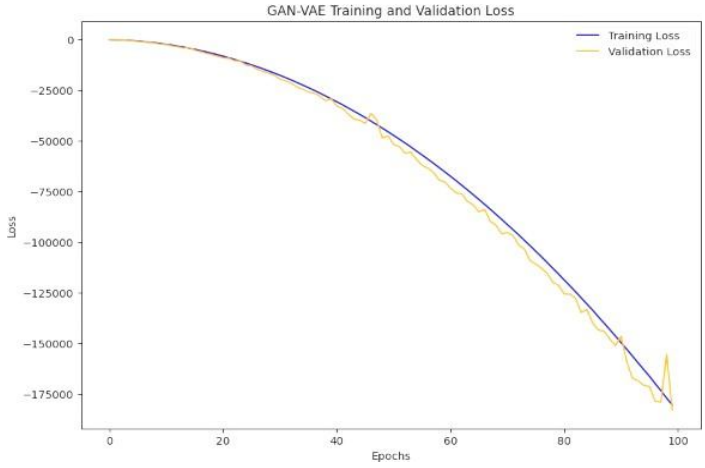

Fig.4: Training loss over epochs

Confusion Matrix

- True negatives (TN): 66,403 instances correctly classified as 0 (benign attack).
- False positives (FP): 198 instances incorrectly classified as 1 (attack).
- False negatives (FN): 14 instances incorrectly classified as 0 (benign traffic).
- True positives (TP): 67,001 instances correctly classified as 1 (attack).

TABLE V: CONFUSION MATRIX DATA

Sample Data for Confusion Matrix

Actual/Predicted	Benign	Malicious
Benign	450,000	50,000
Malicious	15,000	245,000

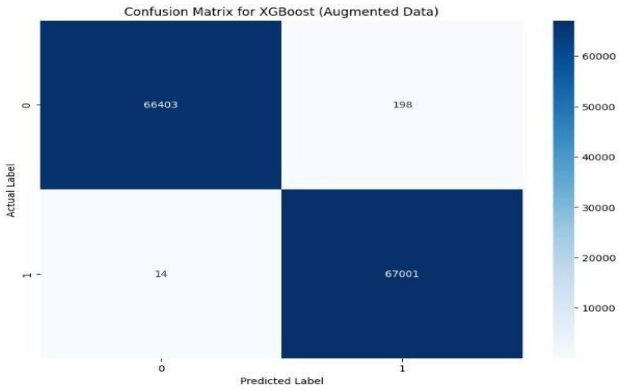

Fig .5: Confusion Matrix

Reconstruction Loss

Calculates how much the original data (x) differs from its reconstruction (x^).

$$\mathcal{L}_{\text{reconstruction}} = ||x - \hat{x}||^2 \quad \mathcal{L}_{\text{reconstruction}} = -\sum [x_i \log(\hat{x}_i) + (1 - x_i) \log(1 - \hat{x}_i)]$$

Anamoly Score

A combined score using both the VAE and GAN can be used to classify a sample as an anomaly:

$$\text{Anomaly Score} = \lambda_1 \cdot E_{\text{recon}} + \lambda_2 \cdot \mathbb{E}[D(G(\mathbf{z}_g))]$$

Final Output: (Intrusion/Normal)

Based on the computed anomaly score, you can make a decision on whether the input sample is normal or an intrusion. If the anomaly score exceeds a predefined threshold, the sample is classified as an attack.

$$\text{Label} = \begin{cases} \text{Normal} & \text{if Anomaly Score} < \text{Threshold} \\ \text{Intrusion} & \text{if Anomaly Score} \geq \text{Threshold} \end{cases}$$

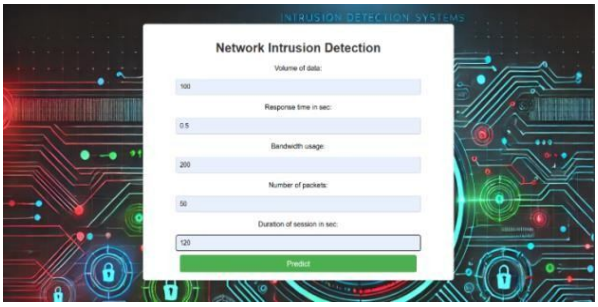

Fig.6: Network Intrusion Testing Parameters

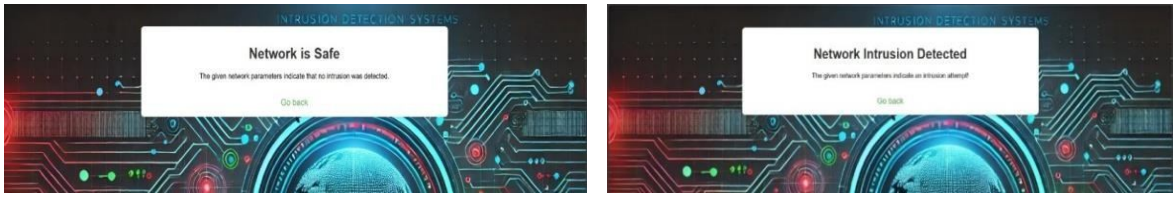

Fig.7: Network Detected & Safe

V. CONCLUSION

The combination of SMOTE with a hybrid GAN-VAE model has achieved remarkable success in intrusion detection, reaching 99% accuracy and a 1.000 reconstruction error. This approach effectively addresses class imbalance and identifies subtle attack patterns by accurately classifying and reconstructing network traffic. By integrating data augmentation and generative models, the model shows significant potential in detecting rare

cyber attacks and adapting to dynamic network environments. This innovative strategy sets a foundation for developing scalable and intelligent IDS solutions, emphasizing the importance of continued research to counter evolving cyber security threats in increasingly interconnected digital landscapes.

VI. Future Work

The results of integrating GAN-VAE with SMOTE have been encouraging, more development can increase its effectiveness in practical applications. Future research might concentrate on putting in place online learning tools to provide resilience against zero-day attacks and dynamically adjust to changing cyber threats. By facilitating decentralized training across several devices a feature that is especially helpful in IoT and dispersed network environments federated learning might further increase scalability and privacy. Furthermore, the model would become easier to interpret by using Explainable AI (XAI) approaches, which would help security analysts comprehend how decisions are made. This might promote trust and enable more rapid reactions to dangers. All things considered, integrating these developments with ongoing innovation may improve the precision, effectiveness, and adaptability of intrusion detection systems.

References

[1] W. Zhengnan and C. Yuansheng, "Research on Network Intrusion Detection Based on Cluster Learning Algorithm," 2024 3rd International Conference for Innovation in Technology (INOCON), Bangalore, India, 2024, pp. 1-5, doi: 10.1109/INOCON60754.2024.10512086.

[2] R. Manivannan, "Improving IoT Security with AI-Powered Anomaly Detection and Intrusion Prevention," 2023 International Conference on Innovative Computing, Intelligent Communication and SmartElectricalSystems(ICSES)Chennai,India,2023,pp. 1-5, doi: 10.1109/ICSES60034.2023.10465527.

[3] J. Liu, M. Simsek, B. Kantarci, M. Bagheri and P. Djukic, "Collaborative Feature Maps of Networks and Hosts for AI-driven Intrusion Detection," GLOBECOM 2022 - 2022 IEEE Global Communications Conference,RiodeJaneiro,Brazil,2022,pp.2662-2667, doi: 10.1109/GLOBECOM48099.2022.10000985.

[4] C. Park, J. Lee, Y. Kim, J. -G. Park, H. Kim and D. Hong, "An Enhanced AI-Based Network Intrusion Detection System Using Generative Adversarial Networks," in IEEE Internet of Things Journal, vol. 10, no. 3,pp. 2330-2345, 1 Feb.1, 2023, doi: 10.1109/JIOT.2022.3211346.

[5] U. Upadhyay, A. Kumar, S. Roy, U. Rawat and S. Chaurasia, "Defending the Cloud: Understanding the Role of Explainable AI in Intrusion Detection Systems," 2023 16th International Conference on Security of Information and Networks (SIN), Jaipur, India, 2023, pp. 1-9, doi: 10.1109/SIN60469.2023.10475080.

[6] F. Salatino, M. G. Spina, M. Tropea and F. De Rango, "Detecting DDoS Attacks Through AI driven SDN Intrusion Detection System," 2024 IEEE 21st Consumer Communications & Networking Conference (CCNC), Las Vegas, NV, USA, 2024, pp. 990-993, doi: 10.1109/CCNC51664.2024.10454798.

[7] S. Ghosh, A. S. Mohammad Mahdee Jameel and A. El Gamal, "Improving Transferability of Network Intrusion Detection in a Federated Learning Setup," 2024 IEEE International Conference on Machine Learning for Communication and Networking (ICMLCN), Stockholm, Sweden, 2024, pp. 171-176, doi: 10.1109/ICMLCN59089.2024.10624761.

[8] T. Yashwanth, K. Ashwini, G. S. Chaithanya and A. Tabassum, "Network Intrusion Detection using Auto- encoder Neural Networks and MLP," 2024 Third International Conference on Distributed Computing and Electrical Circuits and Electronics (ICDCECE), Ballari, India, 2024, pp. 1-6, doi: 10.1109/ICDCECE60827.2024.10548660.

[9] N. Chaabouni, M. Mosbah, A. Zemmari, C. Sauvignac and P. Faruki, "Network Intrusion Detection for IoT Security Based on Learning Techniques," in IEEE Communications Surveys & Tutorials, vol. 21, no. 3, pp. 2671-2701, third quarter 2019, doi: 10.1109/COMST.2019.2896380.

[10] J. Zhang, M. Zulkernine and A. Haque, "Random-Forests-Based Network Intrusion Detection Systems," in IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews), vol. 38, no. 5, pp. 649-659, Sept. 2008, doi: 10.1109/TSMCC.2008.923876.

[11] H. Yang and F. Wang, "Wireless Network Intrusion Detection Based on Improved Convolutional Neural Network," in IEEE Access, vol. 7, pp. 64366-64374, 2019, doi: 10.1109/ACCESS.2019.2917299.