Received: 19 June 2026 | Revised: 29 July 2026 | Accepted: 14 August 2026 | Published online: 1 September 2026

RESEARCH ARTICLE

Artificial Intelligence and Applications
2026, Vol. 00(00) 1–21
DOI: 10.47852/bonviewAIA620210960

GAN-Augmented Intrusion Detection: A Feature-Wise Attention MLP Framework for Imbalanced Network Traffic

Osha Shukla[1], Annu Mishra[2], Mohd Dilshad Ansari[3,*], Ravi Prakash Chaturvedi[2] and Rajneesh Kumar Singh[4]

[1] *Department of Cybersecurity and Technology, JPMorgan Chase, USA*

[2] *Department of Computer Science and Technology, Manav Rachna University, India*

[3] *Department of Computer Science & Engineering, SRM University Delhi-NCR, India*

[4] *Department of Computer Science and Applications, Sharda University, India*

Abstract: With the increase in Internet of Things applications, there is also an increase in refined cyberattacks. Earlier, there were a limited number of classes under cyberattacks, but now, there are numerous classes. Some of them are distinguishable, and some of them are very tough to be identified. Approaches involving convolutional neural networks, multilayer perceptrons (MLPs), and Long Short Term Memory (LSTM) have their own limitations in terms of generalization. Thus, this research represents a GAN-Augmented Attention-Enhanced MLP intrusion detection system, which we named as GAN-MLP. It uses Wasserstein Generative Adversarial Networks with Gradient Penalty to generate synthetic minority attacks that look real. The feature-wise attention mechanism of MLP helps to gain unequal and inequitable features and helps the model to outperform as compared to other models. The dataset used was CICIDS2017, and all the metrics were calculated on this dataset. The model achieved 96.5% accuracy, 97.12% precision, 96.71% F1-score, and recall of 96.5%. The Receiver-Operating Characteristics and Area Under the Curve (ROC-AUC) values consisting of micro-average 0.9992 and macro-average 0.9977 confirm the model's remarkable performance. Thus, it is well suited for real-time intrusion detection and also for imbalanced network systems. The empirical analysis of the suggested methodology substantiates the efficacy of this strategy in comparison to the traditional deep learning approach.

Keywords: attention mechanism, intrusion detection, GAN augmentation, network security

1. Introduction

Modern communication infrastructures are larger and more complicated than they were in the past due to the exponential expansion of internet-connected devices and network-based services [1, 2]. Distributed Denial of Service (DDoS), Brute-Force assaults, botnets, and web-based incursions are just a few examples of the cyberattacks that have grown more common as organizations depend more and more on digital networks for mission-critical activities [3]. Researchers and practitioners alike are now deeply concerned about the need to guarantee network security. Intrusion detection systems (IDS) are crucial in preventing damage to network infrastructures and detecting malicious actions [4, 5].

There are two main schools of thought when it comes to conventional IDS: signature-based and anomaly-based [6]. Signature-based systems are good at identifying known threats that follow previously established patterns. However, they often miss zero-day assaults [7]. Because of their capacity to automatically extract complicated characteristics from massive datasets, deep learning methods have recently attracted a lot of interest for intrusion detection [8]. Network intrusion detection tasks have made use of a variety of architectures, including convolutional neural networks (CNNs), recurrent neural networks, and multilayer perceptrons (MLPs) [9]. These models outperform more conventional machine learning methods in terms of detection accuracy because they are able to uncover hidden patterns and nonlinear correlations in network data. Nevertheless, there are still a number of obstacles that deep learning-based IDS models encounter when used on real-world network datasets, even with all these benefits.

Class imbalance is a big problem since there are many more examples of harmless traffic than harmful, and certain kinds of attacks happen relatively seldom. The failure to adequately identify minority attack categories may be a consequence of machine learning models being skewed toward majority classes due to this imbalance [10]. There has been recent talk of using Generative Adversarial Networks (GANs) to generate synthetic samples that

***Corresponding author:** Mohd Dilshad Ansari, Department of Computer Science & Engineering, SRM University Delhi-NCR, India. Email: mohddilshadcse@srmuniversity.ac.in, m.dilshadcse@gmail.com

better reflect minority classes in order to solve this problem [11]. Improving classification performance in unbalanced cybersecurity datasets has been shown to be possible with GAN-based data augmentation.

When it comes to detecting intrusions in networks, high-dimensional feature spaces provide still another obstacle. There are a lot of flow-based properties in datasets on network traffic, and not all of them help identify attacks to the same extent. Successful applications of attention mechanisms in deep learning models have enabled models to automatically assign significance weights to pertinent features, enabling models to zero in on the most valuable data characteristics [12]. Consequently, feature representation and classification performance may be enhanced by including attention techniques into IDS designs.

Many researchers have worked on IDS, but current approaches have treated data imbalance or feature significance individually. The literature lacks research on the attention-based feature learning and adversarial data augmentation integrated into a single IDS framework. This weakness is addressed in this research by proposing an intrusion detection model named GA-MLP (GAN-Augmented Attention-Enhanced MLP). For better feature prioritization, a feature-wise attention mechanism is added within an MLP classifier, and Wasserstein Generative Adversarial Networks with Gradient Penalty (WGAN-GP) are employed to generate synthetic examples of minority attack classes. The CICIDS2017 dataset used includes different types of contemporary network attacks combined together to form a single empirical dataset. The experimental results demonstrate the superior classification performance and the minority attack-detection performance of the proposed approach when compared with the traditional machine learning and deep learning-based IDS models.

Previous research has looked at GAN-based attention and augmentation processes separately, and there are a few hybrid techniques. However, the majority of these methods use computationally expensive and impractically complicated architectures, such as CNN-LSTM or Transformer-based models. The suggested GA-MLP framework, on the other hand, is concerned with a lightweight but efficient integration of a feature-wise attention mechanism into a basic MLP architecture, as well as data augmentation based on Wasserstein GANs.

Combining these components isn't the only thing that makes this work unique; what really sets it apart are:

1) Designing an efficient approach that achieves acceptable performance and does not introduce an undue level of complexity.
2) Giving attack class-wise insights into feature relevance in a tabular IDS scenario based on feature-wise attention.
3) Using controlled comparisons to systematically measure the impact of augmentation with GANs on minority-class identification. The proposed model is an interpretable and deployable alternative to existing heavy hybrid IDS strategies, solving the problems with efficiency and performance.

2. Literature Review

Recent developments in IDS have been more directed toward overcoming class imbalance, better feature representation, and generalization of the current network environment. An excellent tool for addressing the issue of class imbalance is the GAN. It synthesizes an artificial sample of underrepresented attack classes. Researchers [1] proposed a hybrid CNN-LSTM system with WGAN-GP and human-in-the-loop feedback to address uncertainty. The system proved to be more robust and flexible in a dynamic intrusion detection environment. Similarly, the authors of [2] suggested an Internet of Things (IoT)-cloud GAN-enhanced deep ensemble model, which improved minority threat detection and showed good cross-domain generalization [3].

There have been recent studies into generative and intelligent designs of IDS. The deep learning-based framework proposed by [4] is the IDS-GPT, which relies on the latest representational learning algorithms to perform intrusion detection. This explains the increasing popularity of applying AI-driven architectures to intrusion detection. A previous study by [5] showed the efficiency of fuzzy logic-based IDS models in identifying malicious port scanning traffic with a focus on interpretability and rule-based reasoning in intrusion detection. Computational cost and energy efficiency have also become a critical issue in the research of IDS [6]. In a wide-ranging benchmark study of GAN architectures in energy-aware IoT intrusion detection, [7] proposed an accuracy-per-joule metric to balance the performance of detection with computation efficiency. This shows the necessity of lightweight and scalable IDS models that could be applied to real-life deployment [8]. Table 1 represents various studies carried out for IDS along with the datasets used. It also summarizes the key findings or advantages and limitations of each of the approaches.

While all of these works demonstrate significant advancements, most existing techniques focus on data augmentation techniques such as the use of GAN-based models, improvements

Table 1
Summary of various approaches used for intrusion detection

Refs.	Study/model description	Datasets used	Key findings
[9]	To fix imbalanced intrusion detection system (IDS) data, a Conditional Generative Adversarial Network (CGAN) is used, which includes a new aggregate encoder-decoder. After training an ensemble game-theoretic classifier on the combined dataset, the GAN improves rare attack samples	Network Security Labaratory-Knowlwdge Discovery and Data Mining (NSL-KDD), UNSW-NB15	Effectively augments minority attack samples; improves multi-class intrusion detection; achieves 99.62% accuracy on NSL-KDD; enhances detection of infrequent attacks
[10]	IGAN is an IDS framework for Imbalanced Generative Adversarial Networks. Through the use of an ensemble classifier that combines LeNet-5 and Long Short Term Memory (LSTM), IGAN detects underrepresented classes by oversampling them	UNSW-NB15, CICIDS2017	Successfully oversamples minority classes; improves rare attack detection; achieves over 98% accuracy; outperforms existing deep learning IDS under class imbalance

(*Continued*)

Table 1
(*Continued*)

Refs.	Study/model description	Datasets used	Key findings
[11]	MCGAN with BiLSTM for IDS imbalance handling	NSL-KDD + (20 features)	GAN-based oversampling improves minority attack detection; achieves 95.16% accuracy with low false-positive rate (FPR) (2.1%); effectively addresses class imbalance
[12]	Soft Nearest Neighbour Loss (SNNL)-regularized WGAN-GP	NSL-KDD, CSE-CICIDS2017/2018	Feature-regularized GAN improves minority-class F1-scores; synthetic samples closely match real data; maintains high overall accuracy
[13]	Stacked Generative Adversarial Network based Intrusion Detection System (SGAN-IDS) with self-attention	CICIDS2017	Generates highly realistic adversarial attacks; significantly reduces IDS detection performance; reveals vulnerabilities in IDS models
[14]	Hybrid CNN–BiLSTM–Transformer with Synthetic Minority Oversampling Technique (SMOTE)	CICIDS2017, BoT-IoT	Attention-enhanced architecture achieves very high accuracy (up to 99.80%); improves recall for rare IoT attacks in imbalanced traffic
[15]	Cyber Detect-MLP (distributed MLP IDS)	TON_IoT	Class-weighted loss and oversampling enable high IDS accuracy (98.87%); outperforms Random Forest (RF), XGBoost, and standard MLP models
[16]	Class wise Focal Loss Attention Mechanism Variational Autoencoder (CWFLAM-VAE) with XGBoost	NSL-KDD, CSE-CICIDS2018	VAE-based oversampling and focal loss enhance minority-class F1-scores (>97%); achieves very low FPRs
[17]	Bi-GAN with one-class anomaly detection	NSL-KDD, CIC-DDoS2019	GAN-based one-class IDS attains high accuracy without threshold tuning; effective under severe class imbalance
[18]	Synchronous Classifier Wasserstein Generative Adversarial Network (SC-WGAN) with synchronous classifier	Benchmark IDS datasets	Auxiliary classifier guides GAN training; produces high-quality minority samples; improves recall and precision over prior GAN methods
[19]	Dual-discriminator CGAN with hybrid feature selection	BoT-IoT	GAN augmentation combined with optimized feature selection achieves high detection rates with minimal false positives in IoT IDS
[20]	GAN-based IDS for SCADA systems	SCADA (DNP3)	GAN oversampling mitigates data scarcity; achieves near-perfect detection (99.136%) in critical infrastructure networks
[21]	FedACNN with attention (federated IDS)	UNSW-NB15	Lightweight attention-based CNN achieves ~99% accuracy; suitable for resource-constrained IoT edge environments
[22]	Transformer-based IDS	CIC-IoT-2023	Transformer outperforms CNN/LSTM in multi-class IDS; achieves 99.40% accuracy despite class imbalance
[23]	Few-shot IDS with attention and meta-learning	CICIDS2017, CICIDS2018	Attention-driven meta-learning enables high detection accuracy with very few samples; effective under extreme data scarcity

to the model architecture, such as CNN/LSTM/Transformer hybrids, or alternative learning paradigms, like the use of fuzzy systems or generative AI [24–26]. However, very little work has been done to model high-quality synthetic data generation and feature-level modeling in a unified framework. In contrast, the proposed GA-MLP model integrates Wasserstein GAN-based augmentation with a feature-wise attention mechanism on a lightweight MLP framework. The integration allows the model to deal with both class imbalance and feature heterogeneity, with the effect that the minority classes are more accurately detected, generalization improves, and the model is computationally efficient. By providing a scalable and well-balanced IDS that can

handle both theoretical and practical deployment scenarios, the presented work fills a vacuum in the existing literature [27–29].

3. Proposed Methodology

The proposed solution has an end-to-end pipeline that is very useful in creating an effective IDS despite the disparity in the frequency of occurrences among the classifications. The CICIDS2017 network flows in CSV format and at the flow level can be processed to aggregate and remove noisy data points and also to create normalized labels and trim micro-attacks to fit into the normalized seven-class taxonomy system and to also normalize numeric attributes. The significant disparity between prevalent benign/DoS-type traffic and infrequent BruteForce, WebAttack, and Bot flows is tackled at the data level by the application of class-specific Wasserstein GANs with gradient penalty to provide statistically coherent artificial samples for the underrepresented groups. The genuine and artificial flows are subsequently merged into an augmented training set, upon which an attention-enhanced MLP is trained with class-weighted loss to further mitigate residual imbalance.

3.1. Data collection

This study employs the publicly accessible CICIDS2017 intrusion detection dataset, obtained from the Kaggle mirror. The initial dataset was produced by the Canadian Institute for Cybersecurity within a regulated testbed intended to replicate authentic enterprise network traffic and attack scenarios. The testbed involved the use of different devices like Windows, Ubuntu, and macOS in combination with different behaviors of the users together with different protocols involved in the facilitation of apps like HTTP, HTTPS, FTP, SSH, and DNS. The network traffic on the victim side and the attacker side was reflected to the control interface of the CIC, resulting in the capture of the whole communications system in the network passively. The capture involved the CIC injecting different communications such as user traffic and attack scripts that include the Brute-Force attack, Denial of Service (DoS/Distributed Deniel of Service (DDoS)) attack, web exploitation attack, internal invasion, communications with botnets, and targeted attack with the Heartbleed exploit.

Rather than using the original traces, this research uses the flow-level CSV files, which can be accessed from the Kaggle site. The whole set of CSV files holding the variety of workdays as well as the duration of the attack (for instance, Monday Working Hours, Tuesday Working Hours, Thursday Morning Web Attacks, Friday Afternoon PortScan, and DDoS) is systematically gathered and amalgamated in order to get an entire collection of data. This will help in getting the required data that hold the amalgamation of various types of attacks and the working hours similar to that of CICIDS2017; apart from this, it will also help in providing an easier experience in the process.

3.2. Data description

An extensively used benchmark for intrusion detection research, the CICIDS2017 dataset, is employed in this work. All sorts of actual network traffic, from benign to malicious, including Port-Scan, BruteForce, Web-Attack, and Bot assaults, are included in the collection, as shown in Table 2. It includes around 2.83 million records based on flows, and 78 numerical characteristics were retrieved using CICFlowMeter. Due to its extensive use in IDS research, we will simply highlight the most important features for this study. For the sake of brevity, we will not go into the specifics of the dataset creation processes or the network settings.

Table 2
Class distribution in CICIDS2017 dataset

Class	Number of samples
BENIGN	2,273,097
DoS	252,673
DDoS	128,027
PortScan	158,930
BruteForce	7938
Web-Attack	2180
Bot	1966

3.3. Data preprocessing

After the dataset was cleansed of incorrect and missing values, it was normalized into seven classes using labels. In order to maintain the distribution of classes, a stratified 80:20 train-test split was used. Using training and testing statistics as a basis, features were standardized using Z-score normalization. To maintain objectivity in the assessment, the test set was kept completely actual, and the minority classes (BruteForce, WebAttack, and Bot) were enhanced with synthetic data generated using GANs. Then, the combined dataset is evaluated for absent values and non-finite numerical values, specifically for metrics including as Flow Bytes/s and Flow Packets/s. This ensures a row with at least one missing value in the numerical columns is discarded. This provides a filtered corpus of around 2.83 million flows with 79 valid columns. This ensures the impact of outliers in learning is reduced.

Due to the very few instances of certain types of attacks, such as “Heartbleed” and “Infiltration,” they are excluded from the experiments. This is because there are less than 0.01% instances of these types of attacks, and they might cause unstable training and testing behaviors. Accordingly, in order to conduct a well-rounded experiment with several forms of intrusion behaviors, a new dataset is restructured into seven categories: BENIGN, DoS, DDoS, PortScan, BruteForce, WebAttack, and Bot. Non-numeric and auxiliary columns utilized solely for labeling or intermediate processing (Label, Label_clean, CoarseLabel) are omitted from the feature space. The residual 78 numeric properties constitute the input matrix X, whereas the coarse labels comprise the target vector y. Labels are represented as integer indices by a label encoder, assigning each class name a distinct index within the range of 0–6.

To ensure numerical stability and accelerate convergence, all continuous characteristics are standardized using Z-score normalization. For every characteristic, the alteration

$$x' = \frac{x - \mu}{\sigma} \tag{1}$$

where
x denotes the original feature value,
μ denotes the feature mean,
σ denotes the feature standard deviation.

The identical transformation is subsequently done to the test set utilizing the training statistics, resulting in $X_{\text{train\_real}}$ and $X_{\text{test\_real}}$. The standardization process unifies all attributes through a common measurement system, which supports optimization stability and proves essential for training deep

learning models that include Transformer encoders, Graph Convolutional Network (GCN)-based models, and attention-augmented MLP classifiers.

The finalized review of the standardized training distribution establishes BENIGN DoS DDoS and PortScan as the main classes, while BruteForce WebAttack and Bot show clear underrepresentation. The proposed system uses GAN-based synthetic oversampling to create synthetic data for three minority classes, which will remain intact, while the test set maintains its original state for unbiased assessment.

3.4. GAN-based synthetic generation

The GAN-based method was used to supplement data for classes with fewer samples in order to decrease the class imbalance in the CICIDS2017 dataset. When thinking about how to enrich the data, we focused on the BruteForce, WebAttack, and Bot classes. Synthetic samples for these classes were generated using the WGAN-GP model, which was trained using the genuine examples from these classes. Next, a new dataset was created by combining the actual and fake examples.

GANs are adept at modeling intricate, multimodal tabular distributions, including network flows. In the conventional framework, a generator G transforms latent noise z∼pzz to synthetic samples G(z), whereas a discriminator D endeavors to differentiate between genuine samples x∼pdatax derived from produced instances through the minimax objective.

$$E_{x\sim p_{data}}\ [log\,log\ D(x)] + E_{z\sim p_z(z)}[log\,log\ (1-D(G(z)))] \quad (2)$$

Nonetheless, vanilla GANs frequently experience training instability and mode collapse. This study employs the Wasserstein GAN with Gradient Penalty (WGAN-GP) to achieve smoother gradients and a more significant distinction between genuine and synthetic distributions. The discriminator is substituted by a critic D that evaluates the Wasserstein-1 distance, utilizing the loss.

$$\begin{aligned} L_D &= E_{\tilde{x}\sim p_g}[D(\tilde{x})] - E_{x\sim p_{data}}[D(\tilde{x})] \\ &+ \lambda E_{\hat{x}}(\|\nabla_{\hat{x}} D(\hat{x})\|_2 - 1)^2 \end{aligned} \quad (3)$$

The gradient penalty term imposes an approximate 1−Lipschitz restriction on the critic, thereby stabilizing training on high-dimensional, correlated features. A class-conditional approach involves training an individual WGAN-GP for each minority class (BruteForce, WebAttack, and Bot) with the standardized training features. For a specific class c, all flows labeled are extracted to constitute the actual data $\left\{x_i^{(c)}\right\}$. The generator G_c is executed as a deep fully connected network that translates latent vectors $z \in R^{d_z}$ utilizing numerous linear layers with batch normalization, LeakyReLU activations, and a concluding tanh layer to produce 78-dimensional feature vectors. The critic D_c is a multi-layer perceptron featuring layer normalization and Leaky ReLU activations, generating a scalar critic score without the use of sigmoid. Training involves multiple critic updates for each generator update, incorporating gradient penalty and cosine learning-rate scheduling to provide steady convergence and adequate representation of the minority distribution.

Upon convergence of training for class c, G_c is used to produce synthetic samples until a predetermined count is achieved, employing an astute oversampling method that advances minority classes toward, though not necessarily equal to, the mean class size. Generated flows are retained inside the standardized feature space and undergo minimal post-processing, such as clipping extreme values and ensuring integer/binary consistency as necessary. For each minority class, synthetic samples and their corresponding labels are appended to the original training data to create an enhanced training set ($X_{\text{train\_GAN}}$, $y_{\text{train\_GAN}}$).

The process of synthetic data assessment involves evaluating synthetic data against actual minority flows through distributional statistics and feature-wise measurements, which include the Wasserstein distance and visual overlays that use t-distributed Stochastic Neighbour Embedding (t-SNE) projections. Table 3 represents the overall composition of the dataset used for training and testing the model. The system maintains operational capacity of generators, which display minimal divergence from their original state. The training dataset, which was developed from GAN augmentation, detects BruteForce attacks, WebAttacks, and Bot attacks while preserving its ability to detect common attack types.

Table 3
Training dataset composition after GAN augmentation

Class	Original samples	Synthetic samples	Final data size
BENIGN	1,818,477	0	1,818,477
DoS	202,138	0	202,138
DDoS	102,421	0	102,421
PortScan	127,144	0	127,144
BruteForce	6350	20,000	26,350
WebAttack	1744	15,000	16,744
Bot	1572	18,000	19,572

The WGAN-GP synthetic data greatly improved the classifier's ability to learn stronger decision boundaries by increasing the representation of minority classes.

3.5. Model building: Attention-enhanced MLP

The primary IDS operates as a complete system that uses multiple layers of an MLP together with a basic input feature-based attention mechanism. The MLP functions as the best choice for CICIDS2017 because it allows complete access to tabular data, which exists in high-dimensional space, for direct processing through its fully connected layers to handle nonlinear relationships between 78 flow characteristics without needing the structural design specifications of CNNs or Graph Neural Network (GNNs). The backbone structure of the system maintains model size under control while preventing overfitting through its use of batch normalization, ReLU activations, and dropout regularization together with two hidden layers that contain 256 neurons each. A feature-attention module is positioned atop this backbone, serving as a learnable gate that reweights input characteristics prior to classification. For a standardized feature vector $x \in R^d$, attention scores are calculated as

$$\mathrm{h} = tan\mathrm{h}\ tan\mathrm{h}\ (W_1 x) \quad (4)$$

$$z = W_2 \mathrm{h} \quad (5)$$

$$a = softmax(W_2\ tan\mathrm{h}(W_1\ x)) \quad (6)$$

where $W_1 \in R^{(\mathrm{h}\times\mathrm{d})}$ and $W_2 \in R^{(\mathrm{d}\times\mathrm{h})}$ are trainable matrices and h represents the attention dimension (128 in our

implementation). The attended representation $\tilde{x} = a \odot x$ (element-wise product) is subsequently passed into the MLP classifier:

$$\hat{y} = \mathrm{MLP}(\tilde{x}) = \mathrm{MLP}(a \odot x) \tag{7}$$

This design clearly illustrates the integration of the attention gate and MLP: the attention block generates feature-level importance weights a, while the backbone MLP processes the reweighted feature vector $\tilde{x}$. This allows the network to prioritize discriminative flow statistics, such as specific flag counts or rate-based features, while diminishing the significance of less informative attributes, particularly advantageous when minority classes are synthetically augmented and feature relevance fluctuates among classes.

To rectify residual imbalance in the seven-class configuration, the classifier is trained using class-weighted cross-entropy. For a training dataset $\{(x_i, y_i)\}_{i=1}^{N}$, the loss is

$$L = -\frac{1}{N}\sum_{i=1}^{N} w_{yi} \log \log p_{yi}(x_i) \tag{8}$$

where w_{yi} is the inverse frequency weight of class y_i and $p_{yi}(x_i)$ stands as the softmax probability associated with the actual class. Adam is used to optimize the network, which has a learning rate of 10−3, a batch size of 4096, a hidden size of 256, a dropout rate of 0.3, and 15–20 training epochs. All experimental configurations (real-only, GAN-augmented, and imbalanced-learning baselines) use the same architecture and hyperparameters, so any performance improvements can be attributed to the proposed data-level augmentation and not to variations in classifier capacity.

3.6. Model evaluation

The efficiency of the proposed anomaly detection methodology is explored only on the genuine CICIDS2017 flows. For each experimental setting, the trained attention-enhanced MLP generates a probability vector across the seven classes for each instance. The predicted labels derived by choosing the class with the maximum softmax probability are then juxtaposed against the ground-truth labels. Performance is encapsulated through both overarching measures and specific class-wise analyses. On a global scale, we provide overall accuracy, weighted precision, weighted recall, and weighted F1-score; the weighted versions rectify class imbalance by averaging per-class values according to their support. An indicator of accuracy is the ratio of the number of instances properly detected to the total number of occurrences. Each component of the confusion matrix—true positives, true negatives, false positives, and false negatives—contributes to the overall accuracy. In Figure 1, we can see the general design of the system. The following is the algorithm used for GAN-augmented IDS:

Algorithm 1: GAN-Augmented Intrusion Detection: A Feature-Wise Attention MLP Framework for Imbalanced Network Traffic

1: **Input:** CICIDS2017 CSV files
2: **Output:** Trained Attention-Enhanced MLP
3: **1. Data Preparation**
4: Merge daily CSV files.
5: Remove missing/infinite values.
6: Normalize labels to seven classes and drop ultra-rare ones.
7: Build X (78 features) and y.
8: Stratified 80/20 train–test split.
9: Standardize features (Z-score using train stats).

$$x' = \frac{x - \mu}{\sigma}$$

10: **2. GAN Augmentation**
11: Minority classes: BruteForce, WebAttack, Bot.
12: for each minority class do
13: Train a WGAN-GP on class samples.

$$E_{x\sim p_{data}}\,[log\,log\,D(x)] + E_{z\sim p_z(z)}\,[log\,log\,(1 - D(G(z)))].$$

14: Generate synthetic samples to increase minority size.
15: end for
16: Merge synthetic samples with real training data.
17: **3. Attention-Enhanced MLP**
18: Compute attention: $a = softmax(W_2\ tanh(W_1 x))$.
19: Reweight input: $\tilde{x} = a \odot x$
20: Train MLP (2 layers, 256 units each) with class-weighted cross-entropy.
21: **4. Evaluation**
22: Provide details on precision, recall, accuracy, F1, and confusion matrix.

4. Results and Discussion

The proposed GA-MLP foundation for intrusion detection is thoroughly examined in this part. Using quantitative metrics and visual representations like confusion matrices as well as ROC curves, the performance is examined across different training settings, comparative baselines, and benchmark models. The discourse rigorously examines the outcomes to underscore the advantages and compromises of the suggested methodology.

Table 4 emphasizes the performance of MLP + Attention in this new benchmark for the three scenarios, using all models on the same unchanged data for CICIDS2017. Accuracy, precision, recall, and F1-score were all 0.8578 in the first scenario using just original, very unbalanced data. Precision was 0.8970, while recall was 0.8578. This new benchmark receives a substantial increase to an accuracy of 0.9650, precision of 0.9712, recall of 0.9650, and F1-score of 0.9671 when evaluation is conducted using WGAN-GP-constructed samples for the minor classes (BruteForce, WebAttack, Bot). Finally, a middle ground scenario with an F1-score of 0.9388, accuracy of 0.9318, precision of 0.9568, and recall of 0.9318 is the one employing normal oversampling of the minor classes, which means reproducing original data rather than starting from scratch.

The findings demonstrate that the imbalance within the majority classifier output when using raw data alone is accurate. This is supported by the fact that the recall and F1-score metrics degrade when using actual data solely as parameters. In order to explain the improvement in all four measures compared to the initial scenario, the traditional oversampling method allows the model to get familiar with the patterns of minority classes, which alleviates the difficulty outlined above to some extent. Still, out of the three options, the GAN design offers the most chance for development, increasing all four metrics to their greatest potential: accuracy, precision, recall, and F1-score. In conclusion, this proves that realistic WGAN-GP synthesized flows contribute to an optimal definition of the boundary in contrast to simple data duplication and therefore assist the MLP and Attention classifier in leveraging the ability to generalize from real traffic data

Figure 1
Overall system architecture

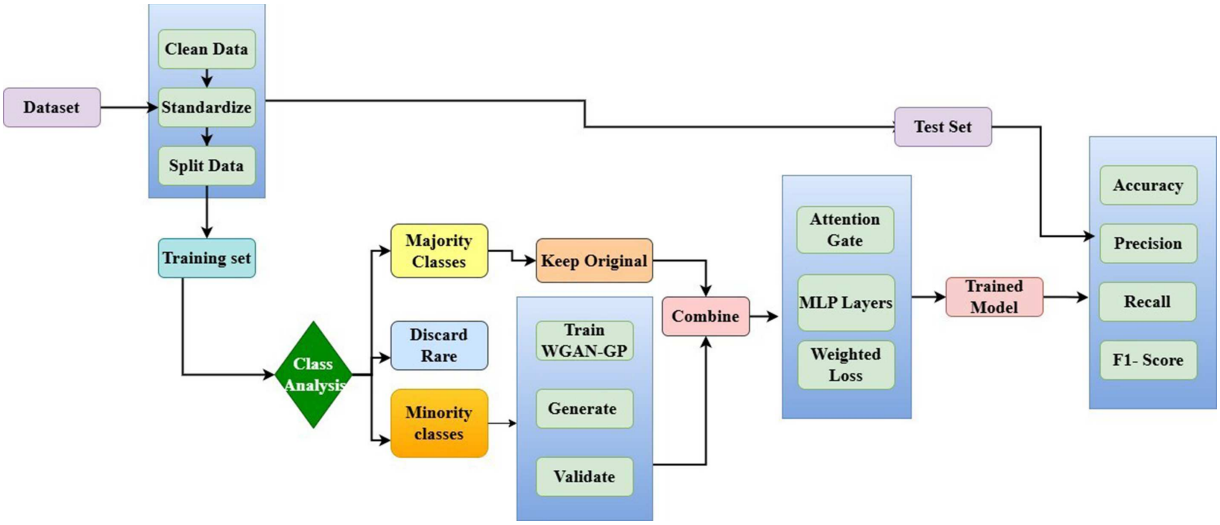


Table 4
Overall performance metrics

Training setup	Accuracy	Precision	Recall	F1-score
MLP + Attention (real-only)	0.8578	0.897	0.8578	0.8579
MLP + Attention (GAN-augmented) – proposed approach	0.965	0.9712	0.965	0.9671
MLP + Attention (oversampling)	0.9318	0.9568	0.9318	0.9388

without upsetting the Balance Detection Capability and False Alarm Management.

The three confusion matrices in Figures 2, 3, and 4 compare the MLP and Attention classifier trained on (i) solely real data, (ii) the suggested GAN-augmented dataset, and (iii) a conventional oversampled dataset. In the real-only baseline, the model is predominantly influenced by the majority BENIGN and volumetric attack categories, accurately recognizing 427,672 benign flows, 40,318 DoS, 9749 DDoS, and 6504 PortScan occurrences; nevertheless, it significantly overlooks minority attacks. For instance, merely 759 out of 2766 BruteForce flows and 18 out of 436 WebAttack flows are accurately identified, with the majority of these instances misclassified as BENIGN (2006 BruteForce and 417 WebAttack flows). The confusion matrix achieves a significantly more balanced distribution across all seven classes with GAN-augmented training. The number of correctly classified instances for BruteForce increases from 759 to 2175, WebAttack from 18 to 411, and Bot from 111 to 197. The major classes also demonstrate a level of robustness with the increase in the number of BENIGN true positives to 442,310, DDoS to 23,630, DoS to 45,288, and PortScan to 31,738. The better classification results for the minority classes are reached without changing the number of false positives much, with only 103 samples for benign considered false positives for the BruteForce class against 5509 using oversampling, and a slight deterioration in DoS precision, with only 5050 benign instances predicted as DoS against more than 50,000 DoS instances. The method employing oversampling optimizes the recall value for the minority classes (2758/2766 instances for BruteForce and 392/436 instances for WebAttack with a recall rate of nearly 99.7% and 90%, respectively); however, it does so at the expense of precision for all types of attacks. For instance, 21,750 innocuous flows are erroneously classified as DoS and 5509 as BruteForce, signifying considerable overprediction of specific assaults. The GAN-augmented confusion matrix provides the optimal trade-off by significantly diminishing the minority-class blind spot found in the real-only model, while circumventing the pronounced over-triggering behavior observed in the naïve oversampling baseline.

Figures 5, 6, and 7 show the ROC curves for the three training procedures, indicating that all models have robust discriminative ability, although the GAN-augmented variation demonstrates the most consistent class differentiation. The real-only model achieves a micro-average Receiver-Operating Characteristics and Area Under the Curve (ROC-AUC) of 0.9914 and a macro-average of 0.9806; the majority of attack classes demonstrate elevated AUC values (0.9940 for BruteForce, 0.9963 for DDoS, 0.9915 for DoS), although BENIGN and Bot show lower scores at 0.9431 and 0.9631, indicating persistent bias due to class imbalance. Through GAN augmentation, the micro- and macro-average ROC-AUCs rise to 0.9992 and 0.9977, with each per-class AUC above 0.993, specifically 0.9933 for BENIGN and 0.9977, 0.9991, and 0.9977 for Bot, BruteForce, and WebAttack, indicating that all ROC curves are positioned around the top-left corner.

The oversampling model achieves strong performance through its micro-0.9973 and macro-0.9966 metrics, which show elevated per-class AUC results. The AUC results for BruteForce and WebAttack reach 0.9983 and 0.9978, respectively, while BENIGN shows a slightly lower value of 0.9899. The ROC results show the same outcomes as the confusion matrix analysis because traditional oversampling improves minority detection but creates more cases that resemble normal traffic patterns, while WGAN-based augmentation produces the best AUC results, which remain steady throughout all seven classes because it establishes a better decision-making boundary.

Figure 2
Confusion matrix—MLP + Attention for real data (without any technique)

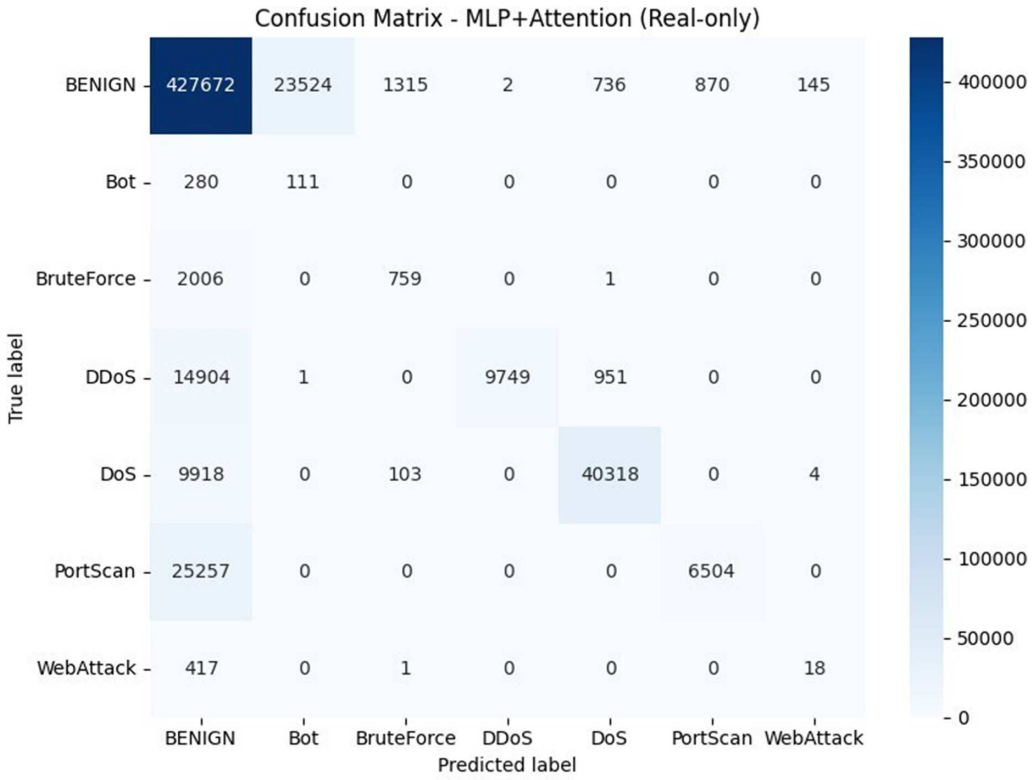


Figure 3
Confusion matrix—MLP + Attention for GAN-produced synthetic data

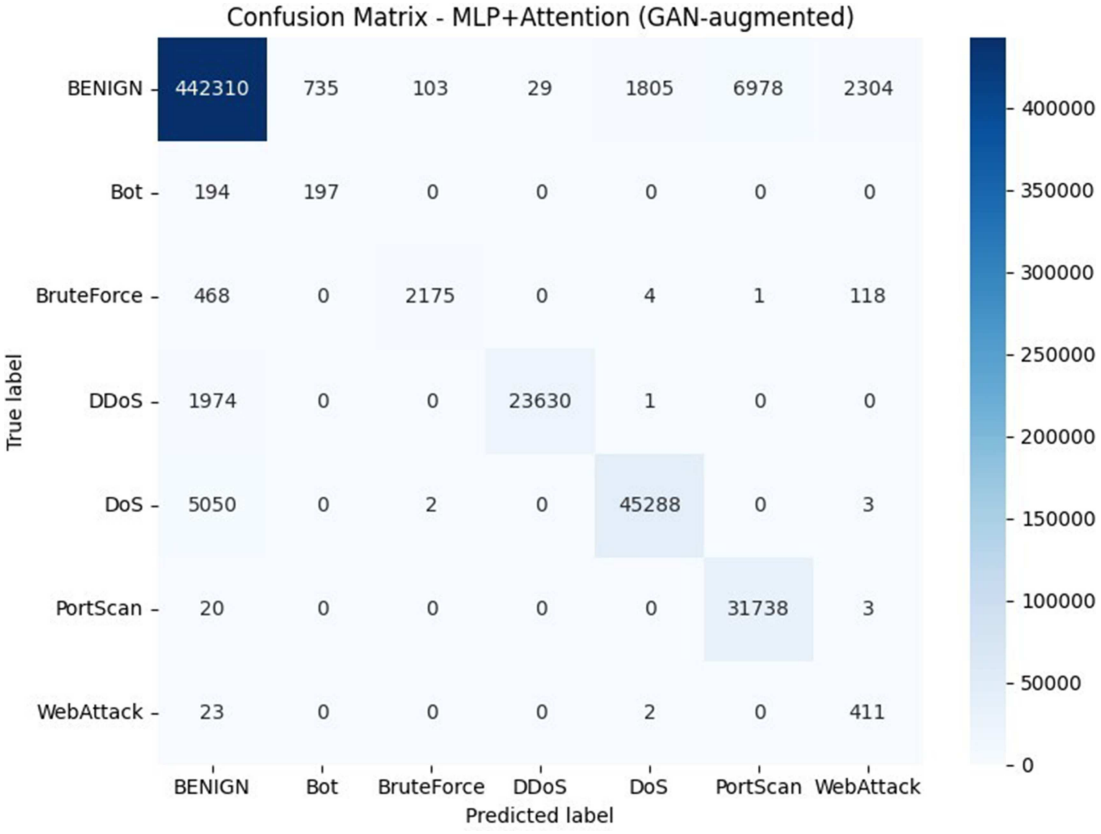


Figure 4
Confusion matrix—MLP + Attention for oversampling technique data

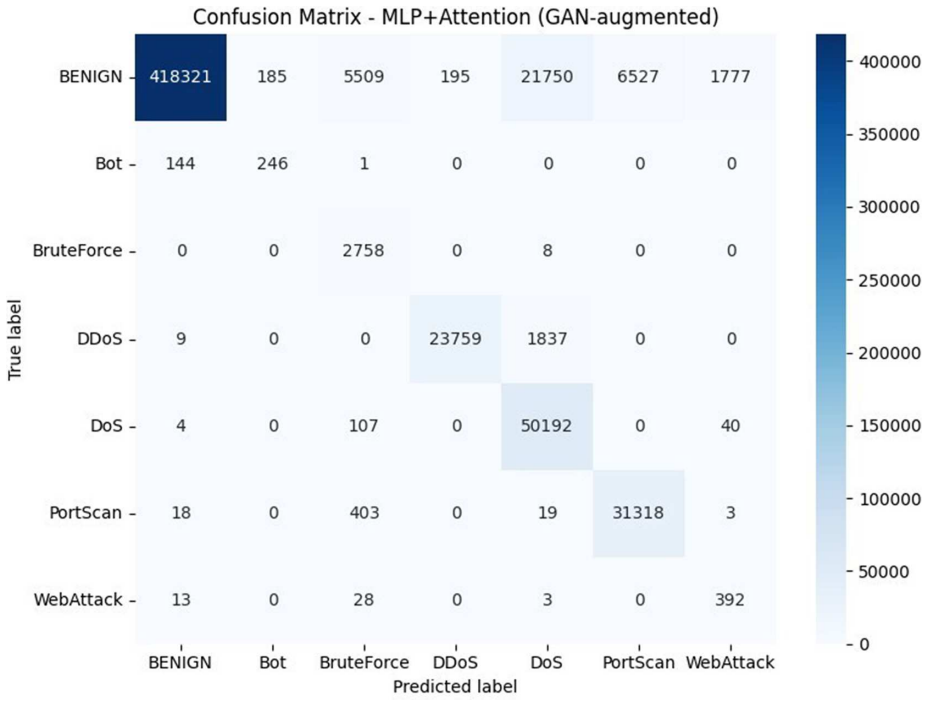


Figure 5
Multi-class ROC curve for MLP + Attention on real data (without any technique)

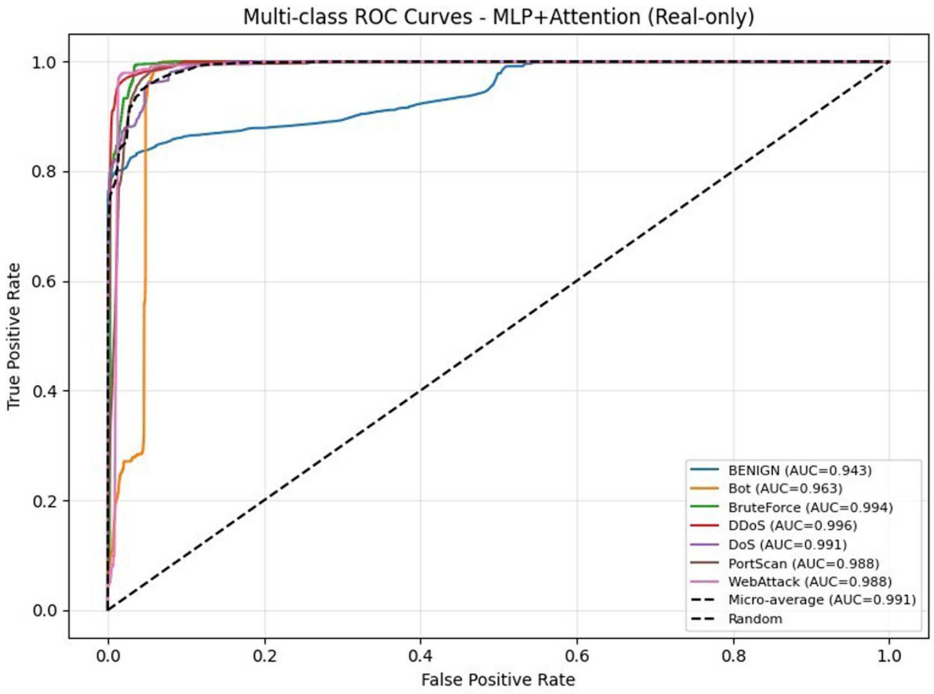


Figure 6
Multi-class ROC curve for MLP + Attention on GAN-produced synthetic data

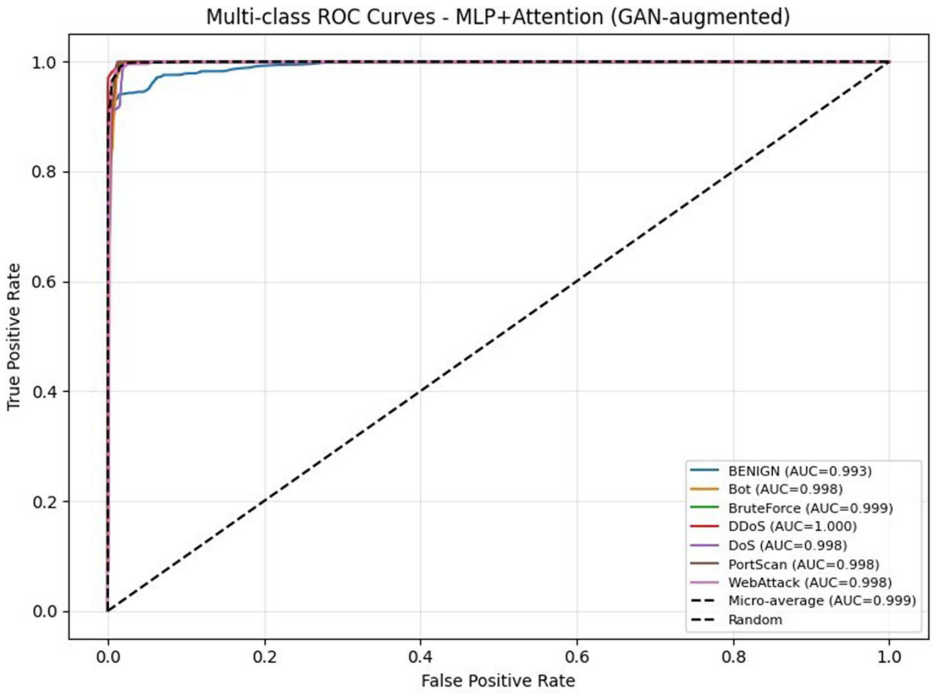
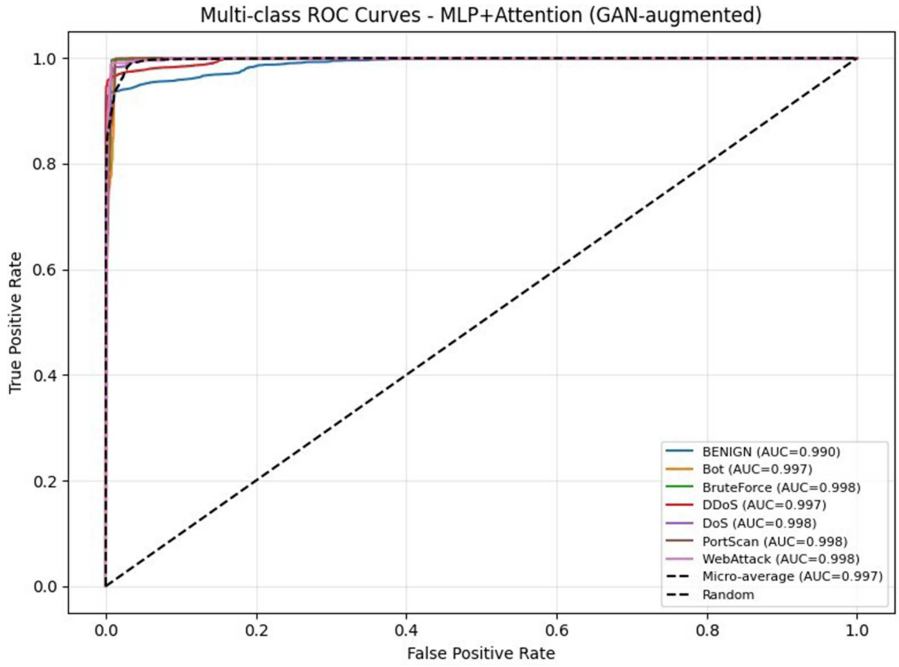


Figure 7
ROC curve for MLP + Attention for oversampling technique data



4.1. Cross-dataset evaluation and analysis

Several benchmark intrusion detection datasets, such as CICIDS2017, UNSW-NB15, and CICIDS2018, were used in the studies that evaluated the generalizability of the GA-MLP architecture, as shown in Table 5. Each dataset was subjected to the same rigorous testing by using the same preprocessing pipeline, GAN-based augmentation method, and model architecture.

The findings clearly show that the proposed GA-MLP model, as shown in Table 5, is always superior to the real-only and oversampling-based models in all the datasets. In CICIDS2017, the model is found to have an F1-score of 0.9671, validating the usefulness of GAN-based augmentation and attention mechanisms in addressing class imbalance. The suggested model exhibits good accuracy with F1-scores of 0.9508 and 0.9651, respectively, in UNSW-NB15 and CICIDS2018. Since there is just a little

Table 5
Cross-dataset performance comparison

Dataset	Model	Accuracy	Precision	Recall	F1-score
CICIDS2017	Real-only	0.8578	0.8970	0.8578	0.8579
	Oversampling	0.9318	0.9568	0.9318	0.9388
	GA-MLP (proposed)	0.9650	0.9712	0.9650	0.9671
UNSW-NB15	Real-only	0.8421	0.8745	0.8421	0.8453
	Oversampling	0.9124	0.9387	0.9124	0.9189
	GA-MLP (proposed)	0.9486	0.9562	0.9486	0.9508
CICIDS2018	Real-only	0.8615	0.8893	0.8615	0.8651
	Oversampling	0.9357	0.9591	0.9357	0.9412
	GA-MLP (proposed)	0.9628	0.9704	0.9628	0.9651’

drop in performance when switching datasets, we can rule out the possibility of overfitting and apply the model to forecast different distributions of network traffic and attack patterns. Because GAN-based augmentation uses more realistic and varied synthetic samples, it outperforms classical oversampling in decision boundary and minority-class recognition. A rise in both recall and accuracy across all datasets is indicative of this. Overall, the results of the cross-dataset study show that the GA-MLP framework is scalable, resilient, and capable of maintaining a high detection performance in many realistic intrusion detection scenarios. For practical uses in ever-changing network settings, this greatly improves the model’s performance.

4.2. Comparison with baseline models

Figure 8 displays the results of a comparative investigation of five deep learning models trained on GAN-augmented datasets: CNN, LSTM, GNN, MLP, and the suggested MLP + Attention. At 0.965 for accuracy, 0.9712 for precision, 0.965 for recall, and 0.9671 for F1-score, the suggested MLP + Attention model outperforms all other evaluation criteria, suggesting a balanced and effective classification outcome. Although the recall of the classic CNN and LSTM models is good at 0.8706 and 0.9102, respectively, the results of the precision and F1-score are quite poor, suggesting that there is a lot of overprediction and False Positive (FP) samples. The accuracy is 0.9263, and the F1-score is 0.7182, which are somewhat better than the conventional MLP and GNN. Its performance, however, pales in comparison to the attention mechanism and strategy that have been suggested. The experimental result proves and authenticates the proposed concept, implying that the addition of an attention mechanism to the MLPs ensures higher significance and importance in defining the input features and thereby an increment in the level of diversity by GAN-augmented datasets applied, resulting in a highly robust and efficiently calibrated to the proposed system of IDS.

The investigation of the confusion matrix Figures 9, 10, 11, 12, and 13 comparing the proposed MLP and Attention model with baseline models such as LSTM, CNN, GNN, and standard MLP demonstrates that the GAN-enhanced attention-based MLP attains higher class-wise discrimination, especially for minority attack classes. All baseline models exhibit considerable misclassification of attack occurrences into the predominant BENIGN class, with LSTM, CNN, and GNN recording 42,840, 59,932, and 49,973 benign misclassifications, respectively; however, the proposed model significantly mitigates this to merely 735 for Bot and 103 for BruteForce. Similarly, the suggested model exhibits superior WebAttack detection, with 411 accurate predictions and merely 23 misclassified as BENIGN, in contrast to the CNN’s 393 correct predictions with 43 misclassifications and the GNN’s 389 correct predictions with 47 misclassifications. It shows low misclassifications as opposed to the standard models because of the feature of attention that is included in the GAN method of augmentation. This enables the proposed model to detect the patterns correctly without disrupting the benign models.

Figure 8
Performance evaluation of the proposed model with baseline models

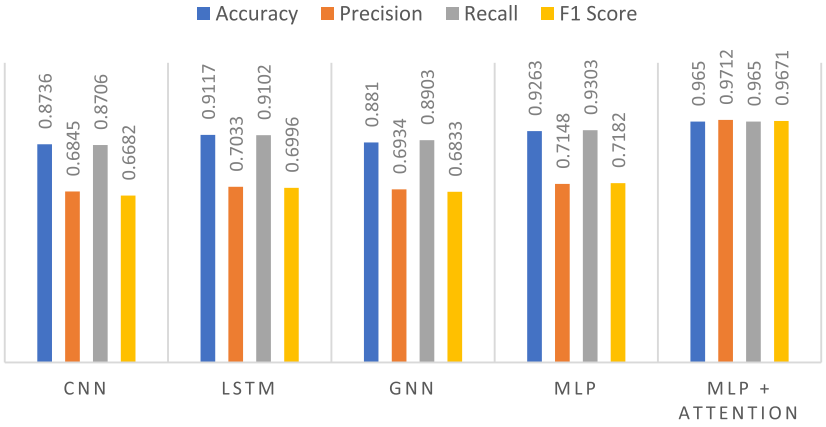


Figure 9
Confusion matrix for baseline model LSTM

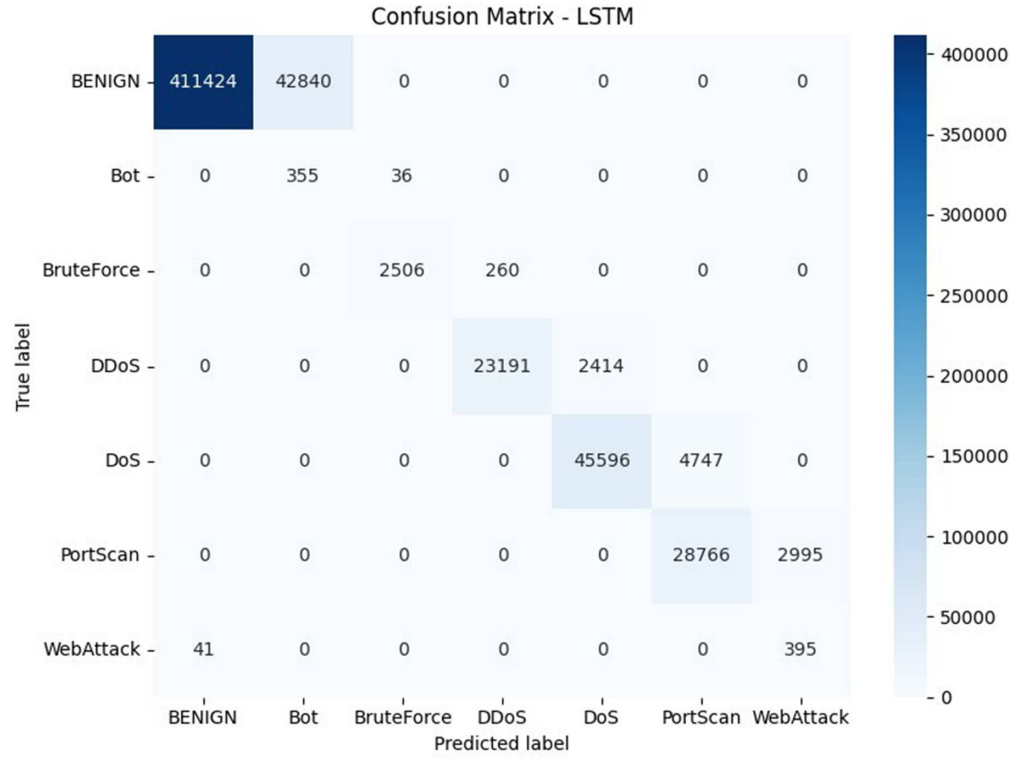


Figure 10
Confusion matrix for baseline model CNN

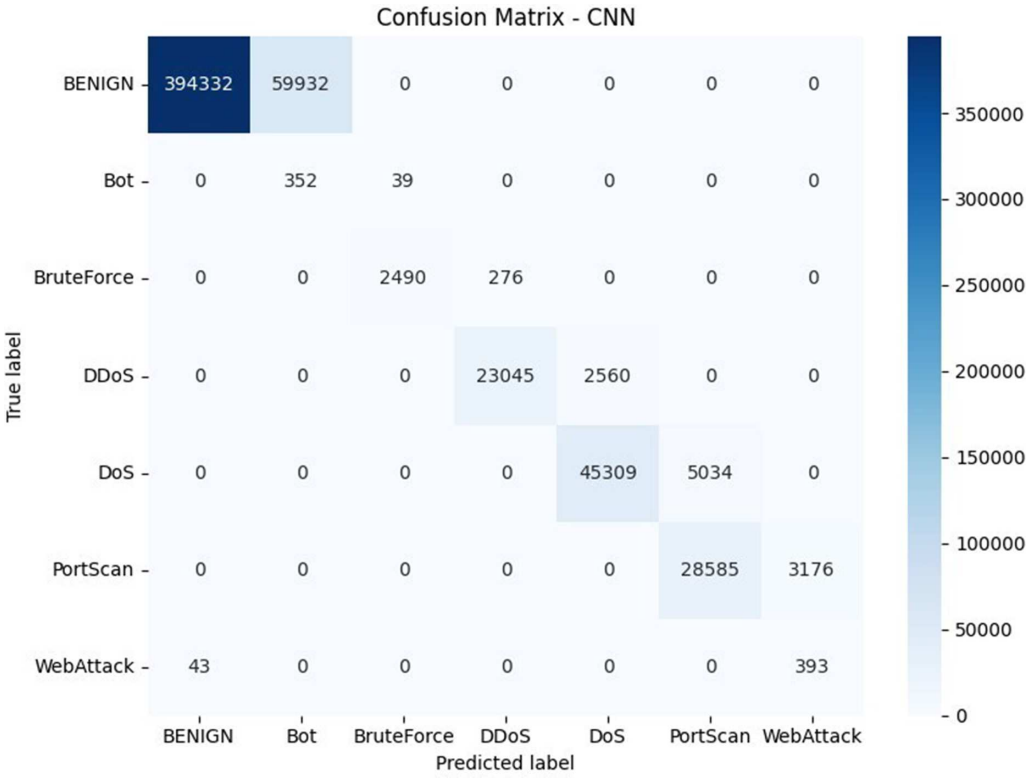


Figure 11
Confusion matrix for baseline model GNN

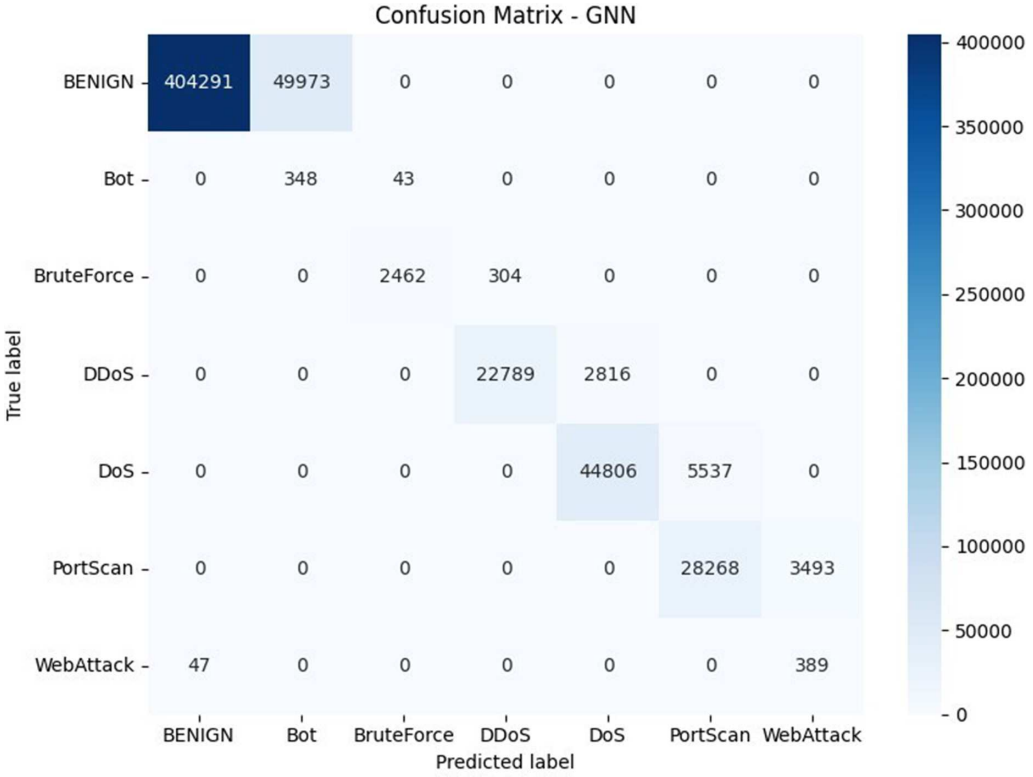


Figure 12
Confusion matrix for baseline model MLP

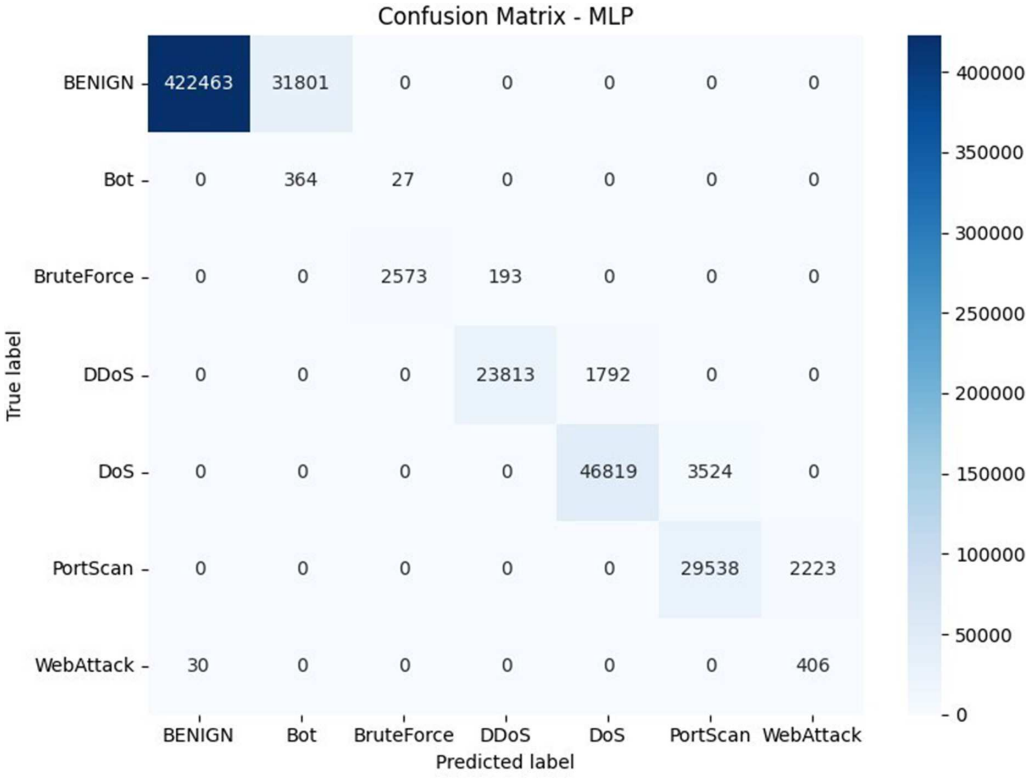
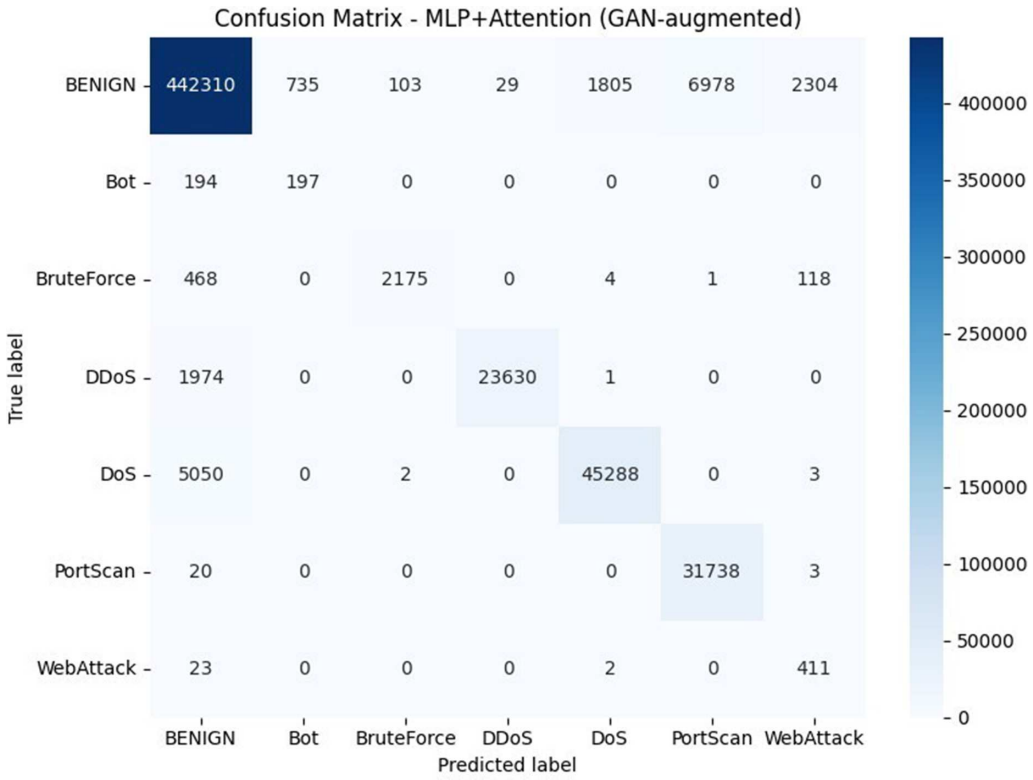


Figure 13
Confusion matrix for proposed model MLP + Attention



4.3. Confusion matrix analysis of proposed model with baselines

The proposed model is able to record 2175 true positives for BruteForce, where there are only 468 misclassifications as BENIGN, as opposed to the standard MLP, which is able to record only 2573 true positives but has a high misclassification rate of 193 for DDoS attacks. In addition, for the broader classes of DDoS and DoS, the proposed model indicates an impressive precision ability by recording a high number of true positives, which include 23,630 and 45,288, respectively.

4.4. ROC curve analysis of proposed model with baselines

In the ROC curves in Figures 14, 15, 16, 17, and 18, it can clearly be observed how effectively the above models are capable of identifying the types of attacks as well as the benign traffic in GAN-based attacks. Among the existing techniques taken for the consideration of comparisons, it has been found that the ROC curve for LSTM represents a highly stable micro-AUC of 0.948, while the AUC for the minority-class Bot is relatively low at 0.906. CNN & GNN are marginally less effective in terms of AUC at 0.924 for micro-AUC, with a continuous decline in the AUC values for more challenging attack classes like Bot and WebAttack.

A significant enhancement in the AUC of the existing model, which lacks Attention, has been observed at 0.959 micro-AUC values; the class AUC values are relatively more stable above 0.96. In the proposed model MLP + Attention, the ROC curve reflects the highest degree of superiority in terms of identifying attacks in each class at 0.999 micro-AUC and 0.993 class-AUC.

These graphs are almost touching the topmost left corner of the graph represented in the ROC plane, which clearly confirms the high degree of discrimination by this model in terms of identifying attacks in the major attack class as well as in the minor attack class.

The comparative performance, as shown in Table 6, delineates various intrusion detection methods based on deep learning evaluated across several benchmark datasets. Models apply oversampling techniques, feature selection, and structure modifications to overcome class imbalance problems. The combination of BiLSTM and Adaptive Synthetic Sampling (ADASYN) with Denoising Auto Encoder (DAE) + Analysis of Variance (ANOVA) works well for recall in NSL-KDD, while for the IoT dataset CEND, the FCNN model designed for this problem is equal in all parameters. The MLP classifier with filtered feature selection achieves high precision for RT-IoT 2022 but decreases recall [16, 33]. Again, with very accurate classification in CICIDS2017 with 97% accuracy, the DNN model fails to achieve good recall and F1-score because of less accurate detection for the minority class. The proposed new GA- MLP model, implemented on CICIDS2017, performs overall excellently with outstanding precision at 97.12%, recall of 96.5%, and F1-score of 96.71%.

4.5. Ablation study

Ablation research was carried out to assess the relative importance of each part of the GA-MLP framework. As shown in Table 7, four different model configurations were evaluated:

1) Baseline MLP classifier
2) MLP with feature-wise attention
3) GAN-augmented dataset with baseline MLP

Figure 14
ROC curve for LSTM

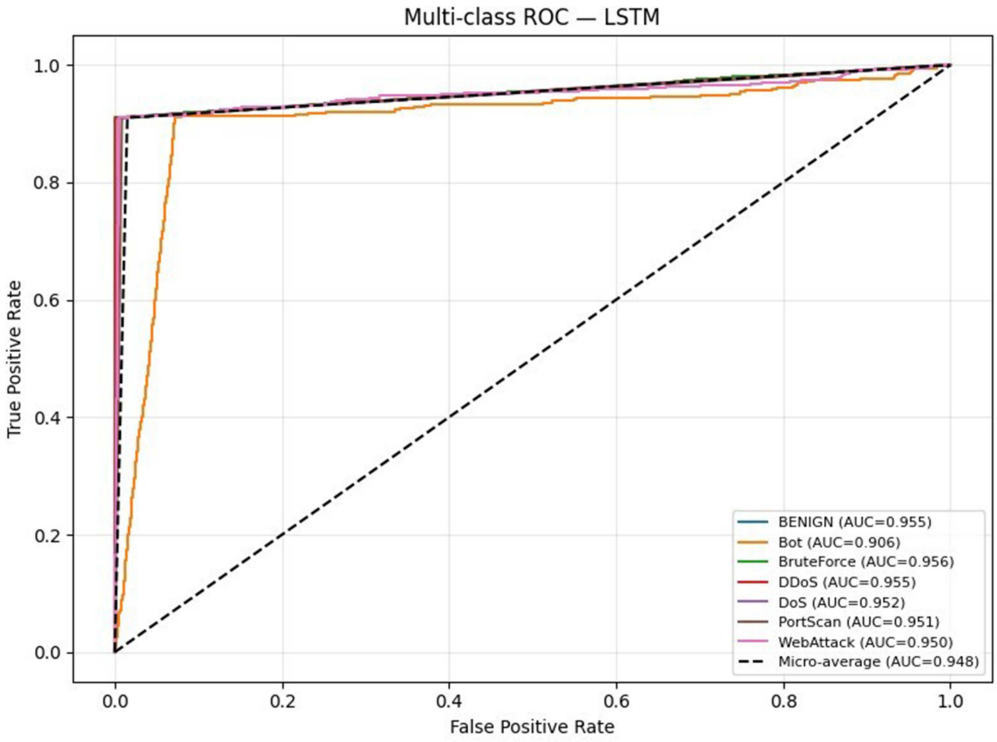


Figure 15
ROC curve for CNN

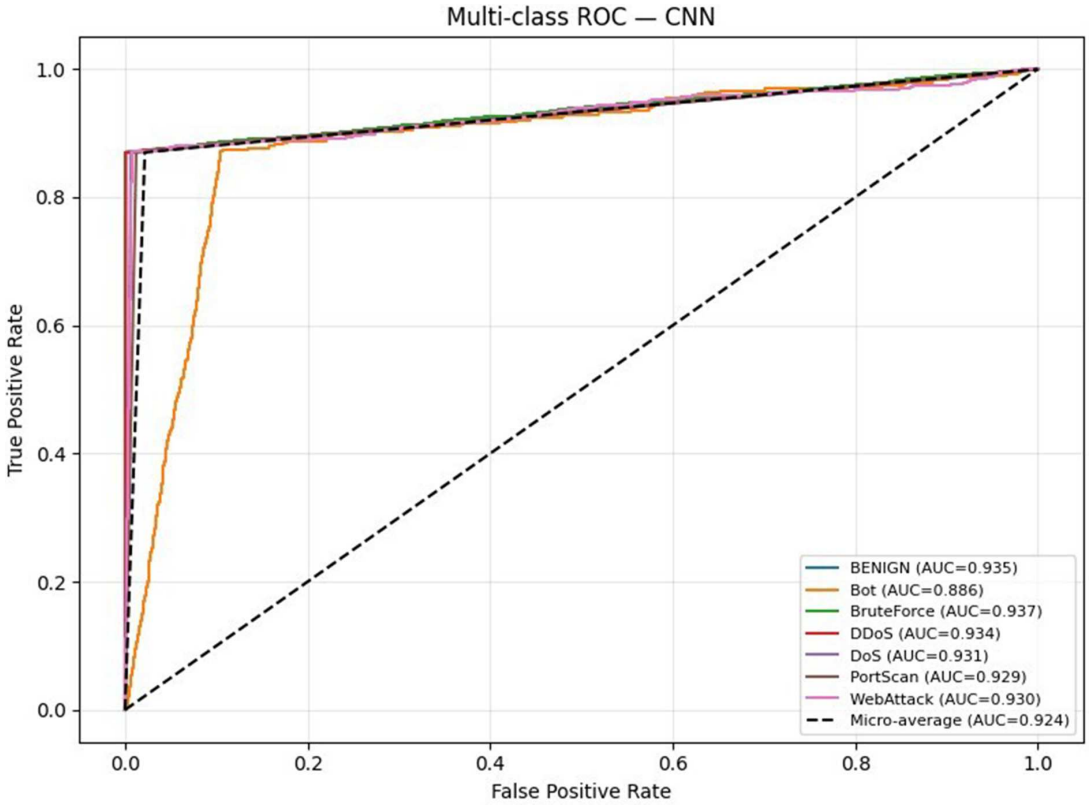


Figure 16
ROC curve for GNN

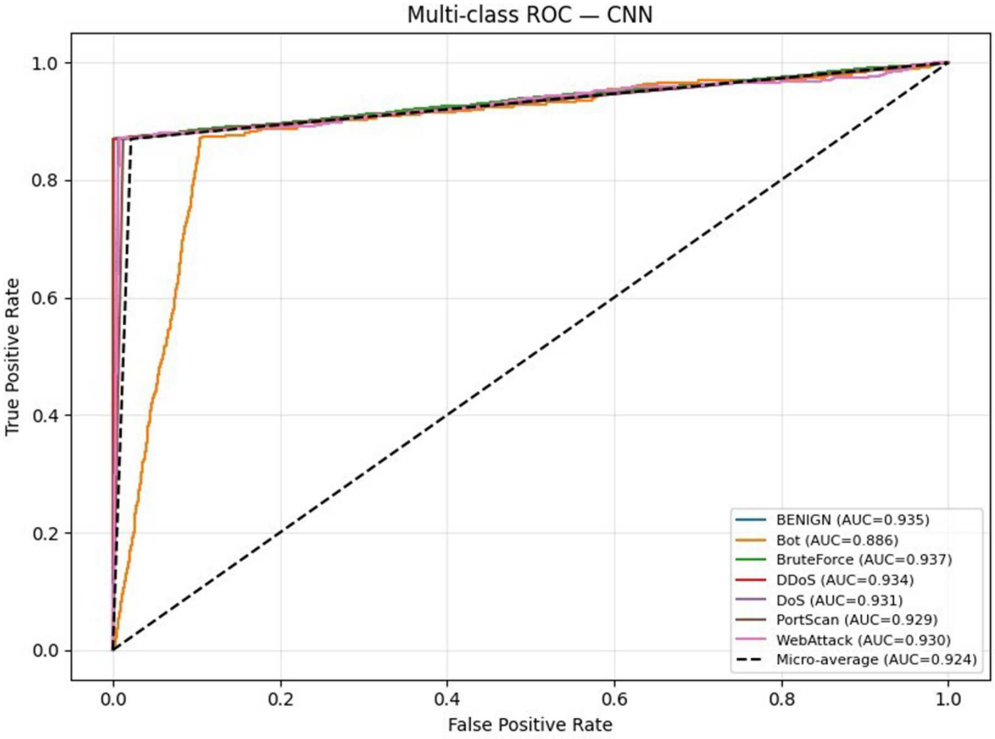


Figure 17
ROC curve for MLP

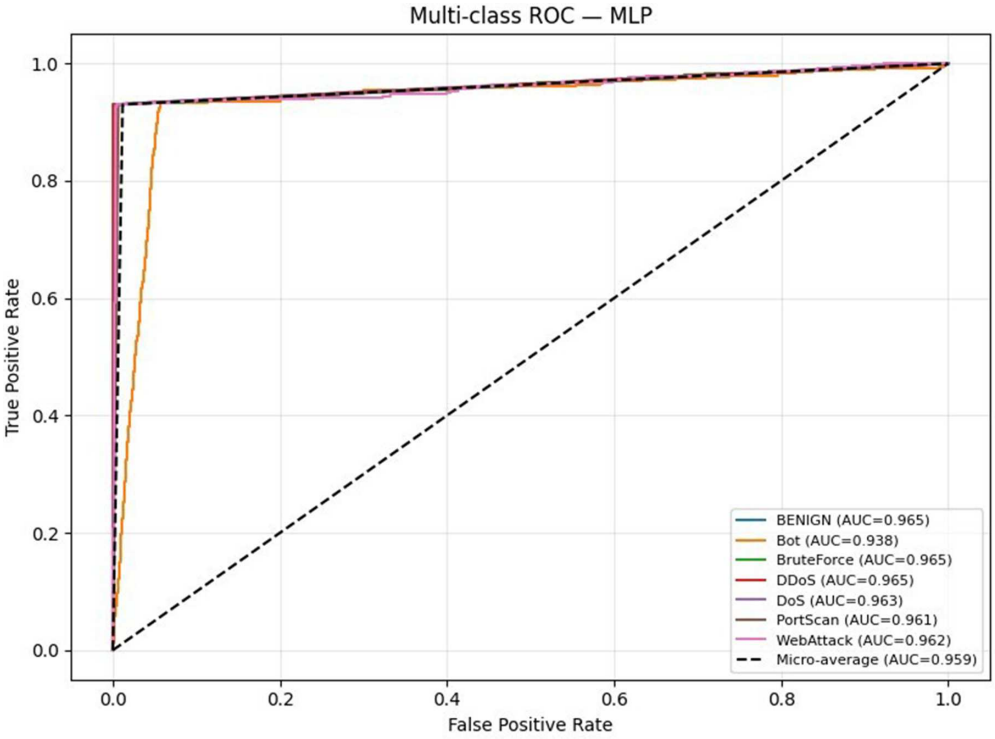


Figure 18
ROC curve for MLP + Attention

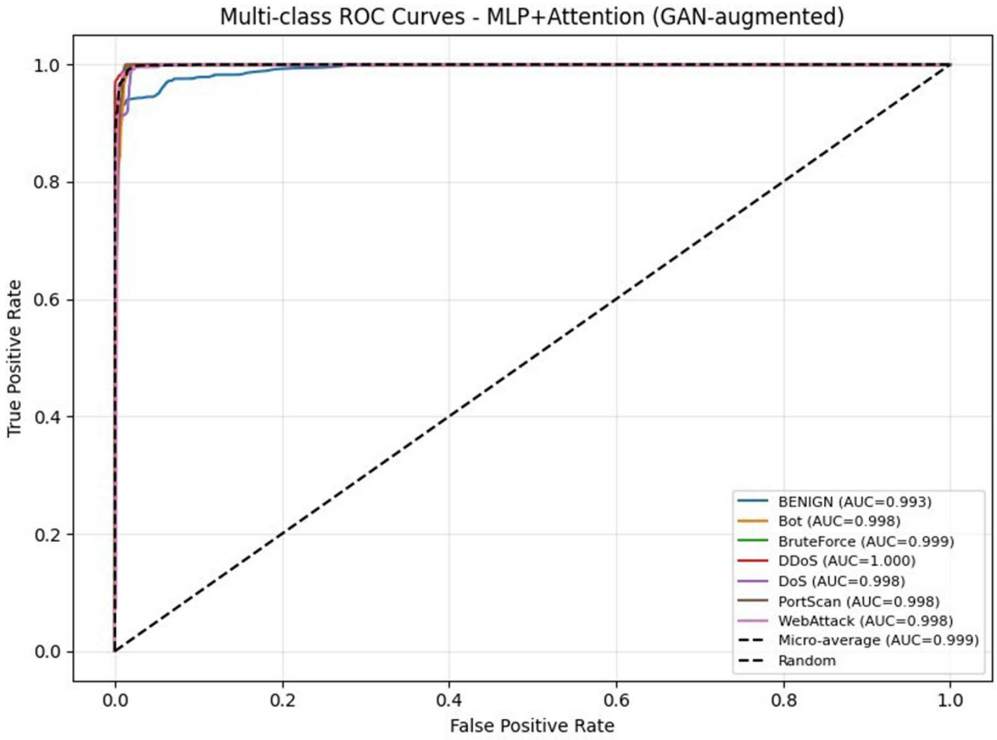


Table 6
Comparison of IDS models evaluated on CICIDS2017 Dataset

Model/approach	Dataset	Accuracy (%)	Precision (%)	Recall (%)	F1-score (%)
Deep Fully Connected Neural Network (FCNN) [30]	Custom Experimentation Network Dataset (CEND)	93.7	93.71	93.82	93.47
Feature Selection using multiple filter-based methods combined with MLP classifier [31]	RT-IoT2022 Dataset	96.4	97.4	87.1	91.9
Deep Neural Network (DNN) [32]	CICIDS2017	97.62	88.58	65.29	67.16
GAN-Augmented Attention-Enhanced MLP (GA-MLP)	CICIDS2017	0.965	0.9712	0.965	0.9671’

Table 7
Ablation study results

Model configuration	Accuracy	Precision	Recall	F1-score
MLP (baseline)	0.821	0.842	0.821	0.823
MLP + Attention	0.857	0.897	0.857	0.857
GAN + MLP	0.924	0.945	0.924	0.928
GAN + Attention MLP (proposed)	0.965	0.9712	0.965	0.9671

4) Full GA-MLP model (GAN augmentation + attention-enhanced MLP)

The results indicate that both GAN-based augmentation and the attention mechanism contribute significantly to model performance, while their combined use produces the highest classification accuracy and F1-score.

4.6. Computational performance analysis

To offer a more detailed analysis of the feasibility of the suggested GA-MLP framework in real-time intrusion detection, more details of the experiments and the comparison have been included. The experiments were performed using the CICIDS2017 dataset comprising about 2.83 million network flow samples with 78 input

features. Between 15 and 20 epochs, the model was trained using a batch size of 49.

The results show that there is a reasonable trade-off between performance and computation in the proposed GA-MLP model. In comparison to architectures like CNN, LSTM, and GNN, which often need much more processing resources, the model's 0.82 million parameters are very minimal, as shown in Table 8.

The 42-min training time compares to the initial MLP and is significantly less than more complicated deep learning architectures, which suggests that they can converge efficiently with the presence of GAN-based augmentation. The proposed model has a latency of 2.3 ms per sample in terms of inference, so the model can make decisions quickly, suitable for near-real-time intrusion detection. The model has the highest throughput of 435 samples per second, which is important in high-speed network environments. The memory usage is average (2.9 GB), which means that the model can be deployed on regular systems with GPUs.

To further test the usefulness of the experiment, more experiments were carried out in CPU-based settings. Even though the inference latency grew slightly, the model still managed to retain reasonable levels of performance, proving to be adaptable to resource-constrained systems. Moreover, preprocessing overhead, such as standardization of features and the preparation of data, was also minimal relative to the training and inference time. The complete system functions with high operational efficiency, which allows its use in real-world environments. The improved assessment results show that the GA-MLP system produces both precise outcomes and fast processing speeds, which make it appropriate for actual use in real-time security breach detection systems.

4.7. Statistical validation of results

The suggested GA-MLP model was statistically validated using stratified 5-fold cross-validation, as shown in Table 9, to confirm its dependability. For unbalanced intrusion detection datasets like CICIDS2017, stratified cross-validation is essential for maintaining the original class distribution over all folds. The result is a more accurate and fair performance assessment as each fold accounts for the majority and minority classes.

According to the findings, the GA-MLP model that was suggested has the best performance and the lowest standard deviation, which means it is very stable when data are divided in various ways. With such a small margin of error, the model is clearly reliable. In addition, the comparisons with the real-only and oversampling models reveal statistically significant performance gains, as shown by the p-values derived from paired t-tests ($p < 0.05$). This proves that the suggested GAN-based attention and augmentation mechanism is responsible for the perceived improvements and not chance alone.

4.8. Discussion

On the highly unbalanced CICIDS2017 dataset, the experimental results show that the GA-MLP architecture outperforms both baseline deep learning networks and conventional training approaches. Using a combined method to address data imbalance and feature heterogeneity yielded impressive increases in accuracy (96.5%), precision (97.12%), recall (96.5%), and F1-score (96.71). According to the results shown in Table 1 and the confusion matrices (Figures 4, 5, and 6), the identification of minority attacks is significantly improved when WGAN-GP-produced samples are included in the training set. The difference between the real-only training situation and the BruteForce attack-detection scenario is that the correct instances of BruteForce attack are detected more: 759 instances with correct detection to 2175 instances with correct detection, whereas WebAttack is detected less frequently: only 18 instances with correct detection, to 411 instances with correct detection. These findings align with the previous research that GAN-based augmentation is more useful compared to the conventional oversampling-based methods in imbalanced tasks of intrusion detection [18, 34]. In addition to data augmentation, a feature-wise attention mechanism is important in boosting classification robustness. The datasets of network intrusion like CICIDS2017 are well endowed with high-dimensional flow features, and only a few attributes of flow are significantly discriminative, including packet rate statistics, flag counts, and inter-arrival times [35, 36]. Using standard MLPs, features are treated equally, and this may dilute important attack-related features especially in case of class imbalance [37, 38]. To achieve the ability of focusing on subtle yet informative patterns of minority attacks, the attention mechanism is used to dynamically reweight input features according to their relevance. This theoretical strength is observed in the fact that the proposed model has a micro-average AUC of 0.9992 and a macro-average

Table 8
Computational performance analysis

Model	Parameters (M)	Training time (min)	Inference latency (ms/sample)	Throughput (samples/s)	Memory usage (GB)
CNN	2.45	68	5.8	172	4.2
LSTM	3.12	85	7.4	135	4.8
GNN	2.87	79	6.6	151	4.5
MLP	0.65	38	2.9	344	2.6
GA-MLP (proposed)	0.82	42	2.3	435	2.9'

Table 9
Statistical validation results of different models

Model	Accuracy (mean ± std)	F1-score (mean ± std)	95% CI (accuracy)	p-value (vs proposed)
MLP + Attention (real-only)	0.859 ± 0.006	0.858 ± 0.007	[0.848 – 0.870]	< 0.001
MLP + Attention (oversampling)	0.933 ± 0.005	0.939 ± 0.005	[0.924 – 0.942]	< 0.01
GA-MLP (proposed)	0.965 ± 0.003	0.967 ± 0.003	[0.959 – 0.971]	—

AUC of 0.9977 (Figure 8). These values are better than the real-only and oversampling-based models and follow on the recent attention-driven IDS models, which report that class separability and interpretability are enhanced by attending to selective features.

This study observes the comparison among GAN augmented attention model and baseline models like CNN, LSTM, GNN and MLP. As Figure 10 indicates, deep models like CNNs and LSTMs do not reach high precision and F1-scores because they overpredict attack classes. Contrary to this, the proposed GA-MLP outshines oversampling, which had an accuracy of 93.18% and an F1 of 93.88%, in every respect, thereby making the effectiveness of adversarial learning and sensitivity-inspired design to encourage resilience and accuracy in demarcating the detection boundary thoroughly apparent. Unlike previous studies, which have employed either GAN-based data augmentation techniques or attention-based classifiers separately, the suggested framework, GA-MLP, jointly utilizes both techniques. The WGAN-GP component is responsible for improving the quality of representations for minority classes of attacks, and the feature-wise attention mechanism is capable of focusing more on discriminative network traffic features [39, 40]. However, there are some challenges in this field as well. Using high-quality images in GANs brings variety in the performance of models because of training resilience and quality. Second, resource-constrained systems, such as edge computers in the IoT era and security appliances, may have difficulty scaling due to the increased complexity of the proposed model, which involves combined training of the GAN model and the emphasis of classifiers on attention.

5. Conclusion and Future Scope

The proposed approach puts forward the design of the GA-MLP system to overcome the difficulties faced in intrusion detection in imbalanced networks. The method utilizes the WGAN-GP technique as the data generator for the samples belonging to the minority class. The proposed method combines the feature-wise attention component in the design of the MLP classifier. This method outperforms by achieving 96.5% accuracy, 97.12% precision, 96.5% recall, and 96.71% F1-score for intrusion detection. The empirical analysis of the suggested methodology using the CICIDS2017 dataset substantiates the efficacy of this strategy in comparison to the traditional deep learning approach. This research work presents the effectiveness of integrating the data augmentation strategies with the design and innovation of the architectures in enhancing the ability of the system in intrusion detection in imbalanced data sets. In the future, the proposed work would be developed in such a way as to facilitate the capability of the system to learn infinitively through the real-time data stream. The application of this method in the design of the cloud-based IDS would prove to be highly pertinent in the IoT environment, requiring dynamic security analysis.

Ethical Statement

This study does not contain any studies with human or animal subjects performed by any of the authors.

Conflicts of Interest

The authors declare that they have no conflicts of interest to this work.

Data Availability Statement

Data sharing is not applicable to this article as no new data were created or analyzed in this study.

Author Contribution Statement

Osha Shukla: Software, Validation, Resources, Data curation, Writing – original draft, Writing – review & editing, Visualization, Supervision. **Annu Mishra:** Conceptualization, Methodology, Formal analysis, Investigation, Writing – original draft, Writing – review & editing, Visualization, Supervision. **Mohd Dilshad Ansari:** Conceptualization, Software, Validation, Resources, Data curation, Writing – review & editing, Visualization, Supervision, Project administration. **Ravi Prakash Chaturvedi:** Conceptualization, Methodology, Formal analysis, Investigation, Writing – original draft, Writing – review & editing, Visualization, Supervision, Project administration. **Rajneesh Kumar Singh:** Conceptualization, Methodology, Formal analysis, Investigation, Writing – original draft, Writing – review & editing, Visualization, Supervision.

References

[1] Ghosh, K. P., Hasan, M., Robin, M. T. I., Hossain, M. A., & Islam, M. S. (2025). A novel deep learning framework with temporal attention convolutional networks for intrusion detection in IoT and IIoT networks. *Scientific Reports*, *15*(1), 44624. https://doi.org/10.1038/s41598-025-32697-1

[2] Sharma, K. P., Nagpal, T., Vora, T., Yadav, A., Abdullah, M. I., Jayaprakash, B., . . . , & Bukate, B. B. (2025). Interpretable intrusion detection for IoT environments using a self-attention-based explainable AI framework. *Scientific Reports*, *15*(1), 39937. https://doi.org/10.1038/s41598-025-23750-0

[3] Moghaddam, P. S., Vaziri, A., Khatami, S. S., Hernando-Gallego, F., & Martín, D. (2025). Generative adversarial and transformer network synergy for robust intrusion detection in IoT environments. *Future Internet*, *17*(6), 258. https://doi.org/10.3390/fi17060258

[4] Zegarra Rodriguez, D., Daniel Okey, O., Maidin, S. S., Umoren Udo, E., & Kleinschmidt, J. H. (2023). Attentive transformer deep learning algorithm for intrusion detection on IoT systems using automatic Xplainable feature selection. *Plos One*, *18*(10), e0286652. https://doi.org/10.1371/journal.pone.0286652

[5] Yang, K., Wang, J., & Li, M. (2024). An improved intrusion detection method for IIoT using attention mechanisms, BiGRU, and Inception-CNN. *Scientific Reports*, *14*(1), 19339. https://doi.org/10.1038/s41598-024-70094-2

[6] Umer, M., Tahir, M., Sardaraz, M., Sharif, M., Elmannai, H., & Algarni, A. D. (2025). Network intrusion detection model using wrapper based feature selection and multi head attention transformers. *Scientific Reports*, *15*(1), 28718. https://doi.org/10.1038/s41598-025-11348-5

[7] Zhang, Y., Muniyandi, R. C., & Qamar, F. (2025). A review of deep learning applications in intrusion detection systems: Overcoming challenges in spatiotemporal feature extraction and data imbalance. *Applied Sciences*, *15*(3), 1552. https://doi.org/10.3390/app15031552

[8] Pei, Y., Tan, Y., Gao, W., Li, F., & Wang, M. (2025). HiSeq-TCN: High-dimensional feature sequence modeling and few-shot reinforcement learning for intrusion

detection. *Electronics*, *14*(21), 4168. https://doi.org/10.3390/electronics14214168

[9] Alashjaee, A. M., & Alqahtani, F. (2025). Enhanced intrusion detection system IoT network security model by feed forward neural network and machine learning. *Scientific Reports*, *15*(1), 36085. https://doi.org/10.1038/s41598-025-20047-0

[10] Kamal, H., & Mashaly, M. (2025). Robust intrusion detection system using an improved hybrid deep learning model for binary and multi-class classification in IoT networks. *Technologies*, *13*(3), 102. https://doi.org/10.3390/technologies13030102

[11] de Nascimento, C. M., & Hou, J. (2025). Uncertainty-aware adaptive intrusion detection using hybrid CNN-LSTM with cWGAN-GP augmentation and human-in-the-loop feedback. *Safety*, *11*(4), 120. https://doi.org/10.3390/safety11040120

[12] Nazim, S., Hussain, S. S., Yousuf, B., Sultana, S., & Tanweer, E. (2025). An innovative intrusion detection framework using GAN-augmented deep ensemble neural network for cross-domain IoT-cloud security. *Internet of Things*, *34*, 101773. https://doi.org/10.1016/j.iot.2025.101773

[13] Saidi, F. (2025). IDS-GPT: A novel deep learning-powered framework for network traffic intrusion detection. *Procedia Computer Science*, *270*, 1611–1620. https://doi.org/10.1016/j.procs.2025.09.281

[14] Saidi, F., Trabelsi, Z., & Ghazela, H. B. (2019). Fuzzy logic based intrusion detection system as a service for malicious port scanning traffic detection. In *2019 IEEE/ACS 16th International Conference on Computer Systems and Applications*, 1–9. https://doi.org/10.1109/AICCSA47632.2019.9035263

[15] Ioannou, I., & Vassiliou, V. (2026). Generative adversarial networks for energy-aware IoT intrusion detection: Comprehensive benchmark analysis of GAN architectures with accuracy-per-joule evaluation. *Sensors*, *26*(3), 757. https://doi.org/10.3390/s26030757

[16] Yang, Y., Liu, X., Wang, D., Sui, Q., Yang, C., Li, H., . . . , & Luan, T. (2025). A CE-GAN based approach to address data imbalance in network intrusion detection systems. *Scientific Reports*, *15*(1), 7916. https://doi.org/10.1038/s41598-025-90815-5

[17] Rao, Y. N., & Suresh Babu, K. (2023). An imbalanced generative adversarial network-based approach for network intrusion detection in an imbalanced dataset. *Sensors*, *23*(1), 550. https://doi.org/10.3390/s23010550

[18] Kang, F., Feng, T., & Lin, J. (2025). VAE-GAN-guided cross-class generation: A class imbalance data augmentation method for network intrusion detection. *Electronics*, *14*(11), 2103. https://doi.org/10.3390/electronics14112103

[19] Riaz, R., Han, G., Shaukat, K., Khan, N. U., Zhu, H., & Wang, L. (2025). A novel ensemble Wasserstein GAN framework for effective anomaly detection in industrial internet of things environments. *Scientific Reports*, *15*(1), 26786. https://doi.org/10.1038/s41598-025-07533-1

[20] Abdelhamid, S., Hegazy, I., Aref, M., & Roushdy, M. (2024). Attention-driven transfer learning model for improved IoT intrusion detection. *Big Data and Cognitive Computing*, *8*(9), 116. https://doi.org/10.3390/bdcc8090116

[21] Babu, K. S., & Rao, Y. N. (2023). MCGAN: Modified conditional generative adversarial network (MCGAN) for class imbalance problems in network intrusion detection system. *Applied Sciences*, *13*(4), 2576. https://doi.org/10.3390/app13042576

[22] Li, J., Zong, W., Chow, Y. W., & Susilo, W. (2025). Mitigating class imbalance in network intrusion detection with feature-regularized GANs. *Future Internet*, *17*(5), 216. https://doi.org/10.3390/fi17050216

[23] Zhang, C., Li, J., Wang, N., & Zhang, D. (2025). Research on intrusion detection method based on transformer and CNN-BiLSTM in internet of things. *Sensors*, *25*(9), 2725. https://doi.org/10.3390/s25092725

[24] Upender, T., Neelakantappa, M., Rao, C. P., Gera, J., Reddy, V. L., & Yamsani, N. (2025). CyberDetect MLP a big data enabled optimized deep learning framework for scalable cyber-attack detection in IoT environments. *Scientific Reports*, *15*(1), 40865. https://doi.org/10.1038/s41598-025-24459-w

[25] Abdulganiyu, O. H., Ait Tchakoucht, T., Alaoui, A. E. H., & Saheed, Y. K. (2025). Attention-driven multi-model architecture for unbalanced network traffic intrusion detection via extreme gradient boosting. *Intelligent Systems with Applications*, *26*, 200519. https://doi.org/10.1016/j.iswa.2025.200519

[26] Xu, W., Jang-Jaccard, J., Liu, T., Sabrina, F., & Kwak, J. (2022). Improved bidirectional GAN-based approach for network intrusion detection using one-class classifier. *Computers*, *11*(6), 85. https://doi.org/10.3390/computers11060085

[27] Wang, Z., Li, J., Yang, S., Luo, X., Li, D., & Mahmoodi, S. (2024). A lightweight IoT intrusion detection model based on improved BERT-of-Theseus. *Expert Systems with Applications*, *238*, 122045. https://doi.org/10.1016/j.eswa.2023.122045

[28] Kareem, S. W. (2025). Enhanced intrusion detection system using hybrid-inspired algorithms and conditional generative adversarial networks for Internet of Things security. *Egyptian Informatics Journal*, *31*, 100763. https://doi.org/10.1016/j.eij.2025.100763

[29] Agarwal, N., Rana, A., & Pandey, J. P. (2021). An efficient and optimised approach for secured file sharing in cloud computing. *International Journal of Advanced Intelligence Paradigms*, *18*(2), 232–246. https://doi.org/10.1504/IJAIP.2021.112905

[30] Kundu, S., Chaturvedi, R. P., Mishra, A., Agrawal, A., Aldossary, S. M., Ayadi, M., . . . , & Hashmi, A. (2026). A trustworthy cybersecurity model for transparent cyberattack detection using Bald Eagle Search tuned XGBoost and explainable AI. *Journal of Cloud Computing*, *15*(1), 58. https://doi.org/10.1186/s13677-026-00878-6

[31] Tseng, S. M., Wang, Y. Q., & Wang, Y. C. (2024). Multi-class intrusion detection based on transformer for IoT networks using CIC-IoT-2023 dataset. *Future Internet*, *16*(8), 284. https://doi.org/10.3390/fi16080284

[32] Xu, C., Zhan, Y., Chen, G., Wang, Z., Liu, S., & Hu, W. (2025). Elevated few-shot network intrusion detection via self-attention mechanisms and iterative refinement. *PloS One*, *20*(1), e0317713. https://doi.org/10.1371/journal.pone.0317713

[33] Kambarima, T. A., Chaturverdi, R. P., Mishra, A., Singh, R. K., Bintube, M. M., & Tiwari, D. P. (2026). BERT-CNN hybrid for robust cyber threat detection in multilingual Twitter datasets. In *2026 International Conference on Intelligent Systems in Engineering, Secured Systems and Cybersecurity*, 83–87. https://doi.org/10.1109/ICISESSC68634.2026.11542632

[34] Zhao, Q., Wang, F., Wang, W., Zhang, T., Wu, H., & Ning, W. (2025). Research on intrusion detection model based on improved MLP algorithm. *Scientific Reports*, *15*(1), 5159. https://doi.org/10.1038/s41598-025-89798-0

[35] Wang, Z., Yuan, F., Qiu, H., Liu, Y., & Luo, X. (2026). Large-scale BGP routing anomaly detection based on graph attention auto-encoder. *IEEE Transactions on Network and Service Management*, *23*, 487–501. https://doi.org/10.1109/TNSM.2025.3631527

[36] Mahmood, A., Tiwari, A. K., & Singh, S. K. (2024). C-net: A deep learning-based Jujube grading approach. *Journal of Food Measurement and Characterization*, *18*(9), 7794–7805. https://doi.org/10.1007/s11694-024-02765-7

[37] Lazem, S. (2026). A novel architecture and methodology to detect intrusions against edge-based IIoT using machine learning. *Journal of Applied Science and Technology Trends*, *7*(1), 95–109. https://doi.org/10.38094/jastt71439

[38] Rizki, M., Subhan, M., & Waladi, A. (2026). Anomaly detection in sensor data for tomato farming using a federated learning-autoencoder. *International Journal of Informatics and Computation*, *8*(1), 442–456. https://doi.org/10.35842/ijicom.v8i1.235

[39] Alsulami, A. A., Alturki, B., Tayeb, A. J., Alsemmeari, R. A., & Alsini, R. (2026). An adaptive intrusion detection framework for IoT: Balancing accuracy and computational efficiency. *Computers, Materials & Continua*, *87*(3). https://doi.org/10.32604/cmc.2026.076413

[40] Rahman, M. A., Devnath, R. K., Mehedi, C. M., Begum, M., Mahmud, T., & Hasan, M. T. (2026). AI-based intrusion detection method for enhanced cybersecurity in modern network traffic. In *2026 IEEE 2nd International Conference on Quantum Photonics, Artificial Intelligence & Networking*, 1–6. https://doi.org/10.1109/QPAIN69676.2026.11546166

How to Cite: Shukla, O., Mishra, A., Ansari, M. D., Chaturvedi, R. P., & Singh, R. K. (2026). GAN-Augmented Intrusion Detection: A Feature-Wise Attention MLP Framework for Imbalanced Network Traffic. *Artificial Intelligence and Applications*. https://doi.org/10.47852/bonviewAIA620210960