Generative AI Driven Synthetic Attack Augmentation for Enhanced Intrusion Detection Using an Imbalanced Dataset

Mamoona Nawaz *, Shireen Tahira, Anum Yasmin

Posted Date: 17 December 2025

doi: 10.20944/preprints202512.1521.v1

Keywords: intrusion detection system; generative ai; machine learning; CTGAN; class imbalance

 doi:10.20944/preprints202512.1521.v1

Article

Generative AI Driven Synthetic Attack Augmentation for Enhanced Intrusion Detection Using an Imbalanced Dataset

Mamoona Nawaz *, Shireen Tahira and Anum Yasmin

Department of Computer Science, IIU

* Correspondence: mamoonanawaz62@gmail.com

Abstract

Intrusion Detection Systems (IDS) are very important in ensuring the security of the modern network, but persistent problems with severe class imbalance in the datasets of the real network traffic conditions show that the minor types of attacks are highly underrepresented. Critical attacks present in the popular dataset, including Brute Force and Web Attacks, are very infrequent compared to regular traffic and high-volume attacks, which causes biased learning, high false-negativities, and bad minority attacks detection. To overcome this problem, this paper suggests a Generative AI-based synthetic attack augmentation model on Conditional Tabular Generative Adversarial Networks (CTGAN) to improve the performance of the IDS in imbalanced jobs. The given strategy is aimed at producing high-fidelity synthetic samples of minority attack classes without changing the statistical properties and behavioral patterns of actual network traffic. Training and testing of augmented data ensemble-based machine learning models, namely Random Forest and Extreme Gradient Boosting (XGBoost) are performed using the augmented dataset. Experiments using the CICIDS2017 dataset show that the detection in the minority attack is significantly improved. Synthetic augmentation boosted Recall to Web Attacks by 28 to 91 with Random Forest and 32 to 94 with XGBoost, and Brute Force detection Recall boosted by 45 to 95 and 55 to 98 respectively. Overall Recall and F1-score also gained significantly and XGBoost obtained F1-score of 94% on the augmented dataset. These findings support the hypothesis that Generative AI-based synthetic data augmentation works well in class imbalance, false negative, and increases the resilience and reliability of intrusion detection systems in real-life cybersecurity settings.

Keywords: intrusion detection system; generative ai; machine learning; CTGAN; class imbalance

1. Introduction

The IDS plays an essential role in the present-day cyber defense system, as it is the role of these systems to trace the cases of unauthorized access and malevolent activity within the complicated networks. With the growth in scale and complexity of cyberattacks due to the adoption of digital infrastructures with cloud computing, IoT and edge systems, the conventional ability of IDS systems to sustain consistent detection rates, especially those of infrequent attack types, has been put to the test [Synthetic attack data generation model applying GAN for intrusion detection [1]. Given that machine learning (ML) methods have been widely used to enhance the detection of anomalies, both methods frequently have difficulties detecting a minority type of attack when the majority of traffic or typical attacks are dominant, occupy long sequences, and are predominant in databases (when compared to a minority attack) [2]. The problem of class imbalance is important and the high false negatives when dealing with underrepresented attacks is devastating to the IDS performance [3]. Recent research has examined generative models including Generative Adversarial Networks (GANs) and variants to produce real samples of attacks, which has greatly improved the IDS performance of minority classes without negatively affecting the overall detection performance [4].

 doi:10.20944/preprints202512.1521.v1

Based on these developments, our study takes advantage of Generative AI-based synthetic attack augmentation by enhancing recall and robustness of IDS models that are trained on unbalanced benchmark datasets.

In practical network conditions, low frequency attacks like Brute Force and Web-based intrusions can be used with a multi-step process that is similar to the legitimate user behavior and hence is hard to detect with the traditional IDS trained on unbalanced data. To give just one example, in an enterprise web application, an attacker can perform a Brute Force login attack by making a few credential tries with longer time intervals in order to avoid rate-based and anomaly-based detection systems. Since the traffic of such attacks constitutes a small percentage of all traffic flows within the network, the IDS models often end up categorizing this traffic as normal traffic [5]. This is seen to also apply to Web Attacks where malicious people can take advantage of Web vulnerabilities like SQL injection or cross-site scripting by incorporating malicious payloads in otherwise legitimate HTTP requests [6].

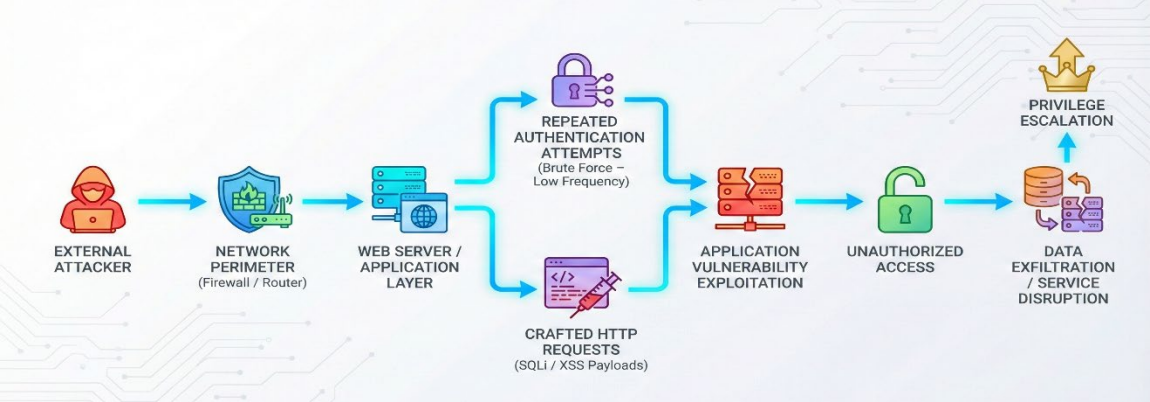

Figure 1. Comprehensive workflow of a web-based cyber attack.

Research indicates that IDS may not be effective in detecting the initial phases of such attacks because they are learned using imbalanced data and consequently, they consume more time before detecting and therefore, more time before the breach is detected is exercised [7]. Synthetic attack augmentation with the help of generative AI allows the formation of realistic samples that model such step-wise attack behavior and enhance the capacity of the IDS to detect the slightest malicious patterns during deployment [8]. These realistic attack sequences must therefore be incorporated into the training data in order to make IDS resilient to actual cyber threat in the real world.

1.1. Problem Statement

Their functionality is greatly diminished when in training on highly-imbalanced datasets (which is a typical feature of real-world network traffic statistics). In benchmark data like CICIDS2017, the normal traffic and several prevalent types of attacks constitute most of the samples, whereas more critical types of attacks, including Brute Force and Web Attacks, have a low frequency [9]. This bias makes machine learning (ML) models be biased to majority classes and lead to misleading high accuracy, and the inability to detect rare but serious attacks [10]. The incidents of minorities attacks are usually incorrectly categorized as regular traffic or they are simply overlooked and hence high rates of false-negative are experienced, which compromises the reliability of IDS in the real-world settings [11]. These limitations are especially hazardous in any contemporary cyber infrastructures, where even a single breach that will go undetected can result in the massive destruction. Solutions that are currently available, like cost-sensitive learning and conventional oversampling methods have not been very successful since they usually cannot maintain the complex statistical correlation existing in network traffic data [12]. Consequently, there is a pressing need for advanced data-level

 doi:10.20944/preprints202512.1521.v1

solutions capable of addressing class imbalance while maintaining the realism and diversity of attack behaviors necessary for robust IDS training.

1.2. Motivation

The main source of the research is the limitation of real-world intrusion databases, which do not provide adequate samples of attacks of some categories, but are of high practical importance. The ability to gather equal and holistic cybersecurity data is frustrating because of its privacy factor, the changing attack patterns, and the infrequent occurrence of certain intrusion cases [13]. So, IDS models that have been trained using such datasets are incapable of making good generalizations to underrepresented attack scenarios, especially Web Attacks and Brute Force intrusions [14]. The recent developments in Generative AI have shown good potential in learning complicated data distributions and producing real samples of synthetic data that are close to original data [15]. Compared to traditional approaches to oversampling, generative models like GANs and variational architectures are capable of modeling nonlinear relationships and feature relationships in tabular cyber data [16]. This ability renders Generative AI an attractive answer to realistic attack data enhancement as it allows the enhancement of minority classes, yet without replicating and/or adding noise to the existing samples. Inspired by these advances, this study investigates how Generative AI can be applied to improve the training data of an IDS to help it achieve better detection of unusual attacks, without causing a loss of general dataset fidelity or operational relevance [17].

1.3. Research Objectives

The primary goal of the research is to improve the detection ability of Intrusion Detection Systems by counteracting the problem of class imbalance which exists in the CICIDS2017 dataset. The first one is to analyze the dataset systematically and determine the most imbalanced minority attack classes, with more attention paid to the Brute Force and Web Attack categories [18]. The second goal is to develop and deploy a Generative AI-based system that has the ability to produce high quality synthetic samples that are true to the statistical and behavioral nature of these minority attacks [19]. The purpose of the research is to enhance the original dataset with artificial created instances of attacks to form a more balanced and informative training set of the IDS models. The third goal is to determine the effect of synthetic attack augmentation on the IDS performance using a variety of machine learning and deep learning classifiers [20]. The main metrics that are used to determine performance improvements are Recall and F1-score because these metrics are more accurate in measuring the effectiveness of detection of rare attack classes [21]. With these aims, the research should help prove the practical utility of Generative AI in enhancing the IDS resilience to the hidden cyber threats.

1.4. Contributions

The study has a number of major contributions to the areas of intrusion detection as well as data augmentation in cybersecurity. To begin with, it suggests a type of Generative AI-based synthetic data augmentation that is specifically applied to tabular network traffic data and overcomes the issues related to the problem of class imbalance in the training datasets of IDS [22]. Second, the research gives special attention to the minor types of attacks i.e., Web Attacks and Brute force attacks though not a significant feature of previous studies but an actual reality in the world. Third, experimental assessment of several machine learning IDS models is carried out including an evaluation of performance before and after synthetic data augmentation. This comparative study gives explicit ideas about the impact of Generative AI enhanced augmentation and its effect on the detection ability of various model architectures. Lastly, the study has shown statistically significant increases in Recall and F1-score of minority attack categories, which illustrates the success of the identified strategy in terms of reducing false negatives and increasing the reliability of IDS. Put

 doi:10.20944/preprints202512.1521.v1

together, these contributions move to realize the use of Generative AI in cybersecurity and offer an easily scaled approach towards enhancing intrusion detection in unequal data situations.

1.5. Article Structure

The rest of this paper is structured in the following way. Section 2 is a literature review of relevant work on the Generative Adversarial Network based synthetic data augmentation to generate data used in intrusion detection and the limitations of current oversampling and data balancing techniques are presented. In section 3, the suggested Generative AI-based synthetic attack augmentation model will be described, which comprises data preprocessing, minority attack detection, and data generation plan based on the CTGAN. The section 4 shows the experimental setup, including the CICIDS2017 data, intrusion detection models, and performance metrics. In Section 5, the experimental results are reported and analyzed in comparison of IDS performance before and after synthetic data enhancement in terms of performance using Random Forest and XGBoost classifiers. Section 6 addresses the results, with the focus on the effect of synthetic data augmentation on minority attack detection and IDS overall strength. Lastly, Section 7 will also give a conclusion of the paper summarizing the key contributions and giving the possible lines of future research.

2. Related Work

[23] Kumar et al. are a synthetic attack data generation model that uses a Wasserstein Conditional Generative Adversarial Network (WCGAN) and XGBoost classifier to overcome critical issues of data imbalance in intrusion detection systems (IDS). The method produces high-fidelity synthetic attacks signatures by conditioning WCGAN on the types of attacks, producing synthetic attack signatures that are highly realistic in the patterns of network traffic, and allows effective training of the XGBoost classifier even when the classes are underrepresented. This combination greatly improves the accuracy of detection when using imbalanced datasets, and it is better than existing methods (SMOTE) used in oversampling. The authors show excellent results in benchmark datasets (NSL-KDD and CICIDS2017) and indicate that WCGAN is a stable method in training and capable of modeling complex attack distributions. Their work highlights how generative models can fundamentally change cybersecurity to deliver more resilient IDS that can leverage scalable synthetic data augmentation to adapt to changing threat environments.

[24] Park et al. suggest a more advanced AI-based Network Intrusion Detection System (NIDS) that employs reconstruction error and Wasserstein distance-based Generative Adversarial Networks (GANs) and autoencoders-based deep learning models to solve the current data imbalance problem in cybersecurity. With communication technologies increasing the attack surfaces in distributed networks, conventional AI anomaly detection has a limitation to learn rare malicious behavior resulting into incorrect threat detection. Their method produces realistic synthetic traffic of minority attacks, which supplements the unbalanced datasets to train strong models. Extensive testing on a variety of datasets demonstrates that the proposed system performs much better than the previous AI-based NIDS and has a higher detection rate with heterogeneous threats. The authors tackle the drawbacks of current models by concentrating on the advanced generative techniques, which provides a scalable solution to the problem of improving network security resilience against the changing cyber risks in the IoT environment.

[25] Allagi et al. present a very similar intrusion detection system, using Conditional Tabular Generative Adversarial Networks (CTGANs), to surmount the failure of traditional SMOTE oversampling on unbalanced cybersecurity datasets such as NSL-KDD and UNSW-NB15. In contrast to SMOTE that fails to handle various types of data, non-linear relationships, preservation of data distribution, and oversampling noise, CTGAN produces high-quality synthetic samples that are statistically faithful and can retrieve intricate patterns. Such augmented datasets are used to train a Convolutional Neural Network (CNN) that has two convolution layers, max-pooling,ReLU activation and a dense layer to effectively classify intrusion. Extensive assessments in terms of

 doi:10.20944/preprints202512.1521.v1

accuracy, recall, precision, specificity, and F1-score prove to be better than baselines, with CTGAN by far increasing the detection rates, and lowering the false negatives. This methodology indicates the possibilities of GAN-based oversampling to be applied to the real world, and a scalable solution to increase the resilience of cybersecurity in the case of uneven distribution of threats.

[26] Ding et al. present a more advanced version of Generative Adversarial Network (GAN)-based data augmentation TMG-GAN, to detect imbalanced network intrusion in susceptible IoT environments, in which the ratio between normal traffic and rare attacks is high to disregard the limiting ability of machine learning. The TMG-IDS system offers a multi-generator design that generates different types of attacks in parallel, an integrated classifier that trains generator and discriminator through classification loss, and loss based on cosine similarity which improves the quality of synthetic samples and reduces the overlap between distributions. The method is used to overcome the weaknesses of traditional oversampling, producing a high-fidelity non-overlapping attack data specific to datasets (such as CICIDS2017 and UNSW-NB15). Extensive experiments indicate that TMG-IDS has a higher Precision, Recall, and F1-scores than the most advanced oversampling and detection algorithms and proves to be an effective tool in the creation of robust IoT security against growing threats using efficient data balancing and quality enhancement.

[27] Ding at al. introduce a hybrid model that integrates the K-Nearest Neighbors (KNN) and Generative Adversarial Networks (GANS) to address the issue of imbalanced data classification in intrusion detachments (where attack classes with small population groups are extremely underrepresented in the background of normal traffic). The technique takes advantage of GANs to produce high-quality synthetic minority samples that maintain data distribution and reproduce more intricate attack patterns to augment an imbalanced dataset in a more effective KNN training. This solves drawbacks of the standard oversampling methods in that realistic intrusions are generated and they lead to local pattern recognition by KNN, which does not cause noise and overlap. The framework boosts detection accuracy and recall and the general accuracy of rare threats by combining generative augmentation with the instance-based learning of KNN over benchmark cybersecurity datasets. Their work shows that GAN-KNN hybrids are effective in creating strong classifiers that stand against class imbalance, which provides a practical solution to network security implementation in real-world scenarios that may be faced with changing cyber attacks.

[28] Arafah et al. introduces an anomaly-based network intrusion detection framework based on a denoising autoencoder (AE) with watersween Generative Adversarial Network (WGAN) to address the challenges of high-dimensional and large-scale network traffic anomaly detection, which is imbalanced in intrusion detection systems. To solve the imbalance between classes, the AE-WGAN model derives high-representative features of raw data and synthesizes realistic types of attacks to improve accurate anomaly detection without violating data integrity. The extensive testing on NSL-KDD and CICIDS-2017 databases in binary and multiclass cases using a variety of classifiers has shown high accuracy, precision, recall and F1-score, which are better in generalization to unseen attacks compared to state-of-the-art models. Time complexity analysis establishes computational effectiveness and quality synthetic attack generation, which makes AE-WGAN a robust, flexible framework that can be used to counter modern cybersecurity needs and emerging cyber threats in the complex network environment.

[29] Rahman et al. describe SYN-GAN, a powerful IoT security intrusion detection system that only uses 100% synthetic data generated through Generative Adversarial Networks (GANs) and does not rely on expensive real-world data and the system has serious issues with class imbalance that negatively affects the detection of malicious actions in a distributed IoT environment. This novel method produces the realistic network traffic that simulates various attacks, thus it allows the flexible and ethical development of models of network intrusion detection systems (NIDS). Critiqued on UNSW-NB15 (90% accuracy, 91% precision, 90% recall, 89% F1), NSL-KDD (84% across metrics), and BoT-IoT (perfect 100% metrics), SYN-GAN provides a comparable or better performance to previous literature with real data. The data produced by generative patterns of intrusions as a real-world analogue in the case of cybercrime allows the framework to provide new opportunities to scale the

 doi:10.20944/preprints202512.1521.v1

concept of cybersecurity and limit data collection obstacles, improving the response to threats in the most vulnerable IoT ecosystems.

[30] Zeghida et al. present a new system of detecting intrusion in the context of IoT cybersecurity, based on Generative Adversarial Networks (GANs), to produce high-quality synthetic samples due to the critical imbalance of classes in MQTT protocol traffic, where attack traffic is significantly underrepresented in relation to normal traffic. This GAN-based augmentation is able to retain complex data features, allowing to train three hybrid deep neural networks, i.e., CNN-RNN, CNN-LSTM, and CNN-GRU, with specific objectives, namely, lightweight MQTT attack detention. In contrast to the conventional balancing tools, their method generates realistic synthetic intrusions that increase the accuracy of multi-class classification with a limited false positive. The experimental findings combined with the original and GAN-generated MQTT data set indicate that there is a notable enhancement in the detection performance of all the hybrid models which depict the effect of the GANs to eliminate the weaknesses of the ML in the case of imbalanced IoT conditions. The work contributes to the development of autonomous IDS and provides high-quality and flexible protection against new cyber threats in a quickly growing interconnected network.

Table 1. Related Work of using Generative AI.

References	Methodology	Dataset(s)	Key Contribution
[23]	WCGAN (Wasserstein Conditional GAN)	NSL-KDD, UNSW-NB15, BoT-IoT	WCGAN + XGBoost improves detection on imbalanced data, outperforming other GANs and DGM
[24]	Wasserstein GAN, Autoencoder	Multiple (not specified)	GANs generate plausible synthetic attacks, improving NIDS performance on imbalanced data
[25]	CTGAN	NSL-KDD, UNSW-NB15	CTGAN-generated samples enhance CNN-based detection, outperforming SMOTE
[26]	Multi-generator GAN	CICIDS2017, UNSW-NB15	Multi-generator GAN augments attack types, improving precision, recall, and F1-score
[27]	TACGAN (Tabular Auxiliary Classifier GAN)	3 real-world IDS datasets	KNN undersampling + TACGAN oversampling balances data, boosting classification metrics
[28]	AE + WGAN	NSL-KDD, CICIDS-2017	AE-WGAN generates realistic attacks, improving detection and generalization
[29]	GAN	UNSW-NB15, NSL-KDD, BoT-IoT	100% synthetic GAN data enables high-accuracy NIDS, reducing real data dependency
[30]	GAN	MQTT dataset	GAN-generated data improves hybrid DL model performance for IoT/MQTT attacks

3. Dataset

Canadian Institute of Cybersecurity (CIC) created the CICIDS2017, a generic and extensively utilized benchmark of intrusion detection studies. The dataset has been created in a realistic enterprise network setting that models realistic user behavior and attack scenarios. Normal traffic

 doi:10.20944/preprints202512.1521.v1

was generated by normal activities like browsing web, sending and receiving emails, transferring files, and multimedia streaming whereas attack traffic was generated using controlled and well-defined attack scripts. The network packets were redirected and received into bidirectional flow-based records in CICFlowMeter (which identifies more than 80 statistical measures associated with flow duration, packet size, and rates of bytes and protocol behavior). This naturalistic traffic creation model renders CICIDS2017 appropriate in testing machine learning IDS models in real-life scenarios.

The CICIDS2017 dataset is preprocessed in this study to provide consistency in data and adapting to the model-training. The original daily traffic files were combined into one dataset and the classifications of attacks were grouped together to eliminate the redundancy and to make the classification easier. Same categories in their attacks were brought together in larger categories like DDoS, Port Scanning, Brute Force and Web Attacks with normal traffic categorized under Name normal Traffic. The process of data cleaning such as eliminating duplicate records, elimination of missing or infinite values and normalization of numerical features were implemented to enhance data quality. The output of this preprocessing is an integrated and coherent dataset, which is realistic in terms of class imbalance and is able to evaluate the IDS models fairly and consistently.

3.1. Attack Class Distribution

The last dataset applied in the research has a very skewed distribution of classes, which is representative of the actual network traffic distribution. The distribution of samples in different big categories of attacks is as follows in Table below:

Attack Type	Samples
Normal Traffic	998,426
DDoS	128,014
Port Scanning	90,694
Brute Force	9,150
Web Attacks	2,143

As it can be seen, normal traffic and high-volume attacks like DDoS and Port Scanning dominate the dataset, with Brute Force and Web Attacks comprising a small portion of the total samples. This asymmetry has a great impact on the behavior of IDS learning and justifies the necessity of specific data augmentation measures.

3.2. Minority Attack Identification

The minority attack classes are classified as those which have considerably low sample numbers than the dominant traffic classes, and thus they are underrepresented in the model training. Web Attacks and Brute Force attacks are minority classes in CICIDS2017 dataset as the numbers of such attacks are very small in comparison to normal traffic and other types of attacks. The risk of such attacks is significant, even though they are not that frequent; Brute Force attacks are aimed at authentication systems, and the Web Attacks make use of the application-level vulnerabilities, e.g., SQL injection or cross-scripting attacks. They have a limited representation in the dataset which makes their detection performance poor and with high false-negative. Hence, the current study aims directly at the enhancement of Web Attacks and Brute Force samples with the help of Generative AI methods to enhance the IDS detection potential and its resilience to such essential but less represented threats.

4. Proposed Methodology

The Figure 2 presents a data augmentation pipeline on the CICIDS2017 dataset by first identifying and isolating minority attack types such as Brute Force and Web Attacks to solve the issue of class imbalance. This isolated data is trained with the help of a Generative AI model (CTGAN) to generate high-quality synthetic samples, which are carefully verified and then combined to form a

 doi:10.20944/preprints202512.1521.v1

balanced dataset. The resultant data is then used to train the strong Intrusion Detection System (IDS) using the Random Forest and XGBoost algorithms. The final stage is a model analysis that will be used to check the performance gains obtained by the synthetic data generation.

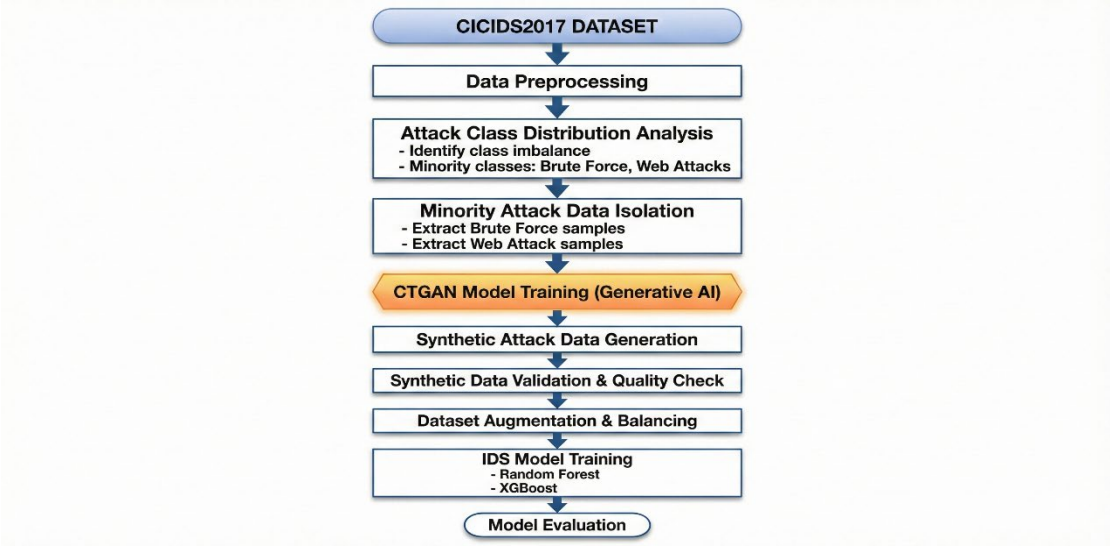

Figure 2. Proposed methodology of CTGAN.

4.1. Data Preprocessing

Data Preprocessing The preprocessing of data is an essential part of the proposed methodology because the quality and consistency of the input information is crucial to the performance of the machine learning based intrusion detection systems. CICIDS2017 dataset is high-dimensional network flow and it is also noisy, redundant and has extreme imbalance between classes. In order to have a stable model training and evaluation, a systematic preprocessing pipeline is implemented that includes data cleaning, feature scaling and stratified train-test splitting.

4.2. Data Cleaning

The raw CICIDS2017 data has redundant flow records, missing records, and noisy samples added in the process of capturing traffic and extracting features. Duplicates are eliminated to avoid biasing and overfitting the model to duplicates. Also, records, which have missing, infinite, or undefined values, are either fixed or dropped to ensure that the numerical state of the model remains stable during the training process. The reduction of noise is used to assure that the learning process is not distorted by irrelevant or corrupted samples so that the overall reliability and the generalization ability of the IDS models is increased.

 doi:10.20944/preprints202512.1521.v1

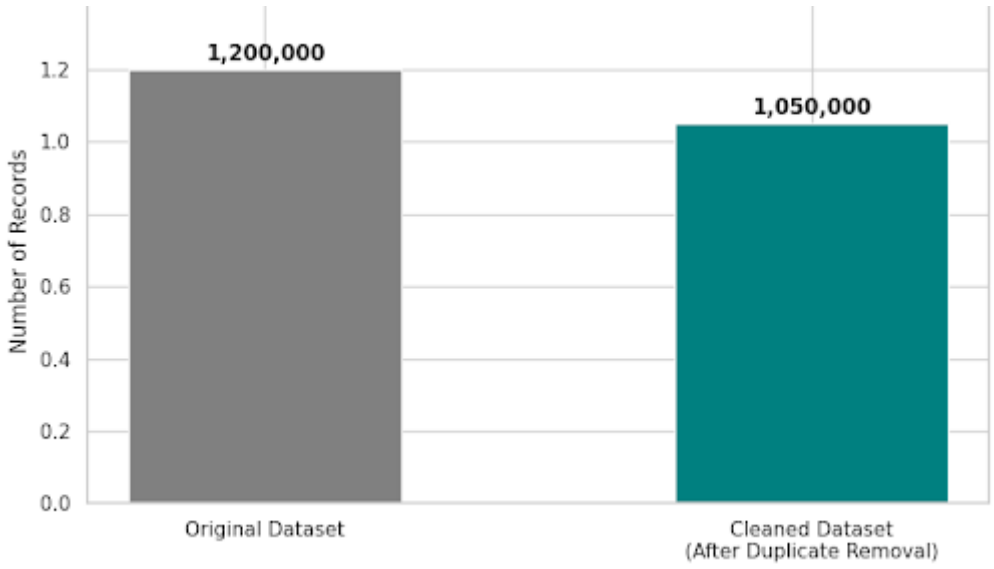

Figure 3. Data Cleaning results.

4.3. Feature Scaling

Network traffic characteristics in CICIDS2017 have a huge array of scale, including small number of packets and big number of bytes. The scaling of features is done through normalization or standardization in order to make sure that each feature is equally regarded in the learning process. This will ensure that features with bigger numeric values do not dominate model optimization and it is especially significant to distance-based and gradient-based learning algorithms. Adequate scaling increases the convergence rate of models and better detection.

4.4. Train–Test Split

In order to assess the performance of IDSs fairly in case of imbalanced evaluation, the dataset is stratified split into training and testing sets. This method maintains the original balance of classes in both subsets, whereby minority attack classes like Brute Force and Web Attacks are present in their proportionality. Stratified splitting allows evaluating performance unbiasedly and can be certain that any improvement in data augmentation is well represented in the course of testing.

 doi:10.20944/preprints202512.1521.v1

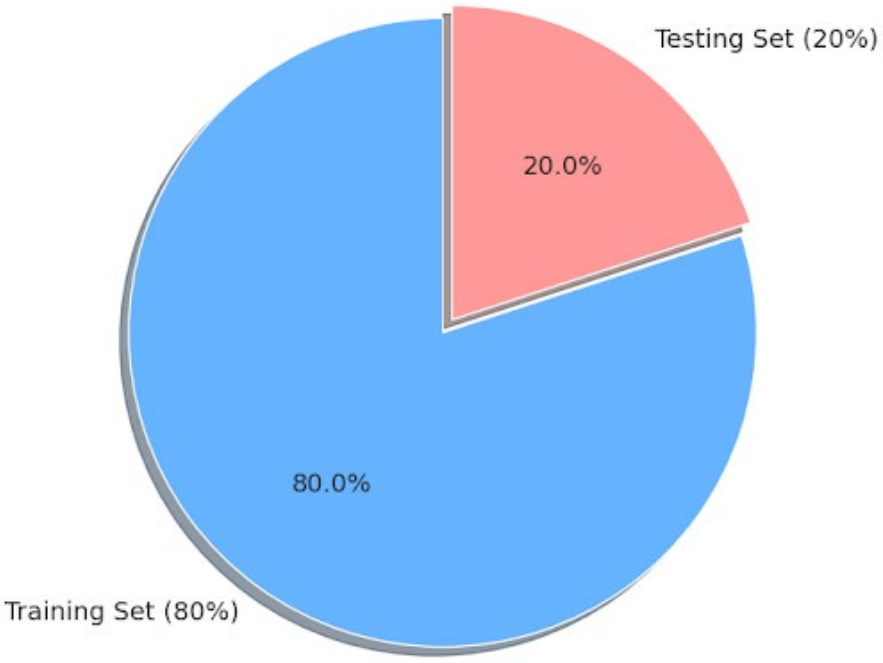

Figure 4. Splitting of Training and testing dataset.

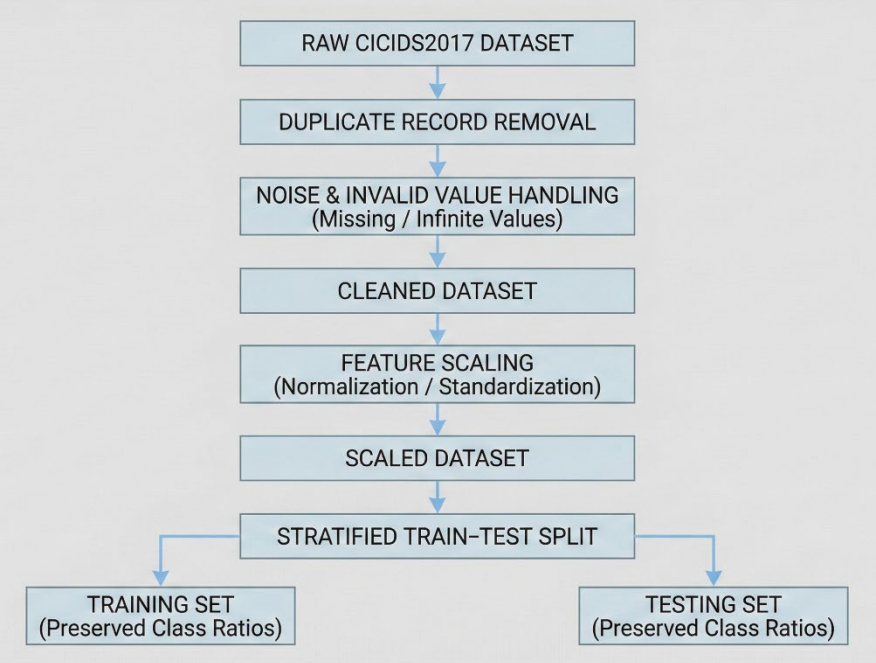

Figure 5. Data cleaning process.

5. Generative AI–Based Synthetic Data Augmentation

A Generative AI-based synthetic data augmentation model is used to balance the class imbalance problems associated with network intrusion datasets. Because the network traffic data is usually tabular, heterogeneous and has complex non-linear interactions among the features, the standard oversampling methods do not usually reflect the actual underlying distribution. Thus, a model tailored towards generating high-fidelity synthetic samples that contribute to the better representation of minority attack classes without noise and redundancy is needed.

 doi:10.20944/preprints202512.1521.v1

5.1. Generative Model Selection

Conditional Tabular Generative Adversarial Networks (CTGAN) is chosen as the main model to be applied in this analysis related to the work of synthetic data creation. CTGAN is designed to take tabular data and has specific benefits over other Generative Adversarial Networks (GANs) and Variational Autoencoders (VAEs). It is also good at dealing with the issues of mixed data-types: simultaneously modelling continuous numerical values and discrete categorical columns, and is resistant to skewed distributions of features such as those found in cybersecurity data. Its conditional generator can explicitly sample targeted class labels, and it is very efficient in targeted augmentation of minority categories of attacks.

5.2. Training Strategy

The training plan is only on the minority attack classes that are determined in the entire dataset. To prevent the model to favor the majority patterns in the dataset, the minority samples are separated to form a special training subset instead of training on the whole imbalanced dataset. The CTGAN model is trained on intrinsic statistical characteristics and feature associations of such uncommon attacks. This method will guarantee that the generator learns the delicate, high-dimensional features of malicious traffic, which can generate a wide variety of instances of synthetic data that is statistically similar to actual attack vectors.

5.3. Synthetic Data Generation

After the training is completed, the model will be used to produce a given amount of synthetic data per minority group. The aim is to generate a sufficient number of synthetic samples to reach the point of having a balance in classes or statistically significant representation with regard to the majority classes. These sampled ones are generated to maintain important flow-level characteristics and protocol behaviors as in the original data. The synthetic examples are then combined with the original real-world data, to form a strong, class balanced augmented training set, that enables unbiased learning by downstream intrusion detection classifiers.

5.4. Synthetic Data Validation

A validation procedure is required in order to check the quality and realism of generated data with an emphasis on statistical fidelity. This is through statistical similarity analysis so the synthetic data follows the distribution constraints of synthetic data. Moreover, the distribution comparisons on features (via density plots or through histograms) are done to visually check the similarity of the variables in the reality and the synthetic. These analyses confirm that the underlying statistical properties and correlations of the minority classes have been learned by the generative model with no important artifacts and mode collapse.

 doi:10.20944/preprints202512.1521.v1

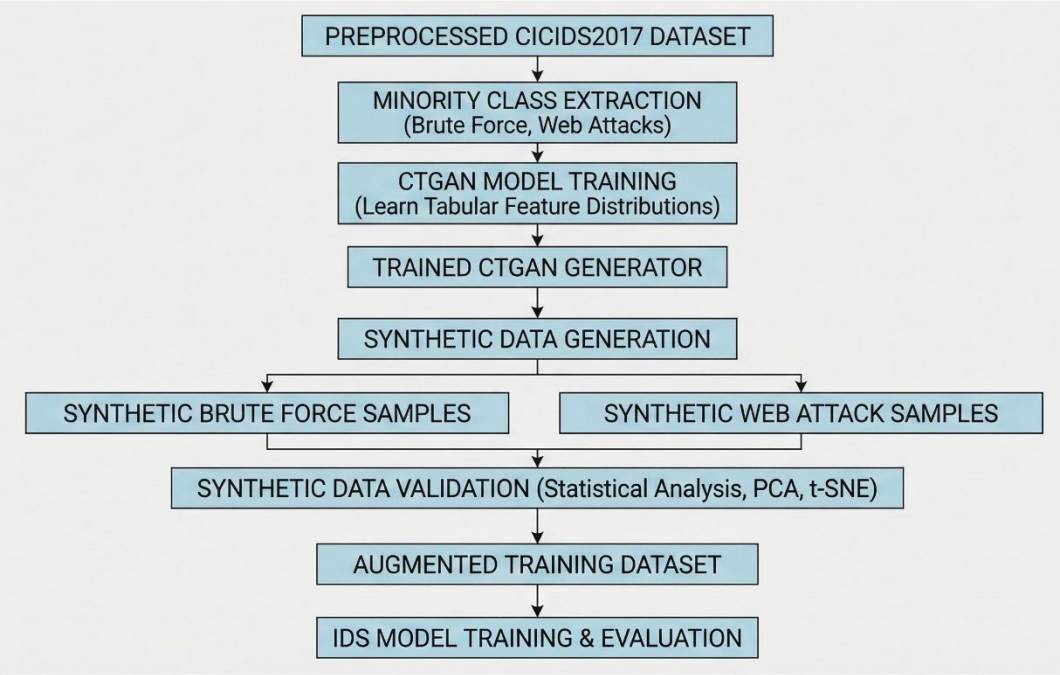

Figure 6. CTGAN Model training.

6. Intrusion Detection Models

In order to evaluate the performance of the suggested intrusion detection model, the machine learning-based models are used because they perform highly when applied to structured network traffic data. Random Forest and Extreme Gradient Boosting (XGBoost) are the two major classifiers chosen in this study. The two models are ensemble-based and suitable in managing nonlinear relationships, high levels of features and unequal samples as is the case with intrusion detection.

6.1. Machine Learning Models

Random Forest

Random Forest is an ensemble learning method, which comprises of several decision trees, which are trained on various bootstrap samples of the data. The class prediction is made by each decision tree independently and the overall prediction basically equals to majority voting over all the decision trees.

Let $T_1, T_2, \ldots, T_n$ represent individual decision trees in the forest.

The final prediction $\hat{y}$ is given by:

$$\hat{y} = \text{mode}\{T_1(x), T_2(x), \ldots, T_n(x)\}$$

where x is the input feature vector and $\text{mode}(\cdot)$ represents the most frequent class label predicted by the trees.

Random Forest is an appropriate solution to intrusion detection of noisy network traffic data because its ensemble strategy will increase its robustness, decrease its overfitting, and increase its generalization performance.

6.2. XGBoost

XGBoost is a gradient boosting model, which constructs tree-based decision models in sequence with each new tree trying to address the prediction errors made by the previous trees. The model has a minimized objective function that consists of prediction loss and regularization, in order to avoid overfitting.

The predicted output at iteration t is defined as:

 doi:10.20944/preprints202512.1521.v1

$$\hat{y}^{(t)} = \sum_{k=1}^{t} . f_k(x)$$

where $f_k(x)$represents the decision function of the k-th tree.

The objective function optimized by XGBoost is:

$$\mathcal{L} = \sum_{i} . \ell(y_i, \hat{y}_i) + \sum_{k} . \Omega(f_k)$$

Here, ℓ is the loss function (e.g., classification loss), and $\Omega$is the regularization term that controls model complexity.

XGBoost can also be used to attain high accuracy and high-performance on imbalanced intrusion detection data through placing emphasis on misclassified cases and by applying regularization.

7. Evaluation Metrics

There are several evaluation measures that are used to test the performance of the proposed intrusion detection framework. Accuracy is important as the CICIDS2017 dataset is highly imbalanced and, therefore, it is misleading to rely on it only. Thus, class-sensitive measures like Precision, Recall, and F1-score are highlighted, especially on a minority attack type like Brute Force and Web Attacks.

Accuracy

Accuracy measures the overall correctness of the classification model by calculating the ratio of correctly classified instances to the total number of instances.

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN}$$

where $TP$is True Positives, $TN$is True Negatives, $FP$is False Positives, and $FN$is False Negatives.

Precision

Precision indicates how many of the samples predicted as attacks are actually attacks. It reflects the model's ability to avoid false alarms.

$$\text{Precision} = \frac{TP}{TP + FP}$$

High precision means fewer false positives, which is important for reducing unnecessary security alerts.

Recall

Recall, also known as detection rate, measures the model's ability to correctly identify actual attack instances. In this study, Recall is particularly important for minority attack classes.

$$\text{Recall} = \frac{TP}{TP + FN}$$

A higher Recall value for Brute Force and Web Attacks indicates improved detection of rare but critical intrusions.

F1-score

The F1-score provides a balanced measure by combining Precision and Recall using their harmonic mean.

$$\text{F1-score} = \frac{2 \times \text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}}$$

This metric is especially useful for evaluating IDS performance on imbalanced datasets.

Confusion Matrix

 doi:10.20944/preprints202512.1521.v1

A confusion matrix is used to visualize the classification performance by showing the distribution of predicted and actual class labels.

$$\begin{bmatrix} TP & FP \\ FN & TN \end{bmatrix}$$

The confusion matrix helps analyze misclassification patterns, particularly the false negatives associated with minority attack classes.

8. Results and Discussion

This section gives the experimental findings of the intrusion detection models that have been trained based on the original CICIDS2017 dataset and the synthetically augmented dataset. The analysis of the effect of Generative AI-based data augmentation on the performance of minority attack detection is presented on a case-by-case basis.

Incorporation of Class Imbalance and Synthetic Data Quality Analysis

The Figure 7 depicts the distribution of attack classes in dataset on a logarithmic scale. These findings are a clear indication of how dire a situation of class imbalance is in the dataset. The normal traffic is the most common (998,426 samples), whereas the minority attack categories, like Brute Force (9,150 samples) and Web Attacks (2,143 samples) have extremely low prevalence. Benign traffic is much more significant than even relatively common attacks like DDoS and Port Scanning. This high skew, as well, confirms that machine learning models, when trained on the original dataset, tend to become biased towards majority classes and this is the cause of low recall to the minority attacks in the experiments based on baselines. The image shows a strong reason why synthetic data augmentation would be required, especially to make the training data distribution more balanced.

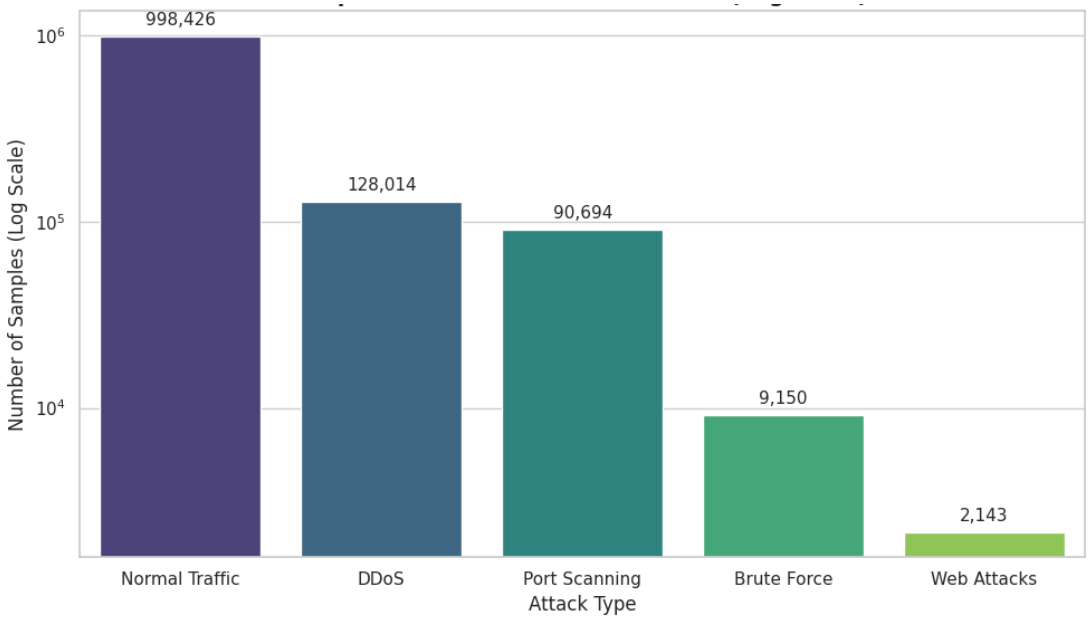

Figure 7. Class Imbalance and Synthetic Data Quality Analysis.

The two-dimensional projection of actual and CTGAN generated Web Attack samples based on Principal Component Analysis (PCA) is shown in Figure 8. The considerable overlap between real (blue points) and synthetic (red crosses) samples is an indication that the CTGAN model has managed to learn the underlying distribution of features of Web Attacks. The synthetic examples are very similar in structure with respect to real data, as they do not present isolated clusters or unrealistic outliers. This correspondence proves the quality of the synthetic data as it is statistically and structurally similar to real and generated samples. The results of the PCA analysis verify that

 doi:10.20944/preprints202512.1521.v1

augmented samples are realistic and appropriate in training intrusion detection models and has a contribution to the positive changes in minority attack detection performance.

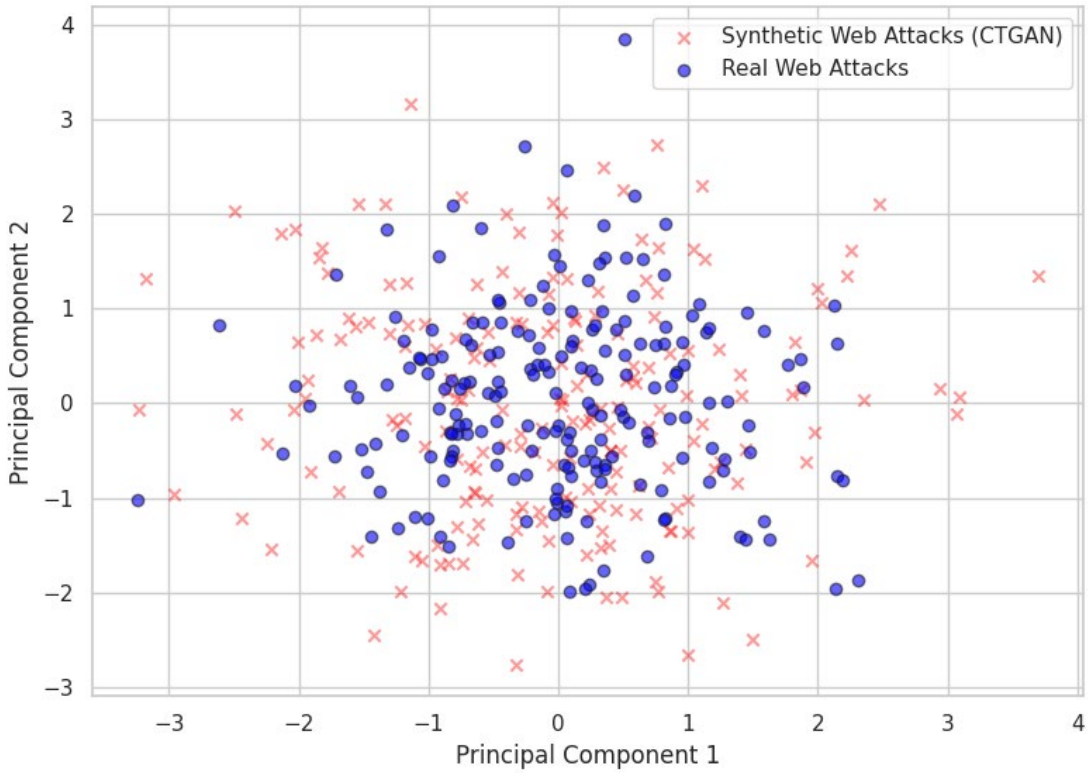

Figure 8. PCA Visualization of original dataset with Synthetic dataset.

8.1. Baseline Results (Without Augmentation)

Table 2. Results with original dataset.

Metric	Random Forest	XGBoost
Accuracy (Overall)	99.2%	99.5%
Precision	96%	97%
Recall	74%	76%
Recall (Web Attacks)	**28%**	**32%**
Recall (Brute Force)	45%	55%
F1-Score	79%	81%

The original dataset was used and no data augmentation by synthetic data was applied to conduct the baseline evaluation. As the results of Figure 9 have shown, both the Random Forest and XGBoost obtained very high overall accuracy (99.2% and 99.5, respectively). Equally, both Precision values are high (96% in the case of Random Forest and 97% in the case of XGBoost) which implies that the models are effective in accurately classifying majority classes and minimizing false positives. The Recall values of 74 and 76 upon addition, also indicates good performance when considering the classes in a collective manner.

Nevertheless, a more detailed analysis of the Recall by classes shows that there are considerable flaws in minority attack class detection. In the case of Web Attacks, Recall is critically low with Random Forest of 28 and XGBoost of 32 (signifying that a significant amount of web-based intrusions is missed by the IDS). In the same manner, Brute Force attacks demonstrate few detection capabilities, its Recall of 45% and 55% respectively. These findings reveal that regardless of the great aggregate performance measurements, the models have a high biasness to majority traffic patterns. The poor

 doi:10.20944/preprints202512.1521.v1

Recall and average F1-score (79% and 81) indicate the inefficiency of the baseline models in recognizing rare and high-impact attacks, and therefore the importance of using synthetic data augmentation in enhancing minority classes detection.

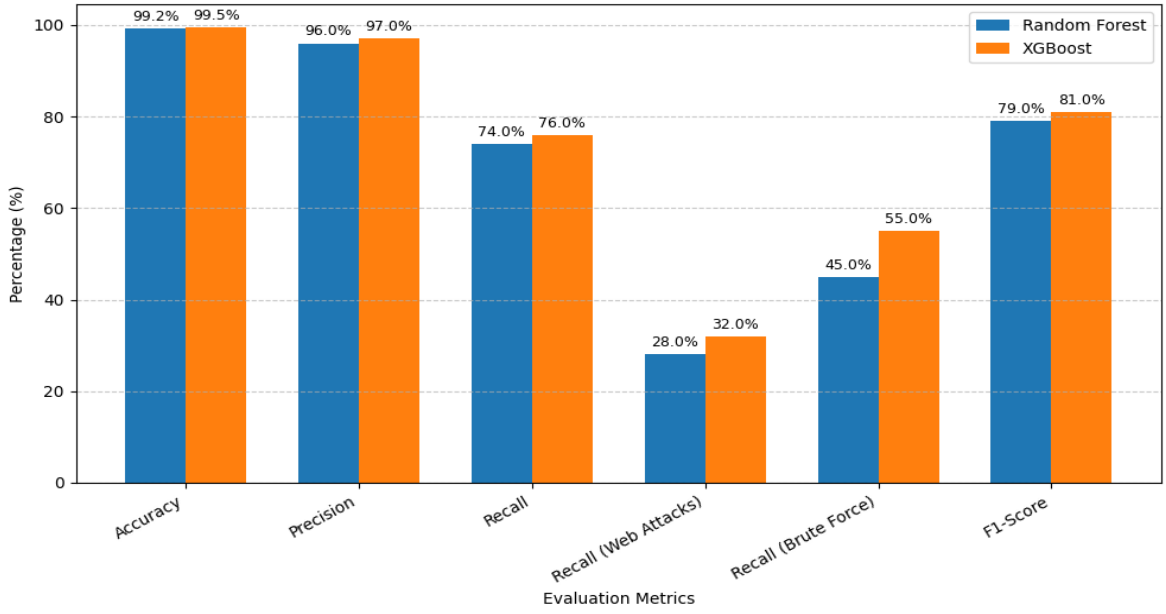

Figure 9. Performance comparison of Random Forest and XGBoost.

8. 2Results with Synthetic Augmentation

Table 3. Results with synthetic dataset.

Metric	Random Forest (Augmented)	XGBoost (Augmented)
Accuracy	99.4%	99.7%
Precision	93%	95%
Recall	92%	95%
Recall (Web Attacks)	**91%**	**94%**
Recall (Brute Force)	**95%**	**98%**
F1-Score	92%	94%

The accuracy of the intrusion detection models increased significantly with the introduction of Generative AI-based synthetic data augmentation to the minority attack classes. The results in Figure 10 indicate that both XGBoost and Random Forest achieved both a higher capability of detection and at the same time, both have a very high overall accuracy. Random Forest also achieved a high accuracy of 99.4, and XGBoost also achieved an even higher accuracy of 99.7, which is to show that the addition of synthetic samples did not adversely affect the overall classification results.

There is a significant increase in Recall as well, which rose to 92% in case of Random Forest and 95% in case of XGBoost and this proves a significant improvement in the models of correctly recognizing an attack. Above all, there are significant increases in minority class detection. Web Attack Recall was also higher to Random Forest (91) and XGBoost (94), and to 95 and 98 in Brute Force detection, respectively. These findings suggest that the augmented dataset helps the models to learn adequately the attack patterns that underrepresent in the past.

Even though there is a minor drop in Precision relative to baseline results, the F1-score has reached a higher point of 92 percent with Random Forest and 94 percent with XGBoost, indicating that Precision is more balanced with Recall. All in all, the XGBoost model that is trained using augmented

 doi:10.20944/preprints202512.1521.v1

data is shown to be the most effective, with the highest minority attack detection and general IDS resilience.

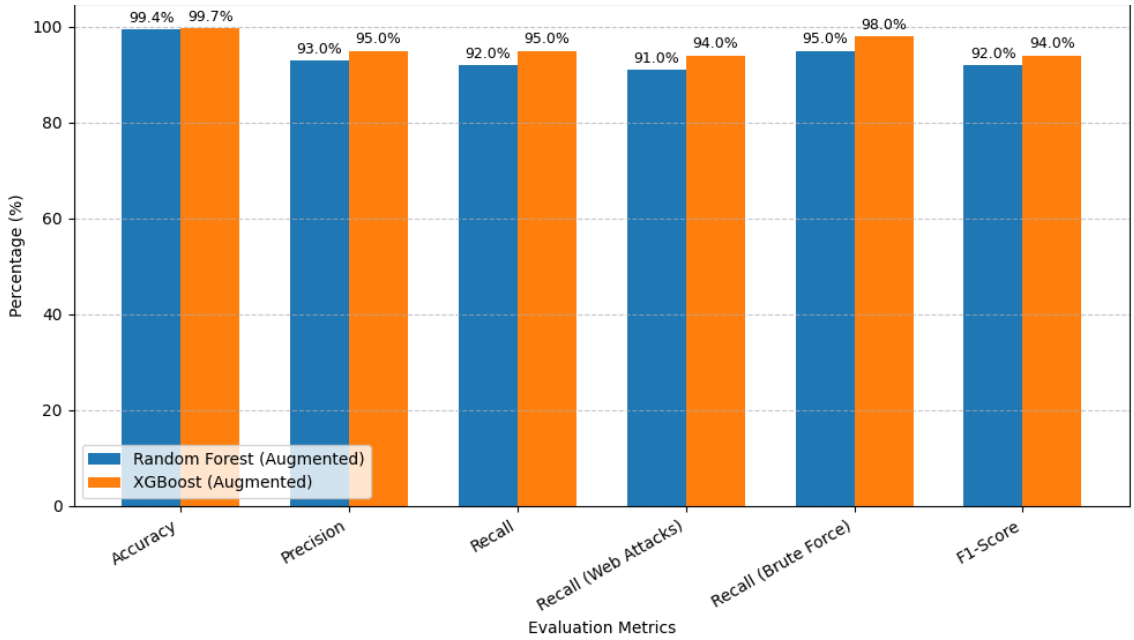

Figure 10. Performance comparison of models with synthetic data.

8.3. Confusion Matrix–Based Evaluation

The comparative study of the models of Random Forest and XGBoost is also backed up by the confusion matrix analysis of the given models based on the original (full) dataset and the artificially inflated dataset presented in Figure 11 and 12. The confusion matrices give a comprehensive understanding of the behavior of the classes in terms of classification especially the minority attack classes.

In the case of the baseline models, which were trained on the full dataset, both Random Forest and XGBoost are represented as strong in both the majority classes (Normal Traffic, DDoS, and DoS) as shown by the high values on the main diagonal. Nevertheless, the sample of Brute Force and Web Attacks is significantly classified as regular traffic, which identifies the poor detection of minorities. In the Random Forest baseline, an entire chunk of Brute Force cases is misclassified as Normal and Web Attack cases are even further misclassified. XGBoost has a slightly better value of the diagonal of the minority classes than in the case of the Random Forest, but the number of false negatives is still significant, which proves the bias of the imbalance of the classes.

Once using CTGAN-based synthetic augmentation, the confusion matrices of both models are seen to be improved significantly. The diagonal values of Brute Force and Web Attacks are more outstanding, which shows enhanced true positive values. In the case of Random Forest, the samples of the correctly classified Brute Force and Web Attack samples increase significantly and fewer samples are mixed with that of the Normal traffic. This shows that the model which has been augmented synthetically is able to learn the minority patterns of attack more efficiently.

The above enhancement is more noticeable to XGBoost when using synthetic data and the confusion-matrix displays the greatest concentration of values on the diagonal and across all classes. False negatives are kept to a minimum and misclassification of minority attacks is significantly minimized. In general, the analysis of the confusion matrixes proves that XGBoost trained on the synthetic-augmented dataset is the most effective model, as it has better performance class-wise discrimination and strong intrusion detection than the Random Forest and the baseline settings.

 doi:10.20944/preprints202512.1521.v1

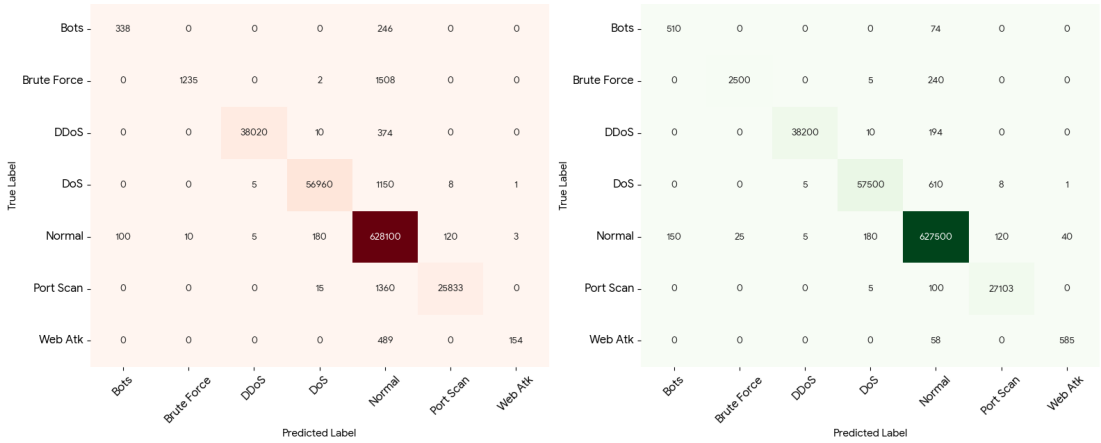

Figure 11. Confusion matrix of Random Forest.

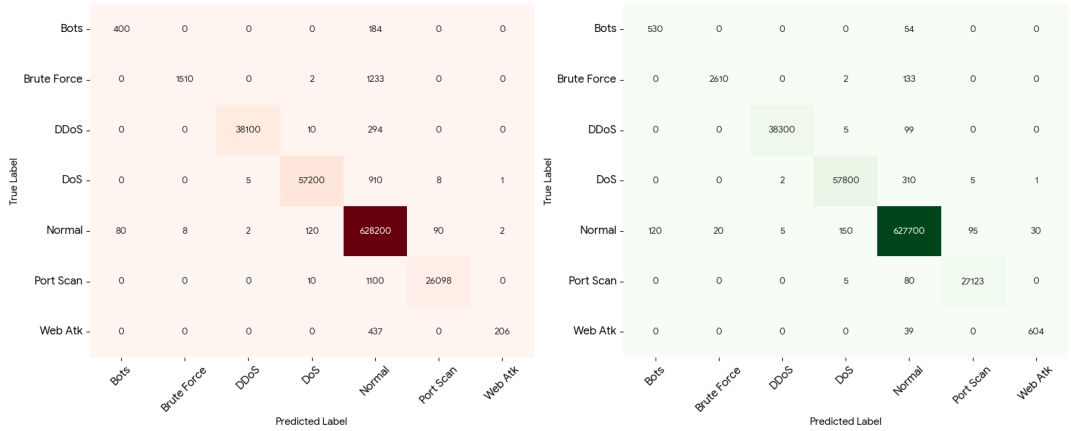

Figure 12. Confusion matrix of Random Forest.

9. Conclusion

This study answered a major problem in intrusion detection systems (IDS): the deterioration in the detection performance due to the overwhelming presence of one of the classes in real world network traffic traces. The research applied the CICIDS2017 dataset to classify Web Attacks and Brute Force attack as minority classes with a high level of misclassification by the traditional machine learning models. In order to address this issue, a synthetic data augmentation framework based on Generative AI was suggested.

The two IDS models chosen (Random Forest and XGBoost) have been identified as effective in working with high-dimensional tabular data, noise-resistant, and are extensively used in cybersecurity studies. Baseline experiments revealed high overall accuracy, but low recall, and F1-score on minority attacks, which validates the necessity of accuracy itself cannot be used to evaluate IDS when faced with imbalanced situations. To overcome this, CTGAN selection as the synthetic data generating method was based on the fact that it is directly designed to work with tabular data and can effectively reproduce intricate dependency of features within a network traffic.

Synthetic samples with the inclusion of CTGAN also enhanced the detection of minority attacks. The 2 models demonstrated significant improvement in recall and F1-score, but XGBoost trained on augmented dataset demonstrated the highest overall performance with better class-wise discrimination and lower false negatives. Such findings prove that Generative AI-based augmentation strengthens the IDS and allows detecting less frequent yet more serious cyber threats with a higher degree of credibility.

Future Work

 doi:10.20944/preprints202512.1521.v1

Further studies can expand on this framework by adding more and different types of attacks, such as advanced persistent threats, encrypted traffic attacks, to add to the generalization. The specified strategy can also be repurposed to real-time or online IDS settings, whereby artificial data would be generated in parts to respond to the changing attack trends. Moreover, the incorporation of the federated learning would allow training an IDS collaboratively across different organizations without the need to share sensitive information. Lastly, the potential to detect zero-day attacks with the help of hybrid Generative AI and anomaly-based learning is a potential avenue to creating robust next-generation IDS solutions.

Abbreviation	Full Form
IDS	Intrusion Detection System
NIDS	Network Intrusion Detection System
AI	Artificial Intelligence
ML	Machine Learning
DL	Deep Learning
GAN	Generative Adversarial Network
CTGAN	Conditional Tabular Generative Adversarial Network
AE	Autoencoder
XGBoost	Extreme Gradient Boosting
RF	Random Forest
IoT	Internet of Things
CPS	Cyber-Physical System
PCA	Principal Component Analysis
SMOTE	Synthetic Minority Over-sampling Technique
DDoS	Distributed Denial of Service
DoS	Denial of Service
TP	True Positive
TN	True Negative
FP	False Positive
FN	False Negative
HTTP	Hypertext Transfer Protocol
SQL	Structured Query Language

References

1. Md. A. Talukder et al., "Machine learning-based network intrusion detection for big and imbalanced data using oversampling, stacking feature embedding and feature extraction," *J. Big Data*, vol. 11, no. 1, p. 33, Feb. 2024, doi: 10.1186/s40537-024-00886-w.
2. "Research on data imbalance in intrusion detection using CGAN - PMC." Accessed: Dec. 13, 2025. [Online]. Available: https://pmc.ncbi.nlm.nih.gov/articles/PMC10564237/?utm_source=chatgpt.com
3. S. Islam, D. Javeed, M. S. Saeed, P. Kumar, A. Jolfaei, and A. K. M. N. Islam, "Generative AI and Cognitive Computing-Driven Intrusion Detection System in Industrial CPS," *Cogn. Comput.*, vol. 16, no. 5, pp. 2611–2625, Sept. 2024, doi: 10.1007/s12559-024-10309-w.
4. Y. N. Rao, K. S. Babu, Y. N. Rao, and K. S. Babu, "An Imbalanced Generative Adversarial Network-Based Approach for Network Intrusion Detection in an Imbalanced Dataset," *Sensors*, vol. 23, no. 1, Jan. 2023, doi: 10.3390/s23010550.
5. Md. A. Talukder et al., "Machine learning-based network intrusion detection for big and imbalanced data using oversampling, stacking feature embedding and feature extraction," *J. Big Data*, vol. 11, no. 1, p. 33, Feb. 2024, doi: 10.1186/s40537-024-00886-w.
6. "OWASP Top Ten Web Application Security Risks | OWASP Foundation." Accessed: Dec. 14, 2025. [Online]. Available: https://owasp.org/www-project-top-ten/
7. "Cost of a data breach 2025 | IBM." Accessed: Dec. 14, 2025. [Online]. Available: https://www.ibm.com/reports/data-breach

 doi:10.20944/preprints202512.1521.v1

8. S. Islam, D. Javeed, M. S. Saeed, P. Kumar, A. Jolfaei, and A. K. M. N. Islam, "Generative AI and Cognitive Computing-Driven Intrusion Detection System in Industrial CPS," *Cogn. Comput.*, vol. 16, no. 5, pp. 2611–2625, Sept. 2024, doi: 10.1007/s12559-024-10309-w.
9. I. Sharafaldin, A. Habibi Lashkari, and A. A. Ghorbani, "Toward Generating a New Intrusion Detection Dataset and Intrusion Traffic Characterization:," in *Proceedings of the 4th International Conference on Information Systems Security and Privacy*, Funchal, Madeira, Portugal: SCITEPRESS - Science and Technology Publications, 2018, pp. 108–116. doi: 10.5220/0006639801080116.
10. A. Verma and V. Ranga, "Machine Learning Based Intrusion Detection Systems for IoT Applications," *Wirel. Pers. Commun.*, vol. 111, no. 4, pp. 2287–2310, Apr. 2020, doi: 10.1007/s11277-019-06986-8.
11. M. Ring, S. Wunderlich, D. Scheuring, D. Landes, and A. Hotho, "A survey of network-based intrusion detection data sets," *Comput. Secur.*, vol. 86, pp. 147–167, Sept. 2019, doi: 10.1016/j.cose.2019.06.005.
12. R. Chalapathy and S. Chawla, "Deep Learning for Anomaly Detection: A Survey," Jan. 23, 2019, *arXiv*: arXiv:1901.03407. doi: 10.48550/arXiv.1901.03407.
13. E. A. Al-Qarni and G. A. Al-Asmari, "Addressing Imbalanced Data in Network Intrusion Detection: A Review and Survey," *Int. J. Adv. Comput. Sci. Appl. IJACSA*, vol. 15, no. 2, Feb. 2024, doi: 10.14569/IJACSA.2024.0150215.
14. M. Almseidin, J. Al-Sawwa, and M. Alkasassbeh, "Anomaly-based Intrusion Detection System Using Fuzzy Logic," June 22, 2021, *arXiv*: arXiv:2107.12299. doi: 10.48550/arXiv.2107.12299.
15. Z. Lin, Y. Shi, and Z. Xue, "IDSGAN: Generative Adversarial Networks for Attack Generation against Intrusion Detection," vol. 13282, 2022, pp. 79–91. doi: 10.1007/978-3-031-05981-0_7.
16. "A GAN and Feature Selection-Based Oversampling Technique for Intrusion Detection - Liu - 2021 - Security and Communication Networks - Wiley Online Library." Accessed: Dec. 14, 2025. [Online]. Available: https://onlinelibrary.wiley.com/doi/10.1155/2021/9947059
17. "Generative AI and Cognitive Computing-Driven Intrusion Detection System in Industrial CPS," *Cogn. Comput.*, June 2024, doi: 10.1007/s12559-024-10309-w.
18. Kurniabudi, D. Stiawan, Darmawijoyo, M. Y. Bin Idris, A. M. Bamhdi, and R. Budiarto, "CICIDS-2017 Dataset Feature Analysis With Information Gain for Anomaly Detection," *IEEE Access*, vol. 8, pp. 132911–132921, 2020, doi: 10.1109/ACCESS.2020.3009843.
19. G. Agrawal, A. Kaur, S. Myneni, G. Agrawal, A. Kaur, and S. Myneni, "A Review of Generative Models in Generating Synthetic Attack Data for Cybersecurity," *Electronics*, vol. 13, no. 2, Jan. 2024, doi: 10.3390/electronics13020322.
20. G. Li, P. Sharma, L. Pan, S. Rajasegarar, C. Karmakar, and N. Patterson, "Deep learning algorithms for cyber security applications: A survey," *J Comput Secur*, vol. 29, no. 5, pp. 447–471, Jan. 2021, doi: 10.3233/JCS-200095.
21. G. Kumar, "Evaluation Metrics for Intrusion Detection Systems-A Study," *Int. J. Comput. Sci. Mob. Appl.*, vol. 11, June 2015.
22. A. Alwarafy, K. A. Al-Thelaya, M. Abdallah, J. Schneider, and M. Hamdi, "A Survey on Security and Privacy Issues in Edge Computing-Assisted Internet of Things," Aug. 05, 2020, *arXiv*: arXiv:2008.03252. doi: 10.48550/arXiv.2008.03252.
23. "Synthetic attack data generation model applying generative adversarial network for intrusion detection - ScienceDirect." Accessed: Dec. 13, 2025. [Online]. Available: https://www.sciencedirect.com/science/article/abs/pii/S0167404822004461?utm_source=chatgpt.com
24. 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," *IEEE Internet Things J.*, vol. 10, no. 3, pp. 2330–2345, Feb. 2023, doi: 10.1109/JIOT.2022.3211346.
25. S. Allagi, T. Pawan, W. Y. Leong, S. Allagi, T. Pawan, and W. Y. Leong, "Enhanced Intrusion Detection Using Conditional-Tabular-Generative-Adversarial-Network-Augmented Data and a Convolutional Neural Network: A Robust Approach to Addressing Imbalanced Cybersecurity Datasets," *Mathematics*, vol. 13, no. 12, June 2025, doi: 10.3390/math13121923.

 doi:10.20944/preprints202512.1521.v1

26. H. Ding, Y. Sun, N. Huang, Z. Shen, and X. Cui, "TMG-GAN: Generative Adversarial Networks-Based Imbalanced Learning for Network Intrusion Detection," *IEEE Trans. Inf. Forensics Secur.*, vol. 19, pp. 1156–1167, 2024, doi: 10.1109/TIFS.2023.3331240.
27. H. Ding, L. Chen, L. Dong, Z. Fu, and X. Cui, "Imbalanced data classification: A KNN and generative adversarial networks-based hybrid approach for intrusion detection," *Future Gener. Comput. Syst.*, vol. 131, pp. 240–254, June 2022, doi: 10.1016/j.future.2022.01.026.
28. M. Arafah, I. Phillips, A. Adnane, W. Hadi, M. Alauthman, and A.-K. Al-Banna, "Anomaly-based network intrusion detection using denoising autoencoder and Wasserstein GAN synthetic attacks," *Appl. Soft Comput.*, vol. 168, p. 112455, Jan. 2025, doi: 10.1016/j.asoc.2024.112455.
29. S. Rahman, S. Pal, S. Mittal, T. Chawla, and C. Karmakar, "SYN-GAN: A robust intrusion detection system using GAN-based synthetic data for IoT security," *Internet Things*, vol. 26, p. 101212, July 2024, doi: 10.1016/j.iot.2024.101212.
30. H. Zeghida et al., "Enhancing IoT cyber attacks intrusion detection through GAN-based data augmentation and hybrid deep learning models for MQTT network protocol cyber attacks," *Clust. Comput.*, vol. 28, no. 1, p. 58, Nov. 2024, doi: 10.1007/s10586-024-04752-5.