Journal of Future Artificial Intelligence and Technologies

E-ISSN: 3048-3719

Research Article

Leveraging GANs for Synthetic Data Generation to Improve Intrusion Detection Systems

Md Abdur Rahman [1,2,*], Guillermo A. Francia [3], and Hossain Shahriar [3]

[1] Department of Mathematics, Jahangirnagar University, Savar, Dhaka, Bangladesh; e-mail : marahmanju@juniv.edu
[2] PhD Student, Dept. of Intelligent Systems and Robotics, University of West Florida, Florida, United States; e-mail : mr252@students.uwf.edu
[3] Center for Cybersecurity, University of West Florida Pensacola, Florida, United States; e-mail : gfranciaiii@uwf.edu; hshahriar@uwf.edu
* Corresponding Author : Md Abdur Rahman

Abstract: This research presents a hybrid intrusion detection approach that integrates Generative Adversarial Networks (GANs) for synthetic data generation with Random Forest (RF) as the primary classifier. The study aims to improve detection performance in cybersecurity applications by enhancing dataset diversity and addressing challenges in traditional models, particularly in detecting minority attack classes often underrepresented in real-world datasets. The proposed method employs GANs to generate synthetic attack samples that mimic real-world intrusions, which are then combined with real data from the UNSW-NB15 dataset to create a more balanced training set. By leveraging synthetic data augmentation, our approach mitigates issues related to class imbalance and enhances the generalization capability of the classifier. Extensive experiments demonstrate that RF trained on the combined dataset of real and synthetic data achieves superior detection performance compared to models trained exclusively on real data. Specifically, RF trained solely on the original dataset achieves an accuracy of 97.58%, whereas integrating GAN-generated synthetic data improves accuracy to 98.27%. The proposed methodology is further evaluated through comparative analysis against alternative classifiers, including Support Vector Machine (SVM), XGBoost, Gated Recurrent Unit (GRU), and related studies in the field. Our findings indicate that GAN-augmented training significantly enhances detection rates, particularly for rare attack types, while maintaining computational efficiency. Furthermore, RF outperforms other classifiers, including deep learning models, demonstrating its effectiveness as a lightweight yet robust classification method. Integrating GANs with RF offers a scalable and adaptable framework for intrusion detection, ensuring improved resilience against evolving cyber threats.

Keywords: Adversarial Learning; Anomaly Detection; Cybersecurity; Cyber Threat Detection; Intrusion Detection System; Random Forest; Synthetic Data Generation.

Received: November, 24th 2024
Revised: February, 24th 2025
Accepted: February, 25th 2025
Published: February, 28th 2025
Curr. Ver.: February, 28th 2025

Copyright: © 2025 by the authors. Submitted for possible open access publication under the terms and conditions of the Creative Commons Attribution (CC BY SA) license (https://creativecommons.org/licenses/by-sa/4.0/)

1. Introduction

Cybersecurity has become a critical concern in today's interconnected world, where technological advancements are rapidly evolving. As the digital landscape expands, so does the potential for cyber threats, making it imperative to advance cybersecurity research[1]. This article aims to explore the primary objectives and necessities for conducting cybersecurity research, referencing relevant publications highlighting this field's significance and depth.

The advent of 5G technology has transformed access environments, establishing distributed networks for diverse data from sensors, computers, and IoT. Despite expanded processing capacities, the increased access points heighten vulnerability to cyber threats[2], [3]. The cybersecurity challenge of effectively detecting network threats drives extensive research in network intrusion detection systems (NIDSs). Initially, research focused on traditional models like decision trees (DTs) [4] and support vector machines (SVMs) [5], evolving to advanced deep learning approaches CNNs, LSTMs, and autoencoders[6], [7]. Despite

remarkable anomaly detection, deploying these methodologies in real systems encounters limitations. This strategy bridges the gap between effective anomaly detection and practical applicability, contributing to the ongoing evolution of AI-based intrusion detection systems and addressing the need for robust solutions in the rapidly advancing network security landscape.

The prevailing issue of data imbalance in network intrusion detection systems (NIDSs) poses a formidable challenge, as the bulk of network flow data constitutes normal traffic, with rare occurrences of malicious behavior that can disrupt services. Specific types of attacks within malicious behavior are exceptionally rare, limiting the learning capacity of AI models in NIDS and leaving network systems susceptible to poor detection performance.

One of the core objectives of cybersecurity research is to identify, analyze, and develop methods to detect and prevent emerging cyber threats[8]. One research work delves into machine learning algorithms' applications in threat detection [9]. Another research paper investigating the application of machine learning algorithms in threat detection states that cybersecurity research aims to assess weaknesses in systems and networks to ensure timely and effective vulnerability management[9], [10]. Another paper provides insights into various network vulnerabilities and mitigation strategies. Research efforts are directed toward improving incident response and recovery mechanisms to minimize the impact of cyber-attacks[11], [12]. Also, Bejtlich et al. [13] offer a comprehensive guide to incident response procedures.

In the past, researchers have investigated a wide range of approaches to detect and mitigate cyberattacks effectively. These efforts have included the use of traditional machine learning[14]–[19], techniques as well as more advanced methodologies, such as deep learning [20]–[22]and quantum machine learning[23], [24], to address cybersecurity challenges. These methods have successfully identified vulnerabilities, detected threats, and prevented malicious activities in various contexts.

Recently, cybersecurity research has increasingly leveraged Large Language Models (LLMs) to enhance the effectiveness of threat detection solutions [25], [26]. LLMs have been shown to process and understand complex patterns in textual data, proving valuable in analyzing code, identifying vulnerabilities, and predicting cyber threats. Prior studies have fine-tuned pre-trained LLMs using supervised learning on task-specific labeled datasets, such as those from deepset, available via Hugging Face[27]. This methodology aligns with previous research, including Rahman et al. [25], which utilized multi-lingual BERT for dataset embedding and achieved a high accuracy of 96.55% using logistic regression. In that study, BERT embeddings significantly improved model performance by capturing nuanced patterns in the dataset.

In previous research[28], BERT has been applied to generate embeddings for code patterns within cybersecurity datasets. These embeddings were then used as inputs for various machine learning models to analyze their effectiveness in detecting cybersecurity vulnerabilities. Furthermore, supervised fine-tuning was employed on a pre-trained LLM, adapting it to specific tasks using labeled data from deepset on Hugging Face. This approach has enhanced the model's ability to understand and predict complex relationships within the data, pushing the boundaries of LLM-based cybersecurity solutions.

Generative Adversarial Networks (GANs) is one of the best approaches, which can significantly contribute to cybersecurity research by enhancing various aspects of threat detection, response, and defense mechanisms [29]. GANs consist of two competing neural networks: a generator, which produces synthetic data, and a discriminator, which evaluates the authenticity of the generated samples[30]. These models have been successfully applied to cybersecurity research, such as generating synthetic malware samples for better detection, simulating normal network behavior for anomaly detection, and conducting adversarial testing to identify weaknesses in IDS. GANs also facilitate malware transformation into different representations to improve classification accuracy. Additionally, they can create adversarial attacks to test the robustness of security systems and generate cybersecurity data visualizations that aid in understanding attack patterns. However, while GANs offer substantial benefits in cybersecurity, they can also be leveraged by malicious actors to generate sophisticated cyber threats, emphasizing the need for continuous evolution in defensive strategies[31].

While deep learning-based models have demonstrated strong performance in IDS applications, they often require substantial computational resources and large-scale labeled datasets. To mitigate these challenges, this study integrates Random Forest (RF) as the primary classifier due to its efficiency, interpretability, and ability to handle high-dimensional data with less risk of overfitting. RF is an ensemble learning method that constructs multiple decision

trees and aggregates their outputs, making it more robust to noise and data variations[32]. Furthermore, RF has shown superior performance in intrusion detection tasks compared to deep learning models when applied to structured datasets with diverse feature distributions. By combining RF with GAN-generated synthetic data, this study aims to enhance the model's detection performance by enriching training data with synthetic attack samples, thereby improving classification accuracy and generalization.

This research proposes an intrusion detection framework integrating GANs for synthetic data generation with RF as the primary classifier. The primary contributions of this study are: Developing a hybrid intrusion detection approach that leverages GANs for data augmentation and RF for classification, improving detection accuracy; Evaluating the impact of synthetic data on model performance by comparing RF trained on real data alone versus RF trained with a combination of real and synthetic data; Conducting a comparative analysis against state-of-the-art methods and other method, such as SVM, XGBoost, and GRU, to validate the effectiveness of the proposed GAN-RF approach. The rest of the paper is structured as follows: Section 2 presents the dataset, preprocessing steps, and the role of GANs in cybersecurity, including their loss functions. Section 3 outlines the proposed method for intrusion detection. Section 4 presents the results and discussion, analyzing performance and limitations. Finally, Section 5 concludes with key findings and future research directions.

2. Literature Review

2.1. Dataset

The University of New South Wales Network-Based 15 (UNSW-NB15) dataset comprises 49 features and 2,540,044 instances, making it a valuable resource for network security research. Each instance represents a snapshot of network activity over a specific time period, capturing both normal and malicious traffic. The dataset includes various types of attacks, such as denial-of-service (DoS), distributed denial-of-service (DDoS), probing, and malware, ensuring a comprehensive representation of cyber threats. Simulating realistic network environments provides an essential benchmark for evaluating the performance of intrusion detection systems (IDSs) performance in practical settings[33].

A notable characteristic of the UNSW-NB15 dataset is its extensive annotation, where each network flow is labeled according to its traffic type, distinguishing between normal and malicious activities. This structured labeling facilitates the training and validation of machine learning models, allowing for effective cybersecurity research. Furthermore, given the substantial network traffic volume, the dataset is a robust benchmark for assessing detection mechanisms, particularly in high-traffic scenarios where scalability and efficiency are crucial.

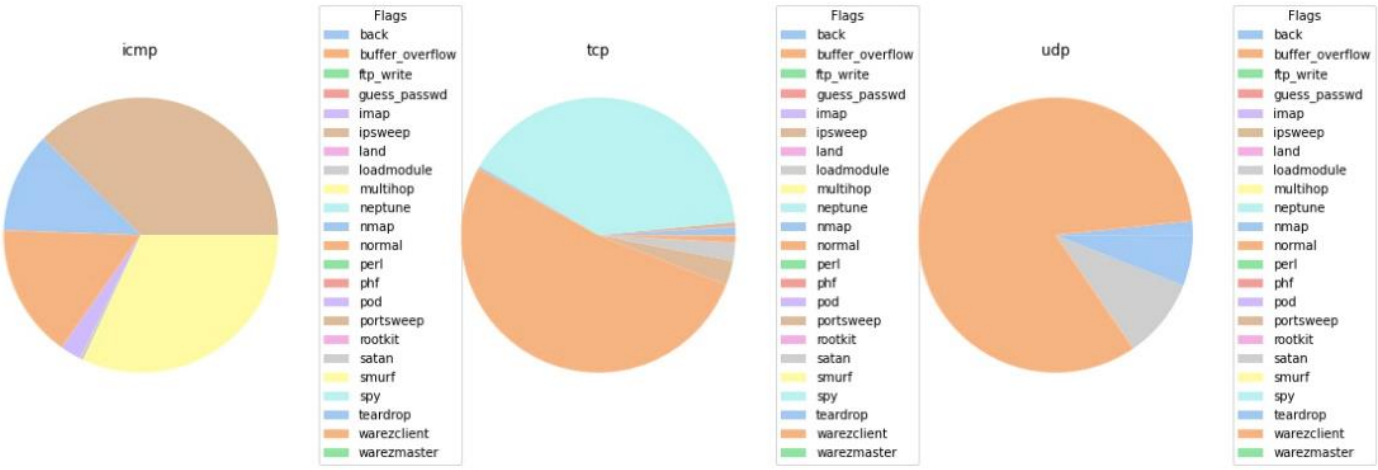

Figure 1. Pie chart for attack by protocol type: network traffic analysis protocol is used to create initial buckets to categorize the data.

Another advantage of this dataset is its protocol diversity, covering multiple communication protocols, including TCP, UDP, and ICMP, enhancing intrusion detection models' generalizability. Given its comprehensive nature, the UNSW-NB15 dataset has been widely

adopted as a benchmark for testing and evaluating various machine learning-based cybersecurity solutions. Figure 1 presents a pie chart illustrating the distribution of attack types based on network traffic protocols, providing an initial categorization framework to aid in identifying and mitigating network security threats.

Beyond its role in intrusion detection, the UNSW-NB15 dataset offers insights into various security threats, deployment strategies, and verification methodologies. It includes attributes related to Internet of Things (IoT) devices, reflecting the increasing integration of IoT within network environments. IoT-specific data is particularly important, as IoT devices introduce unique security challenges that require specialized threat mitigation strategies. By leveraging these critical features, the UNSW-NB15 dataset enables researchers and practitioners to develop more effective vulnerability detection mechanisms, address emerging cybersecurity threats, and strengthen the overall security posture of networked systems.

2.2. Loss Function of GANs

Network intrusion detection systems (NIDS) currently employ various machine learning algorithms that achieve high performance. However, they often face challenges in classifying rare attack classes in large-scale traffic scenarios, and their practical effectiveness against diverse cyber threats remains uncertain[34]. Previous studies have compared the precision and accuracy of different machine learning algorithms on the NSL-KDD dataset, varying in size (full, half, and quarter)[35]. An IoT-tailored IDS research framework has been proposed to address IoT-related challenges, considering the constraints of applying existing techniques to security threats and detection methods[36]. Traditional NIDS research, essential for security in conventional and IoT networks, has often focused on small datasets and classical machine learning approaches.

Recent approaches have leveraged larger datasets incorporating recent cyber threats and deep learning methodologies to overcome these limitations and enhance classification capabilities. Using high-volume traffic datasets, deep learning-based NIDS has demonstrated robust performance against various attack types in real-world scenarios, emphasizing practical applicability. This paper contributes to NIDS research by utilizing extensive datasets and state-of-the-art deep learning techniques, effectively bridging the gap between theoretical advancements and real-world implementation. The objective is to significantly enhance the classification capabilities of NIDS, particularly in large-scale network traffic and evolving cyber threat landscapes.

GANs, introduced by Goodfellow et al. [30], have gained substantial recognition due to their broad applicability in various fields, including cybersecurity. As illustrated in Figure 2, the GAN framework consists of two primary components: a generator network (G) and a discriminator network (D). The generator (G) acts as a data synthesizer, receiving a random noise vector (z) as input, which follows a Gaussian or uniform distribution. Through a learned mapping function, the generator transforms z into synthetic samples, denoted as $G(z)$, which approximate the distribution of real data[37].

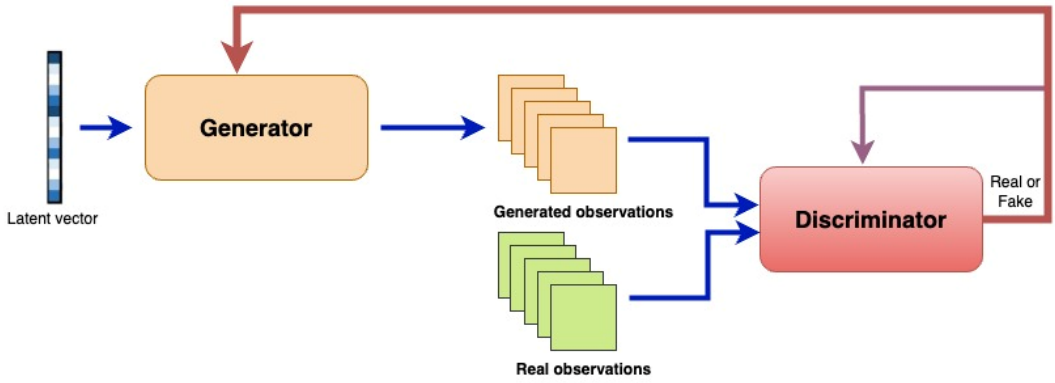

Figure 2. The architecture of a typical GAN model generates a synthetic dataset.

Conversely, the discriminator (D) serves as a binary classifier, distinguishing between real data samples (x) drawn from the true data distribution $p_{data}(x)$ and synthetic samples

generated by the generator $G(z)$. The discriminator estimates the probability that a given sample originates from real data rather than being artificially generated. The generator continuously improves its output to deceive the discriminator, while the discriminator refines its ability to distinguish real from synthetic data. This adversarial interaction drives the GAN to a dynamic equilibrium, where the generator produces highly realistic samples that closely match real data distribution. The training of GANs involves optimizing the following minimax objective function, as formulated in Equation (1).

$$\min_{G}\max_{D} \mathrm{V(D,G)} = \mathbb{E}_{x\sim p_{data}(x)}[\log D(x)] + \mathbb{E}_{2\sim p_z(z)}\left[\log\left(1 - D\left(G(z)\right)\right)\right] \quad (1)$$

Where x represents real observations drawn from $p_{data}(x)$, while z is a latent variable sampled from the prior distribution $p_z(z)$. During training, the generator and discriminator iteratively refine their respective models. The objective function follows a minimax optimization framework, wherein the discriminator maximizes its accuracy in distinguishing real from generated samples. Simultaneously, the generator attempts to minimize $\log\left(1 - D\left(G(z)\right)\right)$, effectively maximizing $D\left(G(z)\right)$ to deceive the discriminator. The discriminator employs cross-entropy loss to improve classification performance, ensuring that the overall objective function $\mathrm{V(D,G)}$ is optimized. During training, the generator is updated with a fixed discriminator to achieve stable convergence.

3. Proposed Method

This research proposes a hybrid methodology to improve intrusion detection performance by addressing the challenges posed by imbalanced cybersecurity datasets. The approach integrates Generative Adversarial Networks (GANs) for synthetic data generation with Random Forest (RF) as the primary classification model. GANs are employed to create synthetic attack samples, which are then combined with real data to improve class balance. This technique reduces data loss and overfitting risks, ensuring better generalization across different attack types.

Figure 2, introduced in Section 2.2, presents the general architecture of the GAN model, while Figure 3 illustrates the overall workflow of the proposed hybrid training process. In this approach, GAN-generated synthetic data is integrated into the RF training pipeline, directly comparing RF trained on real data and RF trained with augmented data to measure the impact of GAN-based oversampling.

The proposed GAN architecture consists of a generator and a discriminator:

- The generator comprises fully connected layers with LeakyReLU activation, generating synthetic attack samples that mimic real-world scenarios.
- The discriminator classifies real versus synthetic data using binary cross-entropy loss, ensuring that only high-quality synthetic samples contribute to dataset augmentation.
- Both networks are trained using the Adam optimizer (learning rate = 0.0002, $\beta1$=0.5) for stable convergence. The GAN model undergoes 10,000 epochs of training with a batch size of 64, ensuring the generation of reliable synthetic samples.
- The final dataset is normalized using min-max scaling to balance feature distributions. The dataset is then split into 70% training and 30% testing subsets to optimize the classification process.

Random Forest (RF) is selected as the primary classifier to assess the impact of synthetic data augmentation. The RF model is configured with 100 estimators, providing a robust and efficient baseline for evaluating the effectiveness of GAN-augmented data. RF was chosen due to its ability to handle high-dimensional data, its resilience against class imbalance, and its efficiency compared to deep learning methods.

Through extensive analysis of cybersecurity datasets, this approach ensures a more balanced data distribution by integrating GAN-generated synthetic samples into the training pipeline. The adversarial learning framework in GANs enhances data diversity, producing more realistic network traffic patterns. This augmentation strategy significantly improves dataset quality, optimizing feature representations before classification.

Integrating GANs with RF introduces a scalable and practical solution for handling imbalanced datasets, particularly in intrusion detection systems. By ensuring that minority attack classes receive adequate representation, the proposed approach enhances dataset preprocessing strategies for cybersecurity applications.

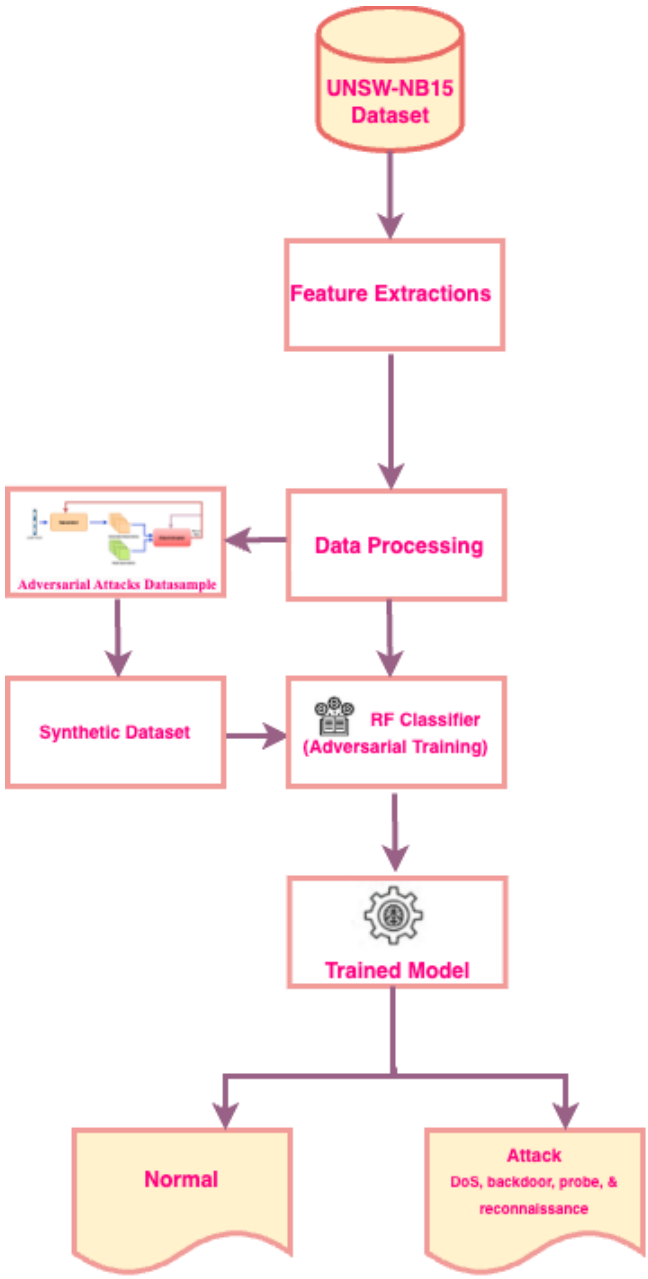

Figure 3. The flowchart illustrates steps to train the proposed model

4. Results and Discussion

The proposed model was implemented using Python with the pandas, TensorFlow, and Keras libraries on the Google Colab platform. The training process involved multiple GAN experiments, varying the number of epochs (100, 1000, and 10000) to analyze how the generator and discriminator loss evolved over time. For the initial 100 and 1000 epochs, training was conducted with a learning rate of 0.0004 and $\beta 1=0.5$, whereas in the final proposed model at 10,000 epochs, a more stable learning rate of 0.0002 was selected to improve convergence. These adjustments aimed to balance generator improvement while preventing mode collapse in longer training phases. Figure 4 presents these experiments, showcasing the learning behavior of GANs under different conditions.

Based on Figure 4, for the initial 100 epochs, the Generator and Discriminator show distinct learning patterns. The generator's loss starts at a high value and gradually decreases as training progresses. The minimum recorded loss for the Generator is 0.33, while the maximum reaches 14.45, indicating that the model is actively adjusting to synthesize realistic data. Meanwhile, the discriminator's loss stabilizes within the range of 0.12 to 2.6, suggesting that it effectively distinguishes between real and synthetic data.

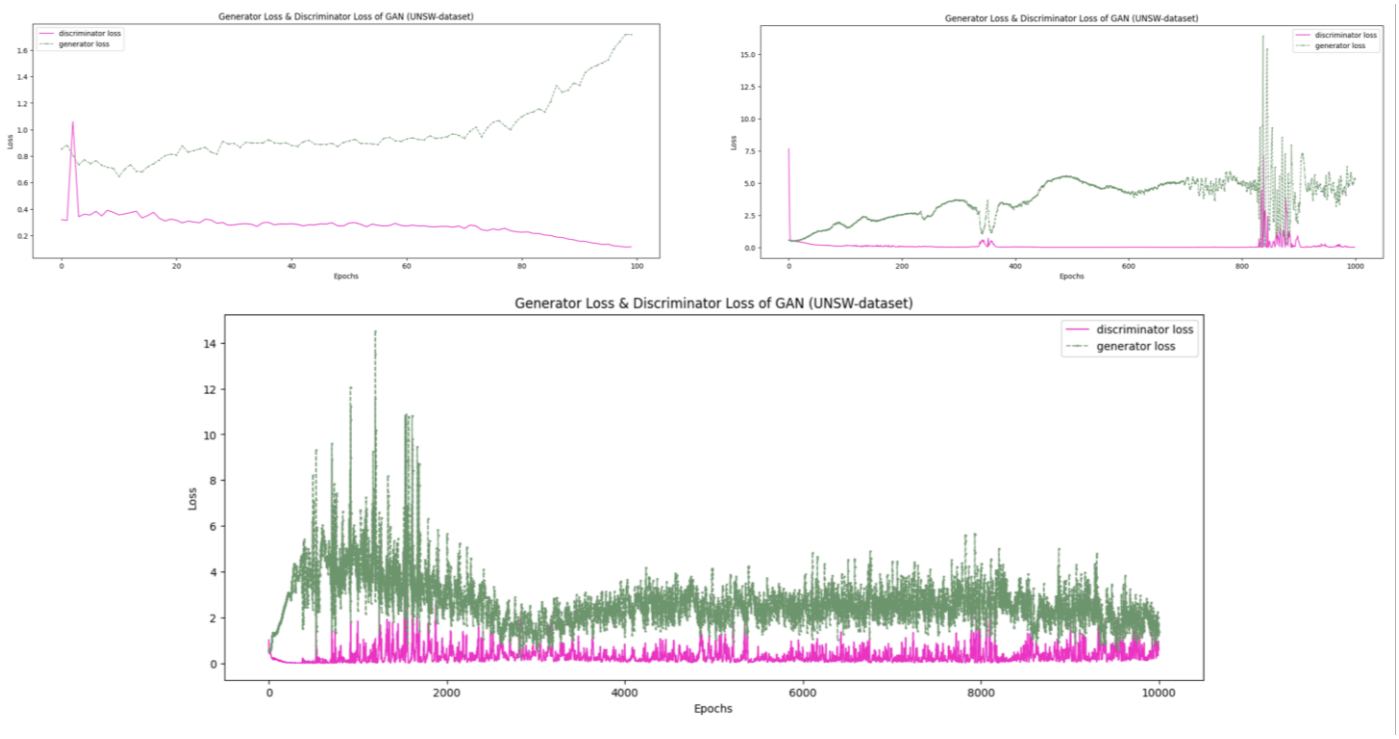

Figure 4. Generator and Discriminator Loss Progression Across Different Epochs for GAN Training on the UNSW-NB15 Dataset.

As training extends to 1000 epochs, the generator's loss gradually decreases, indicating an improvement in synthetic data generation. The discriminator's loss also evolves accordingly, reflecting a more refined capability in classifying real and generated data. The overall trend in Figure 4 suggests that, at this stage, the adversarial training reaches a more structured phase, where the Generator continuously learns to produce more realistic samples while the discriminator adapts to detect them.

By 10,000 epochs, the GAN model exhibits clearer signs of convergence. The Generator loss stabilizes at a minimum of 0.5, indicating that it has successfully learned to generate synthetic samples that closely resemble real data. Meanwhile, the discriminator's loss also stabilizes, implying that it consistently distinguishes between real and synthetic samples without overpowering the Generator. The narrower fluctuation range in the final stage suggests that the adversarial balance between both networks has been achieved, reducing the risk of mode collapse and improving the overall quality of synthetic data generation.

The loss variations recorded in Table 1 highlight the inherent complexity of the UNSW dataset, which presents a broader range of values for both Generator and Discriminator losses. The high variability in the generator's loss at the early training stages indicates that the dataset's diverse attack patterns pose a challenge in generating realistic synthetic samples. However, the gradual stabilization of loss values across different epochs reflects the model's ability to adapt to these complexities over time.

These findings confirm that longer training durations contribute to a more stable adversarial learning process, where the GAN achieves better equilibrium between the Generator and Discriminator. The improved convergence at 10,000 epochs ensures that the generated synthetic data enhances the diversity of the dataset while preserving realistic characteristics necessary for intrusion detection enhancement..

Table 1. Losses by GANs in dataset UNSW.

Type	Epoch	Min. loss	Max. loss
Generator	Real	0.33	14.45
Discriminator	Real	0.12	2.6

To evaluate the effectiveness of the proposed model (GAN+RF) in intrusion detection, a comparative analysis was conducted against the Support Vector Machine (SVM), XGBoost,

and Gated Recurrent Unit (GRU) models. The SVM model utilized an RBF kernel with a penalty parameter of C=1.0, while XGBoost was optimized with a learning rate of 0.1, a maximum depth of 6, and 100 boosting rounds. Additionally, a GRU-based model was included to assess the performance of deep learning methods in handling the UNSW-NB15 dataset. The results are summarized in Table 2.

The findings indicate that the proposed model substantially improves accuracy, particularly when using Random Forest (RF) with GAN-generated data augmentation. Specifically, the RF+GAN model achieved an accuracy of 98.27%, surpassing the previously reported 97.00% accuracy by Belouch et al. [38], using RF alone. Similarly, other conventional machine learning classifiers, such as SVM (92.28%) and Naïve Bayes (74.19%), yielded significantly lower performance, highlighting the limitations of these approaches in complex intrusion detection tasks.

Table 2. Comparison with others methods.

Reference	Method	Accuracy (%)	Dataset
Tang et al. [39]	DNN	75.75	NSL-KDD
Kim et al. [40]	LSTM	96.93	KDDCUP'99
Potluri et al. [41]	DL	97.50	NSL-KDD
Belouch et al. [38]	RF	97.00	UNSW-NB15
	DT	95.82	UNSW-NB15
	SVM	92.28	UNSW-NB15
	NB	74.19	UNSW-NB15
This work	SVM	92.48	UNSW-NB15
	SVM+ GANs	94.11	UNSW-NB15
	GRU	95.53	UNSW-NB15
	GRU+ GANs	96.21	UNSW-NB15
	XGBoost	96.04	UNSW-NB15
	XGBoost + GANs	96.78	UNSW-NB15
	RF	97.58	UNSW-NB15
	RF+ GANs	98.27	UNSW-NB15

The results further demonstrate the impact of synthetic data augmentation in enhancing classifier performance. When trained on the original UNSW-NB15 dataset, RF achieved an accuracy of 97.58%. However, upon integrating GAN-generated synthetic samples, the accuracy increased to 98.27%, reflecting the role of synthetic data in improving feature diversity and classifier robustness. A similar trend is observed for XGBoost, where model accuracy improved from 96.04% to 96.78% after incorporating synthetic data.

From a deep learning perspective, GRU-based models exhibited competitive performance, achieving 95.53% accuracy on the UNSW-NB15 dataset. However, when combined with GAN-generated data, the accuracy increased to 96.21%, reinforcing the role of synthetic augmentation in improving model generalization. The SVM+GAN model also improved over standard SVM (94.11% vs. 92.28%), though it remained less competitive than RF-based approaches.

These results suggest that integrating GAN-generated synthetic data improves the generalization ability of classifiers, particularly for complex datasets such as UNSW-NB15, which contains a diverse range of attack patterns. Moreover, the UNSW-NB15 dataset exhibited higher stability when generating synthetic data. In summary, the findings underscore the efficacy of the proposed RF+GAN approach in intrusion detection, demonstrating superior performance over existing state-of-the-art methods. The integration of synthetic data not only mitigates dataset imbalance but also enhances model robustness, ultimately leading to more reliable cyberattack detection. These results further emphasize the importance of adversarial learning in cybersecurity applications, where traditional classifiers may struggle with underrepresented attack patterns.

5. Conclusions

The research presented in this study focuses on advancing deep learning technologies for network intrusion detection, particularly addressing imbalanced data challenges. The proposed approach leverages RF for classifiers and GANs to generate synthetic data, aiming to enhance the detection performance of intrusion detection systems. The comparative analysis, as depicted in Table II, highlights the efficacy of the proposed model against existing state-of-the-art approaches. Most importantly, our RF model achieved an accuracy of 97.58% when trained on the existing UNSW-NB15 dataset. However, the true strength of our methodology emerges when combining synthetic and real datasets, resulting in a substantial accuracy improvement to 98.27%. This underscores the significance of addressing data imbalance through the integration of synthetic data, showcasing the model's adaptability and improved performance. In contrast to the existing models, our proposed approach consistently outperforms them, surpassing advanced algorithms. The robustness of the proposed model is particularly evident in handling the complex UNSW-NB15 dataset, where traditional machine-learning models exhibit limitations. In summary, the research demonstrates the effectiveness of the proposed RF model with a combined dataset in mitigating imbalanced data issues, thereby contributing significantly to the advancement of intrusion detection systems. The results affirm the potential of incorporating synthetic data in training deep learning models, presenting a promising avenue for future research in the realm of network security.

Author Contributions: Md Abdur Rahman contributed to conceptualization, implementation, and writing under the supervision of Dr. Guillermo A. Francia and Dr. Hossain Shahriar.

Funding: This research received no external funding**.**

Data Availability Statement: Dataset is popular and available in the web.

Acknowledgments: The authors would like to thank editors and reviewers for their support of their valuable feedback.

Conflicts of Interest: The authors declare no conflict of interest.

References

[1] J. P. Ntayagabiri, Y. Bentaleb, J. Ndikumagenge, and H. EL Makhtoum, "A Comprehensive Approach to Protocols and Security in Internet of Things Technology," *J. Comput. Theor. Appl.*, vol. 2, no. 3, pp. 324–341, Feb. 2024, doi: 10.62411/jcta.11660.

[2] R. Khan, P. Kumar, D. N. K. Jayakody, and M. Liyanage, "A Survey on Security and Privacy of 5G Technologies: Potential Solutions, Recent Advancements, and Future Directions," *IEEE Commun. Surv. Tutorials*, vol. 22, no. 1, pp. 196–248, 2020, doi: 10.1109/COMST.2019.2933899.

[3] J. Suomalainen, A. Juhola, S. Shahabuddin, A. Mammela, and I. Ahmad, "Machine Learning Threatens 5G Security," *IEEE Access*, vol. 8, pp. 190822–190842, 2020, doi: 10.1109/ACCESS.2020.3031966.

[4] S. L. Salzberg, "C4.5: Programs for Machine Learning by J. Ross Quinlan. Morgan Kaufmann Publishers, Inc., 1993," *Mach. Learn.*, vol. 16, no. 3, pp. 235–240, Sep. 1994, doi: 10.1007/BF00993309.

[5] N. Cristianini and J. Shawe-Taylor, *An Introduction to Support Vector Machines and Other Kernel-based Learning Methods*. Cambridge University Press, 2000. doi: 10.1017/CBO9780511801389.

[6] A. Kim, M. Park, and D. H. Lee, "AI-IDS: Application of Deep Learning to Real-Time Web Intrusion Detection," *IEEE Access*, vol. 8, pp. 70245–70261, 2020, doi: 10.1109/ACCESS.2020.2986882.

[7] N. Shone, T. N. Ngoc, V. D. Phai, and Q. Shi, "A Deep Learning Approach to Network Intrusion Detection," *IEEE Trans. Emerg. Top. Comput. Intell.*, vol. 2, no. 1, pp. 41–50, Feb. 2018, doi: 10.1109/TETCI.2017.2772792.

[8] M. D. Nguyen, M. T. Nguyen, T. C. Vu, T. M. Ta, Q. A. Tran, and D. T. Nguyen, "A Comprehensive Study on Applications of Blockchain in Wireless Sensor Networks for Security Purposes," *J. Comput. Theor. Appl.*, vol. 2, no. 1, pp. 102–117, Jul. 2024, doi: 10.62411/jcta.10486.

[9] M. S. Akhtar and T. Feng, "Evaluation of Machine Learning Algorithms for Malware Detection," *Sensors*, vol. 23, no. 2, p. 946, Jan. 2023, doi: 10.3390/s23020946.

[10] A. Handa, A. Sharma, and S. K. Shukla, "Machine learning in cybersecurity: A review," *WIREs Data Min. Knowl. Discov.*, vol. 9, no. 4, Jul. 2019, doi: 10.1002/widm.1306.

[11] S. Hore, F. Moomtaheen, A. Shah, and X. Ou, "Towards Optimal Triage and Mitigation of Context-Sensitive Cyber Vulnerabilities," *IEEE Trans. Dependable Secur. Comput.*, vol. 20, no. 2, pp. 1270–1285, Mar. 2023, doi: 10.1109/TDSC.2022.3152164.

[12] H. A. Kholidy, "Autonomous mitigation of cyber risks in the Cyber–Physical Systems," *Futur. Gener. Comput. Syst.*, vol. 115, pp. 171–187, Feb. 2021, doi: 10.1016/j.future.2020.09.002.

[13] R. Bejtlich, *The practice of network security monitoring*. San Francisco, CA: No Starch Press, 2013.

[14] M. S. Akter, H. Shahriar, M. A. Rahman, M. Rahman, and A. Cuzzocrea, "Early Prediction of Cryptocurrency Price Decline: A Deep Learning Approach," in *2023 26th International Conference on Computer and Information Technology (ICCIT)*, Dec. 2023, pp. 1–6. doi: 10.1109/ICCIT60459.2023.10441030.

[15] M. A. Rahman and H. Shahrier, "Towards Developing Generative Adversarial Networks Based Robust Intrusion Detection Systems for Imbalanced Dataset Using Hadoop-PySpark," in *Proceedings of the Third International Conference on Innovations in Computing Research (ICR'24)*, 2024, pp. 449–463. doi: 10.1007/978-3-031-65522-7_40.

[16] Z. S. Dhahir, "A Hybrid Approach for Efficient DDoS Detection in Network Traffic Using CBLOF-Based Feature Engineering and XGBoost," *J. Futur. Artif. Intell. Technol.*, vol. 1, no. 2, pp. 174–190, Sep. 2024, doi: 10.62411/faith.2024-33.

[17] J. P. Ntayagabiri, Y. Bentaleb, J. Ndikumagenge, and H. El Makhtoum, "OMIC: A Bagging-Based Ensemble Learning Framework for Large-Scale IoT Intrusion Detection," *J. Futur. Artif. Intell. Technol.*, vol. 1, no. 4, pp. 401–416, Feb. 2025, doi: 10.62411/faith.3048-3719-63.

[18] A. Çetin and S. Öztürk, "Comprehensive Exploration of Ensemble Machine Learning Techniques for IoT Cybersecurity Across Multi-Class and Binary Classification Tasks," *J. Futur. Artif. Intell. Technol.*, vol. 1, no. 4, pp. 371–384, Feb. 2025, doi: 10.62411/faith.3048-3719-51.

[19] C. S. Htwe, Z. T. T. Myint, and Y. M. Thant, "IoT Security Using Machine Learning Methods with Features Correlation," *J. Comput. Theor. Appl.*, vol. 2, no. 2, pp. 151–163, Aug. 2024, doi: 10.62411/jcta.11179.

[20] D. R. I. M. Setiadi, S. Widiono, A. N. Safriandono, and S. Budi, "Phishing Website Detection Using Bidirectional Gated Recurrent Unit Model and Feature Selection," *J. Futur. Artif. Intell. Technol.*, vol. 1, no. 2, pp. 75–83, Jul. 2024, doi: 10.62411/faith.2024-15.

[21] M. A. Rahman, "Detection of Distributed Denial of Service Attacks based on Machine Learning Algorithms," *Int. J. Smart Home*, vol. 14, no. 2, pp. 15–24, Oct. 2020, doi: 10.21742/IJSH.2020.14.2.02.

[22] M. A. Rahman and H. Shahriar, "Clustering Enabled Robust Intrusion Detection System for Big Data Using Hadoop–PySpark," in *2023 IEEE 20th International Conference on Smart Communities: Improving Quality of Life using AI, Robotics and IoT (HONET)*, Dec. 2023, pp. 249–254. doi: 10.1109/HONET59747.2023.10374747.

[23] M. A. Rahman, H. Shahriar, V. Clincy, M. F. Hossain, and M. Rahman, "A Quantum Generative Adversarial Network-based Intrusion Detection System," in *2023 IEEE 47th Annual Computers, Software, and Applications Conference (COMPSAC)*, Jun. 2023, pp. 1810–1815. doi: 10.1109/COMPSAC57700.2023.00280.

[24] M. A. Rahman *et al.*, "Fine-Tuned Variational Quantum Classifiers for Cyber Attacks Detection Based on Parameterized Quantum Circuits and Optimizers," in *2024 IEEE 48th Annual Computers, Software, and Applications Conference (COMPSAC)*, Jul. 2024, pp. 1067–1072. doi: 10.1109/COMPSAC61105.2024.00144.

[25] M. A. Rahman, H. Shahriar, F. Wu, and A. Cuzzocrea, "Applying Pre-trained Multilingual BERT in Embeddings for Improved Malicious Prompt Injection Attacks Detection," in *2024 2nd International Conference on Artificial Intelligence, Blockchain, and Internet of Things (AIBThings)*, Sep. 2024, pp. 1–7. doi: 10.1109/AIBThings63359.2024.10863664.

[26] M. S. Akter *et al.*, "Authentic Learning Approach for Data Poisoning Vulnerability in LLMs," in *2024 IEEE 48th Annual Computers, Software, and Applications Conference (COMPSAC)*, Jul. 2024, pp. 1504–1505. doi: 10.1109/COMPSAC61105.2024.00210.

[27] M. A. Rahman, F. Wu, A. Cuzzocrea, and S. I. Ahamed, "Fine-tuned Large Language Models (LLMs): Improved Prompt Injection Attacks Detection," *arXiv*. Oct. 27, 2024. [Online]. Available: http://arxiv.org/abs/2410.21337

[28] M. A. Rahman *et al.*, "Embedding with Large Language Models for Classification of HIPAA Safeguard Compliance Rules," *arXiv*. Oct. 27, 2024. [Online]. Available: http://arxiv.org/abs/2410.20664

[29] A. Creswell, T. White, V. Dumoulin, K. Arulkumaran, B. Sengupta, and A. A. Bharath, "Generative Adversarial Networks: An Overview," *arXiv*. Oct. 19, 2017. doi: 10.1109/MSP.2017.2765202.

[30] I. Goodfellow *et al.*, "Generative adversarial networks," *Commun. ACM*, vol. 63, no. 11, pp. 139–144, Oct. 2020, doi: 10.1145/3422622.

[31] C. Yinka-Banjo and O.-A. Ugot, "A review of generative adversarial networks and its application in cybersecurity," *Artif. Intell. Rev.*, vol. 53, no. 3, pp. 1721–1736, Mar. 2020, doi: 10.1007/s10462-019-09717-4.

[32] A. Paul, D. P. Mukherjee, P. Das, A. Gangopadhyay, A. R. Chintha, and S. Kundu, "Improved Random Forest for Classification," *IEEE Trans. Image Process.*, vol. 27, no. 8, pp. 4012–4024, Aug. 2018, doi: 10.1109/TIP.2018.2834830.

[33] N. Moustafa and J. Slay, "UNSW-NB15: a comprehensive data set for network intrusion detection systems (UNSW-NB15 network data set)," in *2015 Military Communications and Information Systems Conference (MilCIS)*, Nov. 2015, pp. 1–6. doi: 10.1109/MilCIS.2015.7348942.

[34] P. Lin, K. Ye, and C.-Z. Xu, "Dynamic Network Anomaly Detection System by Using Deep Learning Techniques," in *Cloud Computing – CLOUD 2019*, 2019, pp. 161–176. doi: 10.1007/978-3-030-23502-4_12.

[35] I. Ahmad, M. Basheri, M. J. Iqbal, and A. Rahim, "Performance Comparison of Support Vector Machine, Random Forest, and Extreme Learning Machine for Intrusion Detection," *IEEE Access*, vol. 6, pp. 33789–33795, Jan. 2018, doi: 10.1109/ACCESS.2018.2841987.

[36] B. B. Zarpelão, R. S. Miani, C. T. Kawakani, and S. C. de Alvarenga, "A survey of intrusion detection in Internet of Things," *J. Netw. Comput. Appl.*, vol. 84, pp. 25–37, Apr. 2017, doi: 10.1016/j.jnca.2017.02.009.

[37] I. Goodfellow *et al.*, "Generative Adversarial Nets," in *Advances in Neural Information Processing Systems*, 2014, vol. 27. [Online]. Available: https://proceedings.neurips.cc/paper_files/paper/2014/file/5ca3e9b122f61f8f06494c97b1afccf3-Paper.pdf

[38] M. Belouch, S. El Hadaj, and M. Idhammad, "Performance evaluation of intrusion detection based on machine learning using Apache Spark," *Procedia Comput. Sci.*, vol. 127, pp. 1–6, 2018, doi: 10.1016/j.procs.2018.01.091.

[39] T. A. Tang, L. Mhamdi, D. McLernon, S. A. R. Zaidi, and M. Ghogho, "Deep learning approach for Network Intrusion Detection in Software Defined Networking," in *2016 International Conference on Wireless Networks and Mobile Communications (WINCOM)*, Oct. 2016, pp. 258–263. doi: 10.1109/WINCOM.2016.7777224.

[40] J. Kim, J. Kim, H. Le Thi Thu, and H. Kim, "Long Short Term Memory Recurrent Neural Network Classifier for Intrusion Detection," in *2016 International Conference on Platform Technology and Service (PlatCon)*, Feb. 2016, pp. 1–5. doi: 10.1109/PlatCon.2016.7456805.

[41] S. Potluri and C. Diedrich, "Accelerated deep neural networks for enhanced Intrusion Detection System," in *2016 IEEE 21st International Conference on Emerging Technologies and Factory Automation (ETFA)*, Sep. 2016, pp. 1–8. doi: 10.1109/ETFA.2016.7733515.