International Journal of Safety and Security Engineering
Vol. 15, No. 1, January, 2025, pp. 189-196
Journal homepage: http://iieta.org/journals/ijsse

Hybrid GAN-LSTM for Enhancing DDoS Detection on Imbalance Dataset

Gregorius Edo, Tohari Ahmad*, Muhammad Aidiel Rachman Putra

Department of Informatics, Institut Teknologi Sepuluh Nopember, Surabaya 60111, Indonesia

Corresponding Author Email: tohari@its.ac.id

https://doi.org/10.18280/ijsse.150120

Received: 14 July 2024
Revised: 27 November 2024
Accepted: 13 December 2024
Available online: 31 January 2025

Keywords:
DDoS attack detection, GANs, LSTM, intrusion detection system, network infrastructure, network security

ABSTRACT

The rapid development of internet network technology has increased the volume of data traffic. This surge in data traffic also raises the risk of Distributed Denial of Service (DDoS) attacks, which pose threats to institutions that rely on complex, interconnected networks for their operations. One of the primary challenges in combating these attacks lies in distinguishing malicious activity from normal traffic, as well as accurately detecting malicious attacks, each of which involves numerous parameters. The complexity of DDoS attacks continues to grow, further complicating detection and mitigation efforts. To address these challenges, more advanced and accurate tools are needed for DDoS attack detection. While significant research has been conducted on DDoS attacks, the use of Generative Adversarial Networks (GANs) for data balancing remains relatively unexplored. This study investigates the impact of dataset imbalances on the accuracy of DDoS attack classification and proposes models that generate synthetic data to address these imbalances. This research includes data collection, preprocessing, synthetic data generation, and performance analysis. GANs are used to generate synthetic data equivalent to the difference between the majority and minority classes in the dataset. A comparison of classification performance between the Long Short-Term Memory (LSTM) method without data balancing and the GAN-augmented model demonstrates improved results. The Hybrid GAN-LSTM model achieves accuracy rates exceeding 98% across all datasets, with F1-scores above 95%. These findings indicate that the Hybrid GAN-LSTM model addresses data imbalance issues and enhances classification accuracy. This study underscores the importance of addressing data imbalances in cybersecurity to improve the detection of DDoS attacks.

1. INTRODUCTION

The development of technology in the Internet network today is fast, as shown by dependence on the Internet, which ranges from the fields of education, economics, and communication [1]. However, this rapid development is accompanied by threats that can disrupt the resilience of network infrastructure, such as Distributed Denial of Service (DDoS) attacks. DDoS attacks threaten institutions, organizations, and enterprises by flooding network resources and disrupting the services used [2]. Although many DDoS attack mitigations are developed in industry as well as in academia, the threat of DDoS attacks can still have a severe impact and increase every year [3]. Distinguishing between DDoS and true or legitimate traffic is a complex task. Therefore, a tool that has a robust DDoS defence mechanism is needed [4].

Currently, the method to detect DDoS is to compare normal traffic statistics with DDoS seen from several parameters such as average of packets per flow (APF), average of bytes per flow (ABF), average of duration per flow (ADF), percentage of pairflow (PPF), grows of single-flow (GSF) and grows of different port (GDP) [5]. Sometimes these methods have difficulty in distinguishing between the two types of network traffic or inaccurate classification (false positive and false negative results) [6]. Better tools in detecting malicious attacks are needed to address the previously mentioned weaknesses to improve cybersecurity. Neural networks excel at identifying these intricate patterns because of their ability to learn from large datasets and recognize subtle relationships between different data points. An example of a DDoS attack happened on Singapore public health services [7], which services, including emails, webpages, and staff tools, were unavailable at that period. DDoS attacks stop legitimate users from accessing websites by flooding them with junk internet traffic. Healthcare and other services like finance, education, vehicles, and all internet-connected devices can also be impacted by this [8]. Since this threat can be felt in various sectors, addressing security concerns is crucial.

In order to address the issue of data imbalance, this research intends to examine classification algorithms by introducing synthetic data. It is done by using imbalance synthetic data on the experiment. The objective of this research is: first, to find out the effect of data imbalance on DDoS attack classification using the Deep Learning Long Short-Term Memory (LSTM) method; and second, to provide insight into the impact of creating synthetic data for balancing using Generative Adversarial Networks on LSTM methods. Specifically, this

research uses the CICDDoS2019 dataset to compare the performance of both test dataset conditions using LSTM.

The LSTM network is great for classifying the type of attack for datasets in the time series format, such as CICDDoS2019. Still, it struggles with imbalance data set, leading to varied performance across the classes. Thus, the model has problems in classifying the imbalanced minor classes and weak performance on the tasks of intrusion detection.

To mitigate this issue, this study looks into the application of Generative Adversarial Networks (GAN) as a technique for data balancing. It can balance the data set and subsequently increase the classification performance for the less dominant types of attacks by creating synthetic samples for these minority classes. The focus is to show that combining GAN to LSTM can increase the performance, ensuring a more reliable and secure intrusion detection system.

The paper is organized to guide the reader to better understand DDoS attacks and their detection. Section I explains the trend of DDoS detection, highlighting the significant problems and challenges they cause. Section II covers existing research on DDoS attack detection and presents a comprehensive survey to put current work in a better perspective. Section III discusses the methods, explaining different approaches for detecting DDoS using comparative analysis such as balancing methods and classification techniques. In Section IV, the evaluation as well as results analysis are presented, aiming at critically assessing how effective this method is. Lastly, Section V presents conclusions to summarize what was discovered from this study and its implications for future research.

2. RELATED WORKS

2.1 GAN for synthetic data generation

GAN [9] can be used as a tool for generating synthetic data. It consists of a network of generators and discriminators that will work with an adversarial gaming perspective [10]. GANs can create synthetic data that is close to real data. Yang and Zhou [11] proposed Imbalance Data Augmentation GAN (IDA-GAN) model used for image generation. In that study, two types of image datasets are used, namely single channel and three channel, both of which are imbalanced. After going through the GAN generation data process, the two datasets were then evaluated using various model. The obtained precision, recall, and F1-score show an improvement. The single-channel dataset received the highest precision, recall, and F1-score in the Modified National Institute of Standards and Technology (MNIST) dataset, which were 88.45%, 83.25%, and 82.5% respectively. The three-channel dataset received the highest precision, recall, and F1-score in the German Traffic Sign Recognition Benchmark (GTSRB) dataset, which were 87.20%, 87.53%, and 86.41% respectively.

2.2 GAN for anomaly detection

GAN is also used for detecting the anomaly of a network flow [12]. In that study, GAN is trained to use recurrent units to create fake data resembling monitored normal networks to deceive discriminator networks using latent noise from uniform distribution. The discriminator networks is then can also detect anomalies from real traffic data without training. For the experiment, a dataset from study the group Orion from the State University of Londrina and CICDDoS2019 are used, which are considered a time series. LSTM, CNN, DNN, GRU, and TCN are used for identification of anomalies or normal network flow, and are applied to discriminator networks. Their unsupervised model was able to achieve an F1-score of up to 99.8% in the first scenario (first dataset) and up to 98.0% in the second scenario (second dataset).

2.3 Gain Ratio

Another method for handling the imbalance dataset is by using Gain Ratio Feature Selection. Gain Ratio makes it possible to remove features that do not have a significant impact [13]. This method normalizes information gain with information entropy. In the study, Random Forest, Extra Tree, Naive Bayes, *k*-Nearest Neighbors, and Support Vector Machine were used to classify malware into 5 classes, namely Adware, Banking Malware, SMS Malware, Riskware, and Benign. The best accuracy is 94.57%, which is better than the previous one without a Gain Ratio of 94.22%. However, the accuracy of the model used is still lower than that of the study using the Extremely Randomized Tree model with more than 97% [14].

2.4 LSTM DDoS detection

LSTM networks can be used to classify time-series datasets on the basis of their temporal dependencies and patterns found in sequential data. In 2022, Gaur et al. [15] proposed the use of LSTM in multi-class classification of the CICDDoS2019 dataset which is a standard benchmark for DDoS attack detection. In order to deal with the problem of data imbalance where some of the classes were heavily undersampled, several of the minority classes were combined into one. While this minimized the factors brought on by data imbalance, it was at the cost of greater loss of classification granularity since the newly formed single class was an amalgamation of many substantially different classes.

Although there is still room for improvement, the results pointed towards the remarkable potential of LSTM when performing time-series classification tasks. These metrics demonstrate the accuracy of LSTM models in recognizing patterns, avoiding false positive targets, and performing moderately well for all classes. The research emphasizes the ability of LSTM as an advanced classifier for more complicated time-series data sets and in the field of security, for example, to detect DDoS attacks. Nevertheless, the modification of minority classes accentuates the problem of finding more sophisticated solutions, like GAN-derived data augmentation or class adaptive optimization techniques, which would enable preserving the characteristics of classes that are in minority in the analyzed datasets.

2.5 GAN-LSTM for imagery generation and classification

LSTM and GAN have proven to be effective in imagery generation and classification tasks. The synergy between the data generation capabilities of GANs and the sequential modeling capabilities of LSTM has shown increases in classification accuracy across diverse domains. For example, Bousmina et al. [16] designed a WGAN-GP and utilized it to generate synthetic features, which allowed them to achieve 97.83% accuracy on the YouTube Aerial Database.

Successfully combining GANs and BiLSTM networks for aerial imagery logging data [17] increased classification accuracy to 90%, highlighting the usefulness of these frameworks in spatial and visual data. These works demonstrate the strength of GAN-LSTM models in generating complex features and boosting performance in multi-faceted datasets.

These applications focusing on imagery have succeeded, but the use of GAN-LSTM in time series data is still unexplored. The datasets like malware detection and CICIDS2017 and CICDDoS2019 have temporal dependencies, which means that the data has interdependency over time and could have benefited from the paired sequential modeling done by LSTM and data augmentation done by GANs. Nevertheless, this hybrid approach may not be researched much in these datasets. While it exists and is successful, employing GAN-LSTM to tackle issues commonly revolving around time series data such as imbalance and noise, still remain relatively unexplored.

3. GAN-LSTM MODEL

The method we use starts with data collection which is a combination of 2-day DDoS records from the dataset. The dataset is then preprocessed by removing unnecessary columns. The GAN will be trained to use a regular neural network to create a fake dataset which is then combined with the original data. The new dataset formed is then divided into training data and test data. LSTM which then extracts the features and provides classification results. The model will be evaluated using graphical loss and validation analysis as well as a classification report. Figure 1 describes the flow of the method and Figure 2 describes its main framework. A more complete explanation is described in detail as follows.

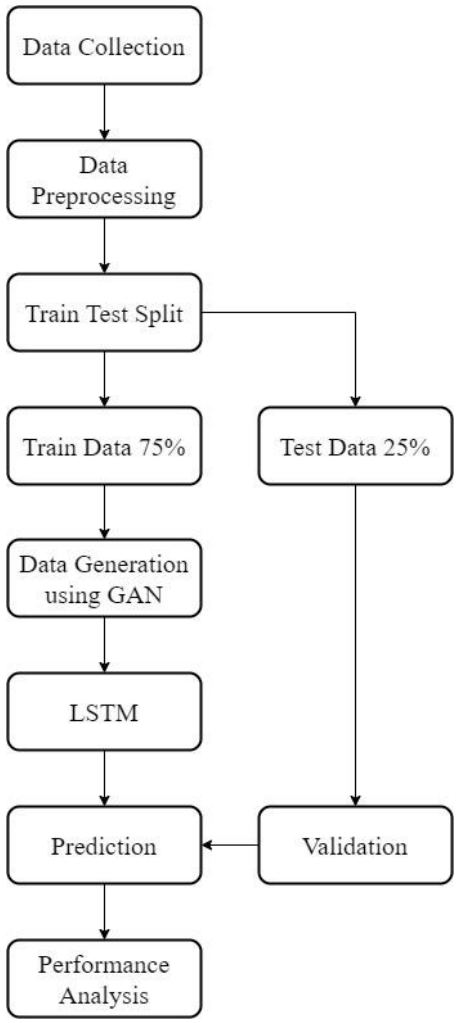

Figure 1. Flow of the method

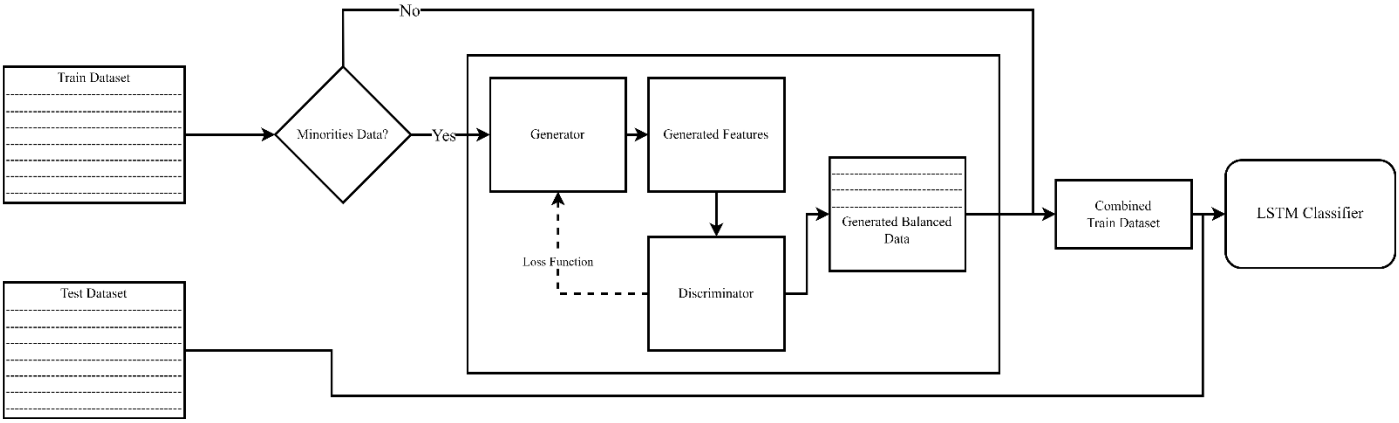

Figure 2. Main framework

3.1 Data preprocessing

The raw data is preprocessed so that the dataset matches the format structure that is suitable for the model. Then it can be proceeded with reducing or removing noise and redundant data [18]. El Sayed et al. [19] stated that removing noisy data can improve the performance of the model. Preprocessing steps used in this study are as follow.

3.3.1 Handling missing value

Missing values exist due to incomplete or inadequate data collection, data collection corruption, and sensor errors on the tool [20]. Missing values will interfere with the understanding and performance of deep learning models; therefore, they need to be removed [21]. The process of removing missing values in the proposed model starts by checking whether the dataset has missing values.

3.1.2 Categorical data

The dataset has some categorical data, such as Unammed: 0, Flow ID, Source IP, Destination IP, Source Port, Destination Port, timestamp, and Label. We dropped categorical data other than Label because it does not provide information that could be used as a classification parameter.

The value in the Label column was converted to a numeric value. This approach can facilitate label manipulation and analysis.

3.1.3 Data characteristics

In this context, Standard Deviation (SD) is used as a reference, which shows the size of variations scattered in data [22]. The higher the SD, the higher the variation in the data. Because of the variations in the data, it is necessary to normalize so that the features in the dataset have a similar scale.

The column with SD equal to zero is deleted. SD equal to zero means that there is no variation in the column so that it cannot provide enough information in the classification process. The dataset at this point has 66 features that have a variety of data and valuable information so that it can be analyzed further.

3.1.4 Normalization

Original data is rescaled using measures of mean and standard deviation so that the resultant feature has unit variance and zero mean [23]. The result of the process is a normalized dataset. Inside the process, the original relationships between data points within the dataset is preserved and the varying scale and distribution is eliminated.

3.1.5 Train-test split

For features that function as characters from the data are entered first into the x variable and for the data the label is entered into the y variable. The two variables were split with a train ratio of 75% and a test of 25%.

3.2 Synthetic data generation

GAN is based on two functions, namely Generator (G) and Discriminator (D). The task of G is to maximize the probability of D making errors in a competitive manner. GAN model is executed for each dataset contained in CICDDoS2019 dataset. Figure 3 describes the GAN architecture used in this study.

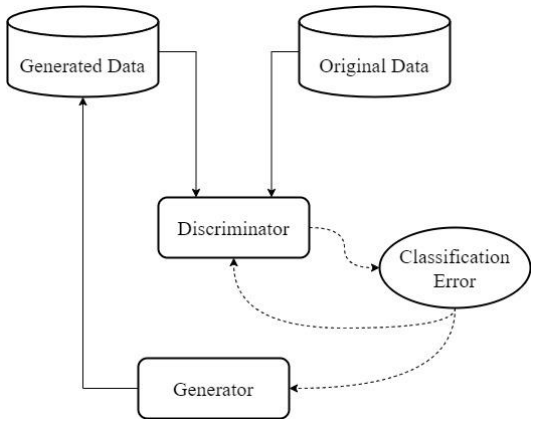

Figure 3. GAN architecture

Original Data: This study uses the CICDDoS2019 dataset, consisting of various network requests to servers. The available request types are normal and DDoS attacks such as UDP floods, Syn floods, and application layer attacks such as HTTP floods. This dataset is a valuable material in testing DDoS attack detection performance using DL and ML techniques [24].

Generator: Generators are trained to create fake data to minimize the difference between real and synthetic data. Noise was generated using gaussian because it has a character that can represent variations that occur under normal conditions.

Discriminator: Discriminator is trained to differentiate between fake and real data. All data, real and fake are then entered into input layer and the discriminator will detect it with output of probability score indicating the likelihood of the input are real or fake.

Classification Error: Binary Cross Entropy is employed as classification error. This method is to give feedback to discriminator and generator as a result it can determine fake and real dataset and also creating a better fake dataset for the generator.

Eqs. (1)-(3) demonstrate the objective function of minimax in GAN:

$$\min_G \max_D V(D,G) = A + B \tag{1}$$

where,

$$A = E_{x\sim p_{data}(x)}\left[\log\left(D(x)\right)\right] \tag{2}$$

and

$$B = E_{z\sim p_z(z)}\left[1 - \log\left(D\left(G(z)\right)\right)\right] \tag{3}$$

The generator is denoted by G, the discriminator is D, real data samples is x, and the noise is z is extracted from the distribution D with the specified standard deviation. The probability of the distribution of random noise is denoted by $p_z(z)$, while the probability of the distribution of real data is denoted by $p_{data}(x)$.

$$Loss_D = -\log\left(D(x_{real})\right) - \log\left(1 - D\left(x_{synthetic}\right)\right) \tag{4}$$

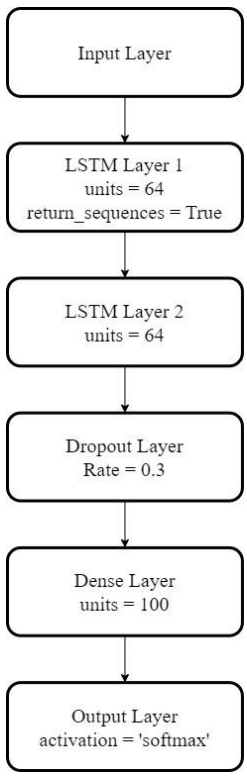

Figure 4. LSTM architecture

One step in the GAN training process is optimizing losses, which is the difference between the target labels and the generated samples. In Eq. (4), the discriminator aims to minimize its own losses and maximize generator loss. The generator is updated throughout training by back-propagating errors using the GAN. The generator's weight is updated using the Adam optimizer, repeating this procedure up to 20,000 epochs until it produces a sample of data that accurately reflects reality.

3.3 LSTM classification

LSTM is a Recurrent Neural Network (RNN) model that can be used on time series data, especially text categorization, sentence generation, and machine translation [25]. Because the data used in this study is time series, LSTM is suitable for use.

The dataset resulting from the GAN generation is then combined with the original dataset, which is then fed into the LSTM model as in Figure 4. The results of the classification are then evaluated.

3.4 Hardware and software

The experiment was done using Python 3.10, where the device has an Intel Core i5-12400F processor, 16GB DDR4 RAM memory with a 12GB Zotac RTX 3060 GPU. This study relies on Tensorflow 2.10. and the excellent scikit-learn 1.3 library in dataset management for classification.

4. RESULT AND DISCUSSION

4.1 Dataset

Figure 5 shows how the raw data CICDDoS2019 dataset is distributed. It is depicted that the combined total number of the UDPLag, Benign and Portmap classes make less than 5% of the whole dataset. This indicates deep underrepresentation of these three classes as compared to the other more dominant classes in the dataset.

In that way, the minority classes (UDPLag, Benign, and Portmap) may not adequately represent the features and behaviors associated with their classes. This makes it difficult to use these classes within a machine learning model because models that are trained on these biases would find problem with extracting robust patterns from these classes. As a result, the imbalance could lead to more optimistically biased predictions, as models are likely to learn more about the dominant classes and ignore the minority ones. This imbalance issue deals with achieving desirable classifications.

Figure 5 shows the dataset before balancing. The dataset looks imbalanced where Syn is the largest class with number of instances of 606,749 or the majority of the data is concentrated in this class. The sum of the other three classes is still less than 50% of the total dataset.

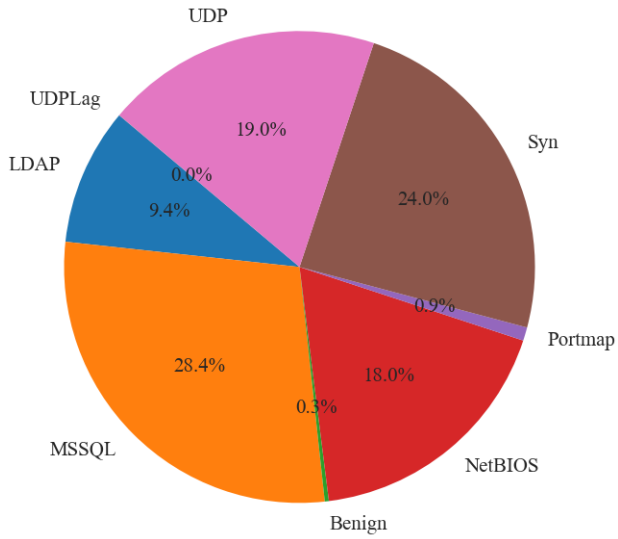

Figure 5. Raw dataset distribution

4.2 Result of data augmentation with GAN

Table 1 shows the number of instances per dataset before and after the GAN generation. GAN showing a great performance by generating a dataset that can mimic the real dataset. Each dataset is given additional synthetic instances as much as the difference in the amount of data in the minority and majority classes. The distribution of the dataset shows a more even distribution of data compared to before augmentation. This balanced data shows that the data can better represent each class because the magnitude of each class is the same. In example, Figure 6 is the representation of UDPLag dataset before GAN. After GAN generation the UDPLag dataset showing a more balanced dataset which showing a better representation of each class. The distribution of UDPLag dataset after synthetic data generation can be seen in Figure 7.

Table 1. Before and after data generation with GAN on train dataset

Dataset	Balancing	Type of Attack							
		Benign	***LDAP***	***MSSQL***	***NetBIOS***	***Portmap***	***Syn***	***UDP***	***UDPLag***
LDAP	Before	3.843	1.428.893	-	152.189	-	-	-	-
	After	1.428.893	1.428.893	-	1.428.893	-	-	-	-
MSSQL	Before	2.096	7.448	4.322.296	-	-	-	-	-
	After	4.322.296	4.322.296	4.322.296	-	-	-	-	-
NetBIOS	Before	991	-	-	2.590.934	-	-	-	-
	After	2.590.934	-	-	2.590.934	-	-	-	-
Portmap	Before	3.551	-	-	-	140.220	-	-	-
	After	140.220	-	-	-	140.220	-	-	-
Syn	Before	26.843	-	-	-	-	3.213.563	-	-
	After	3.213.563	-	-	-	-	3.213.563	-	-
UDP	Before	2.351	-	18.294	-	-	-	2.816.010	-
	After	2.816.010	-	2.816.010	-	-	-	2.816.010	-
UDPLag	Before	3.051	-	-	-	-	455.062	84.356	1.405
	After	455.062	-	-	-	-	455.062	455.062	455.062

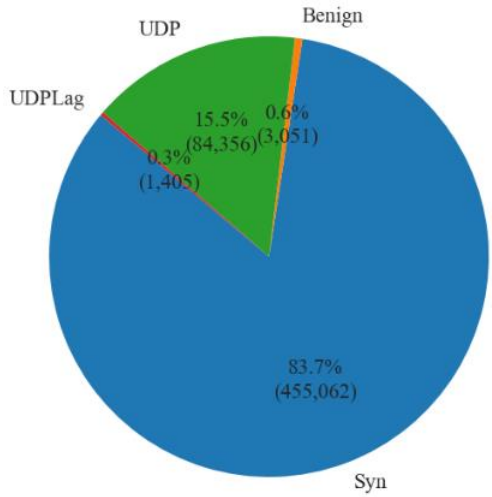

Figure 6. UDPLag raw dataset

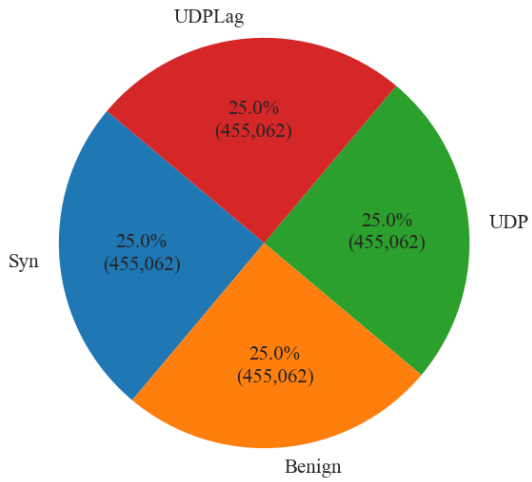

Figure 7. UDPLag dataset after GAN augmentation

4.3 Classification result

After successfully generating data using GAN, the generated and original datasets are combined. The new combined dataset is then split and used for LSTM training. Performance is then evaluated using accuracy and F1-Score values whose results can be seen in Tables 2 and 3.

As it has shown, LSTM is powerful for multi-class classifications because of its accuracy on a variety of datasets. Its flaws include class imbalance problems, which lead to lower F1-Scores in some classes. The class imbalance in the NetBIOS dataset is extreme, for example, where the minor class only has 0.004% of samples. In this regard, some stand alone LSTM models have shown F1-Scores of only 83.77% and 81.89% for MSSQL and UDP datasets, respectively, indicating their struggles with imbalance data.

Using the GAN approach in combination with the LSTM neural network, known as GAN-LSTM, enhances the performance of imbalanced data classification models through the use of synthetic data. As displayed in Table 2, the accuracy of GAN-LSTM is higher than that of LSTM, with all datasets over 99%. It also has high F1-scores, which are more than 99% for each category, meaning that the performance is not biased to majority classes alone. For instance, in the NetBIOS dataset, the F1-score of 84.52% improves to 99.67% after using GAN-LSTM and in both Syn and UDP, the F1-scores are improved by at least 17%. These figures suggest that the GAN-LSTM not only enhances classification performance, but also improves the class imbalance problem, thus making the model more effective and reliable for time-series datasets with bias.

Table 2. Comparison accuracy on models

Dataset	Model	
	LSTM (%)	GAN-LSTM (%)
LDAP	92.37	99.74
MSSQL	87.02	99.88
NetBIOS	95.05	98.67
Portmap	88.03	99.87
Syn	93.56	99.65
UDP	94.31	99.77
UDPLag	97.01	99.76

Table 3. Comparison macro-average F1-Score on models

Dataset	Model	
	LSTM (%)	GAN-LSTM (%)
LDAP	85.46	99.65
MSSQL	83.77	99.64
NetBIOS	84.52	99.67
Portmap	86.71	99.89
Syn	82.44	99.85
UDP	81.89	99.75
UDPLag	82.92	99.73

5. CONCLUSIONS

The Hybrid GAN-LSTM model demonstrates promising results in the detection of DDoS attacks. This approach successfully generates synthetic data of a quality comparable to the original dataset, as evidenced by the improved performance metrics. In this study, LSTM was utilized as a multiclass classification method to assess how the inclusion of synthetic data influences the performance. When the combined dataset, consisting of original and synthetic data, was used, there was an improvement in both accuracy and F1-score metrics.

This augmentation method enables the model to better identify different types of multiclass attacks, as synthetic data generation provides more generalized and balanced data. Consequently, the model benefits from an enhanced ability to distinguish between various attack types, leading to greater overall reliability and robustness.

However, challenges remain for future research, particularly in conducting a more detailed evaluation of the quality of the generated synthetic data. By identifying the specific characteristics of the generated data, it will be possible to design more effective improvement methods, further advancing the efficacy of DDoS attack detection systems.

ACKNOWLEDGMENT

The authors gratefully acknowledge support from the Institut Teknologi Sepuluh Nopember for this work, under project scheme of the Publication Writing and IPR Incentive Program (PPHKI) 2025.

REFERENCES

[1] Ali, T.E., Chong, Y.W., Manickam, S. (2023). Machine learning techniques to detect a DDoS attack in SDN: A systematic review. Applied Sciences, 13(5): 3183. https://doi.org/10.3390/app13053183

[2] Chaganti, R., Boppana, R.V., Ravi, V., Munir, K., et al. (2022). A comprehensive review of denial of service attacks in blockchain ecosystem and open challenges. IEEE Access, 10: 96538-96555. https://doi.org/10.1109/ACCESS.2022.3205019

[3] Abu Bakar, R., Huang, X., Javed, M.S., Hussain, S., Majeed, M.F. (2023). An intelligent agent-based detection system for DDoS attacks using automatic feature extraction and selection. Sensors, 23(6): 3333. https://doi.org/10.3390/s23063333

[4] Acosta-Tejada, D.E., Sanchez-Galan, J.E., Torres-Batista, N. (2023). Analyzing DDoS attack classification with data imbalance using generative adversarial networks. In 2023 IEEE Latin-American Conference on Communications (LATINCOM), Panama City, Panama, pp. 1-6.

[5] Fouladi, R.F., Ermiş, O., Anarim, E. (2020). A DDoS attack detection and defense scheme using time-series analysis for SDN. Journal of Information Security and Applications, 54: 102587. https://doi.org/10.1016/j.jisa.2020.102587

[6] Yousuf, O., Mir, R.N. (2022). DDoS attack detection in Internet of Things using recurrent neural network. Computers and Electrical Engineering, 101: 108034. https://doi.org/10.1016/j.compeleceng.2022.108034

[7] Antoniuk, D. (2023). Singapore public health services hit by DDoS attacks. https://therecord.media/singapore-public-health-services-ddos-attack.

[8] Verma, A., Saha, R., Kumar, N., Kumar, G. (2022). A detailed survey of denial of service for IoT and multimedia systems: Past, present and futuristic development. Multimedia Tools and Applications, 81(14): 19879-19944. https://doi.org/10.1007/s11042-021-11859-z

[9] Goodfellow, I.J., Pouget-Abadie, J., Mirza, M., Xu, B., et al. (2014). Generative adversarial nets. In Proceedings of the 27th International Conference on Neural Information Processing Systems-Volume 2, Montreal, Canada, pp. 2672-2680.

[10] Li, Y., Wang, Q., Zhang, J., Hu, L., Ouyang, W. (2021). The theoretical research of generative adversarial networks: An overview. Neurocomputing, 435: 26-41. https://doi.org/10.1016/j.neucom.2020.12.114

[11] Yang, H., Zhou, Y. (2021). Ida-GAN: A novel imbalanced data augmentation GAN. In 2020 25th International Conference on Pattern Recognition (ICPR), Milan, Italy, pp. 8299-8305. https://doi.org/10.1109/ICPR48806.2021.9411996

[12] Lent, D.M.B., Ruffo, V.G.D.S., Carvalho, L.F., Lloret, J., Rodrigues, J.J., Proença, M.L. (2024). An unsupervised generative adversarial network system to detect DDoS attacks in SDN. IEEE Access, 12: 70690-70706. https://doi.org/10.1109/ACCESS.2024.3402069

[13] Ansori, D.B., Slamet, J., Ghufron, M.Z., Putra, M.A.R., Ahmad, T. (2024). Android malware classification using Gain Ratio and ensembled machine learning. International Journal of Safety and Security Engineering, 14(1): 259-266. https://doi.org/10.18280/ijsse.140126

[14] Nguyen, C.D., Khoa, N.H., Doan, K.N.D., Cam, N.T. (2023). Android malware category and family classification using static analysis. In 2023 International Conference on Information Networking (ICOIN), Bangkok, Thailand, pp. 162-167. https://doi.org/10.1109/ICOIN56518.2023.10049039

[15] Gaur, V., Dogra, A., Gupta, A., Tibrewal, A. (2022). Multiclass classification for DDoS attacks using LSTM time-series model. In 7th International Conference on Computing in Engineering & Technology (ICCET 2022), Online Conference, pp. 135-141. https://doi.org/10.1049/icp.2022.0605

[16] Bousmina, A., Selmi, M., Ben Rhaiem, M.A., Farah, I.R. (2023). A hybrid approach based on GAN and CNN-LSTM for aerial activity recognition. Remote Sensing, 15(14): 3626. https://doi.org/10.3390/rs15143626

[17] Guo, L., Renze, L., Xingyu, L., Juanjuan, T., Lei, C., Yang, Z. (2021). Logging data completion based on an MC-GAN-BiLSTM Model. IEEE Access, 10: 1810-1822. https://doi.org/10.1109/ACCESS.2021.3138194

[18] Maranhão, J.P.A., da Costa, J.P.C., de Freitas, E.P., Javidi, E., de Sousa, R.T. (2020). Noise-robust multilayer perceptron architecture for distributed denial of service attack detection. IEEE Communications Letters, 25(2): 402-406. https://doi.org/10.1109/LCOMM.2020.3032170

[19] El Sayed, M.S., Le-Khac, N.A., Azer, M.A., Jurcut, A.D. (2022). A flow-based anomaly detection approach with feature selection method against DDoS attacks in SDNS. IEEE Transactions on Cognitive Communications and Networking, 8(4): 1862-1880. https://doi.org/10.1109/TCCN.2022.3186331

[20] Rani, P., Kumar, R., Jain, A. (2021). Multistage model for accurate prediction of missing values using imputation methods in heart disease dataset. In Innovative Data Communication Technologies and Application: Proceedings of ICIDCA 2020, pp. 637-653. https://doi.org/10.1007/978-981-15-9651-3_53

[21] Batchu, R.K., Seetha, H. (2022). On improving the performance of DDoS attack detection system. Microprocessors and Microsystems, 93: 104571. https://doi.org/10.1016/J.MICPRO.2022.104571

[22] Makandar, A., Javeriya, S.B. (2023). A comparative analysis of image enhancement techniques for improving blurry identity images. In 2023 IEEE International Carnahan Conference on Security Technology (ICCST), Pune, India, pp. 1-6. https://doi.org/10.1109/ICCST59048.2023.10474273

[23] Singh, D., Singh, B. (2020). Investigating the impact of data normalization on classification performance. Applied Soft Computing, 97: 105524. https://doi.org/10.1016/j.asoc.2019.105524

[24] Ramzan, M., Shoaib, M., Altaf, A., Arshad, S., Iqbal, F., Castilla, Á.K., Ashraf, I. (2023). Distributed denial of service attack detection in network traffic using deep learning algorithm. Sensors, 23(20): 8642. https://doi.org/10.3390/s23208642

[25] Sherstinsky, A. (2020). Fundamentals of recurrent neural network (RNN) and long short-term memory (LSTM) network. Physica D: Nonlinear Phenomena, 404: 132306. https://doi.org/10.1016/j.physd.2019.132306

NOMENCLATURE

G	Generator network
D	Discriminator network
x	Specific data/value from feature
y	Specific data/value from target
z	Noise vector from noise
p	Probability
A	log(D(x)) over the real data distribution
B	(1- log(D(x))) over the real data distribution
Loss	Loss function