Edelweiss Applied Science and Technology
ISSN: 2576-8484
Vol. 9, No. 3, 2993-2999
2025
Publisher: Learning Gate
DOI: 10.55214/25768484.v9i3.5899

Bi-LSTM-XGBoost ensemble-based intrusion detection system: Addressing data imbalance and enhancing minority class performance

Woo-Seong KIM[1], Hyun-Jung KIM[2*]
[1]Konkuk University, The Graduate School of Information & Communication, Department of Convergence Information Technology (Artificial Intelligence Major), Seoul, Republic of Korea.
[2]Konkuk University, Sang-Huh College and The Graduate School of Information & Communication, Department of Convergence Information Technology (Artificial Intelligence Major), Seoul, Republic of Korea; nygirl@konkuk.ac.kr (H.J.K.).

Abstract: Intrusion Detection Systems (IDS) are critical in identifying abnormal network activities and mitigating potential security threats. However, existing IDS solutions struggle with detecting rare attack types, such as Remote-to-Local (R2L) and User-to-Root (U2R), primarily due to data imbalance. To address this challenge, we propose an ensemble model combining Bidirectional Long Short-Term Memory (Bi-LSTM) networks and eXtreme Gradient Boosting (XGBoost). Our model achieves an accuracy of 98.42% on the NSL-KDD dataset, significantly reducing the false positive rates for R2L and U2R classes by approximately 90% and 67%, respectively (p-value < 0.05). Moreover, the proposed model achieves an Area Under the Receiver Operating Characteristic Curve (AUC-ROC) score of 0.89 for R2L detection, outperforming the Bi-LSTM-Random Forest baseline (0.88). For U2R detection, the AUC improved from 0.58 to 0.66. These findings highlight the model's enhanced capability for minority class detection and its potential to mitigate data imbalance issues in IDS. Future work will focus on integrating Conditional Generative Adversarial Networks (Conditional GANs) for data augmentation, optimizing hyperparameters using Particle Swarm Optimization (PSO), and validating the model's generalizability on CICIDS2017 and UNSW-NB15 datasets.

Keywords: *Bi-LSTM, XGBoost, Ensemble Model, Intrusion Detection System, Minority Class Detection.*

1. Introduction

Intrusion Detection Systems (IDS) are vital components of network security, designed to identify anomalous activities and prevent potential security breaches [1, 2]. Despite their importance, IDS performance is significantly hindered when detecting rare attack types, such as Remote-to-Local (R2L) and User-to-Root (U2R), due to inherent data imbalance. This limitation poses critical risks, including data breaches and system outages, emphasizing the urgent need to improve minority class detection in IDS.

Previous studies have explored various approaches to enhance IDS performance. Kasongo and Sun [3] utilized Random Forest-based models to improve detection accuracy; however, the issue of data imbalance persisted Kasongo and Sun [3]. Acharya, et al. [4] leveraged Bi-LSTM to capture temporal dependencies in time-series data, but their approach fell short in addressing minority class detection[4, 5]. Similarly, Rohini, et al. [6] targeted IoT network environments using ensemble models but demonstrated suboptimal performance for R2L and U2R attack types Rohini, et al. [6]. Alsaffar, et al. [7] employed hybrid feature selection and stack ensemble methods, yet practical challenges remained in enhancing detection for minority classes [7].

To overcome these limitations, we propose an ensemble model integrating Bi-LSTM and XGBoost. Bi-LSTM excels at capturing temporal dependencies in sequential data, while XGBoost is effective for

History: Received: 20 January 2025; Revised: 16 March 2025; Accepted: 21 March 2025; Published: 29 March 2025
* Correspondence: nygirl@konkuk.ac.kr

analyzing non-linear features. By combining these strengths, our approach aims to enhance IDS reliability and address the data imbalance issue.

This paper is organized as follows: Section 2 reviews related work and discusses existing approaches to the data imbalance problem. Section 3 details the design and structure of the proposed Bi-LSTM-XGBoost ensemble model. Section 4 presents the evaluation results, and Section 5 discusses the study's contributions, limitations, and future directions.

2. Methodology

In this study, we developed an ensemble model combining Bidirectional Long Short-Term Memory (Bi-LSTM) networks and eXtreme Gradient Boosting (XGBoost) to enhance the detection performance of minority classes in the NSL-KDD dataset. The methodology comprises five key stages: data preprocessing, Bi-LSTM training, XGBoost training, hyperparameter optimization, and ensemble model construction.

2.1. Data Preprocessing

The NSL-KDD dataset was partitioned into training and testing sets and preprocessed to suit the Bi-LSTM and XGBoost frameworks.

Target Variable Conversion: The target variable was converted using one-hot encoding.

Data Reshaping: Input data were transformed into three-dimensional tensors for Bi-LSTM training and two-dimensional arrays for XGBoost.

Feature Selection: Features were evaluated and optimized based on the study by Shin, et al. [8].

2.2. Bi-LSTM Training

The Bi-LSTM model was configured to capture the temporal dependencies in time-series data.

Model Design: The architecture was developed following the guidelines from Padmavathi, et al. [9] incorporating fundamental strategies to address data imbalance [9].

Feature Extraction: A combination of Conv1D and MaxPooling layers was employed to extract critical input features.

Normalization: A dropout rate of 0.5 was applied to mitigate overfitting.

Bidirectional Layers: Two Bi-LSTM layers (64 and 128 units) were utilized to learn bidirectional temporal dependencies.

Output Layer: Multi-class classification was performed using the softmax activation function.

Training Algorithm: The model was trained with the Adam optimizer and categorical cross-entropy loss function.

2.3. XGBoost Training

XGBoost was utilized to learn non-linear characteristics and perform multi-class classification on the NSL-KDD dataset.

- Target Variable Encoding: The target variable was converted using one-hot encoding.
- Model Configuration: The training setup followed the foundational work of Chen and Guestrin [10] leveraging XGBoost's efficiency and capability for non-linear feature analysis [10].
- Hyperparameter Settings:
    - Learning rate: 0.1
    - Maximum depth: 6
    - Minimum split weight: 1
- Evaluation Metric: Multi-class log loss
- Optimization: Hyperparameters were fine-tuned using grid search.

Edelweiss Applied Science and Technology
ISSN: 2576-8484
Vol. 9, No. 3: 2993-2999, 2025
DOI: 10.55214/25768484.v9i3.5899

2.4. Hyperparameter Optimization

To maximize the performance of both Bi-LSTM and XGBoost, hyperparameters were optimized through cross-validation and grid search [11].

- Bi-LSTM:
  - LSTM units: 64, 128
  - Dropout ratio: 0.5
  - Learning rate: Optimized value applied
- XGBoost:
  - Learning rate: 0.1
  - Maximum depth: 6
  - Minimum split weight: 1

2.5. Ensemble Model Design

The prediction outputs of Bi-LSTM and XGBoost were integrated using the hard voting technique to construct the final ensemble model.

Reference Studies: The hybrid approach combining deep learning and machine learning for IDS and minority class detection was inspired bySajid, et al. [12]; Ajeesh and Mathew [13] and Khan and Kim [14].

Combination Methodology: Final classification results were derived using a majority voting scheme that combined the predictions of both models.

3. Experiments and Results

This study quantitatively evaluated the performance of the proposed Bi-LSTM-XGBoost ensemble model against the Bi-LSTM standalone model and the LSTM + Random Forest ensemble model. Performance metrics included accuracy, precision, recall, and F1-score, with a particular focus on improving detection performance for minority classes.

3.1. Performance Metrics

The evaluation of the models was conducted using the following four key metrics:

Accuracy: The proportion of all predictions correctly classified.

Precision: The proportion of predicted positive samples that are truly positive, crucial for reducing false positives.

Recall: The proportion of actual positive samples correctly identified by the model, indicating model stability.

F1-Score: The harmonic mean of precision and recall, particularly valuable for minority class detection performance.

3.1.1. Accuracy

Accuracy serves as the primary indicator of overall IDS performance. The proposed Bi-LSTM-XGBoost ensemble model achieved an accuracy of 98.42%, surpassing both the Bi-LSTM standalone model (98.00%) and the LSTM + Random Forest ensemble model (98.14%).

Table 1.
Accuracy.

Model	Accuracy (%)
Bi-LSTM	98.00
Bi-LSTM + RandomForest Ensemble	98.14
Bi-LSTM + XGBoost Ensemble	98.42

Edelweiss Applied Science and Technology
ISSN: 2576-8484
Vol. 9, No. 3: 2993-2999, 2025
DOI: 10.55214/25768484.v9i3.5899

3.1.2. Precision

Precision is critical for minimizing false positives. The proposed model recorded a precision of 98.53%, exceeding the Bi-LSTM (97.94%) and LSTM + Random Forest (98.23%) models.

Table 2.
Precision.

Model	Precision(%)
Bi-LSTM	97.94
Bi-LSTM + RandomForest Ensemble	98.23
Bi-LSTM + XGBoost Ensemble	98.53

3.1.3. Recall

Recall highlights the model's ability to detect actual positive instances. The proposed Bi-LSTM-XGBoost ensemble model achieved the highest recall of 98.42%.

Table 3.
Recall.

Model	Recall (%)
Bi-LSTM	98.00
Bi-LSTM + RandomForest Ensemble	98.14
Bi-LSTM + XGBoost Ensemble	98.42

3.1.4. F1-Score

The F1-score is particularly relevant for minority class detection. The Bi-LSTM-XGBoost ensemble model achieved the highest F1-score of 98.37%, significantly outperforming the Bi-LSTM model (97.94%) and LSTM + Random Forest ensemble model (98.10%). Statistical significance was confirmed (p-value < 0.05).

Table 4.
F1-Score.

Model	F1-Score (%)
Bi-LSTM	97.94
Bi-LSTM + RandomForest Ensemble	98.10
Bi-LSTM + XGBoost Ensemble	98.37

3.2. Minority Class Detection Performance

3.2.1. R2L and U2R Classes

The proposed Bi-LSTM-XGBoost ensemble model demonstrated significant improvements in detecting minority classes such as R2L and U2R. For the R2L class, false positives decreased by 90%, from 355 cases (Bi-LSTM model) to 36 cases. For the U2R class, false positives decreased by 67%, from 18 cases (Bi-LSTM model) to 6 cases. Additionally, the Area Under the Curve (AUC) values showed marked improvement. R2L AUC increased from 0.88 (LSTM + Random Forest) to 0.89. U2R AUC significantly improved from 0.58 (LSTM + Random Forest) to 0.66. This indicates that the proposed model offers enhanced reliability and performance in detecting minority classes.

Table 4.
AUC Value Table.

Class	Model	False Positives	AUC Value
R2L	Bi-LSTM + RandomForest	74	0.88
	Bi-LSTM + XGBoost	36	0.89
U2R	Bi-LSTM + RandomForest	11	0.58
	Bi-LSTM + XGBoost	5	0.66

Edelweiss Applied Science and Technology
ISSN: 2576-8484
Vol. 9, No. 3: 2993-2999, 2025
DOI: 10.55214/25768484.v9i3.5899

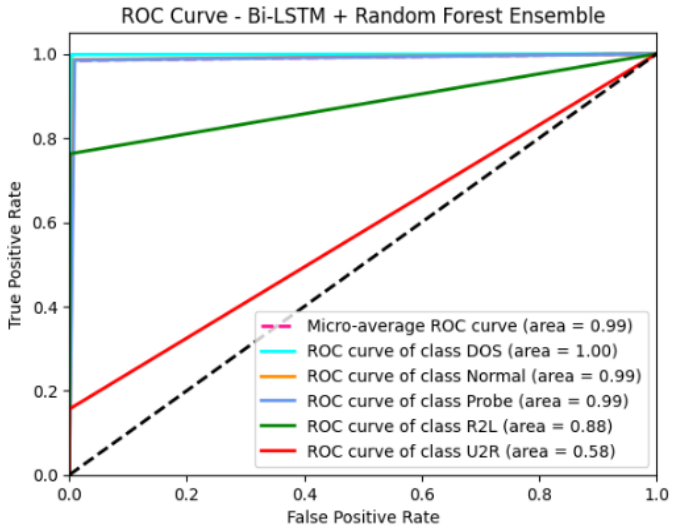

Figure 1.
Bi-LSTM+RF ROC.

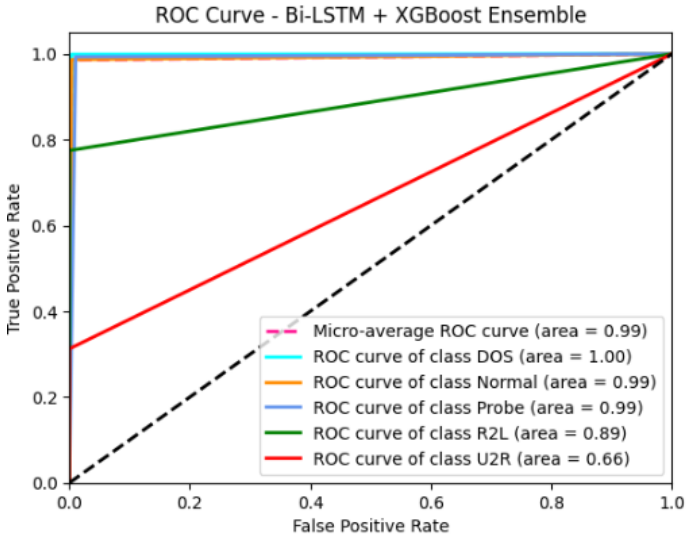

Figure 2.
Bi-LSTM+XGBoost ROC.

Figures 3–5 visually depict the detection performance improvements for R2L and U2R classes, confirming the proposed model's ability to address data imbalance issues.

Edelweiss Applied Science and Technology
ISSN: 2576-8484
Vol. 9, No. 3: 2993-2999, 2025
DOI: 10.55214/25768484.v9i3.5899

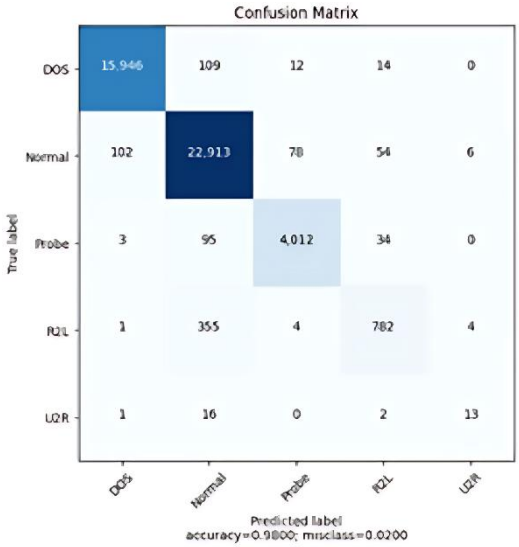

Figure 3.
Bi-LSTM

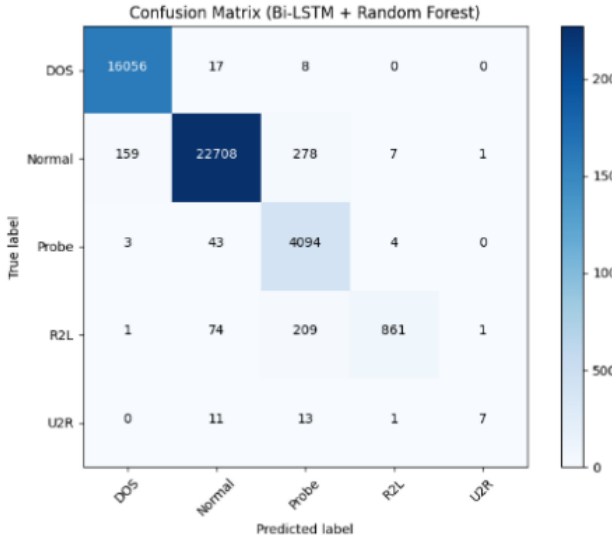

Figure 4.
Bi-LSTM + RF

4. Conclusions

This study proposed an ensemble model combining Bi-LSTM and XGBoost to address the deterioration of detection performance due to data imbalance problems in network intrusion detection systems (IDS). The proposed model, which combines the time series learning strengths of Bi-LSTM with XGBoost's ability to analyze nonlinear characteristics, recorded superior performance over existing models (bi-LSTM and LSTM + Random Forest) with accuracy (98.42%), precision (98.53%), reproducibility (98.42%), and F1-score (98.37%) on the NSL-KDD dataset. In particular, it reduced the false positive rates of the R2L and U2R classes by about 90% and 67%, respectively, and demonstrated that the AUC performance improvement was statistically significant (p-value < 0.05).

Notably, the model reduced false positive rates for R2L and U2R classes by 90% and 67%, respectively, and demonstrated statistically significant improvements in AUC values.

Edelweiss Applied Science and Technology
ISSN: 2576-8484
Vol. 9, No. 3: 2993-2999, 2025
DOI: 10.55214/25768484.v9i3.5899

Further research will explore generalizability using datasets like CICIDS2017 and UNSW-NB15, employ Conditional GANs for data augmentation, and leverage PSO for hyperparameter optimization and model efficiency. These approaches aim to enhance IDS detection performance and reliability, paving the way for advanced network security system designs.

Transparency:

The authors confirm that the manuscript is an honest, accurate, and transparent account of the study; that no vital features of the study have been omitted; and that any discrepancies from the study as planned have been explained. This study followed all ethical practices during writing.

Copyright:

References

[1] S.-C. Lim, "A Real-Time Intrusion Detection based on Monitoring in Network Security," *The Journal of the Institute of Internet, Broadcasting and Communication*, vol. 13, no. 3, pp. 9-15, 2013. https://doi.org/10.7236/jiibc.2013.13.3.9

[2] Y.-J. Kim *et al.*, "A study on the exposures and threats for internet of things (IoT) IP," *The Journal of the Convergence on Culture Technology*, vol. 2, no. 4, pp. 77-82, 2016. https://doi.org/10.17703/jcct.2016.2.4.77

[3] S. M. Kasongo and Y. Sun, "A deep gated recurrent unit based model for wireless intrusion detection system," *ICT Express*, vol. 7, no. 1, pp. 81-87, 2021. https://doi.org/10.1016/j.icte.2020.03.002

[4] T. Acharya, A. Annamalai, and M. F. Chouikha, "Addressing the class imbalance problem in network-based anomaly detection," presented at the In 2024 IEEE 14th Symposium on Computer Applications & Industrial Electronics (ISCAIE) (pp. 1-6). IEEE, 2024.

[5] S. Yoon, "Detecting abnormal human movements based on variational autoencoder," *International Journal of Internet, Broadcasting and Communication*, vol. 15, no. 3, pp. 94-102, 2023.

[6] G. Rohini, C. Gnana Kousalya, and J. Bino, "Intrusion detection system with an ensemble learning and feature selection framework for IoT networks," *IETE Journal of Research*, vol. 69, no. 12, pp. 8859-8875, 2023. https://doi.org/10.1080/03772063.2022.2098187

[7] A. M. Alsaffar, M. Nouri-Baygi, and H. M. Zolbanin, "Shielding networks: Enhancing intrusion detection with hybrid feature selection and stack ensemble learning," *Journal of Big Data*, vol. 11, no. 1, p. 133, 2024. https://doi.org/10.1186/s40537-024-00994-7

[8] Y. Shin, D. Y. Yun, S.-J. Moon, and C.-g. Hwang, "A Research on Accuracy Improvement of Diabetes Recognition Factors Based on XGBoost," *International journal of advanced smart convergence*, vol. 10, no. 2, pp. 73-78, 2021.

[9] B. Padmavathi, A. Bhagyalakshmi, D. Kavitha, and P. Indumathy, "An optimized Bi-LSTM with random synthetic over-sampling strategy for network intrusion detection," *Soft Computing*, vol. 28, no. 1, pp. 777-790, 2024. https://doi.org/10.1007/s00500-023-09483-0

[10] T. Chen and C. Guestrin, "XGBoost: A scalable tree boosting system," in *KDD '16: Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, pp. 785-794. ACM*, 2016.

[11] Y.-S. Lee and P.-J. Moon, "Analysis of open-source hyperparameter optimization software trends," *International Journal of Advanced Culture Technology*, vol. 7, no. 4, pp. 56-62, 2019.

[12] M. Sajid *et al.*, "Enhancing intrusion detection: A hybrid machine and deep learning approach," *Journal of Cloud Computing*, vol. 13, no. 1, p. 123, 2024.

[13] A. Ajeesh and T. Mathew, "Enhancing network security: A comparative analysis of deep learning and machine learning models for intrusion detection," in *Proceedings of the 2024 International Conference on E-mobility, Power Control, and Energy, pp. 1-6. IEEE*, 2024.

[14] M. A. Khan and Y. Kim, "Deep Learning-Based Hybrid Intelligent Intrusion Detection System," *Computers, Materials & Continua*, vol. 68, no. 1, pp. 671-687, 2021. https://doi.org/10.32604/cmc.2021.015647