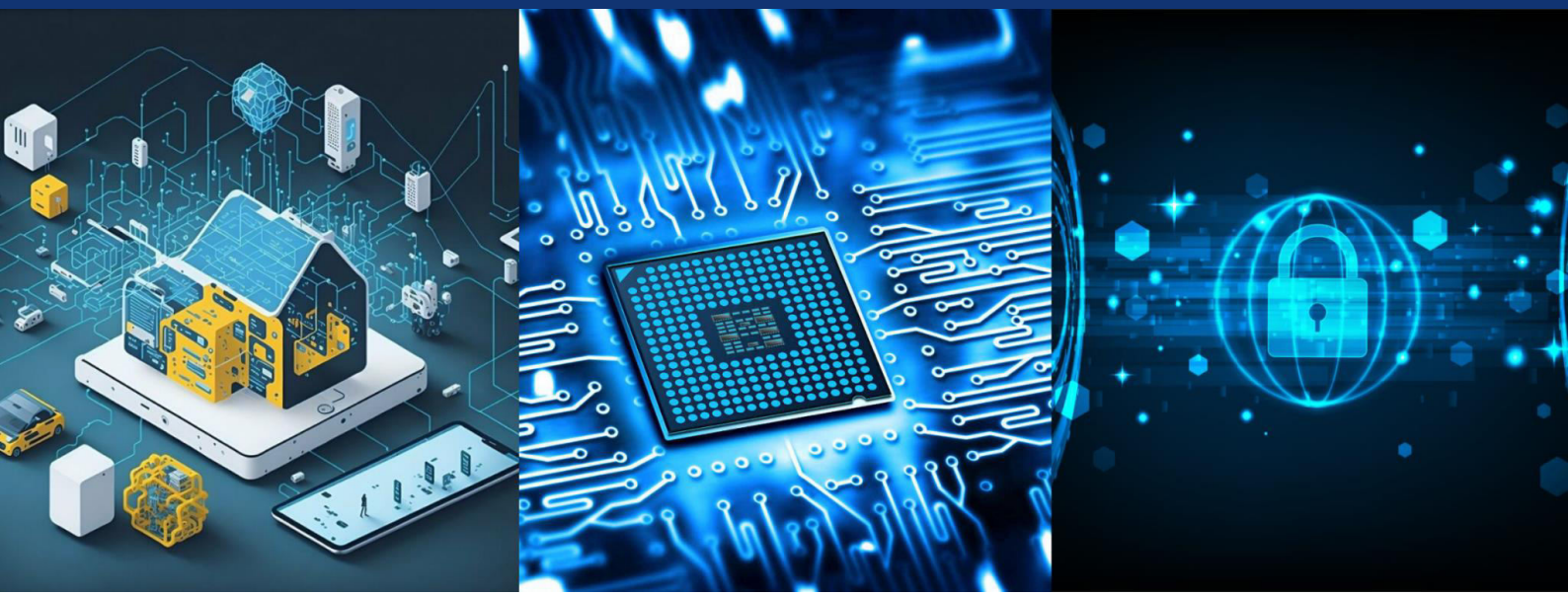
IJIRCCE

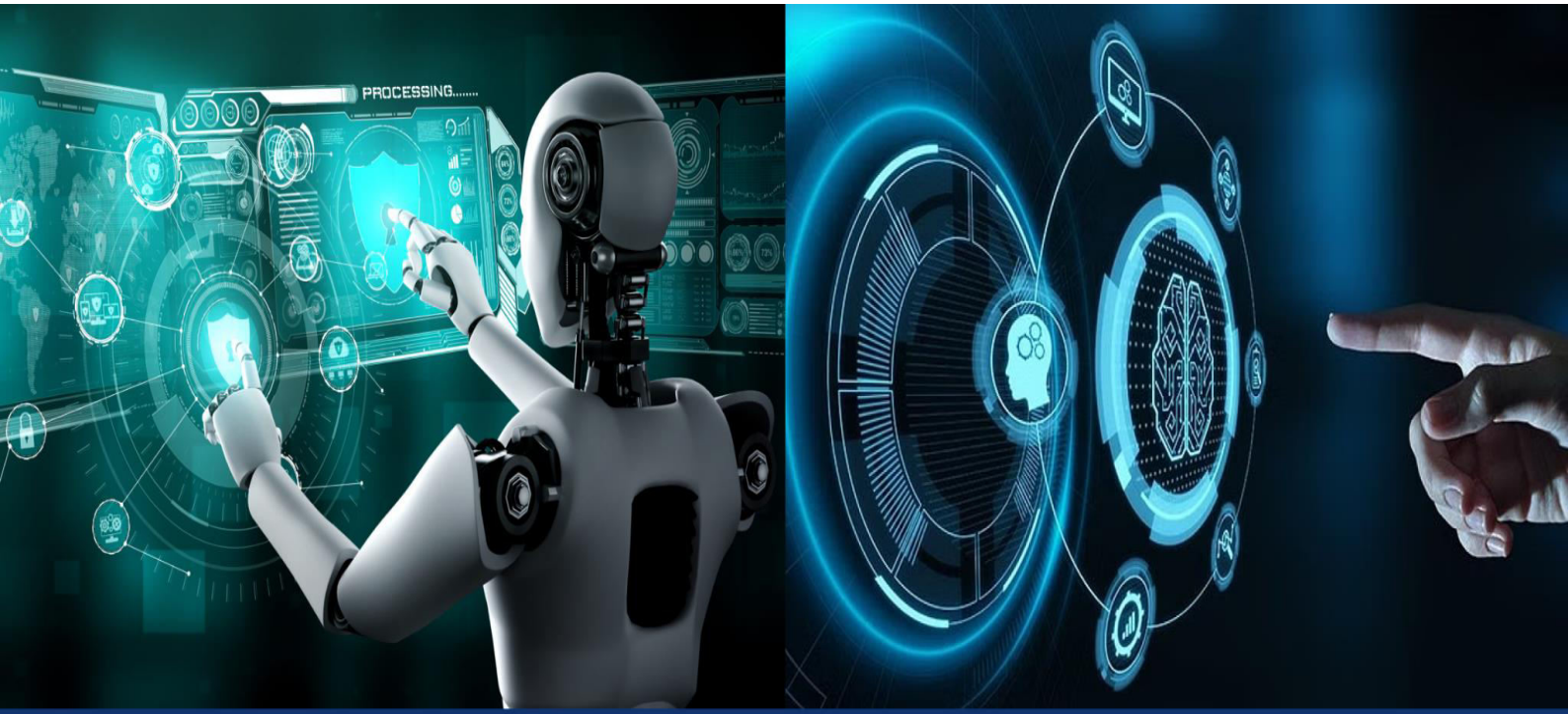
ISSN(O): 2320-9801

ISSN(P): 2320-9798



International Journal of Innovative Research in Computer and Communication Engineering

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)



Impact Factor: 8.771

Volume 14, Issue 4, April 2026

© 2026 IJIRCCE | Volume 14, Issue 4, April 2026 DOI: 10.15680/IJIRCCE.2026.1404069

www.ijircce.com | e-ISSN: 2320-9801, p-ISSN: 2320-9798| Impact Factor: 8.771| ESTD Year: 2013|

International Journal of Innovative Research in Computer and Communication Engineering (IJIRCCE)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

Data Balancing and CNN Based Intrusion Detection System

R.Siva Lakshmi[1], K. Rajeswari[2]

Student, Department of Computer Science and Engineering, Andhra Loyola Institute of Engineering and Technology, ITI Road, Vijayawada, Andhra Pradesh, India[1]

Assistant Professor, Department of Computer Science and Engineering, Andhra Loyola Institute of Engineering and Technology, ITI Road, Vijayawada, Andhra Pradesh, India[2]

ABSTRACT: The help of an automated process that filters and classifies network intrusions is often needed by cyber-security professionals. The classification of the attack type is essential for applying specific preventive measures to secure networks. Numerous Machine Learning (ML) models have been proposed as the foundation for Network Intrusion Detection (NID) systems. Yet, their efficacy varies based on many factors. For instance, an ML model trained on a highly unbalanced dataset may be biased towards over-represented attack types. On the other hand, focusing solely on the ML model's performance in minority classes can have a negative impact on its performance in the majority classes. We proposes a Network Intrusion Detection (NID) system that addresses the issue of imbalanced datasets and uses Convolutional Neural Networks (CNN) to classify different attack types. The performance of the proposed system is compared to other systems that use different techniques such as Random Over-Sampling (ROS), Synthetic Minority Oversampling Technique (SMOTE), Adaptive Synthetic Sampling (ADASYN), and Generative Adversarial Networks (GAN) for data balancing.The NSL-KDD and BoT-IoT datasets are used for benchmarking, and the results show that the proposed system performs well in the minority classes on the binary classification task. Our proposed system scores a good weighted average F1-Score on the multi-class classification task using the BoT-IoT dataset.

I. INTRODUCTION

Cloud computing and Internet of Things (IoT) technologies, and the generations of wireless technologies are advancing expeditiously. With the help of these advanced technologies, millions of users and devices are interconnected. This creates more opportunities for cyber-security attackers to target more victims. Securing users' information and protecting the IoT devices is crucial for the continuation of the communication process. Knowing that some of the targeted systems may have a strong Network Intrusion Detection (NID) system, cyber-security attackers use reformed attack methods. Therefore, a well performing NID system must be able to distinguish new attacks even if it has not seen any or many of them.Many machine learning (ML) based NID systems have been introduced recently. However, while implementing such systems, ML engineers have to address several issues. For instance, fitting models on an imbalanced dataset may result in a high False Alarm Rate (FAR) on the minority classes.

II. LITERATURE SURVEY

Network security has become a critical concern due to the rapid growth of cyber threats and vulnerabilities. Intrusion Detection Systems (IDS) are widely used to monitor network traffic and identify malicious activities. However, one of the major challenges in IDS is the presence of imbalanced datasets, where certain attack classes have significantly fewer samples compared to normal traffic. This study addresses the issue of class imbalance in intrusion detection by using regularized Wasserstein Generative Adversarial Networks (WGAN-IDR). The model generates synthetic samples for minority classes, resulting in a balanced dataset. Experiments conducted on the CICIDS2017 dataset using multiple classification strategies demonstrate improved performance, achieving high F1-scores of 0.99 for binary classification and 0.98 for multiclass classification.

Title: "A Fast Network Intrusion Detection System Using Adaptive Synthetic Oversampling and LightGBM"
This research proposed a hybrid intrusion detection system combining ADASYN oversampling and the LightGBM model. Data preprocessing techniques such as normalization and encoding are applied, followed by synthetic data

© 2026 IJIRCCE | Volume 14, Issue 4, April 2026 DOI: 10.15680/IJIRCCE.2026.1404069

www.ijircce.com | e-ISSN: 2320-9801, p-ISSN: 2320-9798| Impact Factor: 8.771| ESTD Year: 2013|

International Journal of Innovative Research in Computer and Communication Engineering (IJIRCCE)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

generation to address class imbalance. The system achieves high accuracy across multiple datasets, including NSL-KDD, UNSW-NB15, and CICIDS2017, while also reducing computational complexity

III. PROPOSED SYSTEM

We proposes a Network Intrusion Detection (NID) system that addresses the issueof imbalanced datasets and uses Convolutional Neural Networks (CNN) to classify different attack types. The performance of the proposed system is compared to other systems that use different techniques such as Random Over-Sampling (ROS), Synthetic Minority Oversampling Technique (SMOTE), Adaptive Synthetic Sampling (ADASYN), and Generative Adversarial Networks (GAN) for data balancing.To validate the aforementioned contribution, we use two highly imbalanced datasets such as the NSL-KDD dataset and the BoT-IoT dataset.

FUNCTIONAL REQUIREMENTS

1. Data Collection
2. Data Pre-processing
3. Training and Testing
4. Modiling

NON FUNCTIONAL REQUIREMENTS

NON-FUNCTIONAL REQUIREMENT (NFR) specifies the quality attribute of a software system. They judge the software system based on Responsiveness, Usability, Security, Portability and other non-functional standards that are critical to the success of the software system. Example of nonfunctional requirement, “how fast does the website load?” Failing to meet non-functional requirements can result in systems that fail to satisfy user needs. Non- functional Requirements allow you to impose constraints or restrictions on the design of the system across the various agile backlogs. Example, the site should load in 3 seconds when the number of simultaneous users is > 10000. Description of non-functional requirements is just as critical as a functional requirement.

- Usability requirement
- Serviceability requirement
- Manageability requirement
- Recoverability requirement
- Security requirement
- Data Integrity requirement
- Capacity requirement
- Availability requirement
- Scalability requirement
- Interoperability requirement
- Reliability requirement
- Maintainability requirement
- Regulatory requirement
- Environmental requirement

IV. METHODOLOGY

Data Collection

- Use intrusion detection dataset (NSL-KDD / CICIDS2017).

Data Preprocessing

- Clean data, convert text to numbers
- Normalize values
- Split into training and testing data

© 2026 IJIRCCE | Volume 14, Issue 4, April 2026 DOI: 10.15680/IJIRCCE.2026.1404069

www.ijircce.com | e-ISSN: 2320-9801, p-ISSN: 2320-9798| Impact Factor: 8.771| ESTD Year: 2013|

International Journal of Innovative Research in Computer and Communication Engineering (IJIRCCE)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

Data Balancing
- Use SMOTE or sampling methods
- Balance normal and attack data

Model Building (CNN–LSTM)
- CNN extracts important features
- LSTM learns sequence patterns
- Combine both for better detection

Model Training
- Train model using training data
- Use optimizer like Adam

Model Testing
- Test using test data

Classify as Normal or Attack

Evaluation
- Check Accuracy, Precision, Recall

Result
- Improved intrusion detection performance

V. PROPOSED SYSTEM HARDWARE RESULTS

System testing, also referred to as system-level tests or system-integration testing, is the process in which a quality assurance (QA) team evaluates how the various components of an application interact together in the full, integrated system or application. System testing verifies that an application performs tasks as designed. This step, a kind of black box testing, focuses on the functionality of an application. System testing, for example, might check that every kind of user input produces the intended output across the application.

Phases of system testing:
A video tutorial about this test level. System testing examines every component of an application to make sure that they work as a complete and unified whole. A QA team typically conducts system testing after it checks individual modules with functional or user-story testing and then each component through integration testing.

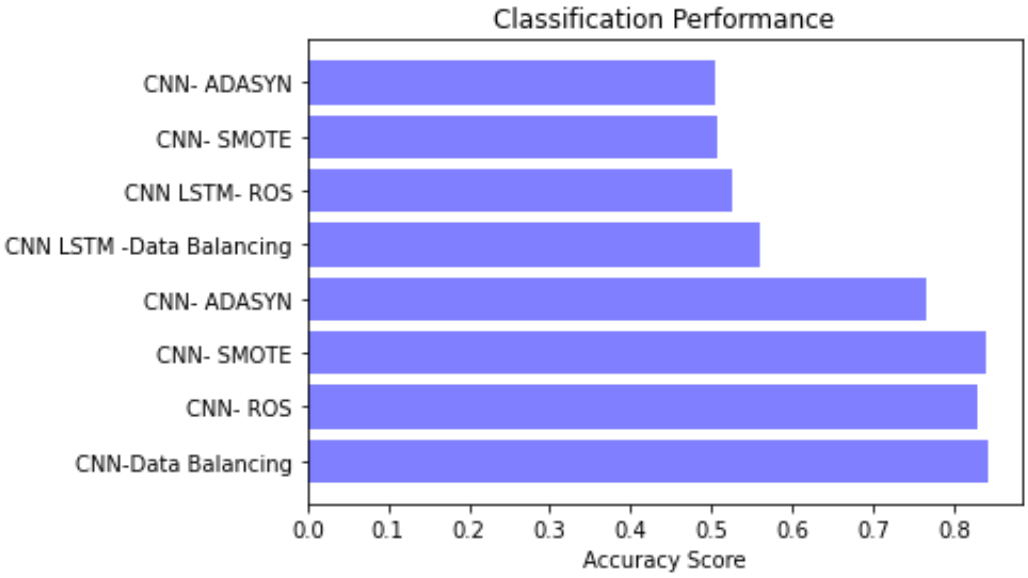
Fig 2.

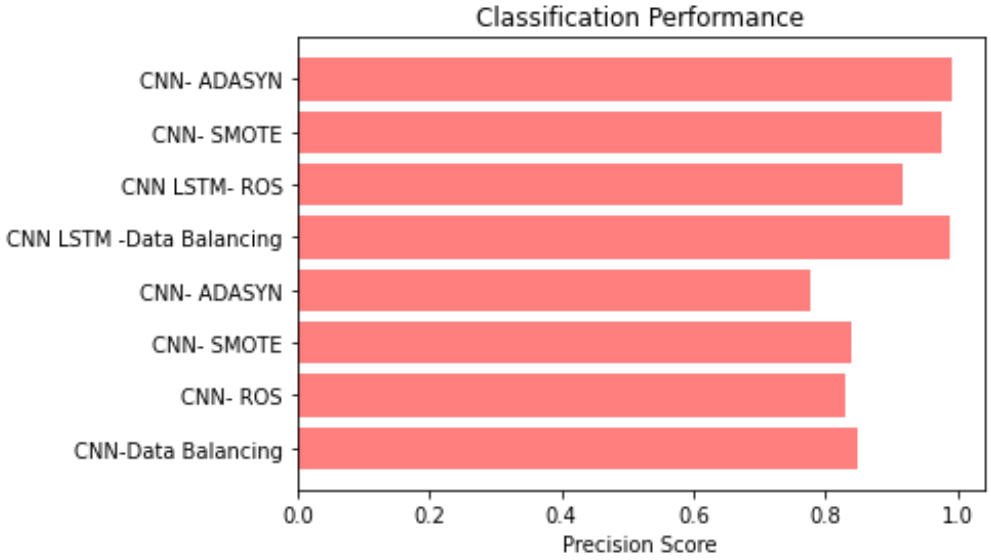
© 2026 IJIRCCE | Volume 14, Issue 4, April 2026 DOI: 10.15680/IJIRCCE.2026.1404069

www.ijircce.com | e-ISSN: 2320-9801, p-ISSN: 2320-9798| Impact Factor: 8.771| ESTD Year: 2013|

International Journal of Innovative Research in Computer and Communication Engineering (IJIRCCE)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)





VI. CONCLUSION

The proposed NID system that addresses the issue of imbalanced datasets and uses Convolutional Neural Networks (CNN) to classify different attack types performs well in the minority classes while maintaining an good recall of on the binary classification task. The proposed system is compared to other systems that use different techniques for data balancing, and the results show that the proposed system outperforms them. Future work can focus on improving the proposed system's performance by using more advanced techniques for data balancing and feature extraction

REFERENCES

[1] Y. Yang, K. Zheng, et al., "Improving the classification effectiveness of intrusion detection by using improved conditional variational autoencoder and deep neural network," Sensors, vol. 19, no. 11, 2019.
[2] A. Fatani, M. Abd Elaziz, et al., "Iot intrusion detection system using deep learning and enhanced transient search optimization," IEEE Access, vol. 9, pp. 123448–123464, 2021.

www.ijircce.com | e-ISSN: 2320-9801, p-ISSN: 2320-9798| Impact Factor: 8.771| ESTD Year: 2013|

International Journal of Innovative Research in Computer and Communication Engineering (IJIRCCE)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

[3] N. Gupta, V. Jindal, and P. Bedi, "Lio-ids: Handling class imbalance using lstm and improved one-vs-one technique in intrusion detection system," Computer Networks, vol. 192, p. 108076, 2021.
[4] K. Jiang, W. Wang, A. Wang, and H. Wu, "Network intrusion detection combined hybrid sampling with deep hierarchical network," IEEE Access, vol. 8, pp. 32464–32476, 2020.
[5] R. Chapaneri and S. Shah, "Enhanced detection of imbalanced malicious network traffic with regularized generative adversarial networks," Journal of Network and Computer Applications, vol. 202, p. 103368, 2022.
[6] H. Ding et al., "Imbalanced data classification: A knn and generative adversarial networks-based hybrid approach for intrusion detection," Future Generation Computer Systems, vol. 131, pp. 240–254, 2022.
[7] X. Zhang, J. Ran, and J. Mi, "An intrusion detection system based on convolutional neural network for imbalanced network traffic," in IEEE 7th International Conference on Computer Science and Network Tech. (ICCSNT), pp. 456–460, 2019.
[8] J. Liu, Y. Gao, and F. Hu, "A fast network intrusion detection system using adaptive synthetic oversampling and lightgbm," Computers & Security, vol. 106, p. 102289, 2021.
[9] B. A. Tama and K. H. Rhee, "An in-depth experimental study of anomaly detection using gradient boosted machine," Neural Computing and Applications, vol. 31, pp. 955–965, 2017.
[10] Y. Yang, K. Zheng, B. Wu, Y. Yang, and X. Wang, "Network intrusion detection based on supervised adversarial variational auto-encoder with regularization," IEEE Access, vol. 8, pp. 42169–42184, 2020.
[11] M. Tavallaee et al., "A detailed analysis of the kdd cup 99 data set," in 2009 IEEE Symposium on Computational Intelligence for Security and Defense Applications, pp. 1–6, 2009.
[12] N. Koroniotis, N. Moustafa, et al., "Towards the development of realistic botnet dataset in the internet of things for network forensic analytics: Bot-iot dataset," CoRR, vol. abs/1811.00701, 2018.
[13] A. Divekar et al., "Benchmarking datasets for anomaly-based network intrusion detection: Kdd cup 99 alternatives," in IEEE 3rd Int. Conf. on Computing, Communication and Security (ICCCS), pp. 1–8, 2018.
[14] S. Huang and K. Lei, "Igan-ids: An imbalanced generative adversarial network towards intrusion detection system in ad-hoc networks," Ad Hoc Networks, vol. 105, p. 102177, 2020.
[15] O. Elghalhoud, K. Naik, et al., "Data balancing and hyper-parameter optimization for machine learning algorithms for secure iot networks," In Proceedings of the 18th ACM Symposium on QoS and Security for Wireless and Mobile Networks (Q2SWinet '22), 2022.
[16] Z. Li, Qin, et al., "Intrusion detection using convolutional neural networks for representation learning," in Neural Information Processing, (Cham), pp. 858–866, Springer International Publishing, 2017.
[17] B. A. Tama, M. Comuzzi, and K.-H. Rhee, "Tse-ids: A two-stage classifier ensemble for intelligent anomaly-based intrusion detection system," IEEE Access, vol. 7, pp. 94497–94507, 2019.

INTERNATIONAL STANDARD SERIAL NUMBER INDIA

INNO SPACE SJIF Scientific Journal Impact Factor

doi crossref

निस्केयर NISCAIR

INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN COMPUTER & COMMUNICATION ENGINEERING

9940 572 462 6381 907 438 ijircce@gmail.com

www.ijircce.com

Scan to save the contact details