Handling Class Imbalance in Binary and Multiclass Intrusion Detection on the UNSW-NB15 Dataset Using Classical Machine Learning

By

Ikramul Hasan Moral (0112230489)
Md. Abu Bakar (0112230200)
Samiur Rahman Omlan (0112230195)
Ahamudul Hasan Prianto (0112230300)
Salman Hossain (0112230328)

Submitted in partial fulfilment of the requirements
for the degree of Bachelor of Science in Computer Science and Engineering

January 15, 2026

Department of Computer Science and Engineering
United International University

Abstract

The abstract should contain a summary of the work presented in this report in a single paragraph. The paragraph should provide a general background of the problem, methodology and results.

Table of Contents

Chapter 1

Introduction

This chapter establishes the foundation of the study by outlining the context of network intrusion detection and the specific challenges addressed in this research.

1.1 Background

Network security has become a critical priority as cyber threats continue to evolve in complexity and frequency. To counter these threats, Machine Learning (ML) is widely employed to design Intrusion Detection Systems (IDS) capable of identifying malicious traffic. The efficacy of these systems relies heavily on the quality of the datasets used for training. While older datasets like KDD99 and NSL-KDD were once standard, they are now considered outdated because they contain duplicate records and lack modern attack patterns [1, 2].

Consequently, the UNSW-NB15 dataset has emerged as a modern benchmark for evaluating IDSs. Developed by the Australian Centre for Cyber Security, this dataset reflects real-world network traffic and includes nine contemporary attack types [3]. Statistical analysis demonstrates that this dataset is significantly more complex and harder to classify than its predecessors due to its non-linear distribution [4].

1.2 Problem Statement

The primary challenge in developing effective IDSs using the UNSW-NB15 dataset is the severe class imbalance [5]. In this dataset, normal traffic heavily outnumbers malicious traffic, and specific attack categories—such as Worms, Shellcode, and Backdoors—appear in extremely small quantities.

Most standard machine learning algorithms are designed to maximize overall accuracy. When applied to imbalanced data, these models tend to bias toward the majority class (Normal traffic) while failing to detect minority attack types [1]. While dimensionality reduction methods like PCA and Autoencoders improve computational efficiency, they do not directly solve this imbalance problem [6]. Furthermore, Cost-Sensitive Learning

provides higher penalties for misclassifying minority classes, but its performance depends heavily on classifier tuning [7].

To address these limitations, synthetic oversampling techniques such as SMOTE have been introduced to generate new minority samples rather than simply duplicating existing ones, making them more effective for imbalanced data [8, 9].

1.3 Motivation

The motivation for this study arises from the limitations observed in current literature. Many recent studies report binary accuracy rates exceeding 99% using ensemble methods. For instance, Primartha and Tama reported high accuracy using Random Forest [10], and Amin et al. achieved 99.28% accuracy in cloud environments [11]. Similarly, More et al. demonstrated state-of-the-art binary accuracy using optimized feature selection [12].

However, these metrics can be misleading. Studies focusing on feature selection often see a significant drop in multiclass performance compared to binary classification [13]. Furthermore, deep learning approaches, while powerful, often fail to report detailed recall rates for rare attack categories [14, 15]. There is a clear need to rigorously investigate advanced imbalance-handling strategies to ensure that modern IDSs can detect rare attack categories effectively [16].

1.4 Objective

The primary objective of this research is to improve the multiclass detection performance of machine learning models on the UNSW-NB15 dataset. Specifically, this study aims to:

- Analyze the impact of class imbalance on the detection of minority attack categories.
- Evaluate and compare the effectiveness of imbalance-handling strategies, specifically Cost-Sensitive Learning and SMOTE [8].
- Establish a reliable baseline for multiclass intrusion detection that prioritizes the identification of rare attacks over simple binary accuracy.

1.5 Contribution

This study contributes to the field of Network Intrusion Detection through the following:

- **Systematic Evaluation of Imbalance Strategies:** We provide a comparative analysis of how different techniques (No Balancing, Class Weighting, and SMOTE) affect model performance on a modern benchmark dataset.
- **Focus on Rare Attack Detection:** Unlike prior works that prioritize binary accuracy, this study explicitly analyzes performance metrics (Precision, Recall, F1-score) for minority classes such as Worms and Shellcode.

- **Adoption of Comprehensive Metrics:** We utilize G-Mean and per-class metrics to provide a fair assessment of model reliability, addressing the pitfalls of using standard accuracy for skewed datasets.

Chapter 2

Literature Review

This chapter provides an overview of the essential concepts, existing research, and gaps that motivate our work.

Older datasets like KDD99 and NSL-KDD are now considered outdated because they contain duplicate records and lack modern attack patterns [3, 17]. For this reason, we selected the UNSW-NB15 dataset, a modern benchmark that includes nine types of contemporary attacks [3]. Statistical analysis shows that this dataset is complex and much harder to classify than KDD99 due to its non-linear distribution [4]. Surveys confirm that UNSW-NB15 is a reliable choice for research, though they also note it is highly imbalanced [2]. Unlike KDD99, it uses consistent attack types in both training and testing sets, ensuring fair results [17]. Recent studies indicate that because of this extreme imbalance, hybrid resampling is needed to detect minority attacks effectively [1].

Across the reviewed studies, class imbalance is identified as the main challenge in network intrusion detection datasets, as minority attacks occur in very small quantities [1]. Dimensionality-reduction methods such as PCA and Autoencoders improve computational efficiency but do not directly solve the imbalance problem [6]. Cost-Sensitive Learning provides higher penalties for misclassifying minority classes, yet its performance heavily depends on classifier tuning and dataset characteristics [7]. Simple oversampling or undersampling often leads to overfitting or information loss, making them less reliable for intrusion detection [1]. SMOTE, introduced as a synthetic oversampling technique, generates new minority samples rather than duplicating or discarding existing ones, making it more effective for imbalanced data [8]. Studies using SMOTE on modern datasets like CSE-CIC-IDS2018 show 4–30 % improvement in minority attack detection [9]. This improvement is especially evident for rare attacks such as infiltration or botnet traffic [9]. SMOTE also integrates well with a wide range of machine-learning models, including Random Forest, Gradient Boosting, and KNN [9]. Compared to CSL and dimensionality-reduction approaches, SMOTE more consistently improves recall and balanced accuracy for IDS tasks [7]. Overall, across all four papers, SMOTE stands out as the most effective

method for handling data imbalance in intrusion detection systems [8].

The UNSW-NB15 dataset has emerged as a standard benchmark for evaluating modern Intrusion Detection Systems (IDS). However, most existing research prioritizes overall binary accuracy rather than the granular detection of individual attack types. Ensemble methods, particularly Random Forest (RF), have consistently demonstrated strong performance in binary classification tasks. For instance, Primartha and Tama(2017) [10] reported that an ensemble of 800 trees achieved an accuracy of 95.5% and a false alarm rate (FAR) of 7.22%. Building on this, Amin et al.(2021) [11] applied Bagging and Random Forest models with ANOVA-based feature selection in cloud environments, achieving a binary accuracy of 99.28%. Despite these promising results, both studies focused exclusively on binary classification (Normal vs. Attack), leaving the poor detection rates of minority attack categories largely unaddressed.

To mitigate the high dimensionality of network traffic features, several studies have shifted toward feature selection rather than data resampling. More et al.(2024) [12] employed correlation-based feature selection and demonstrated that an optimized Random Forest configuration reached a state-of-the-art binary accuracy of 99.45%. Although the study acknowledged the challenges posed by class imbalance, the proposed methodology relied primarily on feature reduction rather than directly addressing skewed class distributions. Kasongo and Sun(2020) [13] further highlighted the limitations of this approach. Their XGBoost-based feature selection improved Decision Tree binary accuracy to 90.85%, yet multiclass performance remained significantly lower at 67.57%. Notably, their analysis showed that models such as Artificial Neural Networks (ANN) performed poorly on minority attack categories—including Worms and Shellcode—and emphasized that future work should integrate synthetic oversampling techniques to enhance minority class detection.

Motivated by these gaps, the present study systematically investigates imbalance-handling strategies explicitly recommended in prior work, aiming to improve multiclass detection performance on the UNSW-NB15 dataset.

Vinayakumar et al. 2019[14] goal is to create Binary and Multi-class classification high accuracy. Resulting in deep learning using many datasets including UNSW NB-15. But it did not deeply analyze imbalanced class nor did show rare attacks. Vibhute et al. 2024 [15] it aims to provide CNN feature and give results on accuracy 99%, precision 99.03% , recall 98.86% , F-1 score 99% and show good architecture performance. It did not published explicit per rare attack recall numbers. Al-Qarni et al. 2024[18] focus on oversampling methods like SMOTE and ADASYN using ML models and it boosted accuracy to ADASYN 92.28% over SMOTE 88.13%.Its limitation is oversampling and unsolved issues as noise and evolving threats. Syed Ibrahim Imtiaz 2022[19] its mainly Focus on binary anomaly vs normal classification by using the UNSW-NB15. DT and RF have 99% accuracy. This relies on the overall accuracy, which can be misleading when minority classes are rare.

Modern benchmarks dataset like UNSW-NB15 are highly valuable for reproducible NIDS research [20] but the persistent issue of class imbalance still remains to some extend, specially in detecting rare attacks such as worms and shellcode which continues to hinder the performance of contemporary intrusion detection systems [5]. Although recent studies increasingly rely on complex deep learning architectures [21], and some work has explored one-class anomaly detection [22], evidence suggests that applying targeted technical optimizations to classical machine learning models can still be highly effective. By rigorously applying and evaluating cost-sensitive learning and resampling techniques, this study establishes an essential baseline for multiclass intrusion detection[16].

Chapter 3

Methodology

This chapter outlines the steps taken to build and evaluate the intrusion detection system. The process begins with selecting the dataset, followed by cleaning and preparing the data, training the models using different strategies to handle class imbalance, and finally evaluating the results.

3.1 Dataset Description

We selected the **UNSW-NB15** dataset for this research [3]. This dataset was created by Moustafa and Slay to address the limitations of older datasets like KDD99. KDD99 is often considered outdated because it contains duplicate records and does not include modern attack patterns [4]. In contrast, UNSW-NB15 includes nine types of contemporary attacks, making it a more realistic benchmark for current network security. Surveys of intrusion detection datasets confirm that UNSW-NB15 is a reliable choice for research, though it is noted to be highly imbalanced, which makes classification difficult [20].

3.2 Data Preprocessing

Raw network data cannot be used directly by machine learning algorithms. We applied a consistent preprocessing pipeline to clean and format the data.

3.2.1 Feature Cleaning and Selection

The first step involved removing data columns that are not useful for detecting attacks. We dropped identifiers such as IP addresses and timestamps. These features are specific to the network setup and do not help the model learn general attack patterns. We utilized feature selection concepts discussed by Janarthanan and Zargari to ensure only relevant information remained [17].

3.2.2 Missing Values

Real-world data often has gaps. To handle this, we checked for missing values in the dataset. For numerical features, we filled missing entries with the median value of that column. For categorical features, we replaced missing entries with a placeholder category labeled "missing".

3.2.3 Labeling (Binary and Multiclass)

We organized the experiment into two distinct tracking tasks:

- **Task A (Binary Classification):** Two labels were used — '0' for Normal traffic and '1' for Attack traffic.
- **Task B (Multiclass Classification):** We retained the detailed labels for all nine attack categories along with Normal traffic.

3.2.4 Data Scaling

Since different features have different ranges, we scaled the numeric features. We used standard scaling to ensure the values fall within a similar range. This step is critical for models like Logistic Regression but less critical for tree-based models.

3.2.5 Splitting Into Train, Validation, and Test

We adhered to the standard training and testing splits provided by the UNSW-NB15 authors. Within the training set, we created a validation split (80% training, 20% validation) to tune model settings before final testing.

3.3 System Flow

The system follows a structured pipeline from raw data to final evaluation. The experimental setup compares different strategies to handle the dataset imbalance.

Experimental Strategies:

1. **S0 (No Balancing):** Baseline model training without modifying class distribution.
2. **S1 (Class Weighting):** Applying higher misclassification penalties for minority classes (cost-sensitive learning) [7].
3. **S2 (Oversampling):** Increasing the number of minority attack samples through resampling techniques [1].

3.4 Technology Used

The implementation was carried out using **Python**. We utilized the **scikit-learn** library for preprocessing, Logistic Regression, and Random Forest models. For gradient boosting, we used **XGBoost**. Oversampling was implemented using the **imbalanced-learn** library.

3.5 Evaluation Metrics

To evaluate the performance of our models, especially for minority attack classes, we used the following metrics:

- **Accuracy:** Overall correctness of predictions.
- **Macro F1-Score:** Average F1 score across all classes, treating each class equally.
- **ROC-AUC:** Ability of the model to distinguish between classes across thresholds.
- **G-Mean:** Balances recall of both majority and minority classes, useful for imbalanced datasets [16].
- **Confusion Matrix:** Summarizes correct and incorrect predictions for each class.

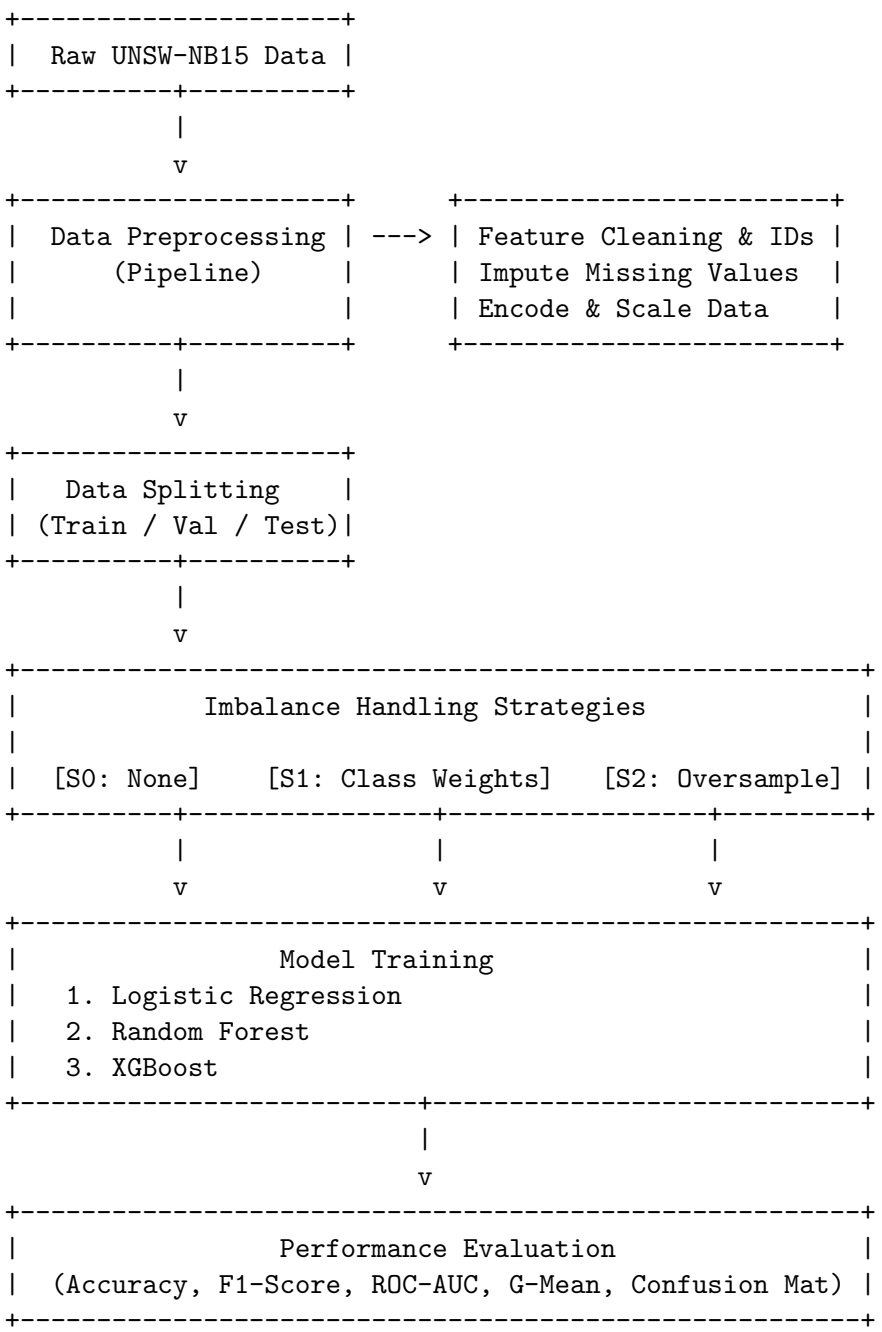

Figure 3.1: System Flow Diagram

Chapter 4

Implementation and Results

4.1 Environment Setup

4.2 Testing and Evaluation

4.3 Results and Discussion

4.4 Summary

Chapter 5

Conclusion

5.1 Summary

5.2 Limitation

5.3 Future Work

References

[1] Sikha Bagui and Kunqi Li. Resampling imbalanced data for network intrusion detection datasets. *Journal of Big Data*, 8(1):6, 2021.

[2] Markus Ring, Sarah Wunderlich, Deniz Scheuring, Dieter Landes, and Andreas Hotho. A Survey of Network-based Intrusion Detection Data Sets. *arXiv e-prints*, page arXiv:1903.02460, March 2019.

[3] Nour Moustafa and Jill 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)*, pages 1–6. IEEE, 2015.

[4] Nour Moustafa and Jill Slay. The evaluation of network anomaly detection systems: Statistical analysis of the unsw-nb15 data set and the comparison with the kdd99 data set. *Information Security Journal: A Global Perspective*, 25(1-3):18–31, 2016.

[5] Vaishnavi Shanmugam, Roozbeh Razavi-Far, and Ehsan Hallaji. Addressing class imbalance in intrusion detection: a comprehensive evaluation of machine learning approaches. *Electronics*, 14(1):69, 2024.

[6] Razan Abdulhammed, Hassan Musafer, Ali Alessa, Miad Faezipour, and Abdelshakour Abuzneid. Features dimensionality reduction approaches for machine learning based network intrusion detection. *Electronics*, 8(3), 2019.

[7] Nguyen Thai-Nghe, Zeno Gantner, and Lars Schmidt-Thieme. Cost-sensitive learning methods for imbalanced data. In *The 2010 International Joint Conference on Neural Networks (IJCNN)*, pages 1–8, 2010.

[8] Nitesh V Chawla, Kevin W Bowyer, Lawrence O Hall, and W Philip Kegelmeyer. Smote: synthetic minority over-sampling technique. *Journal of artificial intelligence research*, 16:321–357, 2002.

[9] Gozde Karatas, Onder Demir, and Ozgur Koray Sahingoz. Increasing the performance of machine learning-based idss on an imbalanced and up-to-date dataset. *IEEE Access*, 8:32150–32162, 2020.

[10] Rifkie Primartha and Bayu Adhi Tama. Anomaly detection using random forest: A performance revisited. In *2017 International conference on data and software engineering (ICoDSE)*, pages 1–6. IEEE, 2017.

[11] Uzma Amin, Aamir S Ahanger, F Masoodi, and AM Bamhdi. Ensemble based effective intrusion detection system for cloud environment over unsw-nb15 dataset. In *Scrs Conf. Proc. Intell. Syst*, pages 483–494, 2021.

[12] Shweta More, Moad Idrissi, Haitham Mahmoud, and A Taufiq Asyhari. Enhanced intrusion detection systems performance with unsw-nb15 data analysis. *Algorithms*, 17(2):64, 2024.

[13] Sydney M Kasongo and Yanxia Sun. Performance analysis of intrusion detection systems using a feature selection method on the unsw-nb15 dataset. *Journal of Big Data*, 7(1):105, 2020.

[14] R. Vinayakumar, Mamoun Alazab, K. P. Soman, Prabaharan Poornachandran, Ameer Al-Nemrat, and Sitalakshmi Venkatraman. Deep learning approach for intelligent intrusion detection system. *IEEE Access*, 7:41525–41550, 2019. Received December 27, 2018, accepted January 3, 2019, date of current version April 11, 2019.

[15] Amol D. Vibhute. Network anomaly detection and performance evaluation of convolutional neural networks on unsw-nb15 dataset. *Procedia Computer Science*, 235:261–268, 2024.

[16] Mithilesh Kumar Choudhary and Atul Kumar Mishra. Review paper on imbalanced network based intrusion detection system using deep learning technique. *International Journal of Advanced Research and Multidisciplinary Trends (IJARMT)*, 2(4):257–264, 2025.

[17] Tharmini Janarthanan and Shahrzad Zargari. Feature selection in unsw-nb15 and kddcup'99 datasets. In *2017 IEEE 26th International Symposium on Industrial Electronics (ISIE)*, pages 1881–1886, 2017.

[18] Elham Abdullah Al-Qarni. Addressing imbalanced data in network intrusion detection: A review and survey. *International Journal of Advanced Computer Science and Applications*, 15(2), 2024.

[19] Syed Ibrahim Imtiaz, Liaqat Ali Khan, Ahmad S. Almadhor, Sidra Abbas, Shtwai Alsubai, Michal Gregus, and Zunera Jalil. Efficient approach for anomaly detection in internet of things traffic using deep learning. *Concurrency and Computation: Practice and Experience*, 34(26):e8266347, 2022.

[20] Markus Ring, Sarah Wunderlich, Deniz Scheuring, Dieter Landes, and Andreas Hotho. A survey of network-based intrusion detection data sets. *Computers & security*, 86:147–167, 2019.

[21] Ghada Abdelmoumin, Jessica Whitaker, Danda B Rawat, and Abdul Rahman. A survey on data-driven learning for intelligent network intrusion detection systems. *Electronics*, 11(2):213, 2022.

[22] Paulina Arregoces, Jaime Vergara, Sergio Armando Gutiérrez, and Juan Felipe Botero. Network-based intrusion detection: A one-class classification approach. In *NOMS*, volume 2022, pages 1–6, 2022.