National College of Ireland

Novel AI-Driven Approach to cloud intrusion detection using Boruta and PSO feature selection

MSc Research Project
MSc in Cybersecurity

Tarankaarthikeshwar Srinivasan
Student ID: 23267941

School of Computing
National College of Ireland

Supervisor: Michael Pantridge

National College of Ireland

MSc Project Submission Sheet

School of Computing

Student Name:	Tarankaarthikeshwar Srinivasan
Student ID:	23267941
Programme:	MSc in Cybersecurity **Year:** 2025
Module:	MSc Research Project
Supervisor:	Michael Pantridge
Submission Due Date:	15/09/2025
Project Title:	Novel AI-Driven Approach to Cloud Intrusion Detection Using Boruta-PSO Feature Selection
Word Count:	4280 **Page Count** 16

I hereby certify that the information contained in this (my submission) is information pertaining to research I conducted for this project. All information other than my own contribution will be fully referenced and listed in the relevant bibliography section at the rear of the project.
ALL internet material must be referenced in the bibliography section. Students are required to use the Referencing Standard specified in the report template. To use other author's written or electronic work is illegal (plagiarism) and may result in disciplinary action.

Signature:	Tarankaarthikeshwar Srinivasan
Date:	15/09/2025

PLEASE READ THE FOLLOWING INSTRUCTIONS AND CHECKLIST

Attach a completed copy of this sheet to each project (including multiple copies)	□
Attach a Moodle submission receipt of the online project submission, to each project (including multiple copies).	□
You must ensure that you retain a HARD COPY of the project, both for your own reference and in case a project is lost or mislaid. It is not sufficient to keep a copy on computer.	□

Assignments that are submitted to the Programme Coordinator Office must be placed into the assignment box located outside the office.

Office Use Only	
Signature:	
Date:	
Penalty Applied (if applicable):	

Novel AI-Driven Approach to Cloud Intrusion Detection Using Boruta-PSO Feature Selection

Tarankaarthikeshwar Srinivasan
23267941

Abstract

Nowadays, cloud computing is essential part of modern digital infrastructure, it is targeted by many different types of cyber-attacks. For these environments to be secure, intrusion detection systems (IDS) must be quick and effective. This study presents a scalable and reliable IDS architecture that uses sophisticated feature engineering and data simplification to improve detection accuracy. To decrease complexity, the datasets of CICIDS 2017 which contains comprehensive assault labels, are first categorized into more general, understandable groups. To save only the most pertinent features, a hybrid feature selection method that combines Boruta and Particle Swarm Optimization (PSO) is utilized. The Synthetic Minority Over-Sampling Technique (SMOTE) is used to address the issue of class imbalance. Several machine learning models, such as Random Forest, LightGBM, and XGBoost, are subsequently trained using the improved dataset.

1 Introduction

These days, cloud computing is essential for various service-oriented organizations. Nowadays firms gain improved accessibility, flexibility, resources, and cost savings by storing essential services and data storage to the cloud. On the contrary, the cloud paradigm presents new security issues. Cloud infrastructures, unlike conventional systems, are distributed, dynamic, and commonly prone to advanced cyber threats such as insider assaults, zero-day exploits, and data exfiltration. Ensuring strong cloud security is crucial for financial institutions, which handle highly sensitive information such as client identities, transaction histories, and regulatory documents.

Cloud computing is now the foundation of advanced digital infrastructure, and security in this enhanced environment is more crucial and challenging than ever. Because cloud platforms store enormous volumes of data, apps, and services, they are often the focus of various cyberthreats. Cloud-based Intrusion Detection Systems (IDS) must be able to swiftly and precisely detect and categorize threats to protect these dynamic and dispersed environments. Cloud security is complicated by complexity and depth of raw security logs and data. The logs often contain various attack labels, like Web Attack SQL Injection, DoS Hulk, and FTP-Patator. Despite being complex, these labels lead to comprehensive or duplicate categories which reduce the amount of effective analysis.

Existing intrusion detection systems mainly depend on static rules and signatures. This results in the failure to identify new threats, particularly in settings with lots of traffic and an extensive number of protocols. In addition, these systems often do not have the agility

necessary for real-time operations to process and evaluate data. The integration of artificial intelligence into cloud security frameworks has generated a lot of interest as a solution to these constraints.

An AI-enhanced intrusion detection and response system designed for cloud ecosystems is put forward in the present research. The proposed system implements enhanced feature selection, data balancing, and multi-model learning techniques to improve detection rates and response efficiency. The input feature set using Boruta and PSO is improved, which mitigates imbalance in attack data through SMOTE, this improvised model enhances learning accuracy and reduces computational overhead. The use of CICIDS 2017 dataset ensures realistic evaluation conditions, hence it simulates modern traffic patterns and attack scenarios. The proposed framework concerns to bridge the gap between the need for security robustness and agility of financial cloud services.

The proposed approach provides a Systematic classification of cyber security attacks in cloud systems is study. Combining fine-grained threat labels into more generic categories such as Brute Force, DoS/DDoS, or Web Attack makes threat detection processes more transparent and effective. This provides support to cloud security practitioners, analysts, and machine learning models, focusing on important traits of threats instead of becoming slowed down by their varied terminology.

2 Related Work

2.1 ML and DL in IDS

Artificial Intelligence (AI) has become equipped well to enhance modern intrusion detection systems (IDS). Conventional machine learning algorithms, such as Decision Trees, Random Forests, and Support Vector Machines (SVMs), have been applied widely due to their ease of interpretation and low computational demands. These models have shown satisfactory performance in recognizing well-known attack patterns, but they often struggle with high-dimensional data and tend to underperform when faced with imbalanced datasets, where malicious activities are significantly outnumbered by benign traffic (Diro & Chilamkurti 2018; Tavallaee et al. 2009).

Deep learning methods, particularly Convolutional Neural Networks (CNNs) and Long Short-Term Memory (LSTM) networks, have shown increased effectiveness in modelling nonlinear and complex patterns present in network traffic. These models can detect subtle anomalies that traditional techniques may miss. However, for higher performance, they require large volumes of labelled data and high computational resources, which limit their deployment in real-time or resource-constrained environments, such as financial cloud systems (Moustafa & Slay 2015).

2.2 Justification for Boruta and PSO

Distribution Shift is the most concerning issue in AI implementation for IDS, in practical use of network traffic, benign traffic is large, but types are diverse and noisier than the attack signature set. The imbalanced information causes classification methods to be biased to detect more frequent attack types, while missing rare yet important ones in practice. The presence of redundant or irrelevant features is another issue in traffic datasets, which can plummet the performance of a model and keep it sticking to training for longer. In addition, because a live cloud environment requires scalability, low latency for notification and the ability to handle regulatory compliance requirements, many existing models were not suitable to be directly integrated mainstream environments (Heaton 2017; Vinayakumar et al. 2019).

Techniques for feature selection are widely studied to overcome these restrictions. The Boruta method, based on Random Forest classifiers, finds pertinent characteristics by comparing the significance of each feature to its shuffled counterparts. Although Boruta is quite good at finding powerful predictors, it is not always optimized for classifier performance (Kursa & Rudnicki 2010). Particle Swarm Optimization (PSO) is used successfully to enhance feature sets by optimizing a specified fitness function and is often linked to a model's accuracy or F1-score. PSO was inspired by the swarm behavior of fish and birds. The advantages of both methods are capitalized in a hybrid strategy that combines Boruta and PSO is suggested in several studies, providing improved performance and decreased dimensionality (Kennedy & Eberhart 1995).

Benchmark for testing intrusion detection programs used these algorithms on the CICIDS 2017 dataset in several recent research. In (Alsaffar, Nouri-Baygi & Zolbanin 2024), the authors used a hybrid feature selection framework that uses Mutual Information and Boruta, followed by stacked ensemble classifier, to achieve accuracy levels above 99% with lower false positive rates. Further works also offered a multi-phase selection pipeline based on Fisher score filtering, clustering, and a PSO-Grey Wolf Optimization wrapper, which considerably decreased computation costs while maintaining accuracy. Despite encouraging results, most of the assessed works lack a comprehensive framework that integrates effective treatment of class imbalance, deployment-ready design, and effective feature selection. The Synthetic Minority Over-Sampling Technique (SMOTE), which is essential, lack class imbalance, is underutilized in most research, particularly when identifying minority class threats such as botnet traffic or infiltration. Additionally, there hasn't been much focus on incorporating these strategies into a cloud-native, modular IDS system that can be used in industries that have stringent compliance requirements (Ali 2024).

In order to overcome these drawbacks, this study suggests a hybrid Boruta-PSO feature selection approach that is tested across several classifiers using the CICIDS 2017 dataset and backed by SMOTE for class rebalancing. The proposed system also introduces a scalable architecture tailored for real-time intrusion detection in cloud-based environments including that of financial systems.

3 Research Methodology

3.1 Description

The proposed study provides an AI-driven intrusion detection framework specifically addressing the needs of cloud environments. This framework contains modular approach, consisting of multiple key stages: data pre-processing, class balancing, hybrid feature selection, model training, and performance evaluation. Dataset CICIDS 2017 provides legitimate and malicious network activities, which makes it an ideal dataset for training and testing cloud-based security models. Data is prepared for modeling by cleaning to remove missing or null entries. Attributes related to categories like protocol or attack type are converted into numerical values by label encoding. This transformation is needed for machine learning algorithms to process data effectively. Additionally, numeric features are normalized using Min-Max scaling, ensuring all values fall within the range of 0 to 1. Performance and convergence speed of models depend on distance measures and gradient descent, like KNN and ensemble classifiers.

3.2 Experimental Setup

The fundamental challenges in training an intrusion detection system are addressing the class imbalance commonly present in real-world network traffic datasets. Most of the records are labelled as normal traffic, while the actual attacks—particularly stealthy or novel (zero-day) attacks are underrepresented. The Synthetic Minority Over-Sampling Technique (SMOTE) generates synthetic samples for minority classes by interpolating between existing instances. This creates a more balanced dataset, enabling the models to better detect infrequent but critical attack behaviors. After the dataset is balanced and pre-processed and it is split into training and testing sets. Model training is conducted using 10-fold cross-validation, which partitions the data into ten subsets using nine for training and one for validation in each iteration. This ensures robustness in the model evaluation process and reduces the risk of over fitting.

3.3 Feature Selection and Model Training

Feature selection plays an important role in improving model performance by reducing dimensionality and minimizing noise. They can remove irrelevant features that reduce accuracy and increase the computational time for training the model. A two-stage hybrid feature selection approach is implemented. First, the Boruta algorithm—built as a wrapper around a Random Forest classifier, is applied to identify and retain only the most relevant features. It is done by comparing importance of original features with shadow (randomized) features. An original feature is removed when it consistently performs worse than the shadow feature. This ensures that only the most informative attributes are preserved. Particle Swarm Optimization (PSO) is used to refine the selected features after following Boruta, PSO is a modern optimization algorithm inspired by the collective behavior of bird flocks or fish

schools. A particle is a subset of features that explores the feature space. The fitness of each particle is obtained using classification accuracy or F1-score from cross-validation. Positions of Particles are updated based on their personal best and the global best position found by the swarm. PSO converges on an optimal or near-optimal subset of features through iterative updates.

Several machine learning classifiers, including Extreme Gradient Boosting (XGBoost), Random Forest (RF), K-Nearest Neighbors (KNN), Support Vector Machine (SVM), CatBoost, LightGBM, and Extra Trees, are trained after selecting the most effective features. Robustness refers to the ability of a model to identify complex patterns, and accurately classify anomalies, these attributes support in selecting an appropriate model for an intrusion detection system to be used in. The performance is optimized when hyperparameters of each model, like the number of trees, learning rate, kernel type, and tree depth, are fine-tuned using grid search and random search. The LightGBM model achieved the highest accuracy (99.84%), followed by Random Forest (99.78%), XGBoost (99.70%), CatBoost (99.66%), and Extra Trees (99.18%). These results prove the effectiveness of ensemble-based classifiers in the context of cloud intrusion detection.

3.4 Evaluation Methodology

Objective assessment of model performance is carried out using an independent test set, separated before training begins. This eliminates the risk of data leakage and provides an unbiased performance metric. Several evaluation indicators are employed, including accuracy, precision, recall, F1-score, and the Area under the Receiver Operating Characteristic Curve (ROC-AUC). These metrics provide a comprehensive view of model efficacy, particularly in environments where detecting both common and rare attacks is crucial. Special attention is given to minimizing false positives and false negatives, which is especially important in financial cloud systems where detection failures can lead to significant economic and reputational loss. Proposed methodology integrates effective data pre-processing, intelligent feature selection, and powerful classification models in a robust evaluation framework. This ability of the system to detect advanced intrusion patterns is improved while maintaining scalability, interpretability, and real-time responsiveness required for practical deployment.

4 Design Specification

Hybrid intrusion detection framework is provided in the proposed architecture provides for enhancing the security of cloud computing environments. It highlights the integration of BORUTA for feature selection and Particle Swarm Optimization (PSO) for parameter tuning, and Synthetic Minority Over-sampling Technique (SMOTE) for addressing class imbalance. The system begins by collecting raw network traffic data, which includes both normal and malicious activities. Data is preprocessed through cleaning, normalization, and encoding to make it suitable for training. Balanced Dataset is obtained through SMOTE synthetically

generates samples for minority attack classes like user-to-root (U2R) and remote-to-local (R2L). After preprocessing, BORUTA identifies the most significant features, removing irrelevant or redundant data that could negatively impact detection accuracy. PSO is used parallel to optimize the hyperparameters of the machine learning model, improving its overall performance in detecting a wide range of intrusions.

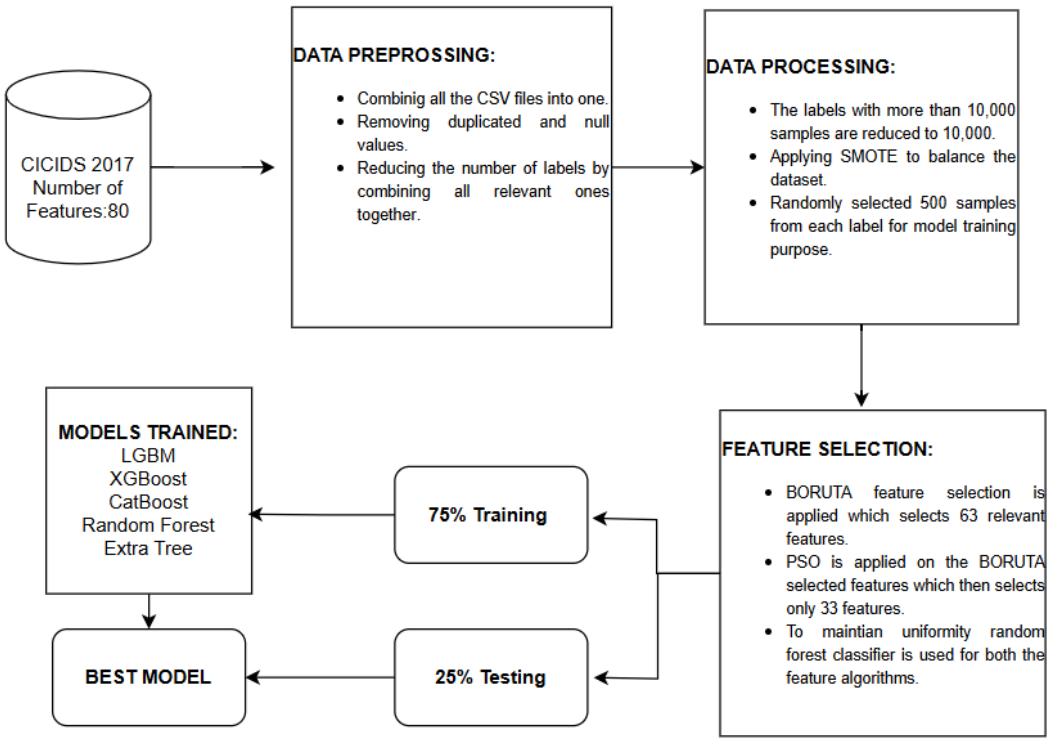


Fig 4.1 Architecture Diagram

The model is deployed in interface that accepts real-time or offline input from users or automated systems. The trained model segregates the incoming data into normal or attack categories. The system identifies specific types of attack , like denial of service (DoS), U2R, R2L, or probe, allowing for a more granular response to threats. The capability of the model multi-level makes it highly effective in managing several types of intrusions in a dynamic cloud environment. The integration of intelligent feature selection, parameter optimization, and data balancing provides improved detection accuracy, reduced false positives, and enhanced adaptability.

The novelty of the project lies within the application of hybrid feature selection and SMOTE along with combining similar labels together. Multiple attack vector labels are combined and brought under the same attack type umbrella, then SMOTE is applied to balance the dataset. Finally, BORUTA and PSO are applied on this balanced dataset which efficiently only selects relevant features with which multiple ML models are trained which makes the adaptable to cloud-native environments quickly with minimal reconfiguration.

4.1 Requirements

Hardware requirements: All the experiments for computing capabilities for model training and evaluation were implemented on the system which has Intel Core i7 processor and 16 GB of RAM.

Software: Scikit-learn, Pandas, Matplotlib, and Seaborn for machine learning, data manipulation, visualization and exploratory data analysis were the libraries used respectively from Python 3.11 version.

Development tools: Scripting, visualization, and model tuning are all executed using Jupyter Notebook. Simplification of data, class balance, and intelligent feature selection are executed in this organized implementation pipeline to achieve a reliable intrusion detection system that suits the cloud architecture.

5 Implementation

The proposed architecture is implemented with machine learning algorithms to build intrusion detection system that categorizes network traffic into normal or attack categories. The process starts with pre-processing steps, like cleaning, normalization, label encoding, and label consolidation. The cleaned dataset is used to build a machine learning model that identifies and differentiates normal behavior and potential hazards. Users can input existing or new network data using the system. When an attack is identified, it is classified into one of the predefined groups.

A clean, balanced, and consistent dataset is achieved through these pre-processing procedures, which improves the overall performance and training effectiveness of intrusion detection models. Hybrid feature selection (Boruta and PSO) combined with SMOTE Some of the attack types, like Heart bleed and Infiltration, were left unnoticed due to the class imbalance. Synthetic Minority Over-sampling Technique (SMOTE) was used to reduce this. By analyzing pre-existing examples, SMOTE creates synthetic samples for minority classes. This prevents an imbalance between classes and improves the distribution of classes.

Feature selection was done in two stages to minimize noise and reduce dimensionality. Boruta algorithm as a wrapper method identifies all statistically significant features by iteratively comparing them to randomized shadow features. Particle Swarm Optimization (PSO) was used to further probe the chances once Boruta had shortlisted key attributes. PSO simulates a swarm of candidate solutions that iteratively modify their placements based on both global and personal bests.

Several machine learning models were used for training and assessing a balanced and processed dataset. Out of them, Light Gradient Boosting Machine (LightGBM) is proven to be the most successful classifier for multi-class intrusion detection. It provides quick training speed and minimal memory usage for handling large amounts of data.

Robustness is assured through stratified 5-fold cross-validation for evaluating the models. Standard criteria like accuracy, precision, recall, F1-score, and AUC-ROC arrived to evaluate performance. LightGBM performed best overall across all metrics, particularly in recognizing complex and uncommon attack types.

6 Evaluation

6.1 Catboost

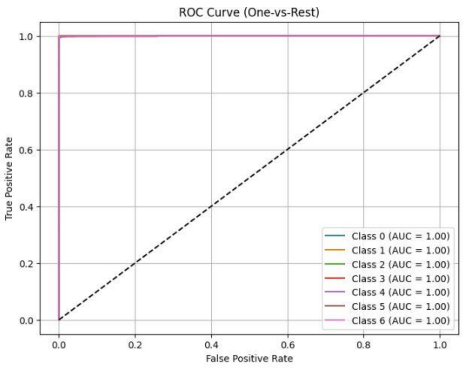


Fig.6.1 CatBoost ROC Curve

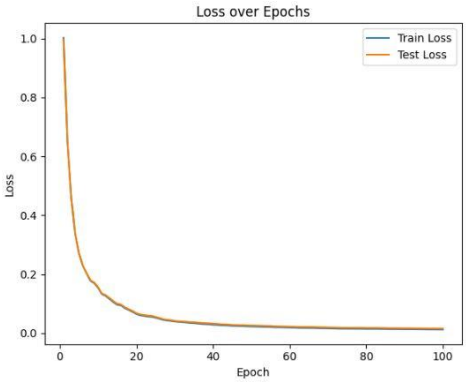


Fig.6.2 CatBoost Loss over Epochs

Classification Report:

	precision	recall	f1-score	support
0	0.99	1.00	1.00	2487
1	1.00	1.00	1.00	2488
2	0.99	1.00	1.00	2508
3	1.00	1.00	1.00	2535
4	1.00	0.99	0.99	2435
5	1.00	0.99	1.00	2523
6	0.99	1.00	1.00	2524
accuracy			1.00	17500
macro avg	1.00	1.00	1.00	17500
weighted avg	1.00	1.00	1.00	17500

Fig.6.3 CatBoost classification report

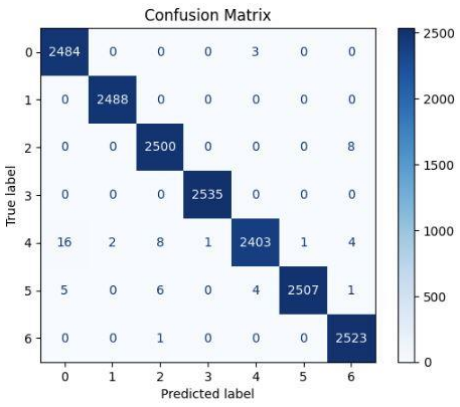


Fig.6.4 CatBoost Confusion Matrix

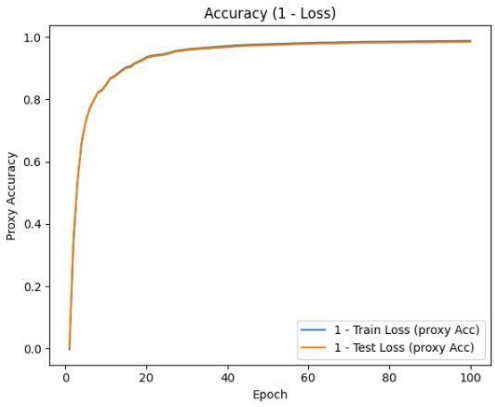


Fig.6.5 CatBoost Accuracy over epochs

The results strongly prove the effectiveness of CatBoost classifier within proposed intrusion detection framework. Both training and testing accuracy curves gives perfect convergence, demonstrating ability of CatBoost to generalize effectively without overfitting. Quick decline and stabilization of loss values further indicate strong learning process of models. Confusion matrix reveals high precision of classifiers , with very low misclassification rates for all attack types. Perfect precision, recall, and F1-scores for all classes is achieved by CatBoost's . Additionally, ROC curve with an AUC of 1.00 for each class shows excellent performance of c in multi-class classification. Its built-in support for categorical variables and advanced gradient boosting techniques greatly enhance its robustness and predictive accuracy in cloud-based intrusion detection systems.

6.2 Extra Tree

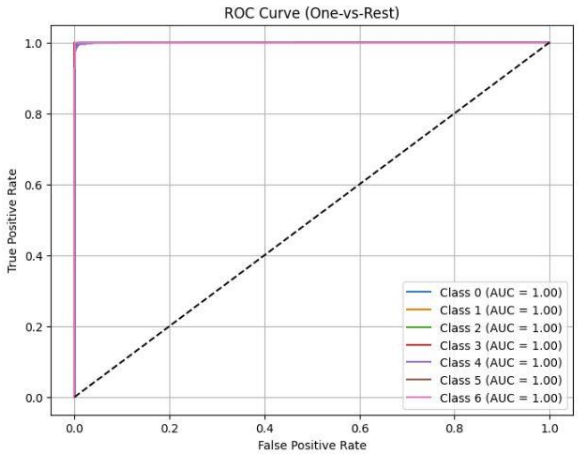


Fig 6.6 Extra Tree ROC Curve

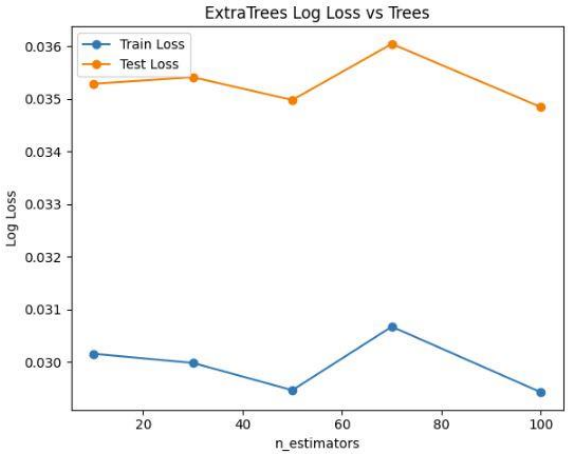
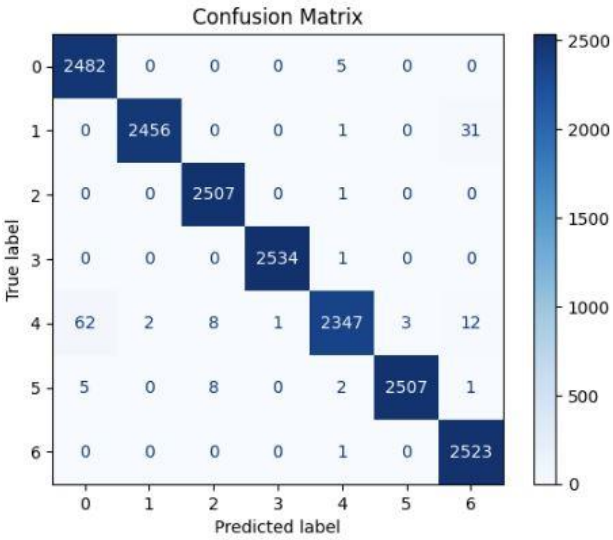


Fig.6.7 Extra Tree log loss vs trees

Classification Report:

	precision	recall	f1-score	support
0	0.97	1.00	0.99	2487
1	1.00	0.99	0.99	2488
2	0.99	1.00	1.00	2508
3	1.00	1.00	1.00	2535
4	1.00	0.96	0.98	2435
5	1.00	0.99	1.00	2523
6	0.98	1.00	0.99	2524
accuracy			0.99	17500
macro avg	0.99	0.99	0.99	17500
weighted avg	0.99	0.99	0.99	17500

Fig.6.8 Extra Tree classification report



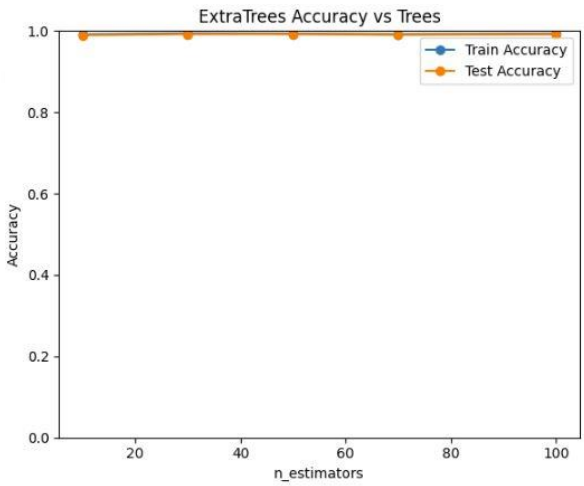
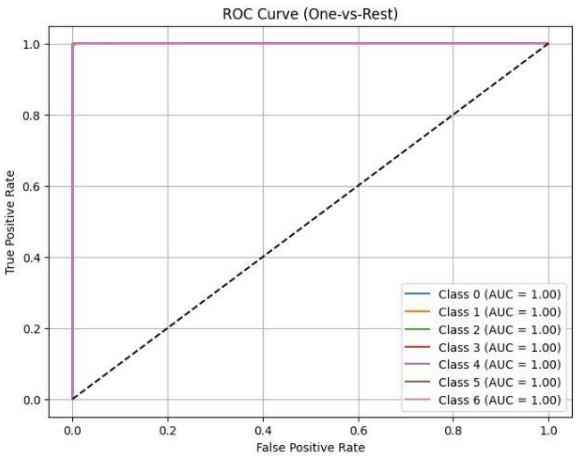


Fig. 6.9 Extra Tree Confusion Matrix **Fig.6.10 Extra Tree accuracy over Trees**

High accuracy, precision, recall, and F1-scores achieved through Extratree for all classes imply that it provides high performance. Regardless of the number of trees, both training and test accuracy were consistently high. Strong generalization indicates low misclassification of confusion matrices. Low and consistent log loss values confirms that Well-calibrated predictions were attained . ROC curves of all classes have an AUC of 1.00 shows superiority in class separation. No data leakage or preprocessing problems, these results demonstrate resilience and appropriateness of the model for the classification task.

6.3 LGBM



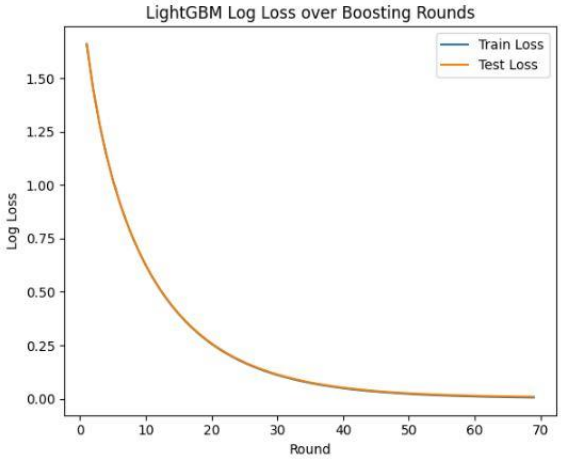


Fig6.11 LGBM ROC Curve **Fig. 6.12 LGBM log loss over boosting rounds**

Classification Report:

	precision	recall	f1-score	support
0	0.99	1.00	1.00	2487
1	1.00	1.00	1.00	2488
2	1.00	1.00	1.00	2508
3	1.00	1.00	1.00	2535
4	1.00	0.99	1.00	2435
5	1.00	1.00	1.00	2523
6	1.00	1.00	1.00	2524
accuracy			1.00	17500
macro avg	1.00	1.00	1.00	17500
weighted avg	1.00	1.00	1.00	17500

Fig6.13 LGBM classification report

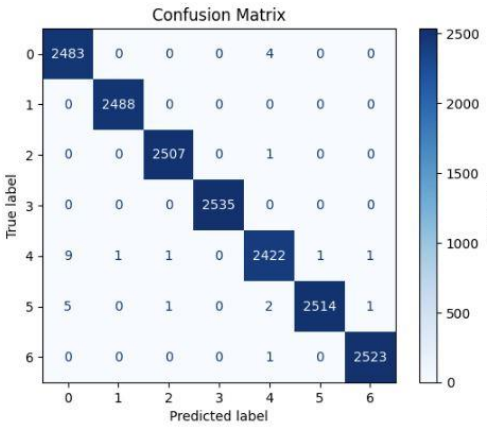


Fig. 6.14 LGBM confusion matrix

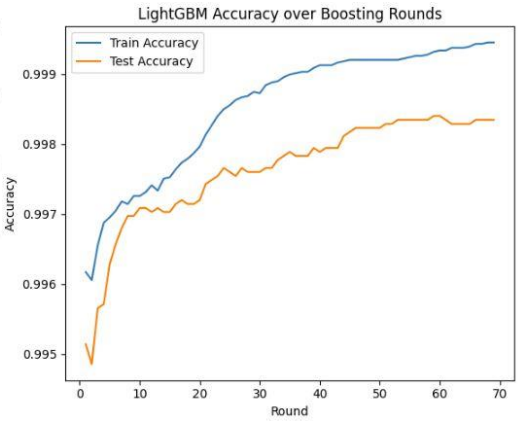


Fig. 6.15 LGBM accuracy over boosting rounds

Light GBM model showed outstanding classification performance over all evaluation metrics. After boosting rounds, the training and test accuracies reached near-perfect values, which replicates the steady improvement in accuracy. The log loss consistently declined, indicating highly confident predictions. The confusion matrix showed minimal misclassification, and the classification report reflected precision, recall, and F1-scores close to or equal to 1.00 for all classes. The ROC curves confirmed perfect separability, with an AUC of 1.00 for each class, demonstrating excellent model discrimination and calibration.

6.4 Random Forest

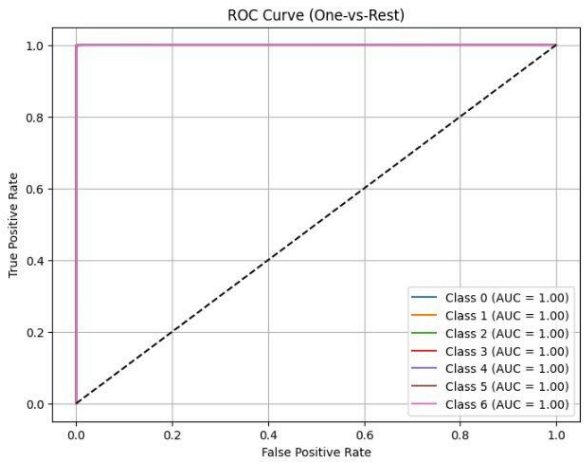


Fig. 6.16 Random Forest ROC curve

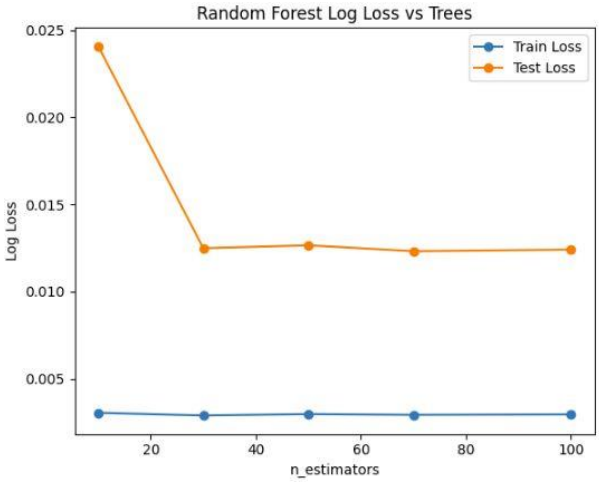


Fig. 6.17 Random Forest log loss vs Trees

Classification Report:

	precision	recall	f1-score	support
0	0.99	1.00	1.00	2487
1	1.00	1.00	1.00	2488
2	1.00	1.00	1.00	2508
3	1.00	1.00	1.00	2535
4	1.00	0.99	0.99	2435
5	1.00	0.99	1.00	2523
6	1.00	1.00	1.00	2524
accuracy			1.00	17500
macro avg	1.00	1.00	1.00	17500
weighted avg	1.00	1.00	1.00	17500

Fig. 6.18 Random Forest classification report

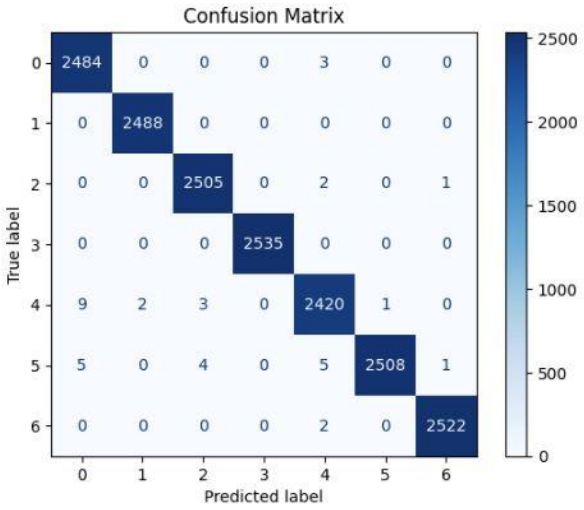


Fig. 6.19 Random Forest Confusion Matrix

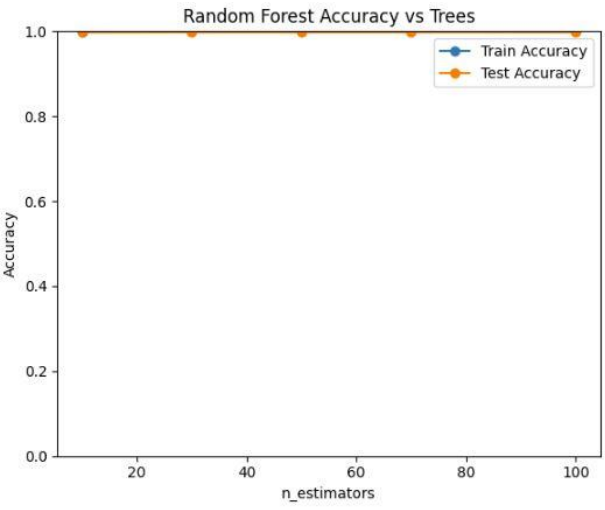


Fig. 6.20 Random Forest Accuracy vs Trees

Consistent high performance in all evaluation metrics is provided by Random Forest model . For both training and test sets, regardless of the number of trees used accuracy remains near perfect . Classification report shows precision, recall, and F1-scores are close to 1.00 for classes. Confusion matrix confirms minimal misclassification and log loss values are low & stable which indicates confident and well-calibrated predictions. ROC curves for all classes visualize AUC of 1.00 which confirms excellent class separation.

6.5 XGBoost

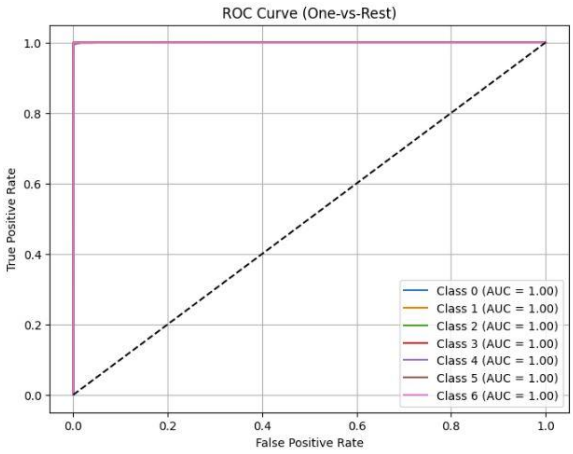


Fig. 6.21 XGBoost ROC curve

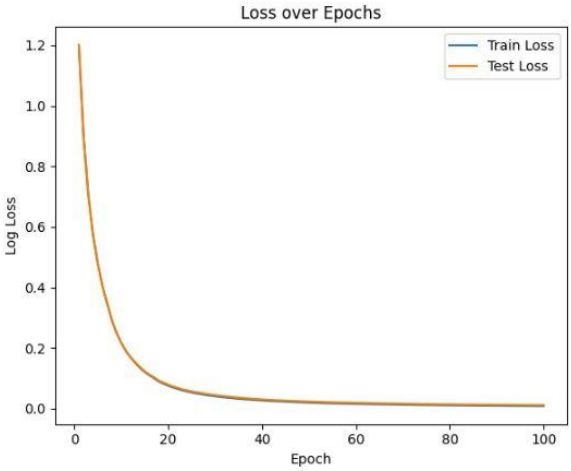
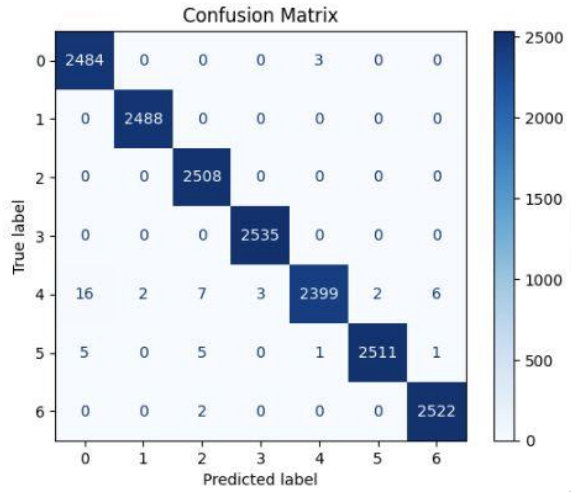


Fig.6.22 XGBoost loss over epochs

Classification Report:

	precision	recall	f1-score	support
0	0.99	1.00	1.00	2487
1	1.00	1.00	1.00	2488
2	0.99	1.00	1.00	2508
3	1.00	1.00	1.00	2535
4	1.00	0.99	0.99	2435
5	1.00	1.00	1.00	2523
6	1.00	1.00	1.00	2524
accuracy			1.00	17500
macro avg	1.00	1.00	1.00	17500
weighted avg	1.00	1.00	1.00	17500

Fig. 6.23 XGBoost Classification Report



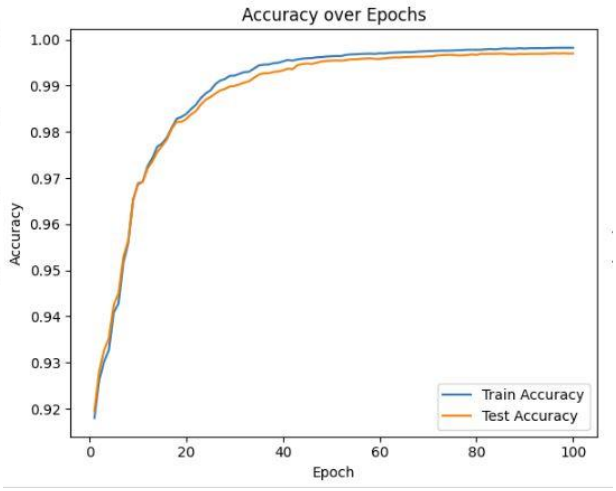


Fig. 6.24 XGBoost Confusion Matrix

Fig. 6.25 XGBoost Accuracy over Epochs

Outstanding results are achieved by XGBoost model. For both training and test sets, accuracy constantly increased over epochs. The loss curves dropped rapidly and stabilized close to zero, which indicates confident predictions are high . The confusion matrix showed very few errors across all classes and the classification report reflected near-perfect precision, recall, and F1-scores. Additionally, the ROC curves confirmed excellent separation with an AUC of 1.00 for every class. Overall, the model demonstrated exceptional reliability and effectiveness for the multi-class classification task.

6.6 Discussion

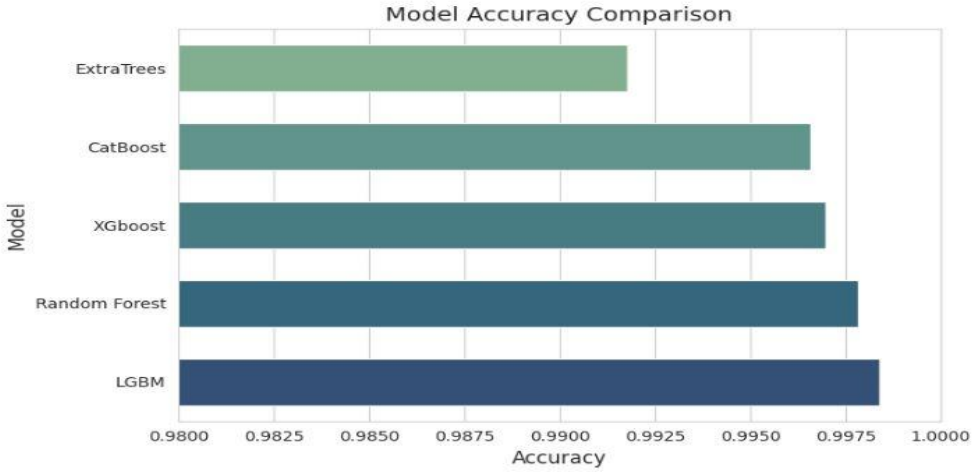


Fig. 6.26 Model Accuracy Comparison

Comparative analysis of several machine learning classifiers shows that for detecting intrusions within cloud environments, the trained models possess high accuracy levels, each exceeding 98%. LightGBM performed better than the other classifiers with highest accuracy, revealing its high ability to identify intrusions. Good performances are also delivered by Random Forest and XGBoost through reliable detection capabilities which stems from their

ensemble learning approaches. Competitive results are provided by CatBoost, while Extra Trees has slightly lower accuracy. CatBoost and Extra Trees classifiers maintain robust classification performance, these results highlight that ensemble and boosting algorithms are particularly adapted well for intrusion detection in cloud settings, offering effective and reliable protection against potential threats.

7 Conclusion and Future Work

Proposed hybrid intrusion detection system effectively enhances cloud security with the integration of BORUTA for optimal feature selection, PSO for efficient model parameter tuning, and SMOTE to handle class imbalance within the dataset. Accuracy and robustness of detection model are improved because of this integration and also helps in enabling it to identify and classify various types of attacks with minimal false positives. The layered classification approach of the system assures perfect recognition of specific intrusion types. For future work, the model can be extended to use deep learning architectures that adapt to evolving attack patterns in real time. While deploying the system in distributed cloud platforms with real-world traffic data it accurately responds to the scalability and high traffic conditions. Enhancing ability of to self-update the model with minimal supervision can further strengthen its resilience against zero-day threats and novel intrusion techniques. If more time were allocated for this project, I would have used different classifiers within the hybrid feature selection algorithms (BORUTA and PSO) to test/evaluate which combination provided the least and the greatest number of relevant features. Furthermore, deep learning models could have been used to facilitate a more advanced classifier for malicious activities, as Deep Learning models are more efficient than Machine Learning models.

References

Ali 2024, 'Exploring Lightweight Deep Learning Techniques for Intrusion Detection Systems in IoT Networks: A Survey', *Deleted Journal*, vol. 20, no. 4s, pp. 1944–1958.

Alsaffar, AM, Nouri-Baygi, M & Zolbanin, HM 2024, 'Shielding networks: enhancing intrusion detection with hybrid feature selection and stack ensemble learning', *Journal of Big Data*, vol. 11, no. 1.

Diro, AA & Chilamkurti, N 2018, 'Distributed attack detection scheme using deep learning approach for Internet of Things', *Future Generation Computer Systems*, vol. 82, pp. 761–768.

Heaton, J 2017, 'Ian Goodfellow, Yoshua Bengio, and Aaron Courville: Deep learning', *Genetic Programming and Evolvable Machines*, vol. 19, no. 1-2, pp. 305–307.

Kennedy, J & Eberhart, R 1995, 'Particle swarm optimization', *Proceedings of ICNN'95 - International Conference on Neural Networks*, vol. 4, pp. 1942–1948.

Kursa, MB & Rudnicki, WR 2010, 'Feature Selection with theBorutaPackage', *Journal of Statistical Software*, vol. 36, no. 11.

Moustafa, N & Slay, J 2015, *UNSW-NB15: a comprehensive data set for network intrusion detection systems (UNSW-NB15 network data set)*, IEEE Xplore, pp. 1–6, viewed 24 May 2020, <https://ieeexplore.ieee.org/document/7348942>.

Pan, X, Lei, M, Sun, J, Wang, H, Ju, T & Bai, L 2023, 'An evolutionary feature selection method based on probability-based initialized particle swarm optimization', *Research Square (Research Square)*, Research Square (United States).

T. Sowmya & E.A. Mary Anita 2023, 'A comprehensive review of AI based intrusion detection system', *ScienceDirect*, vol. 28, pp. 100827–100827.

Tavallaee, M, Bagheri, E, Lu, W & Ghorbani, AA 2009, 'A detailed analysis of the KDD CUP 99 data set', *2009 IEEE Symposium on Computational Intelligence for Security and Defense Applications*.

Vinayakumar, R, Alazab, M, Soman, KP, Poornachandran, P, Al-Nemrat, A & Venkatraman, S 2019, 'Deep Learning Approach for Intelligent Intrusion Detection System', *IEEE Access*, vol. 7, pp. 41525–41550.

Zhang, X, Lin, Q, Mao, W, Liu, S, Dou, Z & Liu, G 2021, 'Hybrid Particle Swarm and Grey Wolf Optimizer and its application to clustering optimization', *Applied Soft Computing*, vol. 101, Elsevier BV, pp. 107061–107061.