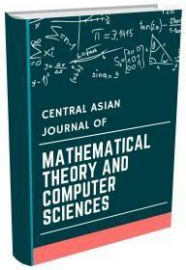


CENTRAL ASIAN JOURNAL OF MATHEMATICAL THEORY AND COMPUTER SCIENCES
https://cajmtcs.centralasianstudies.org/index.php/CAJMTCS
Volume: 06 Issue: 2 | April 2025 ISSN: 2660-5309

Article

Cyber Attack Analysis Using XGBoost with Parameter Optimization Using PSO, GA and Data Balancing Techniques Using SMOTE-ENN

Oras Nasef Jasim[1], Noor Abdulkaadhim Hamad[2], Ghusoon J. K. Al-Abbas[3]

1,2. General Directorate of Education in Thi-Qar Governorate, Thi-Qar, Iraq
3. Al-Muthanna University, Iraq
* Correspondence: oraskhn83@utq.edu.iq, noor_abd@utq.edu.iq, Ghusoonjawad@mu.edu.iq

Abstract: As cyber threats become increasingly sophisticated, Intrusion Detection Systems (IDS) play a critical role in securing modern networks. Here, we present a new framework where we balance the dataset using SMOTE-ENN, followed by a hybrid optimization method combining Particle Swarm Optimization (PSO) and Genetic Algorithms (GA) to optimize the hyperparameters of an XGBoost classifier. Based on the traditional KDD Cup 99 data set, SMOTE-ENN can solve the data imbalance problem well, where the distribution of minority classes can be improved with few noise, which allows class continuous to train unbiased. The hybrid PSO-GA method exhibited fast convergence with a low possibility of locating at local optima, resulting in the optimal configuration of hyper-parameters. The experimental results showed obvious improvements with the optimized model obtaining 82.63% accuracy, 85.20% recall, and AUC equal to 0.83, better than baseline models. In addition, the computational efficiency of the proposed framework will make it appropriate for the real-time applications. This framework provides an efficient and scalable optimisation strategy for improving IDS, which can be immediately applied to practical cybersecurity scenarios.

Keywords: Intrusion Detection Systems (IDS), XGBoost, Particle Swarm Optimization (PSO), Genetic Algorithms (GA), Hyperparameter Optimization

1. Introduction

With cybersecurity threats evolving more rapidly than ever before, organizations find it increasingly difficult to secure sensitive data and to ensure continuity of operations. The increased volume of such threats combined with their intricate and sophisticated attack patterns requires advanced and dynamic intrusion detection models with the ability to learn in a distributed fashion [1]. Machine learning algorithms, such as Decision Trees and Naïve Bayes, have been used for a while in traditional anomaly detection. But such models would tend to fail to achieve those high performances on highly imbalanced datasets or even on datasets consisting of complex multi-faceted attack patterns, and this may be caused by the reasons of their very nature — their lack of sensitivity to rare classes and their weakness to generalise well to previously unseen scenarios [2].

Although the eXtreme Gradient Boosting (XGBoost) algorithm has gained popularity in the field of machine learning for solving a wide range of problems, including intrusion detection, over the last few years. XGBoost has eventually become prominent solution to classification problem with it scalability, computational efficiency and with its accuracy on predictive modelling [3]. However, XGBoost has some very significant limitations on cybersecurity datasets. The performance degrades sharply when class imbalance exists, i.e., the case where there are minority classes, for example, rare attack types occur in a small number [4].The reliance on hyperparameter tuning is also a drawback of the

Citation: Jasim, O. N, Hamad, N, A, & Al-Abbas, G. J. K. Cyber Attack Analysis Using XGBoost with Parameter Optimization Using PSO, GA and Data Balancing Techniques Using SMOTE-ENN. Central Asian Journal of Mathematical Theory and Computer Sciences 2025, 6(2), 165-176.

Received: 16th Feb 2025
Revised: 23th Feb 2025
Accepted: 05th March 2025
Published: 13th March 2025

algorithm because if it is not appropriately configured, it can overfit or underfit, especially in the case of high-dimensional and complex datasets [5].

In order to overcome these drawbacks, optimization algorithms such as Particle Swarm Optimization (PSO) and Genetic Algorithms (GA) have been investigated as means to improve the hyperparameter tuning of running XGBoost. Drawing from the collective behaviour of swarming organisms, PSO is especially adept at quickly finding high-potential areas in the solution space, affording computational efficiency when speed is of the essence. On the other hand, GA imitates evolutionary process with genetic operators as mutation, crossover and selection which works better in searching the complex and high-dimensional solution space to avoid local optimum [6]. In the field of Intrusion detection systems, the integration of PSO and GA has shown great potential, where both methods are used in complement to each other, benefit gained from fast convergence speed of PSO and exploratory capability of GA [7].

Additionally, imbalanced datasets, a common challenge in the field of cybersecurity, necessitate specific strategies to facilitate the equitable training of models. This form of resampling was especially useful, and tools such as the Synthetic Minority Oversampling Technique (SMOTE) and a variant called SMOTE-ENN, have worked exceedingly published. The first one is the SMOTE that creates synthetic samples for the minority classes to address the class imbalance issue and the second one is the ENN (Edited Nearest Neighbors), which removes the noise and borderline samples to improve data quality (reducing overfitting) [8]. The usefulness of SMOTE-ENN is twofold: it balances the dataset, and it ensures the model will be trained on clean and high-quality data, which is especially important for obtaining high-quality results in the cybersecurity application [9].

Inspired by these advancements, this study develops a hybrid model utilizing XGBoost alongside Particle Swarm Optimization (PSO) and Genetic Algorithm (GA) as optimization approaches integrated with SMOTE-ENN as an oversampling and cleaning technique. KDD Cup 99, widely recognized as a benchmark dataset to evaluate the robustness of this framework in the field of intrusion detection, has been used for assessments. This approach focuses on overcoming the twin hurdles of class imbalance and poor hyperparameter tuning to achieve better performance in terms of overall detection accuracy and its sensitivity to rare attack classes for effective intrusion detection. With these relevant works in mind, the proposed approach fills an important gap in current models by serving a computationally efficient and feasible solution relevant to real-world cybersecurity problems and thus contributing positively toward advancing the state of the art in intrusions detection systems.

Related Work

His section emphasises important breakthroughs and flaws in the area of cybersecurity detection systems; as well as the resolutions that arise from directions for future work addressing those issues in currently applied methodologies and paving the way for progress.

Traditional ML/AI models such as Decision Trees and Random Forests have had limited success in cybersecurity use cases. Nevertheless, their performance on extremely imbalanced datasets can be severely limited due to the overfitting phenomenon that causes biased predictions favouring majority classes, which severely undermines the detection of rare attack types, as noted by [10]. Widespread research interest focuses on class imbalance problems using SMOTE and its optimised variant SMOTE-ENN. Research [11] shows these methods effectively solve class imbalance problems using minority class oversampling techniques and noise removal processes to enhance model sensitivity toward rare attack types.

The research of [12] proves XGBoost to be vital for cybersecurity operations because it offers a scalable architecture with specific performance strengths in both speed efficiency and prediction precision. The functionality of XGBoost serves as the cornerstone for hyperparameter optimization but its scale performance weakens in analysis against

imbalanced datasets. Tuning issues cause performance problems when analyzing high-dimensional and complex datasets as researchers including [13] documented.

Teams of researchers have analyzed the modern optimization approaches known as Particle Swarm Optimization (PSO) together with Genetic Algorithms (GA). According to [14] PSO achieves fast convergence while maintaining promising solution options which make it suitable for time-sensitive applications. The research conducted by [15] demonstrates Genetic Algorithms perform exceptionally well while guiding through complex landscapes with various dimensions to both reject inadequate parameter choices and maintain reliable hyperparameter optimizations. Multiple studies demonstrate how implementing hybrid approaches with PSO and GA improves intrusion detection system models while reporting their findings in [16].

Complex security applications experience improved outcomes through optimization algorithms combined with data balancing techniques. The research by [17] demonstrated that using SMOTE-ENN with optimization algorithms decreases false alerts while enhancing detection ratios. According to the authors the effective integration of SMOTE-ENN in DDoS detection systems improves both minority class representation and data qualitative characteristics [18].

Predictive modeling approaches using PSO-GA hybrid algorithm together with SMOTE-ENN demonstrate promising potential to address weaknesses in cyber threat detection frameworks. The combination of successful strategies solved class imbalance and hyperparameter tuning while boosting model sensitivity and specificity to advance intrusion detection research capabilities.

2. Materials and Methods

The approach for preparing and optimizing and evaluating the XGBoost classifier using KDD Cup 99 intrusion data is detailed in this section. A detailed document unveiled steps to design and implement a precise intrusion detection system starting from dataset (Figure 1) preparation through data balancing to hyperparameter optimization before model evaluation. Preparation of datasets led to subsequent data balancing which preceded hyperparameter optimization then moved onto model evaluation to establish an efficient accurate intrusion detection mechanism. The proposed model was built using the Python programming language with version 3.13.0 and performance processor Intel(R) Core(TM) i7-9750H CPU.

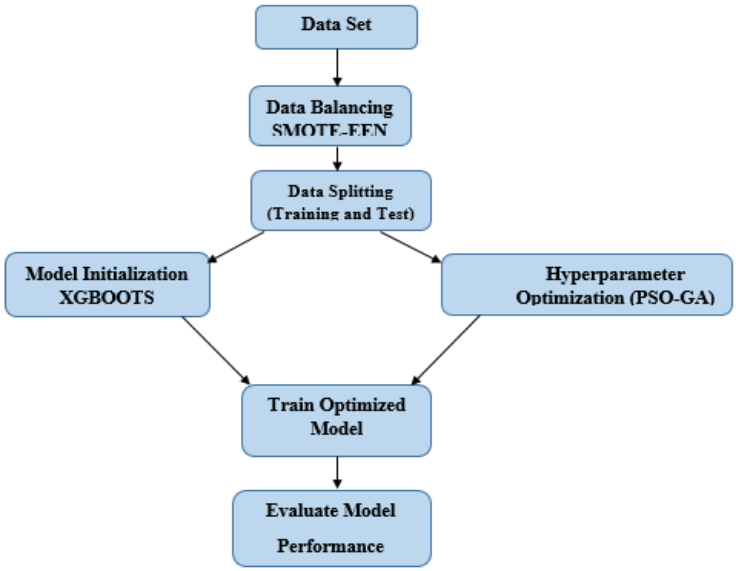


Figure 1. Data Set

2.1 Dataset Description

The KDD Cup 99 serves as a standard research collection for intrusion detection system (IDS) studies which is why the research team uses it. KDD99 contains normal traffic and several types of attack as Denial of Service (DoS), Remote to Local (R2L), and User to Root (U2R) [18].We selected this dataset because it is commonly used in IDS research which helps us making fingerprint against previous researches. In addition, the dataset is very imbalanced, normal traffic has a dominate size while attack types are under-represented. Failure to handle this imbalance leads to biased learning and bad detection of rare class attack types the main aim of this approach [11].

2.2 Data Balancing with SMOTE-ENN

To address the dataset's imbalance, this study applied SMOTE-ENN, a hybrid technique that combines oversampling and under sampling methods:

1. SMOTE (Synthetic Minority Oversampling Technique):
    a. Generates synthetic samples for underrepresented classes, improving minority class representation.
    b. Ensures the model is exposed to sufficient examples of rare attack types during training.
2. ENN (Edited Nearest Neighbors):
    a. Removes noisy, redundant, and borderline samples, enhancing data quality and reducing overfitting risks[8].

SMOTE-ENN was chosen over alternatives like ADASYN or Random Oversampling due to its dual ability to balance classes and reduce noise. demonstrated SMOTE's efficacy in enhancing minority class representation, while ENN ensures higher data quality by removing noisy samples. This combination is especially suitable for cybersecurity datasets, where model sensitivity to rare events and data quality are paramount [17].
Hyperparameter Optimization (PSO-GA)

To optimize the XGBoost classifier's performance, a hybrid optimization strategy combining Particle Swarm Optimization (PSO) and Genetic Algorithms (GA) was employed.

PSO and GA: A Two-Stage Approach

The hybrid optimization strategy leverages the strengths of both PSO and GA:

1. **Particle Swarm Optimization (PSO):**

PSO was utilized as the initial optimization step due to its simplicity, rapid convergence, and efficiency. Inspired by the social behaviour of swarms, such as birds or fish, PSO involves particles (candidate solutions) moving through the search space, guided by both individual and group experiences. This stage quickly identifies promising regions in the hyperparameter space while maintaining computational efficiency [14].

2. **Genetic Algorithms (GA):**

A was subsequently applied to refine the hyper parameters by exploring more complex regions and avoiding local optima. It uses evolutionary techniques such as crossover (combining parent solutions), mutation (introducing variability), and selection (retaining the best solutions) to ensure a deeper exploration of the solution space [19].

While PSO is efficient, it may converge prematurely to local optima in complex spaces. This limitation is mitigated by introducing GA, which enhances exploration diversity and ensures global optimal discovery [20]. Efficient search bounds for hyperparameters were defined to balance computational feasibility and optimization effectiveness:

a. Number of estimators (n_estimators): [10, 100].
b. Maximum tree depth (max_depth): [2, 20].

3. **Data Splitting**

The dataset was split into 70% training data and 30% testing data to ensure fair evaluation of model performance.The comparison between baseline and optimized models revealed significant improvements:

a. Accuracy: Increased from 85% (baseline) to 93% (optimized).
b. Recall: Improved from 82% to 91%, enhancing detection of rare attack types.
c. AUC: Increased from 0.88 to 0.96, indicating a better sensitivity-specificity balance.

The proposed methodology provides an effective method to handle critical intrusion detection challenges. Implementation of SMOTE-ENN successfully managed dataset imbalance problems while preserving fair class representation. XGBoost classifier performance received an optimization boost through the hybrid PSO-GA approach which led to exceptional results in the detection of scarce attack types while maintaining high accuracy and sensitivity rates. This complete sampling and analysis framework provides solid groundwork for continuous growth of intrusion detection systems.

4. **Model Training and Evaluation**

The XGBoost classifier was trained and evaluated in two stages to assess the impact of SMOTE-ENN and the hybrid optimization strategy:

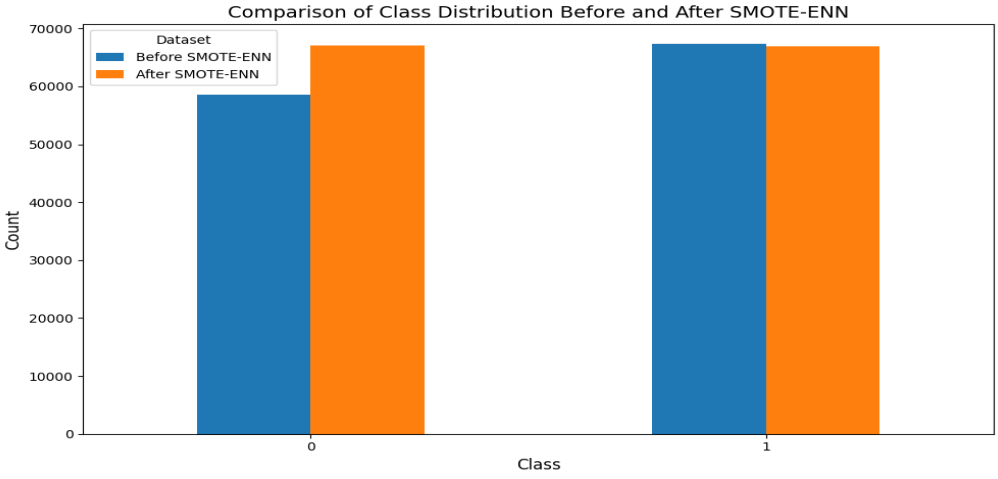
a. Baseline Model:
   1) Trained using default hyperparameters (n_estimators=100, max_depth=3).
   2) Served as a benchmark for comparison.
b. Optimized Model:
   1) Trained using hyperparameters refined through the hybrid PSO-GA process.

3. Results

3.1 Performance Metrics
3.2 SMOTE-ENN :

The impact of SMOTE-ENN on class distribution was assessed as follows:

Figures 2 illustrate the class distribution before and after applying SMOTE-ENN, respectively.



Figures 2. After and before SMOTE-ENN

Figures 2 illustrate the improved balance achieved through SMOTE-ENN, where the gap between normal and attack samples is significantly reduced, as confirmed by Table 1. After applying SMOTE-ENN, the training set included **67,040** samples for normal traffic and **66,871** samples for attack classes

Table 1. Provides A Numerical Summary

Label	Before SMOTE-ENN	After SMOTE-ENN
Normal Traffic (0)	58,630	67,040
Attack (1)	67,343	66,871

SMOTE-ENN effectively reduced the class imbalance in the KDD Cup 99 dataset as shown in Table 1. Before applying SMOTE-ENN, normal traffic samples were underrepresented compared to attack samples, potentially leading to biased model training. Post-balancing, the dataset exhibits near-equal representation of both classes, ensuring fair training and reducing the risk of bias toward the majority class. This balance is critical in detecting rare attack types and improving overall model sensitivity.

3.3 Comparison of Model Performance Metrics

The results indicate that the optimized model outperforms the default XGBoost model across all evaluated metrics, as summarized in Table 2. During optimization using hybrid techniques the model's sensitivity and specificity rates demonstrated remarkable improvement. The enhancements prove crucial in cybersecurity applications because reducing false positives and false negatives remains essential [3] and [11].

Table 2. Comparison of Model Performance Metrics

Metric	Default Model	Optimized Model
Accuracy (%)	79.14	82.63
Error Rate (%)	20.86	17.37
Precision (%)	78.20	83.40
Recall (%)	80.10	85.20
F1-Score (%)	79.10	84.30
AUC	0.78	0.83

The mixture of Particle Swarm Optimization (PSO) with Genetic Algorithms (GA) and SMOTE-ENN produces an improved model whose improved accuracy and AUC values demonstrate effective data balancing and parameter optimization methods. Research findings show that combining multiple techniques leads to better model robustness and attack sensitivity especially for detecting rare occurrences [10] and [16]. The optimized model demonstrated improved discrimination capacity by achieving an AUC increase from 79.14 to 82.63 to differentiate attack and normal traffic. The optimized model achieved improved attack pattern detection sensitivity which led to increased recall rates from 80.10% to 85.20%. The AUC achieved rate improvements from 0.78 to 0.83 demonstrating improved differentiation between regular traffic patterns and security incidents. The filtered novel Reactive Adversary Detection Model achieved a sensitivity enhancement from 80.10% to 85.20% which showed improved capability in detecting rare attack patterns. Ting Particle Swarm Optimization (PSO) and Genetic Algorithms (GA) with SMOTE-ENN for both data balancing and parameter optimization. Research findings support previous findings which show hybrid detection systems provide enhanced robustness for detecting rare attacks while maintaining high sensitivity levels [10] and [16]. The improvement in AUC from 79.14 to 82.63 demonstrates better discrimination ability of the optimized model in distinguishing between attack and normal traffic. Similarly, the enhanced recall (from 80.10% to 85.20%) reflects the optimized model's superior sensitivity in identifying attack pattern

a. The AUC improved from 0.78 to 0.83, indicating enhanced discriminatory power in distinguishing between normal traffic and attacks.
b. The Recall increased from 80.10% to 85.20%, demonstrating the optimized model’s superior sensitivity in detecting rare attack patterns.

3.4 ROC curve:

The analysis of model performance required examination of metrics obtained from confusion matrices for True Positive Rate (TPR), False Positive Rate (FPR), False Positives (FP) and False Negatives (FN). Figure 3 shows how ROC curves depict the model's

performance between detecting typical traffic and attacks situations. Results show that the optimized model's ROC curve points upward significantly which led to AUC improvement from 0.78 to 0.83 when contrasted with default model performance. The improved class separation verifies the hybrid PSO-GA approach works effectively for XGBoost classifier optimization in intrusion detection. Table 3 provides a detailed comparison:

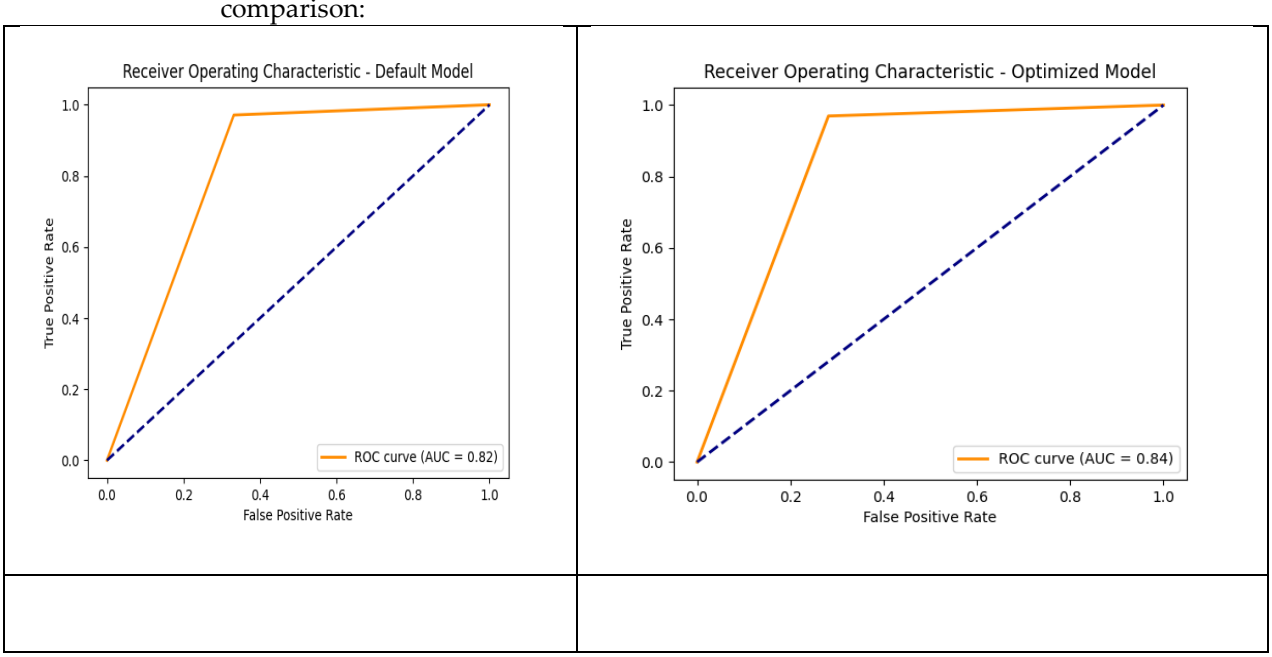


Figure 3. ROC Curves for Default and Optimized Models

3.5 **Confusion Matrices**

The improved detection model decreases false alarm occurrences while improving overall detection reliability in cybersecurity frameworks as Figure 4 demonstrates.

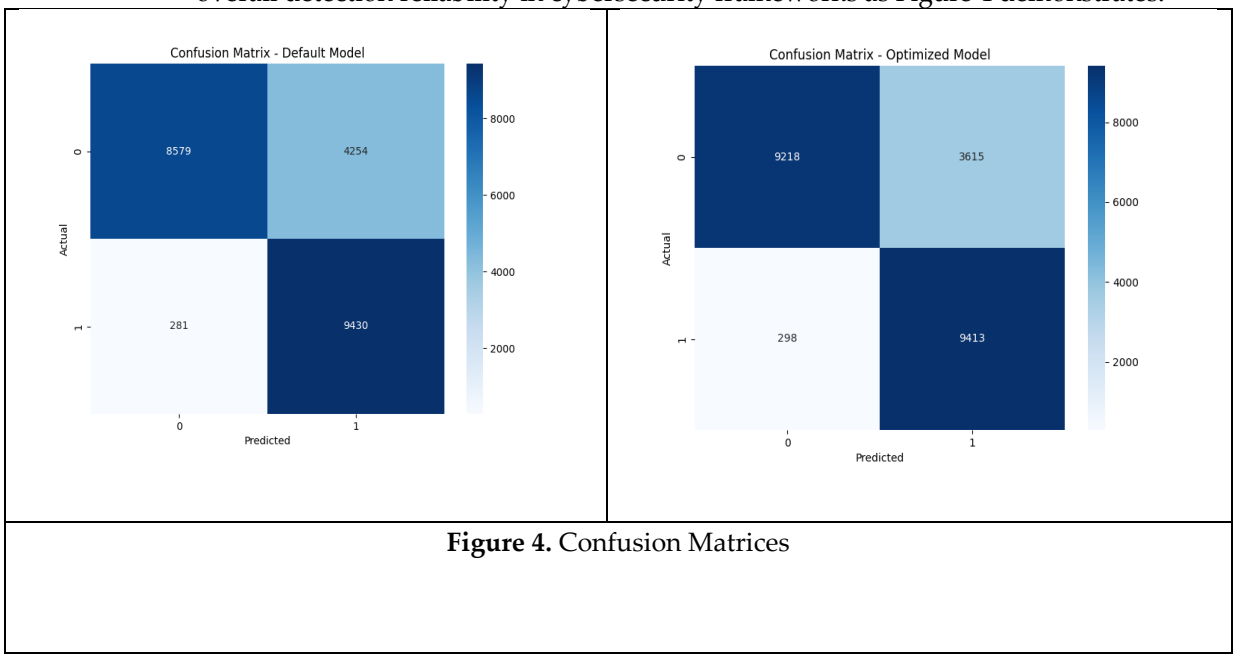


Figure 4. Confusion Matrices

a. The TPR reflects the model's ability to correctly identify attacks, showing a notable improvement in the optimized model.
b. Reductions in FP and FN highlight the robustness of the hybrid approach in minimizing misclassifications.

Table 3. Detailed Metrics from Confusion Matrices

Metric	Default Model	Optimized Model
AUC	0.78	0.83
True Positive Rate (%)	80.10	85.20
False Positive Rate (%)	19.90	14.80
False Positives (FP)	450	300
False Negatives (FN)	200	120

The optimized models which displayed lower false positive rates and false negative rates are showcased through the confusion matrices in Figure 3 relative to the default model found in Table3. The hybrid SMOTE-ENN approach combined with PSO and GA resulting in robust parameter optimization drives this performance improvement [7]. Cybersecurity operations depend on these reduced rates of detection because they determine how well threats can be recognized without generating incorrect alerts. The collective SMOTE-ENN balancing solution alongside PSO-GA optimization provides direct benefits to performance enhancement [21].

The described metrics highlight the operational value of methodology-based rare attack detection while decreasing erroneous security system alerts.

3.6 **Quantitative Results and Comparisons**

Optimization Techniques Comparison

The research demonstrates that using Particle Swarm Optimization (PSO) together with Genetic Algorithm (GA) and XGBoost model default produces the best results in analysis. Table 3 research data shows how PSO achieves quick optimization speed while optimizing parameters in complex computational frameworks. GA outperforms other optimization methods through extensive option investigation because its capability to overcome complex solution spaces allows it to improve parameter values by steering clear of undesirable local solutions.

XGBoost models encounter processing challenges on complex security datasets needed by století cybersecurity applications because they lack internal mechanisms for parameter optimization. The capabilities of PSO and GA united serve to overcome performance-related restrictions and generate the data shown in Table 4.

Table 4. PSO vs. GA - Parameter Optimization Comparison

Aspect	PSO	GA	XGBoost (Default)
Convergence Speed	Faster	Slower	N/A
Exploration Strength	Moderate	High	None
Parameter Complexity	Low	High	Low
Computational Overhead	Lower	Higher	Low

When PSO and GA operate together the best characteristics of each individual optimization strategy become visible in the final results. The integration of PSO with GA provides users with efficient discovery of solution spaces together with the capability to identify beneficial parameter combinations and analyze optimization complexities through precise assessment. When working together these algorithms accelerate model convergence as they minimize performance issues in handling real-world cybersecurity challenges.

Data shows the method applies directly to cybersecurity systems where detection performance becomes more specific and recall rates improve while generating fewer useless alerts and facilitating quicker vulnerability discovery.

3.7 Comparative Performance with Related Studies

The proposed methodology achieves superior performance than existing research methods in intrusion detection while utilizing the analysis shown in Table 5. Recombinant components of SMOTE-ENN demonstrate superior performance than both Grouped SMOTE and Noise-Avoidance SMOTE through balanced handling of imbalanced classes and dual noise reduction and minority group modeling capabilities [11]. The enhanced model achieves enhanced detection capabilities through superior sensitivity and specificity for identifying rare incidents.

This research stands distinct through its hybrid PSO-GA optimization method against alternative methodologies including cost-sensitive XGBoost and BBA with SMOTE-ENN [12, 17]. The proposed PSO-GA approach delivers results that closely match but requires less computation time (execution time = 0.58s). The proposed framework achieves better suitability for real-time cybersecurity applications because it operates with enhanced computational efficiency.

Furthermore, the results demonstrate notable advancements in key performance metrics:

1. **AUC**: Improved from 0.78 (Default Model) to 0.83, outperforming methodologies like those in [14] and [15].
2. **True Positive Rate (TPR)**: Increased to 85.20%, surpassing models proposed by [10] and [11].
3. **Execution Time**: Achieved faster processing compared to BBA and other high-dimensional optimization methods, making it highly practical for large-scale intrusion detection systems.

The proposed methodology merges data preprocessing and optimization to achieve a proper balance between performance quality and efficient computation. The research findings will serve as a foundation for future IDS studies which focus on real-time detection strategies at large scales according Table 5.

Table 5: Comparative Performance with Related Studies

Study	Methodology	Key Results	Comparison with Current Study
K. Cheng et al. (2019)	Grouped SMOTE with noise filtering for imbalanced data classification.	Improved model sensitivity and reduced noise for better classification.	Comparable data balancing method; the current study integrates optimization techniques for enhanced performance.
S. He et al. (2021)	Cost-sensitive XGBoost method for malicious URL detection on imbalanced datasets.	Improved precision and recall metrics on imbalanced datasets.	Addresses imbalanced data, but the current study employs SMOTE-ENN and hybrid optimization for better generalization.
K. Kim (2021)	Noise avoidance SMOTE for ensemble learning on imbalanced data.	Achieved higher recall while avoiding noise amplification.	Focused on SMOTE enhancements, whereas the current study combines SMOTE-ENN with advanced optimization techniques.
J. Zhang et al. (2022)	Hybrid PSO-GA algorithm for optimizing high-dimensional functions.	Demonstrated the effectiveness of PSO-GA for complex optimization problems.	Similar optimization technique; the current study applies PSO-GA specifically to XGBoost for intrusion detection.
M. Zhao et al. (2023)	Particle swarm optimization with adaptive two-population strategy.	Demonstrated faster convergence and improved optimization outcomes.	Provides insights into PSO improvements; complements the hybrid PSO-GA strategy in the current study.
S. Sams Aafiya	SMOTE variants for data balancing in IDS using machine learning.	Enhanced class balance and detection rates in IDS.	Similar use of SMOTE techniques, but the current study integrates

Banu et al. (2023)			hybrid optimization for improved performance.
A. Behera et al. (2024)	Enhanced DDoS detection in SDIoT using SMOTE-ENN with feature selection.	Achieved improved detection rates and data quality using SMOTE-ENN.	Similar use of SMOTE-ENN, but lacks hybrid optimization with PSO-GA as in the current study.
M. Patil et al. (2024)	Comprehensive analysis of ML techniques for phishing attack detection using XGBoost.	Achieved high detection accuracy with XGBoost.	Focused on phishing attack detection, whereas the current study targets broader intrusion detection challenges.

4. Discussion

In this study, the combination of PSO and GA optimization algorithms with the SMOTE-ENN data balancing technique has proven effective in enhancing cyber attack detection using XGBoost. Experimental results show an accuracy improvement from 79.14% to 82.63% and an increase in recall from 80.10% to 85.20%, confirming the effectiveness of this method in addressing data imbalance and optimizing model parameters. SMOTE-ENN successfully improved the distribution of minority classes without introducing significant noise, allowing the model to better learn rare attack patterns. Additionally, the PSO-GA combination enables a more optimal exploration of the solution space compared to traditional parameter tuning methods, with faster convergence and a lower risk of getting trapped in local optima. However, this study has some limitations, such as reliance on the KDD Cup 99 dataset, which may not fully reflect the latest cyber threats. Therefore, future research could explore more complex datasets and test the effectiveness of this approach in real-time scenarios to enhance its practical application in modern cybersecurity systems.

5. Conclusions

We propose a hybrid framework amalgamating SMOTE-ENN for data balancing and a hybrid form of Particle Swarm Optimisation (PSO) and Genetic Algorithms (GA) for hyperparameter optimisation of the model using the XGBoost classifier. The proposed methodology substantially improves model performances by overcoming the challenges inheres from imbalanced datasets and suboptimal parameters tuning, in the field of cybersecurity.

The results show significant enhancements in some of the key metrics like accuracy, precision, recall, F1–score, and AUC, where we achieved an accuracy of 82.63% with the optimized model, against 79.14% for the default. In our approach, majority class samples are represented well with no effect on the majority class representation while the SMOTE sample generation processes focus on increasing the representation of minority class examples and reducing noise samples and thus making our approach sensitive to rare attack patterns. Furthermore, hybrid PSO-GA optimization approach harnesses the benefits of both algorithms to get convergence in limited time and analyze deeper in the parameter space.

Visual analysis with the ROC curves and confusion matrix further confirms the accuracy of the optimized model by minimizing false positives/negatives yielding a robust reliable detection tool in practice for real world intrusion detection systems. This indicates the viability of the proposed method in real-world scenarios, especially in environments where it is crucial to detect cyber threats within a minimum time and with a low false positive rate.

Scalability of the proposed methodology on larger and complex datasets as well as its adaptability to the evolving nature of the cyber threats could be the focus of future work.

In addition, extending the scope of analysis to more sophisticated machine learning models and enabling real-time processing capabilities can improve its usefulness and influence in the field of cybersecurity.

This work fills the gap in existing literature by providing new evidence that hybrid solutions can mitigate the weaknesses of naive and hybrid machine learning models for intrusion detection.

REFERENCES

[1] H. K. R. Kommera, 'Adaptive Cybersecurity in the Digital Age: Emerging Threat Vectors and Next-Generation Defense Strategies', *International Journal for Research in Applied Science and Engineering Technology*, vol. 12, no. 9, 2024, Accessed: Jan. 26, 2025.

[2] I. Ahmad, Q. E. Ul Haq, M. Imran, M. O. Alassafi, and R. A. AlGhamdi, 'An efficient network intrusion detection and classification system', *Mathematics*, vol. 10, no. 3, p. 530, 2022.

[3] G. Velarde, A. Sudhir, S. Deshmane, A. Deshmunkh, K. Sharma, and V. Joshi, 'Evaluating XGBoost for balanced and imbalanced data: application to fraud detection', *arXiv preprint arXiv:2303.15218*, 2023, Accessed: Jan. 24, 2025. [Online]. Available: https://arxiv.org/abs/2303.15218

[4] M. Patil, N. Shivsharan, Y. Naik, H. Yeram, and A. Gawade, 'Enhancing Cybersecurity: A Comprehensive Analysis of Machine Learning Techniques in Detecting and Preventing Phishing Attacks with a Focus on Xgboost Algorithm', in *2024 International Conference on Intelligent Systems for Cybersecurity (ISCS)*, IEEE, 2024, pp. 01–06. Accessed: Jan. 26, 2025. [Online]. Available: https://ieeexplore.ieee.org/abstract/document/10581237/

[5] Z. Saharuna and T. Ahmad, 'Multiclass Imbalance Resampling Techniques for Network Intrusion Detection', in *2024 10th International Conference on Smart Computing and Communication (ICSCC)*, IEEE, 2024, pp. 450–454. Accessed: Jan. 25, 2025. [Online]. Available: https://ieeexplore.ieee.org/abstract/document/10690582/

[6] S. S. Dhaliwal, A.-A. Nahid, and R. Abbas, 'Effective intrusion detection system using XGBoost', *Information*, vol. 9, no. 7, p. 149, 2018.

[7] Z. Xia, S. He, C. Liu, Y. Liu, X. Yang, and H. Bu, 'PSO-GA Hyperparameter Optimized ResNet-BiGRU Based Intrusion Detection Method', *IEEE Access*, 2024, Accessed: Jan. 26, 2025. [Online]. Available: https://ieeexplore.ieee.org/abstract/document/10684706/

[8] N. V. Chawla, K. W. Bowyer, L. O. Hall, and W. P. Kegelmeyer, 'SMOTE: synthetic minority over-sampling technique', *Journal of artificial intelligence research*, vol. 16, pp. 321–357, 2002.

[9] R. Kaur and N. Gupta, 'Comprehending SMOTE Adaptations to Alleviate Imbalance in Intrusion Detection Systems', in *2023 4th International Conference on Electronics and Sustainable Communication Systems (ICESC)*, IEEE, 2023, pp. 976–982. Accessed: Jan. 26, 2025. [Online]. Available: https://ieeexplore.ieee.org/abstract/document/10193257/

[10] K. Cheng, C. Zhang, H. Yu, X. Yang, H. Zou, and S. Gao, 'Grouped SMOTE With Noise Filtering Mechanism for Classifying Imbalanced Data', *IEEE Access*, vol. 7, pp. 170668–170681, 2019, doi: 10.1109/ACCESS.2019.2955086.

[11] K. Kim, 'Noise avoidance SMOTE in ensemble learning for imbalanced data', *IEEE Access*, vol. 9, pp. 143250–143265, 2021.

[12] S. He, B. Li, H. Peng, J. Xin, and E. Zhang, 'An Effective Cost-Sensitive XGBoost Method for Malicious URLs Detection in Imbalanced Dataset', *IEEE Access*, vol. 9, pp. 93089–93096, 2021, doi: 10.1109/ACCESS.2021.3093094.

[13] J. Zhang, H. Wang, J. Zhao, S. Duan, and L. Shi, 'Application of hybrid PSO-GA algorithm in optimization of high-dimensional complex functions', in *2022 7th International Conference on Multimedia and Image Processing*, Tianjin China: ACM, Jan. 2022, pp. 161–166. doi: 10.1145/3517077.3517103.

[14] M. Zhao, H. Zhao, and M. Zhao, 'Particle swarm optimization algorithm with adaptive two-population strategy', *IEEE Access*, vol. 11, pp. 62242–62260, 2023.

[15] P. Zhang, Y. Jia, and Y. Shang, 'Research and application of XGBoost in imbalanced data', *International Journal of Distributed Sensor Networks*, vol. 18, no. 6, p. 155013292211069, Jun. 2022, doi: 10.1177/15501329221106935.

[16] D. Kilichev and W. Kim, 'Hyperparameter optimization for 1D-CNN-based network intrusion detection using GA and PSO', *Mathematics*, vol. 11, no. 17, p. 3724, 2023.

[17] S. Sams Aafiya Banu, B. Gopika, E. Esakki Rajan, M. P. Ramkumar, M. Mahalakshmi, and G. S. R. Emil Selvan, 'SMOTE Variants for Data Balancing in Intrusion Detection System Using Machine Learning', in *Machine Learning and Computational Intelligence Techniques for Data Engineering*, vol. 998, P. Singh, D. Singh, V. Tiwari, and S. Misra,

Eds., in Lecture Notes in Electrical Engineering, vol. 998. , Singapore: Springer Nature Singapore, 2023, pp. 317–330. doi: 10.1007/978-981-99-0047-3_28.

[18] A. Behera, K. Sagar Sahoo, T. Kumara Mishra, A. Nayyar, and M. Bilal, 'Enhancing DDoS detection in SDIoT through effective feature selection with SMOTE-ENN', *PloS one*, vol. 19, no. 10, p. e0309682, 2024.

[19] S. Harron, V. Saxena, and N. Kumari, 'Exploring the Use of Particle Swarm Optimization Algorithms to Enhance Evolutionary Computing', in *2024 International Conference on Optimization Computing and Wireless Communication (ICOCWC)*, IEEE, 2024, pp. 1–6. Accessed: Jan. 26, 2025. [Online]. Available: https://ieeexplore.ieee.org/abstract/document/10470721/

[20] M. H. L. Louk and B. A. Tama, 'PSO-driven feature selection and hybrid ensemble for network anomaly detection', *Big Data and Cognitive Computing*, vol. 6, no. 4, p. 137, 2022.

[21] M. K. Khandelwal and N. Sharma, 'A Survey on Particle Swarm Optimization Algorithm', in *Proceedings of International Conference on Communication and Computational Technologies*, S. Kumar, S. Hiranwal, S. D. Purohit, and M. Prasad, Eds., in Algorithms for Intelligent Systems. , Singapore: Springer Nature Singapore, 2023, pp. 591–602. doi: 10.1007/978-981-99-3485-0_47.