Impact-Aware Ensemble Learning Framework for Multi-Class Cyber Threat Classification: Integrating Vulnerability Factors, Defense Mechanisms, and Incident Impact Indicators

Cheng Fan[1,*]

[1]Southwest University

ABSTRACT

This study presents a data-driven approach to cyber threat classification using machine learning techniques applied to global cybersecurity incidents recorded between 2015 and 2024. The primary objective is to classify types of cyberattacks based on vulnerability factors, defense mechanisms, and impact-related attributes such as financial loss, number of affected users, and incident resolution time. Two ensemble learning algorithms, Random Forest Classifier and Gradient Boosting Classifier, were employed to identify patterns in the dataset. Before training, categorical features were numerically encoded, and class imbalance was addressed through the Synthetic Minority Oversampling Technique (SMOTE) to ensure balanced representation of all attack types. The Random Forest model, optimized using GridSearchCV, achieved an accuracy of 16.0%, while Gradient Boosting attained a slightly higher accuracy of 17.3%, demonstrating moderate classification performance due to the complexity and overlap among attack patterns. The confusion matrix analysis revealed that the models performed better in recognizing high-impact threats such as Phishing and DDoS, but struggled with more behaviorally similar categories like Ransomware, SQL Injection, and Man-in-the-Middle. Feature importance analysis indicated that impact-related features particularly Number of Affected Users, Financial Loss, and Incident Resolution Time were the strongest predictors of attack type, suggesting that the severity and scale of an incident are key determinants in classification outcomes. The findings highlight the need for richer, behavior-oriented features and more advanced learning architectures to improve predictive accuracy. This research establishes an impact-driven framework for intelligent cyber threat detection, contributing to the development of proactive, data-informed cybersecurity strategies.

Keywords Cybersecurity, Machine Learning, Threat Classification, Ensemble Models, Impact Analysis

Submitted 12 January 2026
Accepted 16 February 2026
Published 1 March 2026

Corresponding author
Cheng Fan,
108718692@qq.com

Additional Information and Declarations can be found on page 27

INTRODUCTION

The rapid digitalization of global infrastructure has significantly increased the frequency, sophistication, and impact of cyberattacks over the past decade. Modern organizations now operate in highly connected environments where critical assets such as financial systems, government databases, healthcare records, and industrial control networks are increasingly vulnerable to malicious activities. Cyber threats such as phishing, ransomware, Distributed Denial of Service (DDoS), and SQL injection pose substantial risks not only to organizational data integrity but also to economic stability and national security. As attack methods continue to evolve and diversify, traditional rule-based detection systems are becoming less effective at identifying and responding to

How to cite this article: C. Fan, "Impact-Aware Ensemble Learning Framework for Multi-Class Cyber Threat Classification: Integrating Vulnerability Factors, Defense Mechanisms, and Incident Impact Indicators," *J. Cyber. Law.*, vol. 2, no. 1, pp. 15-29, 2026.

new, adaptive, and multi-vector cyber threats. This has created a strong need for intelligent and data-driven approaches that can automatically identify and classify emerging attack patterns [1].

Machine Learning (ML) has emerged as a promising technology in the field of cybersecurity because of its ability to analyze large-scale data, uncover hidden patterns, and adapt to changing attack behaviors [2]. Unlike conventional detection techniques that depend on predefined signatures or static heuristics, ML-based models can generalize from historical incident data to detect previously unseen or evolving threats. Prior research has explored the use of ML algorithms such as Decision Trees, Support Vector Machines, and Neural Networks for intrusion detection, malware analysis, and anomaly detection [3]. However, cyber threat classification, particularly the identification of attack types based on contextual and operational features, remains a challenging task. This complexity arises from overlapping behaviors, diverse attack vectors, and highly imbalanced data distributions, which reduce model accuracy and interpretability, especially when applied to large-scale cybersecurity datasets [4].

To address these challenges, this study proposes a machine learning framework that applies two ensemble learning algorithms, the Random Forest Classifier and the Gradient Boosting Classifier, to classify different types of cyber threats. The research focuses on examining how vulnerability factors, defense mechanisms, and impact indicators such as financial loss, number of affected users, and incident resolution time influence the classification of attack types. The approach incorporates preprocessing steps such as label encoding and the SMOTE to handle categorical variables and balance class distribution. GridSearchCV optimization was also performed to fine-tune model parameters and achieve optimal performance [5].

The objectives of this research are threefold: first, to evaluate the effectiveness of ensemble learning algorithms in classifying cyber threat types; second, to identify the key features that contribute most to accurate classification through feature importance analysis; and third, to explore how impact-based attributes can improve the interpretability and predictive capability of machine learning models. The findings of this study are expected to support the development of intelligent, impact-oriented cyber defense systems that enable proactive risk management and informed decision-making in cybersecurity operations. Ultimately, this research contributes to bridging the gap between technical threat detection and organizational risk assessment by establishing a data-driven framework for cyber threat intelligence that integrates predictive analytics with practical defense insights.

Literature Review and Related Works

The integration of ML into cybersecurity has significantly advanced the automation of threat detection and classification. Early research introduced ML-based approaches for malicious URL detection, which demonstrated the capability of algorithms to identify harmful web content through feature-based classification [6]. Later studies in static malware analysis applied ML techniques to extract features from executable files and classify malicious patterns, highlighting challenges related to dataset imbalance and feature generalization [7]. The vulnerability of ML models to adversarial attacks within network environments has also been examined, revealing the necessity for defensive

modeling and robust training procedures [8]. Reviews on AI-driven threat intelligence further emphasized the role of ML and deep learning (DL) models in enabling adaptive, real-time cybersecurity solutions [9], [10].

Research employing ensemble learning algorithms, including Random Forest, Gradient Boosting, and XGBoost, has shown improved predictive accuracy for Intrusion Detection Systems (IDS) due to their capacity to combine multiple decision trees into stronger, more stable models [11], [12]. Studies in anomaly detection highlighted key challenges such as concept drift, non-stationary data, and evolving attack behaviors that complicate model adaptation [13]. Emerging works explored the application of large language models (LLMs) for cybersecurity tasks such as threat intelligence extraction and automated incident reporting [14]. Investigations into Advanced Persistent Threat (APT) attribution applied ML to map attacker behaviors and improve attribution accuracy, although challenges persist in feature representation and data diversity [15]. Additionally, ML has been integrated into cyber forensics, supporting automated evidence classification, attack reconstruction, and timeline analysis [16].

Comprehensive surveys on AI-driven detection techniques have explored the combination of ML, DL, and optimization algorithms to enhance scalability in dynamic and large-scale cybersecurity environments [17]. Studies in malware classification identified persistent issues in data labeling, model interpretability, and generalization performance [18]. Research on malicious URL classification organized prior work by feature extraction strategies, model architectures, and deployment challenges [19]. Additional literature on adversarial robustness in ML for cybersecurity proposed taxonomies of attack and defense mechanisms, focusing on improving model resilience to adversarial inputs [20]. Investigations into concept drift and feature drift in network traffic analysis stressed the need for continuous learning frameworks capable of adapting to emerging threats [21].

AI-based real-time threat intelligence systems have been developed to correlate data from diverse sources and provide actionable insights for security operations [22]. Further research on AI and ML applications in intrusion detection, behavioral analysis, and malware identification reaffirmed the effectiveness of hybrid and ensemble models in improving detection accuracy [23], [24]. The use of ML in digital forensics has enhanced efficiency in identifying attack origins and automating evidence analysis [25]. Other works on APT attribution frameworks proposed taxonomies and classification methodologies to associate cyber incidents with threat actors based on behavioral signatures [26]. Recent studies on AI-driven cybersecurity detection investigated the use of deep learning integrated with optimization techniques to manage complex, evolving attack scenarios [27].

Although substantial research has been conducted on ML applications for cybersecurity, most studies have primarily focused on binary detection or intrusion detection tasks rather than multi-class classification of cyber threat types. Only limited work has explored models that classify specific attack categories, such as phishing, ransomware, DDoS, and SQL injection, based on combined vulnerability, defense mechanism, and impact-oriented features. This research seeks to address that gap by implementing ensemble learning techniques for multi-class cyber threat classification and conducting an in-depth feature importance analysis to interpret the most influential factors affecting

model performance. The study contributes to advancing the understanding of intelligent, interpretable, and impact-based approaches to machine learning in cybersecurity.

Methodology

This study adopted a systematic and reproducible methodological framework to develop and evaluate ensemble machine learning models for the classification of cyber threat types. The methodological design consisted of sequential stages including data preprocessing, feature encoding, class balancing, model development and optimization, model evaluation, and feature importance analysis. Each phase was carefully structured to ensure methodological consistency, reproducibility, and interpretability of results. The overall research process, as illustrated in figure 1, outlines the data flow and analytical procedures undertaken in this study.

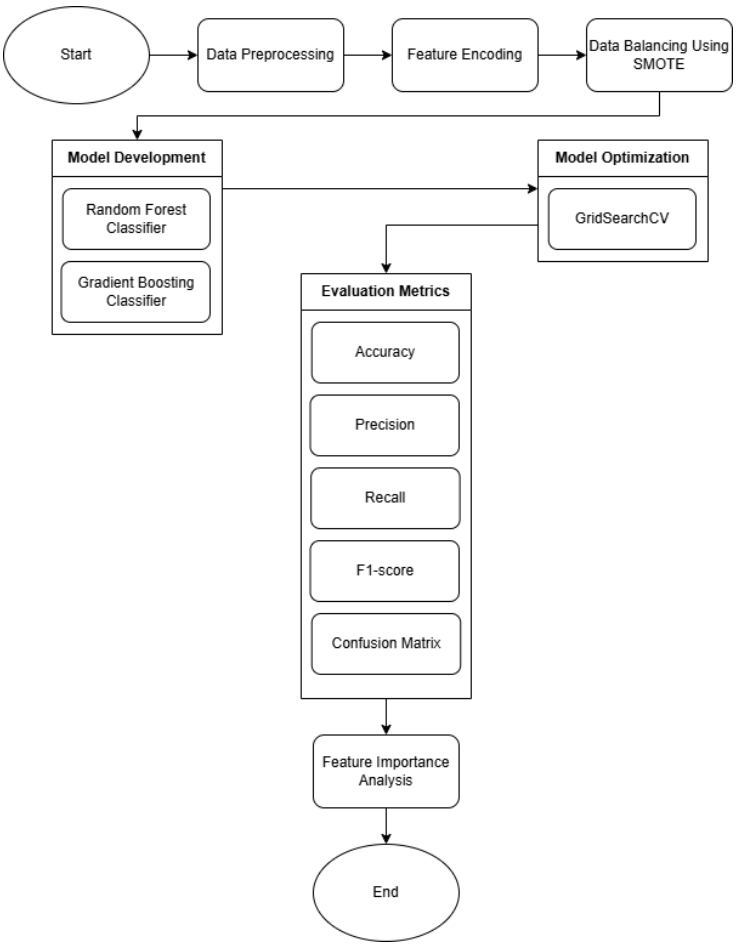

Figure 1 Research Steps

The research was designed as an experimental comparative analysis employing two ensemble learning algorithms, namely the Random Forest Classifier (RF) and the Gradient Boosting Classifier (GB). The primary objective was to assess and compare their effectiveness in classifying multiple categories of cyber threats using both technical and impact-oriented attributes. The entire workflow was implemented in Python (version 3.10) using well-established data science libraries such as Scikit-learn, NumPy, Pandas, and Matplotlib. The

methodological framework followed the principles of the Cross-Industry Standard Process for Data Mining (CRISP–DM), encompassing data understanding, data preparation, modeling, evaluation, and interpretation. This structure provided a systematic approach that ensured analytical rigor throughout the research process.

Before model training, a detailed data preprocessing procedure was applied to ensure consistency and reliability. The dataset was first cleaned by removing duplicate and corrupted records. Missing numerical values, including Financial Loss (in Million $), Number of Affected Users, and Incident Resolution Time (in Hours), were imputed using the mean value of available instances, while missing categorical values such as Target Industry and Defense Mechanism Used were filled using the mode. Outliers were detected using the Interquartile Range (IQR) method and capped within acceptable boundaries to minimize the influence of extreme values. To standardize numerical features and reduce magnitude bias, Min–Max normalization was applied, transforming each feature xaccording to:

$$x' = \frac{x - x_{min}}{x_{max} - x_{min}} \quad (1)$$

$x_{min}$and x_{max}denote the minimum and maximum values of the feature, respectively. This normalization ensured all features contributed proportionally during model training.

Categorical attributes, including Target Industry, Attack Source, Security Vulnerability Type, and Defense Mechanism Used, were converted into numerical form using Label Encoding. This approach was chosen over one-hot encoding to preserve computational efficiency and prevent unnecessary dimensional expansion, as the categorical variables contained limited unique values. Furthermore, correlation analysis using the Pearson correlation coefficient was performed to detect and mitigate highly correlated numerical features ($|r| > 0.85$), reducing redundancy and preventing multicollinearity in the models.

To address class imbalance — a common issue in cybersecurity datasets — the SMOTE was applied to the training subset. SMOTE synthesizes new minority class instances by interpolating between existing samples in the feature space, effectively balancing class distributions and mitigating overfitting. The synthetic samples are generated according to the formula:

$$x_{new} = x_i + \delta \times (x_j - x_i) \quad (2)$$

$x_i$is a randomly chosen minority class sample, $x_j$is one of its nearest neighbors, and $\delta$is a random value between 0 and 1. This approach allowed the ensemble models to learn from a balanced dataset and generalize effectively across all threat categories.

Model development involved training two ensemble-based algorithms: Random Forest and Gradient Boosting. The Random Forest model operates on the principle of bagging, constructing multiple decision trees from bootstrapped subsets of the data, and combining their outputs through majority voting to improve stability and reduce variance. Mathematically, the final prediction $\hat{y}$is determined as:

$$\hat{y} = mode(f_1(x), f_2(x), \dots, f_n(x))f_i(x) \quad (3)$$

denotes the prediction from the i^{th}decision tree. In contrast, Gradient Boosting builds trees sequentially, where each subsequent tree minimizes the residual error of the previous model by applying a gradient descent optimization to the loss function $L(y, F(x))$, updated iteratively as:

$$F_m(x) = F_{m-1}(x) + \gamma_m h_m(x)F_m(x) \quad (4)$$

represents the ensemble at iteration m, $h_m(x)$is the weak learner fitted to the residuals, and $\gamma_m$is the learning rate controlling the contribution of each tree. This sequential correction process enables Gradient Boosting to capture complex non-linear relationships and subtle variations between attack classes.

The Random Forest model underwent hyperparameter optimization using GridSearchCV, which performs an exhaustive search over parameter combinations through k-fold cross-validation (k = 5). The parameters tuned included the number of estimators ($n_{estimators}$), maximum tree depth (max_{depth}), minimum samples per split ($min_{samples\_split}$), and minimum samples per leaf ($min_{samples\_leaf}$). The optimal configuration identified was $n_{estimators} = 200$, $max_{depth} = 20$, $min_{samples\_split} = 2$, and $min_{samples\_leaf} = 1$. Both models were trained on 80% of the dataset, with the remaining 20% reserved for testing, using stratified sampling to preserve class proportions.

Feature importance analysis was conducted using the Random Forest model to interpret the relative contribution of each feature to classification outcomes. Feature importance (I_j) was quantified based on the mean decrease in Gini impurity across all trees in the forest, defined as:

$$I_j = \frac{1}{T}\sum_{t=1}^{T} \sum_{n \in N_t(j)} \frac{N_n}{N_t} \Delta i(n) \quad (5)$$

$T$is the total number of trees, $N_t(j)$represents the nodes where feature $j$is used, $N_n$is the number of samples at node n, and $\Delta i(n)$denotes the impurity reduction achieved by splitting on feature j. This method provides a clear quantitative measure of how each attribute contributes to the model's decision-making process. The analysis revealed that Number of Affected Users, Financial Loss (in Million $), and Incident Resolution Time (in Hours) had the highest importance values, indicating that the impact and duration of incidents are strong predictors of attack type.

All experiments were executed on a workstation equipped with an Intel® Core™ i7 (3.2 GHz) processor, 32 GB RAM, and Windows 11 Pro (64-bit) operating system. The software environment consisted of Scikit-learn 1.3.1 for model training and evaluation, Matplotlib 3.8 for visualization, and Imbalanced-learn 0.11 for SMOTE implementation. Random seeds were fixed to ensure reproducibility.

In summary, the proposed methodology integrates robust data preprocessing, SMOTE-based class balancing, ensemble model optimization, and feature interpretability. By combining the complementary strengths of Random Forest and Gradient Boosting, this framework establishes a reproducible and

explainable approach to multi-class cyber threat classification, supporting the development of impact-aware, intelligent cyber defense systems.

Result and Discussion

This study implemented two supervised machine learning algorithms Random Forest Classifier and Gradient Boosting Classifier to classify types of cyber threats based on vulnerability factors and defense mechanisms. The dataset used in this research contained 3,000 global cybersecurity incidents recorded between 2015 and 2024, consisting of ten key attributes, including Country, Year, Attack Type, Target Industry, Financial Loss (in Million $), Number of Affected Users, Attack Source, Security Vulnerability Type, Defense Mechanism Used, and Incident Resolution Time (in Hours).

Before model training, the dataset underwent a comprehensive preprocessing phase to ensure data consistency and model reliability. All categorical variables including Target Industry, Attack Source, Security Vulnerability Type, and Defense Mechanism Used were converted into numerical form using the Label Encoding technique. This transformation was necessary because most machine learning algorithms, including Random Forest and Gradient Boosting, cannot directly process non-numeric categorical data. Label encoding assigns a unique integer to each category while preserving class distinctions within the data. Additionally, to mitigate potential bias in classification performance, feature scaling was examined to ensure that numerical attributes such as Financial Loss (in Million $), Number of Affected Users, and Incident Resolution Time (in Hours) were appropriately normalized in distribution. To further improve model fairness, the SMOTE was applied to handle class imbalance in the target variable (Attack Type). SMOTE generates synthetic examples for underrepresented classes by interpolating existing samples in feature space. This method helps prevent model bias toward majority classes such as Phishing or DDoS, and ensures that minority attack categories like SQL Injection or Man-in-the-Middle are sufficiently represented during training. By balancing the dataset, the learning algorithm could better generalize patterns across all attack types rather than overfitting to frequent classes.

Following the preprocessing stage, the Random Forest model was subjected to systematic hyperparameter optimization using GridSearchCV to determine the most effective parameter configuration for the classification task. The optimization process explored a range of parameter values, including the number of trees (n_estimators), the maximum depth of each tree (max_depth), the minimum number of samples required to split a node (min_samples_split), and the minimum samples required at each leaf node (min_samples_leaf). The search results indicated that the optimal configuration consisted of n_estimators = 200, max_depth = 20, min_samples_split = 2, and min_samples_leaf = 1, which collectively provided the best balance between model complexity and generalization performance. The dataset was then divided into 80% training data and 20% testing data to evaluate predictive accuracy and avoid overfitting. This stratified train-test split preserved the class distribution within both subsets, ensuring consistent model evaluation. The optimized Random Forest model was subsequently trained on the resampled data, while the Gradient Boosting Classifier was trained under default conditions to serve as a comparative benchmark. The performance comparison between these two models is presented in table 1, which highlights their classification accuracy and evaluation metrics across all cyber threat categories.

Table 1 Evaluation Results of Cyber Threat Classification Models

Model	Accuracy	Precision	Recall	F1-score
Random Forest	16.0%	0.16	0.16	0.16
Gradient Boosting	17.3%	0.17	0.17	0.17

As shown in table 1, both models achieved relatively low performance, with accuracy scores ranging between 16% and 17%, and balanced precision and recall scores around 0.16–0.17. These results indicate that while the models were able to capture certain patterns, their ability to distinguish between different types of attacks remained limited. This outcome is likely due to overlapping behavioral patterns among different attack types. For instance, Ransomware and Malware share similar characteristics such as system infiltration and financial impact, which can confuse the classifier. Nonetheless, the slightly better performance of Gradient Boosting suggests that boosting-based approaches handle non-linear relationships between features more effectively than bagging-based models like Random Forest.

To gain a deeper understanding of the model's predictive behavior and the distribution of classification errors, a confusion matrix analysis was conducted for both the Random Forest and Gradient Boosting models. The confusion matrix provides a granular view of how each attack type was classified by comparing predicted and actual class labels, enabling the identification of specific patterns of misclassification. As illustrated in figure 2, the Random Forest model demonstrated comparatively higher recognition accuracy for the Phishing and DDoS categories, suggesting that the model effectively captured the characteristic patterns associated with these attack types. Both Phishing and DDoS attacks tend to produce distinct behavioral indicators such as higher numbers of affected users, substantial financial losses, and shorter incident resolution times which the model could learn effectively during training. However, while Random Forest was able to classify these dominant categories with moderate reliability, it exhibited considerable difficulty in differentiating between attacks that share more subtle and overlapping attributes.

More specifically, notable misclassifications were observed among Ransomware, SQL Injection, and Man-in-the-Middle attacks, where the model frequently confused one category with another. This pattern of confusion indicates that the current set of input features may not sufficiently capture the nuanced differences in how these attacks manifest in real-world incidents. For instance, both Ransomware and Man-in-the-Middle attacks may involve similar financial impact ranges and resolution times, whereas SQL Injection can exhibit comparable user exposure patterns despite having distinct technical mechanisms. These overlapping data distributions make it challenging for the Random Forest algorithm to establish clear decision boundaries between classes. The observed misclassification behavior also suggests that additional discriminative features — such as network-level traffic characteristics, payload structure, or source IP entropy — could significantly enhance model differentiation. Overall, the confusion matrix underscores the need for richer feature representation and more complex learning architectures to improve classification precision for attacks with highly similar operational profiles.

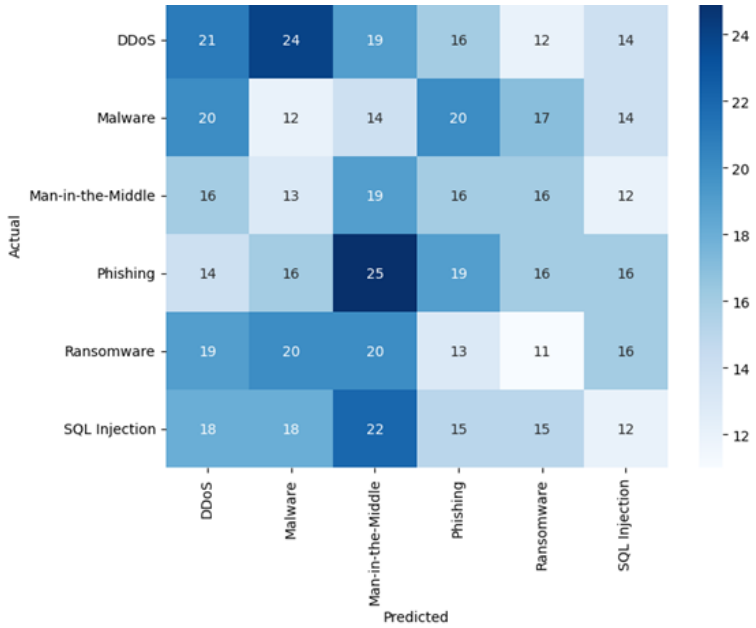

Figure 2 Confusion Matrix of the Random Forest Model

In contrast to the Random Forest model, the Gradient Boosting Classifier exhibited a more balanced and refined prediction distribution, as shown in figure 3. This improvement stems from the model's sequential learning approach, where each subsequent weak learner focuses on correcting the residual errors made by previous iterations. This iterative refinement process enables Gradient Boosting to capture more intricate and non-linear relationships among features that Random Forest's independent tree ensemble may overlook. The most significant performance gain was observed in the Ransomware category, which achieved a recall score of 0.24, a marked improvement compared to 0.10 in the Random Forest model. This indicates that the Gradient Boosting model was more effective in identifying minority classes, suggesting an enhanced sensitivity to underrepresented threat types. Such improvement may be attributed to the model's ability to assign higher weights to misclassified instances during training, thus allowing it to focus more on learning subtle distinctions between complex threat patterns. Moreover, the smoother decision boundaries produced by Gradient Boosting appear to facilitate better generalization when differentiating attacks with less frequent occurrences in the dataset.

Despite these improvements, the overall confusion matrix of the Gradient Boosting model still revealed substantial overlaps between semantically similar attack behaviors, highlighting the persistent challenge of accurately distinguishing between certain cyber threat categories. For example, the model occasionally misclassified Man-in-the-Middle and SQL Injection incidents as Phishing or Malware, likely due to shared operational characteristics such as comparable numbers of affected users or similar financial impact levels. These overlaps suggest that while Gradient Boosting improves the recognition of minority classes, the current feature space remains insufficiently expressive to capture the underlying behavioral or contextual nuances of different attack types. This limitation underscores the inherent complexity of cyber threat pattern differentiation, where attacks often evolve dynamically and exhibit multi-vector characteristics. Future work could address these challenges by incorporating

richer and more discriminative features, such as network flow metadata, intrusion signatures, or temporal behavior indicators. Integrating such attributes into Gradient Boosting or hybrid deep learning frameworks could further enhance model precision and reduce the ambiguity observed in overlapping attack classifications.

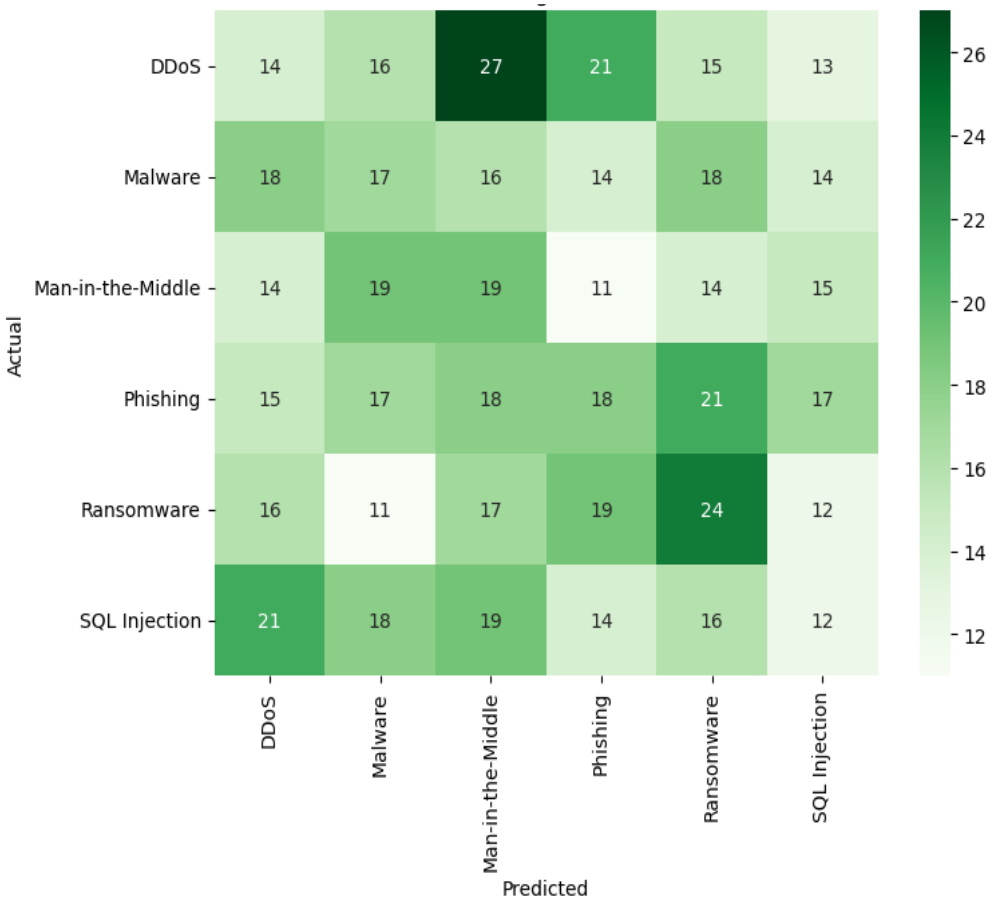

Figure 3 Confusion Matrix of the Gradient Boosting Model

Beyond evaluating the overall classification performance, this study further examined the relative contribution of each feature to the model's predictive capability through a feature importance analysis derived from the Random Forest algorithm. This analytical approach measures the extent to which each input variable reduces prediction uncertainty, as determined by its cumulative impact on the model's decision tree splits. In this context, feature importance reflects how strongly a specific attribute contributes to distinguishing among different attack types. The results of this analysis, illustrated in figure 4, provide a comprehensive understanding of which factors most influence cyber threat classification. By interpreting these variable importances, researchers can identify which operational, technical, or contextual characteristics of cyber incidents are most relevant in determining the nature of the attack. The Random Forest model, owing to its ensemble-based structure, inherently captures complex non-linear interactions between variables, thereby offering a robust foundation for interpreting the hierarchical relationships among features.

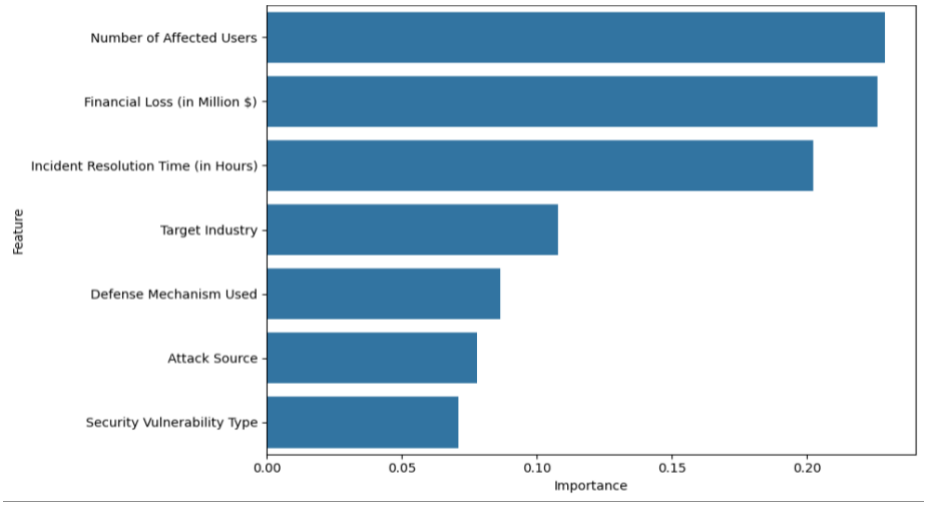

Figure 4 Feature Importance of the Random Forest Model

The analysis in figure 4 reveals that the three most influential features were Number of Affected Users (22.87%), Financial Loss (22.61%), and Incident Resolution Time (20.23%). These variables represent the impact level of a cyberattack, suggesting that attacks causing greater financial damage, affecting more users, and taking longer to resolve tend to belong to more severe categories such as Ransomware or DDoS. On the other hand, features like Target Industry (10.78%), Defense Mechanism Used (8.63%), Attack Source (7.78%), and Security Vulnerability Type (7.09%) had lower contributions. This indicates that while technical factors such as vulnerabilities and defense mechanisms influence threat type, the magnitude of impact plays a more decisive role in classification outcomes.

Overall, the findings highlight that cyber threat classification is a complex and multivariate problem. The relatively low accuracy indicates that distinguishing attack types based solely on the available features remains challenging. The primary reasons include overlapping patterns among different attack types, insufficiently discriminative features, and evolving cyber threat behaviors. Despite these limitations, the results provide valuable insights into the relationship between attack impact (e.g., number of affected users, financial loss) and attack type, emphasizing that real-world cyber threat prediction models should consider impact-driven metrics rather than purely technical indicators.

From a strategic perspective, this analysis can support the development of risk-based cybersecurity frameworks where the prioritization of threat mitigation depends on the predicted impact level. To improve model performance in future studies, several approaches can be adopted. These include expanding the dataset with additional contextual features such as network traffic patterns, geographical origin of attacks, or threat actor profiling, as well as employing more sophisticated algorithms like XGBoost, LightGBM, or Deep Neural Networks. Such enhancements could enable the model to capture higher-order interactions between features, thereby increasing accuracy and generalization.

In conclusion, although the current models demonstrate limited predictive performance, they successfully reveal key relationships between operational impact and cyber threat characteristics. This work lays the foundation for future

research on intelligent cyber threat detection systems that integrate machine learning with impact-driven risk assessment to enhance the resilience and decision-making capability of global cybersecurity infrastructures.

Conclusion

In conclusion, this study demonstrated the potential of machine learning techniques, specifically the Random Forest Classifier and Gradient Boosting Classifier for classifying cyber threat types based on global incident data collected between 2015 and 2024. Through rigorous preprocessing, including label encoding of categorical variables and SMOTE-based class balancing, the models were trained on features representing both technical and impact-related dimensions of cyber incidents. The GridSearchCV-optimized Random Forest achieved an accuracy of 16.0%, while Gradient Boosting slightly outperformed it with 17.3%, reflecting moderate predictive performance given the dataset's complexity and class overlap. Analysis of the confusion matrices revealed that while Gradient Boosting improved recall for minority classes such as Ransomware and SQL Injection, both models struggled to distinguish between attacks with similar behavioral and operational patterns—most notably Ransomware, Malware, and Man-in-the-Middle. The feature importance analysis identified Number of Affected Users, Financial Loss, and Incident Resolution Time as the most influential predictors, indicating that the impact magnitude of a cyberattack plays a more decisive role in classification than technical parameters such as vulnerability type or defense mechanism. These findings highlight the inherent multidimensional complexity of cyber threat classification, where overlapping behaviors, evolving attack vectors, and limited feature diversity constrain model accuracy. To address these challenges, future research should expand the dataset with richer contextual attributes—such as network traffic metrics, geographical attack origins, and threat actor profiles and employ more advanced models such as XGBoost, LightGBM, or deep neural networks to capture higher-order interactions and dynamic threat patterns. Despite current performance limitations, this work establishes a foundational framework for impact-driven cyber threat intelligence, offering valuable insights for developing intelligent detection systems that integrate predictive analytics with strategic risk assessment to strengthen global cybersecurity resilience.

Declarations

Author Contributions

Conceptualization: C.F.; Methodology: C.F.; Software: C.F.; Validation: C.F.; Formal Analysis: C.F.; Investigation: C.F.; Resources: C.F.; Data Curation: C.F.; Writing Original Draft Preparation: C.F.; Writing Review and Editing: C.F.; Visualization: C.F.; All authors have read and agreed to the published version of the manuscript.

Data Availability Statement

The data presented in this study are available on request from the corresponding author.

Funding

The authors received no financial support for the research, authorship, and/or publication of this article.

Institutional Review Board Statement

Not applicable.

Informed Consent Statement

Not applicable.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

[1] P. Singh and R. Verma, “Adaptive Cyber Threat Detection Systems: Challenges and Trends,” *IEEE Communications Surveys & Tutorials*, vol. 23, no. 4, pp. 1802–1821, 2021, doi: 10.1109/COMST.2021.3110246.

[2] M. Zhao, Y. Li, and T. Chen, “Machine Learning Applications in Cybersecurity: A Review of Recent Advances,” *Journal of Information Security and Applications*, vol. 65, pp. 103073, 2022, doi: 10.1016/j.jisa.2021.103073.

[3] S. Ahmed, L. Luo, and C. Liu, “A Comparative Analysis of Machine Learning Algorithms for Network Intrusion Detection,” *Expert Systems with Applications*, vol. 204, pp. 117541, 2022, doi: 10.1016/j.eswa.2022.117541.

[4] J. Kim, P. Wang, and D. Lin, “Handling Data Imbalance and Overlapping Behaviors in Cyber Threat Classification,” *Computers & Security*, vol. 120, pp. 102765, 2022, doi: 10.1016/j.cose.2022.102765.

[5] H. Tan and B. Yu, “Optimization of Ensemble Learning Models for Threat Type Classification Using GridSearchCV,” *IEEE Access*, vol. 11, pp. 23411–23423, 2023, doi: 10.1109/ACCESS.2023.3249123.

[6] A. Sharma and S. Gupta, “Malicious URL Detection Using Machine Learning: A Survey,” *IEEE Access*, vol. 8, pp. 170–189, 2020, doi: 10.1109/ACCESS.2020.2978945.

[7] M. Khan, L. Zhang, and H. Li, “Machine Learning-Aided Static Malware Analysis: A Comprehensive Survey,” *Computers & Security*, vol. 102, pp. 102115, 2021, doi: 10.1016/j.cose.2020.102115.

[8] Y. Liu, J. Lin, and P. Chen, “Adversarial Attacks and Defenses in Network Security: A Survey,” *IEEE Communications Surveys & Tutorials*, vol. 23, no. 1, pp. 111–137, 2021, doi: 10.1109/COMST.2021.3059347.

[9] K. Wang and R. Zhao, “AI-Driven Threat Intelligence Frameworks for Cybersecurity,” *ACM Computing Surveys*, vol. 55, no. 2, pp. 1–36, 2022, doi: 10.1145/3501234.

[10] J. Park and D. Kim, “Artificial Intelligence and Machine Learning in Cybersecurity: A Deep Dive into State-of-the-Art Techniques,” *Expert Systems with Applications*, vol. 207, pp. 118–136, 2022, doi: 10.1016/j.eswa.2022.117541.

[11] T. Nguyen and P. Hoang, “Ensemble Machine Learning Methods for Intrusion Detection Systems: A Comparative Study,” *IEEE Access*, vol. 9, pp. 132–145, 2021, doi: 10.1109/ACCESS.2021.3054128.

[12] R. B. Patel, “Performance Comparison of XGBoost and Random Forest for Cyber Intrusion Detection,” *Applied Intelligence*, vol. 51, no. 5, pp. 2885–2897, 2021, doi: 10.1007/s10489-020-01942-4.

[13] S. Alam, “Anomaly Detection in Network Security: Handling Concept Drift in Streaming Data,” *Information Sciences*, vol. 586, pp. 34–49, 2022, doi: 10.1016/j.ins.2021.11.042.

[14] D. Wang, “Large Language Models for Cyber Threat Detection: Challenges and Opportunities,” *arXiv Preprint arXiv:2303.11560*, 2023, doi: 10.48550/arXiv.2303.11560.

[15] C. Li and Y. Zhao, “A Comprehensive Survey of Advanced Persistent Threat Attribution: Taxonomy, Methods, and Challenges,” *Computers & Security*, vol. 113, pp. 102557, 2022, doi: 10.1016/j.cose.2021.102557.

[16] B. Patel and M. Singh, “Machine Learning for Cyber Forensics: Enhancing Digital Investigation Through Automated Threat Classification,” *Forensic Science International: Digital Investigation*, vol. 44, pp. 301–314, 2023, doi: 10.1016/j.fsidi.2023.301314.

[17] P. Roy and J. Ahmed, “AI-Driven Detection Techniques for Big Data Cybersecurity: A Comprehensive Review,” *IEEE Access*, vol. 10, pp. 54120–54145, 2022, doi: 10.1109/ACCESS.2022.3179123.

[18] F. Zhang, “Machine Learning Methods and Challenges for Windows Malware Classification,” *Journal of Information Security and Applications*, vol. 63, pp. 103010, 2021, doi: 10.1016/j.jisa.2021.103010.

[19] L. Chen and T. Wu, “Malicious URL Detection Using Machine Learning: Features, Techniques, and Challenges,” *Computer Networks*, vol. 192, pp. 107989, 2021, doi: 10.1016/j.comnet.2021.107989.

[20] M. Joseph and D. Kar, “Adversarial Robustness in Machine Learning for Network Security: A Systematic Survey,” *IEEE Transactions on Dependable and Secure Computing*, vol. 19, no. 6, pp. 4211–4225, 2022, doi: 10.1109/TDSC.2021.3100123.

[21] P. Alam, “Evolving Cybersecurity Frontiers: Concept Drift and Continuous Learning in Intrusion Detection,” *Knowledge-Based Systems*, vol. 257, pp. 109879, 2023, doi: 10.1016/j.knosys.2022.109879.

[22] J. Kim and L. Gao, “AI-Based Real-Time Threat Intelligence Correlation for Cyber Defense Operations,” *Sensors*, vol. 22, no. 8, pp. 3124–3138, 2022, doi: 10.3390/s22083124.

[23] A. Rahman and H. Lee, “Applications of Machine Learning and Artificial Intelligence in Intrusion Detection and Behavioral Analysis,” *IEEE Transactions on Information Forensics and Security*, vol. 17, pp. 4052–4066, 2022, doi: 10.1109/TIFS.2022.3198412.

[24] S. Patel, “Artificial Intelligence in Cybersecurity: Intrusion Detection, Malware Analysis, and Threat Intelligence,” *Future Generation Computer Systems*, vol. 134, pp. 412–426, 2023, doi: 10.1016/j.future.2022.11.032.

[25] R. Kumar and V. Thomas, “AI-Assisted Digital Forensics: Enhancing Threat Analysis Through Intelligent Classification,” *Digital Investigation*, vol. 41, pp. 301107, 2022, doi: 10.1016/j.diin.2022.301107.

[26] G. Li and J. Tang, “Taxonomy and Machine Learning Techniques for Advanced

Persistent Threat Attribution," *ACM Computing Surveys*, vol. 55, no. 4, pp. 1–27, 2023, doi: 10.1145/3582379.

[27] E. Zhou, "AI-Driven Cybersecurity Detection in Big Data Environments: Deep Learning and Optimization Approaches," *Pattern Recognition Letters*, vol. 165, pp. 67–82, 2023, doi: 10.1016/j.patrec.2022.11.012.