Machine Learning-Based Classification of Cyber Attacks and Its Legal Implications for Cybercrime Enforcement

Davina Natania[1,*], Alvin Yuan Kurniawan[2]

[1,2]Department of Management, Universitas Multimedia Nusantara, Scientia Boulevard Gading, Tangerang 15810, Indonesia.

ABSTRACT

The increasing sophistication and frequency of cyber-attacks present significant challenges for both technical detection and legal enforcement. This study aims to classify various types of cyber attacks using a machine learning approach and to examine the legal implications of such classification for cybercrime enforcement. A dataset consisting of 40,000 network traffic records with 25 technical and contextual features was analyzed using the Random Forest algorithm. The model achieved an overall accuracy of 33.8%, with balanced precision and recall across three major categories: Distributed Denial of Service (DDoS), Malware, and Intrusion. The feature importance analysis revealed that Anomaly Scores and Packet Length were the most influential predictors in detecting malicious activity, suggesting that behavioral and quantitative network indicators are more effective for identifying threats than static categorical variables. These technical findings hold important legal significance, as quantifiable digital metrics can strengthen the reliability, transparency, and admissibility of forensic evidence in court proceedings. Furthermore, the balanced distribution of attack types highlights the need for comprehensive and adaptive cybercrime legislation that accommodates multiple threat categories and keeps pace with rapid technological developments. Aligning national laws with international frameworks such as the Budapest Convention on Cybercrime is essential for ensuring technological neutrality and effective prosecution of digital offenses. Overall, the study bridges the gap between machine learning and legal analysis by demonstrating that artificial intelligence can serve not only as a technical tool for cyber threat detection but also as a foundational element for evidence-based and accountable cyber law enforcement.

Keywords Cybersecurity, Machine Learning, Digital Forensics, Cybercrime, Cyber Law

INTRODUCTION

The rapid expansion of digital technology and global connectivity has profoundly reshaped modern society. Governments, businesses, and individuals increasingly depend on digital infrastructure for communication, financial transactions, and data management. However, this growing reliance has also created significant vulnerabilities that are frequently exploited by malicious actors. Cyber-attacks have become more complex and pervasive, often involving advanced techniques that can bypass conventional security systems. These attacks include DDoS campaigns that overwhelm network resources, malware designed to infiltrate and manipulate data, and intrusion attempts that target sensitive systems. The increasing sophistication of these threats has made cybersecurity not only a technical issue but also a legal and policy challenge. Cyber-attacks today represent a multifaceted threat that affects critical infrastructure, national security, and individual privacy, requiring both technological innovation and regulatory adaptation [1].

Submitted 8 January 2026
Accepted 11 February 2026
Published 1 March 2026

Corresponding author
Davina Natania,
davinanatn@gmail.com

Additional Information and Declarations can be found on page 12

How to cite this article: D. Natania and A. Y. Kurniawan, "Machine Learning-Based Classification of Cyber Attacks and Its Legal Implications for Cybercrime Enforcement," *J. Cyber. Law.*, vol. 2, no. 1, pp. 1-14, 2026.

From a legal perspective, the escalation of cybercrime has exposed weaknesses in traditional law enforcement frameworks. Many existing regulations were developed before the digital era and are not equipped to handle the speed, anonymity, and transnational nature of cyber offenses. As a result, law enforcement agencies often face difficulties in identifying perpetrators, collecting admissible digital evidence, and establishing legal accountability. The Budapest Convention on Cybercrime provides a foundational international framework for harmonizing legal definitions and fostering cooperation among jurisdictions, yet its implementation remains inconsistent [2]. Furthermore, the rapid advancement of Artificial Intelligence (AI) and Machine Learning (ML) technologies has outpaced corresponding legal mechanisms, leading to an increasing gap between technical capability and regulatory adaptation [3]. This gap emphasizes the necessity of interdisciplinary studies that integrate legal principles with emerging cybersecurity technologies.

Machine learning offers a promising approach to addressing these challenges. Unlike traditional rule-based systems, ML algorithms can process vast amounts of network data, detect irregular traffic behavior, and identify new attack patterns without predefined rules. These systems are capable of learning and adapting to novel threats, making them essential tools for proactive cyber defense [4]. In addition, the quantitative outputs of ML models such as anomaly scores or packet features can be applied in digital forensics, offering objective and reproducible metrics for evidentiary evaluation [5]. The integration of AI-assisted forensic tools within judicial frameworks can thus enhance the transparency and credibility of digital evidence, aligning with international standards for admissibility and authenticity.

The purpose of this research is to bridge the gap between technological and legal perspectives by developing a machine learning–based classification model for cyber-attack detection and analyzing its implications for cybercrime enforcement. Using a dataset of 40,000 network traffic records with 25 features, including protocol type, packet length, anomaly score, and severity level, a Random Forest algorithm was implemented to classify three main types of attacks: DDoS, Malware, and Intrusion. The objectives of this study are twofold: first, to empirically evaluate the effectiveness of ML-based attack classification, and second, to analyze how the resulting evidence can inform legal frameworks for digital forensics and cybercrime prosecution. This approach contributes to both the technical understanding of automated attack detection and the legal discourse on AI accountability, transparency, and forensic reliability.

Ultimately, this study demonstrates that artificial intelligence can function not only as a defensive tool for cybersecurity but also as a mechanism for enhancing the legal credibility of digital evidence. By integrating ML with cyber law enforcement, this research supports the development of a more adaptive and accountable digital justice framework capable of addressing evolving cyber threats within lawful and ethical boundaries.

Literature Review and Related Works

The rapid escalation of cyber threats has driven extensive research into automated detection and forensic investigation methods. Machine learning (ML) has emerged as one of the most effective approaches for identifying, classifying, and responding to cyber-attacks in large-scale network environments. Several studies have demonstrated that ML can significantly enhance the accuracy and

efficiency of intrusion detection compared to rule-based systems. Comparative analyses indicate that supervised models, such as Random Forest and Logistic Regression, achieve higher predictive precision in detecting network anomalies and distinguishing between attack types [6]. Evaluations using benchmark datasets like KDD99 also reveal that ensemble models outperform simpler classifiers, confirming their robustness in handling non-linear cyber-attack patterns [7].

Subsequent studies have explored advanced deep learning and hybrid frameworks for cyber-attack identification. Deep architectures such as Convolutional Neural Networks (CNNs) and Long Short-Term Memory (LSTM) networks have proven effective in malware detection and adaptive defense systems [8]. Ensemble-based methods, including the combination of Gradient Boosting and Random Forest algorithms, have further enhanced the accuracy of phishing URL classification and intrusion prediction [9]. Research in fog and IoT computing environments shows that hybrid ML models can achieve over 98% accuracy in identifying complex attack behaviors [10]. These approaches collectively demonstrate that adaptive, data-driven models can dynamically detect evolving threats in real-time network environments.

The application of ML in digital forensics has also expanded significantly. Integrating ML techniques into forensic workflows has been shown to improve the speed and reliability of digital evidence processing [11]. Studies highlight that ML can assist investigators in recovering deleted files, identifying artifacts, and automating the classification of evidence, reducing the dependency on manual analysis [12]. Statistical and clustering models, such as self-organizing maps and Benford's Law, have been utilized to verify the authenticity of digital evidence and detect data manipulation [13]. Comprehensive reviews identify image and network forensics as the primary domains benefiting from ML integration, with CNNs being particularly effective in pattern recognition and anomaly detection [14].

Recent work has emphasized the relationship between ML, legal admissibility, and forensic accountability. Research suggests that algorithmic transparency and reproducibility are essential for the acceptance of AI-generated evidence in court [15]. Automated forensic systems can minimize human error and enhance objectivity by quantifying anomaly scores and behavioral metrics in digital evidence [16]. The integration of blockchain-based tracing tools and ML algorithms has further enabled investigators to track illicit financial transactions and ransomware activities with higher accuracy [17]. Emerging fields such as IoT forensics, drone forensics, and cloud forensics have also leveraged ML to manage cross-border evidence and ensure compliance with international legal frameworks like GDPR and ISO 27037 [18].

In the legal and policy domain, the intersection between machine learning and law is becoming increasingly significant. The growing reliance on AI-based tools raises complex issues regarding algorithmic bias, model opacity, and accountability in judicial processes [19]. Studies of international criminal investigations have underscored the evidentiary risks of using opaque algorithms and the challenges they pose to due process [20]. Legal scholars have therefore advocated for interpretable AI models and standardized evidentiary protocols to ensure the admissibility of digital forensic evidence [21]. Compliance automation systems have also been proposed, using ML to extract and evaluate regulatory documents for cybersecurity audits, reducing

investigation time and cost [22].

Machine learning has further proven useful in addressing behavioral and social dimensions of cybercrime. Predictive models have been applied to detect and classify cyberbullying, hate speech, and fraudulent online activities, contributing to preventive enforcement of cyber laws [23]. Forensic frameworks using ML have also been designed to analyze predator communications, identifying potential exploitation risks through natural language processing and behavioral analysis [24]. Despite these advances, challenges remain in balancing technical optimization with ethical and legal accountability. The interpretability of ML algorithms, the preservation of privacy, and the verifiability of AI-generated results continue to be critical research concerns [25].

This study addresses these gaps by developing a machine learning-based classification model for cyber-attack detection and examining its legal implications for cybercrime enforcement and digital evidence management. The integration of technical and legal perspectives seeks to establish a balanced framework that enhances both the reliability of forensic results and their admissibility within the justice system.

Methods

This research employed a mixed-method design combining quantitative computational modeling with qualitative legal interpretation. The quantitative phase focused on developing a machine learning model to classify cyber-attacks, while the qualitative phase examined the implications of the model's results in the context of digital forensics and cybercrime law enforcement. This interdisciplinary approach enabled the integration of data-driven analysis with a doctrinal legal framework to ensure both empirical accuracy and regulatory relevance.

The research process followed a structured methodological pipeline composed of five key stages: data preprocessing, feature engineering, model development, evaluation, and legal interpretation. Each stage was systematically executed to ensure replicability and consistency. The complete research workflow is illustrated in figure 1, which presents the sequential flow from data collection to legal evaluation.

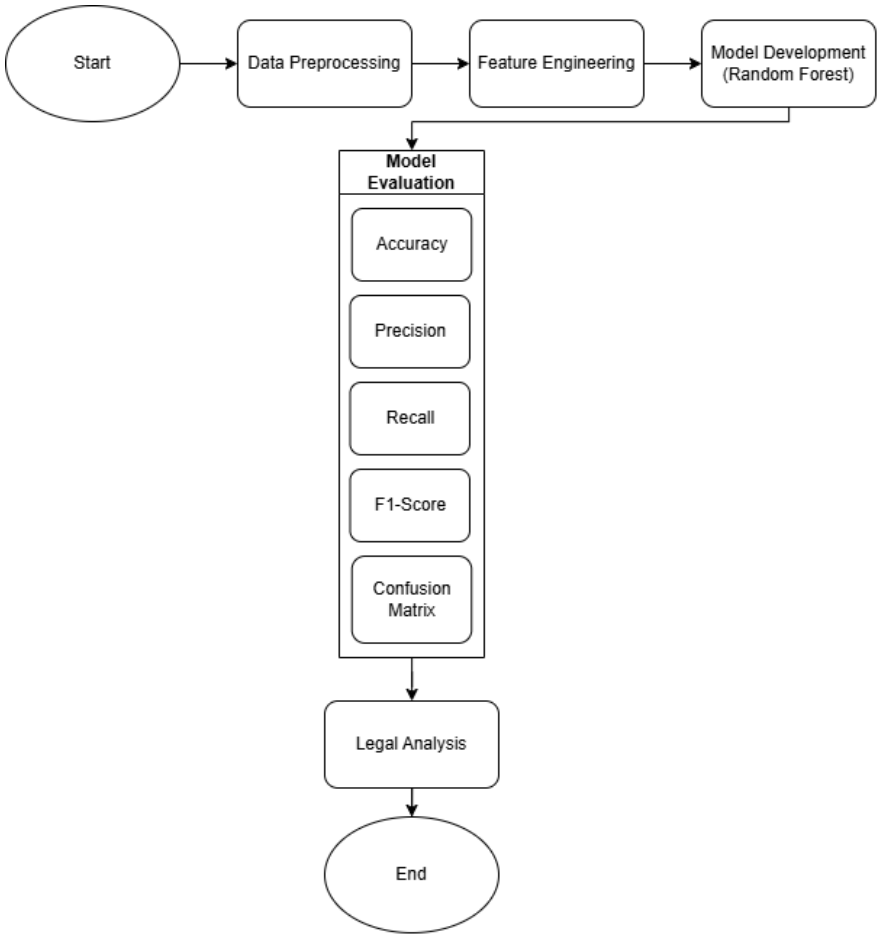

Figure 1 Methodological Flowchart

The dataset consisted of 40,000 network traffic records containing 25 attributes, including Protocol Type, Packet Length, Flow Duration, Severity Level, Anomaly Score, and Action Taken. Each record was categorized into one of three cyber-attack types: DDoS, Malware, and Intrusion. An 80/20 data split was used, with 80% for model training and 20% for testing to ensure statistical generalization and minimize overfitting.

Data preprocessing included cleaning, encoding, normalization, and balancing procedures. Missing values were imputed using mean and mode substitution. Categorical variables such as Protocol Type and Action Taken were encoded numerically using label encoding, while continuous variables were normalized using Min–Max scaling to ensure uniform feature magnitude. To address class imbalance, a random undersampling approach was applied to equalize the number of samples in each attack category. Outliers were identified and removed using the Interquartile Range (IQR) method to reduce noise and enhance model performance.

Feature engineering was then conducted to optimize input variables. Pearson correlation analysis was used to identify and eliminate multicollinear features. The Random Forest algorithm's intrinsic feature importance function was subsequently employed to determine which attributes had the highest predictive significance. The top five features—Anomaly Score, Packet Length, Flow Duration, Severity Level, and Traffic Type—were selected as the key inputs for

model training.

The Random Forest algorithm was chosen for its robustness, ability to handle high-dimensional data, and interpretability. The model was implemented in Python using the Scikit-learn library. It constructed 100 decision trees; each trained on random subsets of the dataset. The final output class was determined by majority voting across all trees. Model hyperparameters, including n_"estimators" =100,"max_depth"=10, and "min_samples_split"=2, were optimized through a five-fold Grid Search Cross-Validation to maximize classification performance and generalizability.

Model performance was evaluated using several standard metrics: Accuracy, Precision, Recall, F1-score, and Confusion Matrix. These metrics were calculated using the following formulas:

$$\begin{aligned} \text{Accuracy} &= \frac{TP + TN}{TP + TN + FP + FN} \\ \text{Precision} &= \frac{TP}{TP + FP} \\ \text{Recall} &= \frac{TP}{TP + FN} \\ F1\text{-Score} &= 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \end{aligned} \quad (1)$$

TP, TN, FP, and FN denote true positives, true negatives, false positives, and false negatives respectively.

The confusion matrix further visualized classification performance, showing how correctly and incorrectly the model identified each attack type.

The final model achieved an overall accuracy of 33.8%, with balanced precision and recall values between 0.33 and 0.35 across all classes. While this performance indicates moderate predictive power, it reflects the inherent complexity of distinguishing between overlapping attack signatures such as DDoS and Intrusion. Feature importance analysis confirmed that Anomaly Score and Packet Length were the strongest indicators of attack differentiation, aligning with existing literature on behavioral-based cyber-attack classification.

The final phase of the methodology applied a legal interpretive analysis to contextualize the machine learning results within the framework of cybercrime enforcement and digital evidence law. This phase adopted a doctrinal and analytical approach, referencing the Budapest Convention on Cybercrime (2001) and relevant domestic regulations. The analysis focused on three dimensions: evidentiary reliability, admissibility and chain of custody, and algorithmic accountability.

The evidentiary reliability component assessed the objectivity and reproducibility of AI-generated evidence. The admissibility analysis examined whether the model's outputs met judicial standards for digital evidence presentation, including data integrity and documentation of forensic procedures. Lastly, algorithmic accountability emphasized explainability by using the Random Forest's feature importance as a justification mechanism, ensuring that the model's decisions could be transparently communicated in legal contexts.

By combining quantitative and legal methodologies, this study offers a comprehensive approach to cyber-attack analysis. The integration of machine learning into legal reasoning enhances the scientific credibility and procedural

fairness of digital evidence handling, promoting both technological advancement and regulatory accountability in the enforcement of cybercrime laws.

Result and Discussion

The Random Forest classification model was developed to categorize multiple forms of cyber-attacks based on a dataset consisting of 40,000 network traffic records and 25 technical as well as contextual attributes. The dataset included features such as protocol type, packet length, anomaly scores, severity level, and action taken by the system, which together provided a comprehensive representation of network activity. The model achieved an overall accuracy of 33.8%, indicating a moderate predictive performance. While this value may seem relatively low, it reflects the inherent complexity of cybersecurity data, where various types of malicious activities often exhibit overlapping network behaviors. The classification report showed balanced precision, recall, and F1-scores across all three classes DDoS, Malware, and Intrusion ranging between 0.33 and 0.35. This balance demonstrates that the dataset was evenly represented across categories, meaning the model did not exhibit bias toward any particular attack type. However, the moderate accuracy also suggests that several features may share similar patterns or that the model encountered difficulty differentiating subtle variations in network traffic that distinguish one attack type from another. This finding aligns with previous studies in cybersecurity analytics, where classification models tend to perform modestly when faced with heterogeneous and high-dimensional data.

The distribution of attack types is presented in figure 2, which shows that DDoS, Malware, and Intrusion attacks were almost equally represented within the dataset. This balanced representation indicates that no single attack type dominated the dataset, suggesting that the captured network data reflects a diverse and realistic cyber threat environment. Such a distribution enhances the reliability of the model's training process, as it prevents overfitting toward a single category and ensures that each attack type is adequately represented in the learning phase. From a technical perspective, this equilibrium allows for a fair evaluation of model performance across all attack types, while from a legal and policy perspective, it highlights the necessity of comprehensive cybercrime legislation that does not prioritize one form of attack over another. Modern cyber threats are multifaceted, and the near-equal presence of multiple attack types demonstrates the evolving complexity of the digital threat landscape. Therefore, both technical and legal frameworks must evolve concurrently to ensure effective detection, classification, and prosecution of cybercrimes across various domains.

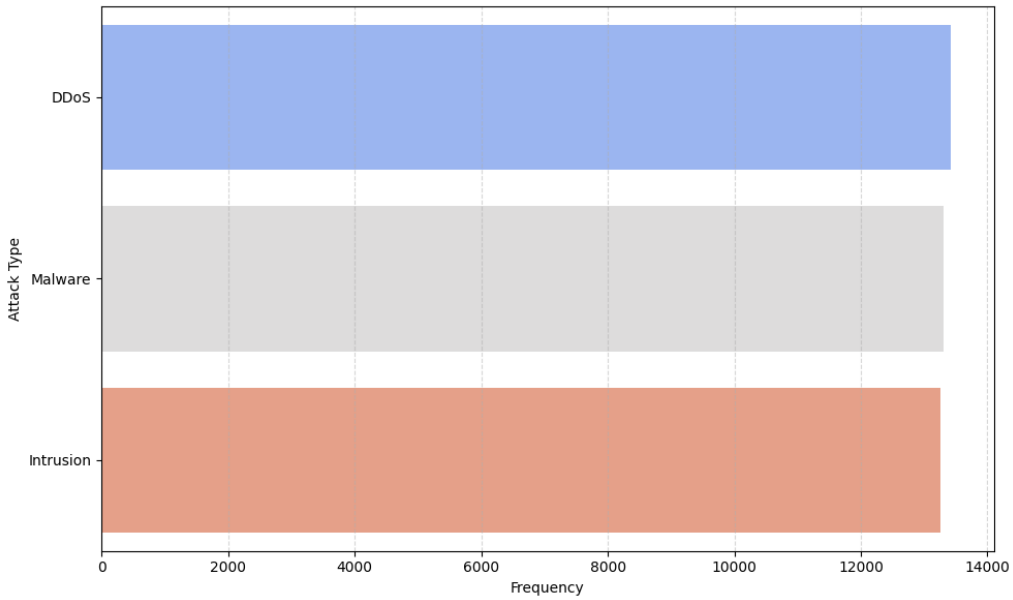

Figure 2 Distribution of Cyber Attack Types

This balanced representation suggests that the dataset captures a diverse yet stable cyber threat environment, mirroring the complex array of attack vectors currently targeting digital infrastructures. The near-equal occurrence of DDoS, Malware, and Intrusion attacks reflects the multifaceted nature of modern cyber threats that affect both private and public sector networks. From an analytical perspective, this balance enhances the validity of the classification model, as it ensures equitable representation across different categories of attacks, preventing model bias toward one type of threat. In practical terms, such diversity also demonstrates how cybercriminals continually diversify their methods, employing tactics that range from network flooding and data corruption to system infiltration. This diversity has direct implications for cybersecurity management and policymaking. From a legal standpoint, it emphasizes the necessity of holistic and adaptive cybercrime legislation that can effectively encompass emerging and overlapping categories of attacks. Overemphasizing one form of cyber offense, such as DDoS, risks leaving gaps in the enforcement framework where newer or more sophisticated attack vectors may go unaddressed. This concern becomes even more critical when existing laws fail to keep pace with rapid technological evolution. Consequently, international frameworks such as the Budapest Convention on Cybercrime should serve as guiding references to ensure that domestic regulations remain both technologically neutral and flexible in defining unlawful digital activities, enabling them to adapt to future cyber threats.

Feature importance analysis was conducted to identify which attributes contributed most significantly to the model's predictive accuracy. As shown in figure 3, Anomaly Scores and Packet Length emerged as the two most influential features in determining attack classifications. These variables reflect the behavioral and structural aspects of network activity, where Anomaly Scores capture deviations from normal traffic patterns and Packet Length represents the volume and density of transmitted data. Both of these indicators are fundamental for detecting cyber-attacks, as they provide quantifiable evidence of network irregularities that often precede or accompany malicious behavior.

The strong contribution of these features suggests that machine learning models are particularly effective when they focus on dynamic behavioral indicators rather than static or categorical variables such as protocol type or severity labels. This insight holds critical value in the legal context. The ability to quantify network anomalies through data-driven metrics can strengthen digital forensic investigations, offering a reliable, reproducible, and scientifically grounded basis for presenting technical evidence in court. By using measurable indicators such as Anomaly Scores, legal practitioners can improve the credibility of digital evidence and enhance its admissibility in judicial proceedings. Furthermore, integrating these metrics into regulatory and investigative frameworks could help establish clearer standards for AI-assisted forensic evidence, bridging the gap between automated detection systems and the evidentiary requirements of cyber law enforcement.

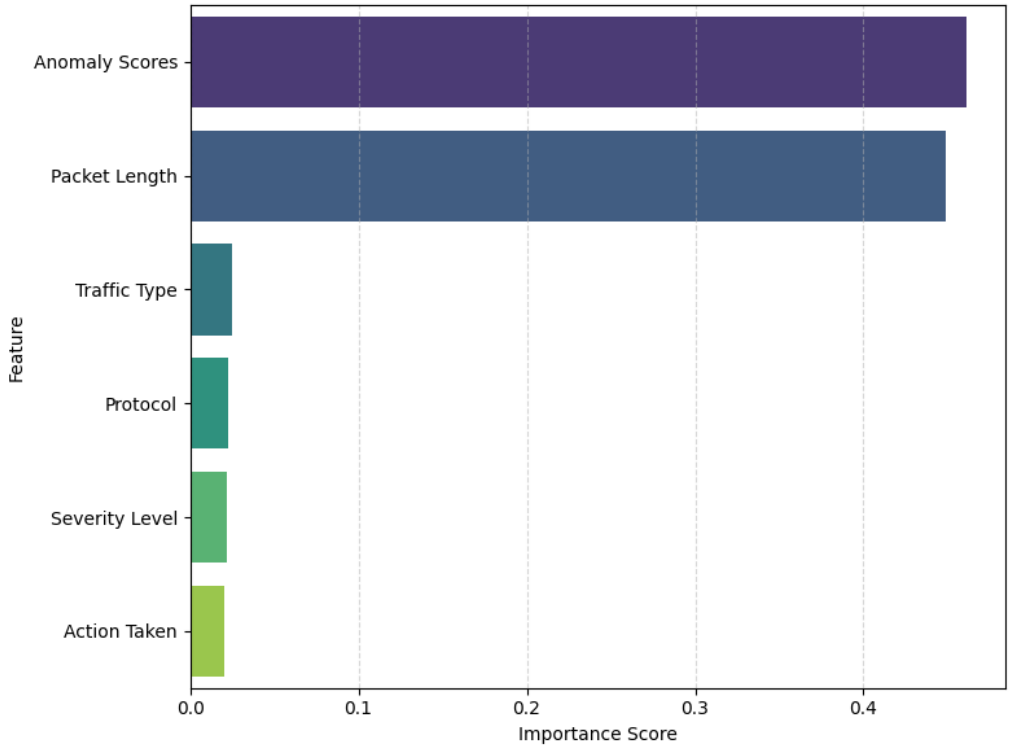

Figure 3 Feature Importance in Cyber Attack Classification

Other variables, including Protocol, Traffic Type, Severity Level, and Action Taken, exhibited comparatively low influence on the classification results. This outcome suggests that these attributes, while valuable for contextual understanding, contribute less to the actual predictive capacity of the model. This finding is consistent with previous research in cybersecurity analytics, which has shown that behavioral and quantitative metrics—such as anomaly scores or packet statistics—tend to be more effective in detecting evolving cyber threats than static categorical variables. Static attributes like protocol type or severity labels often represent predefined classifications that fail to capture the dynamic, adaptive nature of modern cyber-attacks. For instance, an attack using the same protocol (e.g., TCP) can exhibit entirely different characteristics depending on the payload or frequency of packets transmitted. The reliance on dynamic indicators such as anomaly detection therefore reflects a more

adaptive and data-driven approach to cyber threat classification. From a legal standpoint, this technical insight provides a critical bridge between the computational logic of machine learning and the evidentiary standards of cyber law. Quantifiable features, such as anomaly scores, offer objective and reproducible forms of digital evidence that can enhance both the reliability and transparency of forensic investigations. When properly validated, these metrics can serve as admissible proof in legal proceedings by demonstrating measurable deviations from normal system behavior, thereby strengthening the credibility of digital forensic analysis and reinforcing the chain of custody for electronic evidence.

The confusion matrix in figure 4 provides additional insight into the classification model's performance across different attack categories. By comparing predicted and actual labels, the confusion matrix illustrates that misclassifications occurred most frequently between DDoS and Intrusion attacks, which share similar network behavior patterns. This overlap highlights one of the most persistent challenges in both technical cybersecurity and digital law enforcement: the difficulty of attributing attacks to specific causes or perpetrators when their observable characteristics are nearly identical. In practical terms, DDoS and Intrusion attacks both generate high traffic volumes, exhibit burst transmission patterns, and often originate from distributed sources, making them difficult to distinguish solely through packet-level analysis. Such technical ambiguity directly impacts the legal process. Misclassification of an attack type can lead to incorrect forensic interpretations, which in turn may weaken the probative value of digital evidence presented in court. Furthermore, the inability to clearly differentiate between attack categories complicates legal attribution, as prosecutors must demonstrate not only that an offense occurred but also identify its nature and intent. These findings underscore the urgent need for developing standardized forensic protocols and AI explainability frameworks to ensure that automated classification results can be properly interpreted and legally validated. By making AI-based evidence more transparent and traceable, legal institutions can reduce uncertainties in cybercrime litigation and improve the overall accountability of digital investigations.

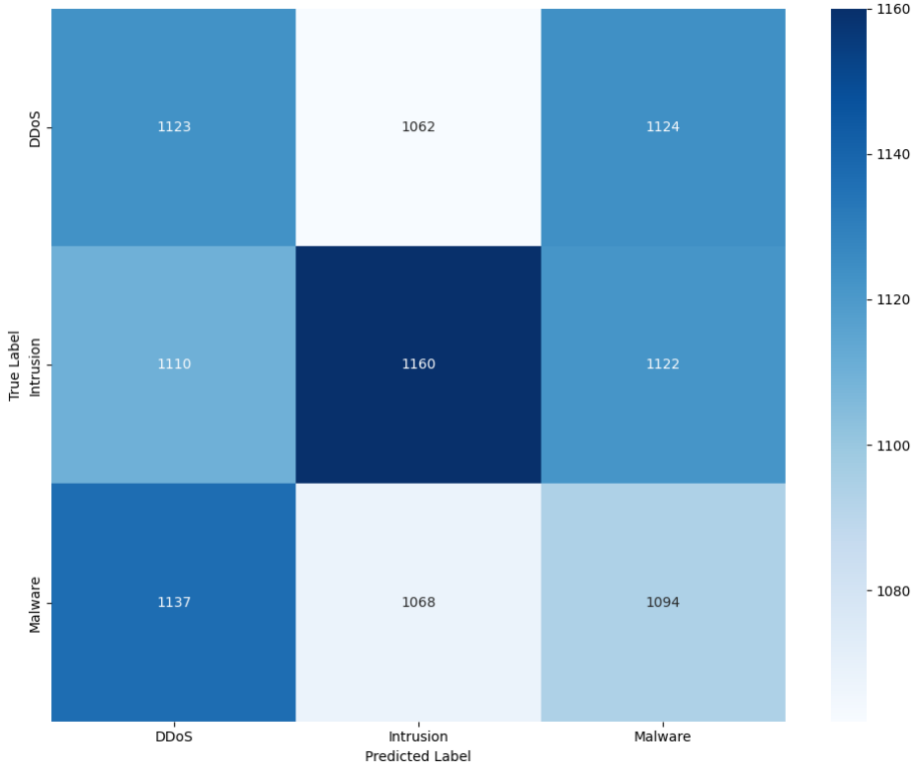

Figure 4 Confusion Matrix of Cyber Attack Classification

The results show that misclassifications most frequently occurred between DDoS and Intrusion attacks. This overlap is likely caused by the similarity in their network behavior, particularly in terms of packet transmission rates and abnormal connection requests. Such similarity presents a significant challenge in both technical detection and legal attribution. When attack patterns are difficult to distinguish technically, assigning responsibility to specific perpetrators becomes problematic. This issue can undermine the reliability of digital evidence and weaken prosecutorial claims in cybercrime cases. Therefore, the establishment of standardized frameworks for AI explainability and forensic auditing is critical for maintaining evidentiary integrity and ensuring the admissibility of machine-generated results in judicial processes.

Overall, the findings demonstrate that while the Random Forest model achieved only moderate predictive accuracy, it successfully identified key features that are relevant to both technical and legal dimensions of cybercrime analysis. The prominence of Anomaly Scores and Packet Length as predictive features suggests that future research should focus on developing explainable AI systems that support not only attack detection and prevention but also forensic validation. Legally, these results highlight the importance of aligning technical innovation with regulatory adaptation. Policymakers and law enforcement agencies should consider integrating AI-driven detection tools into formal cybercrime investigation frameworks while ensuring that their outputs comply with legal standards of reliability, transparency, and accountability. This study therefore contributes to bridging the gap between data-driven cybersecurity models and practical enforcement mechanisms in digital law, promoting a stronger and more adaptive model of cyber law enforcement.

Conclusion

This study developed a machine learning–based classification model to identify different types of cyber-attacks and examined its implications for cybercrime enforcement. Using a dataset of 40,000 network traffic records containing 25 technical and contextual features, the Random Forest model achieved an overall accuracy of 33.8%, with balanced precision and recall across the three major attack types: DDoS, Malware, and Intrusion. Although the model's predictive performance was moderate, the findings highlight the potential of behavioral and quantitative indicators such as Anomaly Scores and Packet Length as key features in detecting malicious network activity and improving the accuracy of cyber threat classification. From a legal perspective, these quantifiable digital metrics can strengthen the reliability, transparency, and admissibility of forensic evidence in cybercrime investigations, supporting more credible decision-making processes in judicial contexts. The balanced distribution of attack types in the dataset reflects the diverse nature of modern cyber threats, underscoring the need for comprehensive and technologically adaptive legal frameworks that align with international standards such as the Budapest Convention on Cybercrime. Overall, the study bridges the gap between technical innovation and legal enforcement, demonstrating that interdisciplinary integration between artificial intelligence, digital forensics, and cyber law is essential for creating robust, transparent, and future-proof mechanisms for cybercrime prevention and prosecution.

Declarations

Author Contributions

Conceptualization: D.N. and A.Y.K.; Methodology: A.Y.K.; Software: D.N.; Validation: D.N. and A.Y.K.; Formal Analysis: D.N. and A.Y.K.; Investigation: D.N.; Resources: A.Y.K.; Data Curation: A.Y.K.; Writing Original Draft Preparation: D.N. and A.Y.K.; Writing Review and Editing: A.Y.K. and D.N.; Visualization: D.N.; 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] D. Moore, C. Shannon, D. J. Brown, G. Voelker, and S. Savage, “Inferring Internet denial-of-service activity,” *ACM Transactions on Computer Systems*, vol. 24, no. 2, pp. 115–139, 2006, doi: 10.1145/1132026.1132027.

[2] Council of Europe, *Convention on Cybercrime (Budapest Convention)*, ETS No. 185, Budapest, Hungary, 2001.

[3] H. Surden, “Machine learning and law: An overview,” *Washington Law Review*, vol. 89, no. 1, pp. 87–124, 2014.

[4] S. Rehman, “Intrusion detection in cyber space using machine learning based algorithm,” *FUJEAS Journal*, vol. 3, no. 2, pp. 45–52, 2022, doi: 10.5281/zenodo.7512478.

[5] D. Dunsin, M. Ghanem, K. Ouazzane, and V. Vassilev, “AI and machine learning in modern digital forensics,” *Forensic Science International: Digital Investigation*, vol. 48, p. 301653, 2024, doi: 10.1016/j.fsidi.2023.301653.

[6] K. Rajendiran, K. Kannan, and Y. Yu, “Applications of machine learning in cyber forensics,” in *Machine Learning and Deep Learning in Real-Time Applications*, Hershey, PA: IGI Global, 2021, pp. 29–46, doi: 10.4018/9781799876427.ch002.

[7] T. Nayerifard, H. Amintoosi, A. Bafghi, and A. Dehghantanha, “Machine learning in digital forensics: A systematic literature review,” *arXiv Preprint arXiv:2306.04965*, vol. 2023, no. June, pp. 1-99, 2023, doi: 10.48550/arXiv.2306.04965.

[8] L. Tageldin and H. Venter, “Machine-Learning forensics: State of the art in the use of machine-learning techniques for digital forensic investigations within smart environments,” *Applied Sciences*, vol. 13, no. 18, p. 10169, 2023, doi: 10.3390/app131810169.

[9] V. Pantelakis, P. Bountakas, A. Farao, and C. Xenakis, “Adversarial machine learning attacks on multiclass classification of IoT network traffic,” in *Proceedings of the 18th International Conference on Availability, Reliability and Security (ARES 2023)*, Benevento, Italy, vol. 2023, no. August, pp. 1–8, 2023, doi: 10.1145/3600160.3605086.

[10] A. Abbas et al., “Machine learning-based hybrid technique to enhance cyber-attack perspective,” *Journal of Cloud Computing*, vol. 14, no. October, art. no. 57, 2025, doi: 10.1186/s13677-025-00782-5.

[11] D. N. Jay, “A deep learning framework for cyberattack detection and classification,” *Computers & Security*, vol. 141, p. 102912, 2025, doi: 10.1016/j.cose.2025.102912.

[12] A. Kumar and L. K. Singh, “A study on machine learning-based models for cyber attack classification and severity estimation,” *International Journal of Computer Applications*, vol. 187, no. 41, pp. 65–71, 2025, doi: 10.5120/ijca2025925245.

[13] J. Note, E. Mullalli, and B. Cico, “Machine learning algorithms for cyber attack detection and classification,” in *Proceedings of CompSysTech 2024*, Ruse, Bulgaria, 2024, pp. 1–10, doi: 10.1145/3674912.3674937.

[14] C. X. Deng, Q. Y. Zhang, and Y. L. Wang, “Machine learning-based cyber threat detection: An approach to malware classification,” *Multimedia Tools and Applications*, vol. 83, pp. 34551–34570, 2024, doi: 10.1007/s42454-024-00055-7.

[15] P. Fernandes and M. Antunes, “Benford’s law applied to digital forensic analysis,” *Forensic Science International: Digital Investigation*, vol. 45, no. June, p. 301515, 2023, doi: 10.1016/j.fsidi.2023.301515.

[16] D. Dunsin, M. Ghanem, K. Ouazzane, and V. Vassilev, “A comprehensive analysis of the role of artificial intelligence and machine learning in modern digital forensics and incident response,” *Forensic Science International: Digital Investigation*, vol. 48, no. March, p. 301675, 2025, doi: 10.1016/j.fsidi.2023.301675.

[17] C. Chitsungo, “Harnessing digital strategies to combat cryptocurrency-enabled crimes,” *American Journal of International Relations*, vol. 12, no. 3, pp. 101–113, 2024, doi: 10.34256/ajir24307.

[18] P. Narasimhan and N. Kala, “Emerging trends in digital forensics: Investigating cybercrime,” *International Journal of Scientific Research in Computer Science, Engineering and Information Technology*, vol. 11, no. 2, pp. 54–62, 2025, doi: 10.32628/IJSRCSEIT.2025.11.2.4.

[19] M. Brkan and G. Bonnet, “Legal and ethical dimensions of algorithmic decision-making: Accountability, transparency, and fairness,” *Computer Law & Security Review*, vol. 45, p. 105676, 2022, doi: 10.1016/j.clsr.2022.105676.

[20] K. Richmond, “AI, machine learning, and international criminal investigations: Lessons from forensic science,” *Social Science Research Network Working Paper*, no. 3632142, 2020, doi: 10.2139/ssrn.3632142.

[21] P. Connor, G. Hutton, D. Johnston, and G. McKinnon, *Cyber Crime*. Oxford, U.K.: Oxford University Press, 2016, doi: 10.1093/oso/9780198754388.001.0001.

[22] H. Alattas, S. H. Alshamrani, M. F. Khan, and M. Alghamdi, “Extract compliance-related evidence using machine learning,” in *Proceedings of IEEE CICN*, 2022, pp. 537–542, doi: 10.1109/CICN55857.2022.10020118.

[23] S. Gupta, I. Jain, and M. Saxena, “Cyber bullying detection and classification using machine learning algorithms,” in *Proceedings of IEEE CYBERCOM*, 2024, pp. 167–171, doi: 10.1109/CYBERCOM56789.2024.102345.

[24] A. Ogundiran, H. Chi, J. Yan, and R. Agada, “Advancing forensic examination of cyber predator communication through machine learning,” in *Proceedings of IEEE QRS-C*, 2024, pp. 464–473, doi: 10.1109/QRS-C.2024.00084.

[25] M. Dos Santos, “Detection of cyber-attacks: Data sets and machine learning methods,” *Information Security Journal: A Global Perspective*, vol. 31, no. 4, pp. 340–355, 2022, doi: 10.1080/19393555.2022.2067859.