Kinetik: Game Technology, Information System, Computer Network, Computing, Electronics, and Control
Journal homepage: http://kinetik.umm.ac.id
ISSN: 2503-2267
Vol. 10, No. 4, November, Pp. 593-602

XGBoost-powered ransomware detection: A gradient-based machine learning approach for robust performance

Wildanil Ghozi[1], Heru Lestiawan[1], Ramadhan Rakhmat Sani[1], Jassim Nadheer Hussein[2], Fauzi Adi Rafrastara[*1]
Faculty of Computer Science, Universitas Dian Nuswantoro, Indonesia[1]
School of Business, Management and Technology, Alfa University College, Malaysia[2]

Article Info
Keywords:
Ransomware, Machine Learning, XGBoost, Ensemble Learning, Classification

Article history:
Received: July 08, 2025
Accepted: September 12, 2025
Published: November 01, 2025

Cite:
W. Ghozi, H. Lestiawan, R. R. Sani, J. N. Hussein, and F. A. Rafrastara, "XGBoost-Powered Ransomware Detection: A Gradient-Based Machine Learning Approach for Robust Performance", *KINETIK*, vol. 10, no. 4, Nov. 2025.
https://doi.org/10.22219/kinetik.v10i4.2405

*Corresponding author.
Fauzi Adi Rafrastara
E-mail address:
fauziadi@dsn.dinus.ac.id

Abstract
Ransomware remains a rapidly evolving cyber threat, causing substantial financial and operational disruptions globally. Traditional signature-based detection systems are ineffective against sophisticated, zero-day attacks due to their static nature. Consequently, machine learning-based approaches offer a more effective and adaptive alternative. This study proposes an approach utilizing XGBoost for highly effective ransomware detection. We conducted a rigorous comparative analysis of prominent ensemble learning algorithms—XGBoost, Random Forest, Gradient Boosting, and AdaBoost—on the RISS Ransomware Dataset, comprising 1,524 instances. Our experimental results unequivocally demonstrate XGBoost as the superior ensemble model, achieving an impressive 97.60% accuracy and F1-Score. This performance surpassed Gradient Boosting (97.20%), Random Forest (96.94%), and AdaBoost (96.50%). Furthermore, this study benchmarked XGBoost against established state-of-the-art (SOTA) methods, including Support Vector Machine (SVM) and the SA-CNN-IS deep learning approach. The comprehensive results underscore the core contribution of this study: by applying XGBoost with a carefully structured machine learning pipeline, our approach consistently outperforms two state-of-the-art methods (SVM and SA-CNN-IS) as well as other ensemble algorithms. This highlights the critical role of methodological precision in maximizing detection performance against evolving ransomware threats.

1. Introduction

Ransomware is a rapidly evolving cybersecurity threat, with increasingly sophisticated attack techniques and strategies. It is a type of malware designed to encrypt user data and demand a ransom for its recovery [1], [2], [3]. These attacks typically target computers, servers, or networks with security vulnerabilities that can be exploited by attackers. Cybercriminals use ransomware as a financial tool to extort victims by leveraging their critical data [4], [5]. In recent years, ransomware has evolved rapidly and has become one of the most significant threats in cybersecurity [6], [7].

The impact of ransomware extends beyond individuals, posing serious risks to businesses, governments, and critical infrastructure [8], [9], [10], [11]. The financial losses from ransomware attacks continue to rise annually, amounting to billions of dollars globally [8], [12], [13]. In addition to financial damage, ransomware attacks can severely tarnish the reputation of affected organizations. The inability to restore data in a timely manner may trigger a domino effect, disrupting essential operations [14], [15].

Traditional signature-based antivirus systems have fundamental weaknesses in detecting ransomware [16], [17]. Ransomware continuously evolves and has the potential to execute Zero-Day Attacks, where vulnerabilities are exploited before security vendors have developed patches or solutions [18], [19]. Due to its polymorphic nature, ransomware can dynamically change its structure, making it difficult to detect using signature-based methods [20]. Consequently, many ransomware variants can bypass conventional security measures and inflict significant damage on unprotected users.

To address these limitations, more adaptive approaches, such as machine learning-based detection, are required [20], [21], [22]. This approach enables ransomware identification based on attack patterns and characteristics rather than relying solely on specific signatures. Machine learning technology offers a significant advantage in recognizing emerging threats and detecting ransomware with higher accuracy compared to traditional methods [23].

One of this study's key contributions is a rigorous comparative analysis of various machine learning algorithms for ransomware detection. Beyond a general comparison with existing state-of-the-art methods, we specifically evaluate and compare XGBoost against other prominent ensemble learning techniques, including Random Forest, Gradient Boosting (GBM), and AdaBoost. We aim to address the existing gap in consistently high-performing and adaptive

detection methods by demonstrating that XGBoost significantly outperforms these established ensemble approaches and other state-of-the-art solutions in terms of detection performance, efficiency, and adaptability against the evolving landscape of ransomware threats. Our comprehensive experiments will unequivocally establish XGBoost as the most effective algorithm for this critical task.

As information technology advances, ransomware attacks have evolved in infiltration methods, data encryption techniques, and extortion strategies targeting victims [24]. Previous studies have examined the economic, political, and security implications of ransomware [25], [26], [27]. These studies highlight that ransomware has become one of the most damaging cyber threats, disrupting business operations, threatening governmental systems, and posing risks to critical infrastructure.

Signature-based detection systems, despite their widespread use, have significant limitations in combating modern malware, especially oligomorphic, polymorphic, metamorphic ransomware, and those exploiting zero-day vulnerabilities [18], [19], [21]. Zero-day attacks occur when attackers exploit security flaws that are unknown to vendors or the cybersecurity community, leaving no effective defense available. These attacks allow ransomware to infiltrate systems undetected, posing a much greater threat compared to variants that can be identified through conventional signature-based methods. Research conducted by cybersecurity experts indicates that signature-based detection methods cannot keep up with the rapid evolution of malware. Therefore, more adaptive solutions are required for effective ransomware detection [20].

In recent years, machine learning has emerged as a promising solution for ransomware detection. Various studies have proposed algorithms such as Random Forest, Support Vector Machine (SVM), and deep learning as alternative methods for analyzing and identifying malware attacks, including ransomware. Literature reviews indicate that these algorithms have advantages in recognizing attack patterns, enabling more accurate ransomware detection compared to traditional approaches [14], [28], [29], [30].

In [30], the authors introduced a novel framework for pre-encryption detection of crypto-ransomware, called RENTAKA. They compared SVM with four other classifiers, Random Forest, Naïve Bayes, k-Nearest Neighbor (kNN), and J48. As a result, Support Vector Machine (SVM) achieved the highest accuracy score of 97.05%. The dataset used was RISS Ransomware Dataset. However, the accuracy still needs improvement as false positives and false negatives on ransomware detection can negatively impact users.

The study by [31] focuses on early detection of zero-day ransomware using Portable Executable (PE) Header features through the SA-CNN-IS method, a deep learning approach. The research utilized the RISS Ransomware Dataset, which included various ransomware samples, including zero-day variants that traditional security systems have not documented. Results indicate that this method achieved 96.31% accuracy, demonstrating the effectiveness of machine learning in identifying ransomware attack patterns. The primary advantage of this approach lies in its ability to analyze PE Header structures, commonly used by ransomware to mask its activities.

Comparing machine learning algorithms is a crucial step in ensuring optimal performance in ransomware detection. Although previous studies have explored various algorithms, their performance remains suboptimal—particularly in achieving zero tolerance—highlighting the need for a more rigorous comparison among advanced ensemble methods. Given the promising results of SVM [30] and the SA-CNN-IS deep learning approach [31] reported in recent literature for ransomware detection, this study aims to enhance existing research by conducting a more rigorous and extensive comparative analysis. We will specifically evaluate and contrast the performance of XGBoost against other prominent ensemble learning techniques, namely Random Forest, Gradient Boosting (GBM), and AdaBoost, alongside the aforementioned state-of-the-art methods. By doing so, this research seeks to provide new, data-driven insights into selecting the most effective and adaptable algorithm for robust ransomware detection, ensuring superior protection against future cyber threats. However, the performance of algorithms used in both state-of-the-art approaches remains suboptimal and requires further enhancement to achieve zero tolerance, given the potentially catastrophic impact of ransomware attacks.

2. Research Method

In this study, selecting the appropriate hardware and software is crucial to ensuring a smooth research process until completion. The hardware used consists of a computer equipped with an Intel Xeon E5620 processor, 16 GB RAM, 2 GB VGA, and 1 TB SSD as the storage medium. Meanwhile, for data processing, modeling, and evaluation, the study utilizes Python programming language along with Google Colab, a cloud-based environment for executing code and performing model analysis.

The research process is divided into several stages, as illustrated in Figure 1, with the first stage being Data Preparation. The dataset employed in this study is the RISS Ransomware Dataset [32], [33], a public dataset containing 1,525 instances and a total of 30,969 features. The target variable in this dataset consists of two classes, with 1 representing ransomware and 0 representing goodware. The class distribution in this dataset is 1:1.62, with 582 instances classified as ransomware and 942 instances categorized as goodware. This distribution is slightly imbalanced but still tolerable, meaning that the dataset does not require significant adjustments for machine learning model training.

Further details on the dataset are presented in Table 1. In preparing this dataset, two features were removed—feature number 1 (ID) and feature number 3 (Ransomware Family). Meanwhile, feature number 2 was designated as the label or target variable. As a result, 30,967 features remain, which will be further processed in the subsequent analysis.

The second stage is Preprocessing, where features with constant values are removed. Constant-value features do not contribute to detection accuracy; instead, they add unnecessary computational overhead. Therefore, these features must be eliminated. After verification, 7,351 constant-value features were identified and subsequently removed. As a result, 23,616 features remain, which will be used for modeling.

In the modelling stage, four ensemble algorithms, Random Forest, AdaBoost, Gradient Boosting, and XGBoost, are applied. Other machine learning algorithms, such as k-Nearest Neighbors (kNN), Decision Tree, Naïve Bayes, and Support Vector Machine (SVM), have already been tested in previous studies [30]. However, these algorithms were still evaluated in this experiment, and the accuracy scores obtained did not differ significantly. Ultimately, the four selected ensemble algorithms used in this modeling phase will be compared with the performance scores of state-of-the-art models [30], [31].

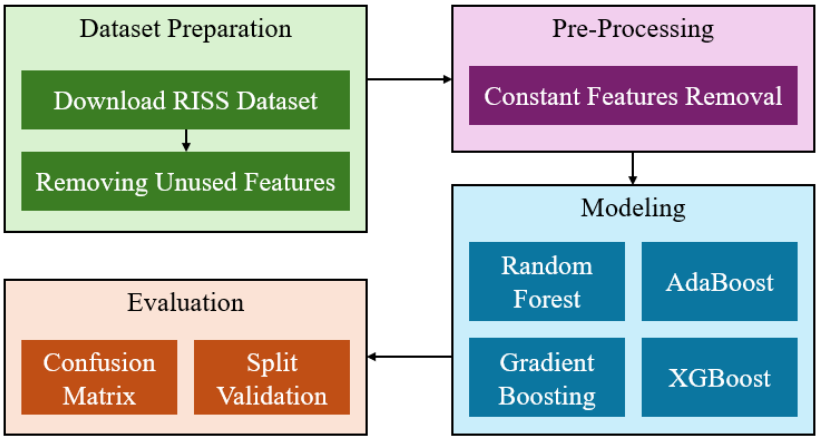

Figure 1. Research Stages: Dataset Preparation, Pre-Processing, Modeling, and Evaluation

Table 1. Details of RISS Ransomware Dataset

Dataset Name	RISS Ransomware Dataset
Number of Records	1524 (Ransomware: 582 and Goodware: 942)
Number of Features	30,969
Number of Classes	2 (Ransomware and Goodware)
Missing Values	None

Ensemble learning algorithms enhance model performance by combining multiple "weak" learners, often decision trees, to create a stronger and more robust predictive model. This section details the mechanisms of Random Forest, AdaBoost, Gradient Boosting, and XGBoost, highlighting their unique approaches to boosting and aggregation, which are crucial for effective ransomware detection.

Random Forest is an ensemble learning method that operates by constructing a multitude of decision trees during training and outputting the mode of the classes (for classification) or mean prediction (for regression) of the individual trees [1]. Unlike boosting algorithms that build trees sequentially to correct prior errors, Random Forest builds trees in parallel. Each tree in the forest is trained on a random subset of the training data (bootstrap sampling) and, crucially, considers only a random subset of features at each split. This dual randomness—in data and features—ensures that the individual trees are diverse and decorrelated. This decorrelation is vital for reducing variance and preventing overfitting, making Random Forest highly robust to noisy data and capable of handling high-dimensional datasets without extensive feature engineering. Its parallelizable nature also contributes to its efficiency [34]. The block diagram of Random Forest can be seen in Figure 2.

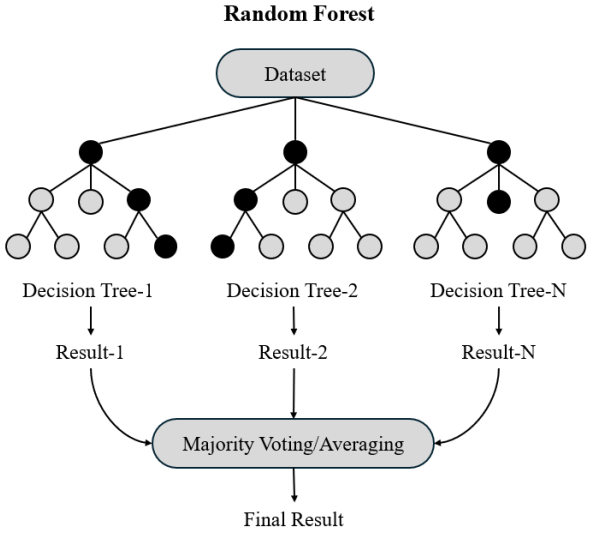

Figure 2. Random Forest Block Diagram (adapted from [35])

AdaBoost (Adaptive Boosting) is an ensemble learning algorithm that improves model performance by combining multiple weak learners, primarily focusing on correctly classifying difficult instances [36]. Instead of optimizing a loss function directly, AdaBoost iteratively assigns higher weights to misclassified instances. Each subsequent weak learner's contribution is then adjusted based on its success in correctly classifying these weighted data points. This emphasis on difficult-to-classify examples makes AdaBoost sensitive to noisy data and outliers. Because AdaBoost typically relies on many simple weak learners (often decision stumps), it tends to be less prone to overfitting compared to more complex boosting methods, although this can vary depending on the data characteristics. AdaBoost assigns weights to misclassified instances and updates them iteratively. The formula for the weight update for the weak classifier is shown in Equation 1, where α_t represents the weight assigned to the weak classifier at iteration t, and e_t denotes the error rate of that weak classifier. The block diagram of AdaBoost can be seen in Figure 3.

$$\alpha_t = \frac{1}{2} ln \frac{1 - e_t}{e_t} \tag{1}$$

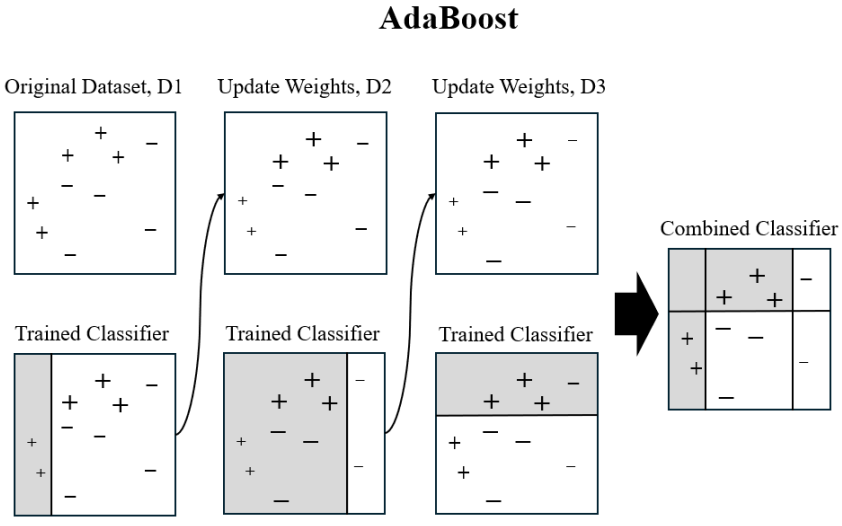

Figure 3. AdaBoost Block Diagram (adapted from [35])

Gradient Boosting optimizes the model by minimizing a loss function, adjusting tree weights iteratively based on residual errors [37][38]. Gradient Boosting builds trees sequentially, where each tree corrects the mistakes of the previous one by focusing on the negative gradient of the loss function (residuals). This allows Gradient Boosting to be more flexible, making it better at capturing complex patterns and handling overfitting through techniques like

regularization [39]. The update rule for the model is shown in Equation 2, where $F_m(x)$ is the updated model, $h_m(x)$ is the weak learner, and γ is the learning rate. The block diagram of Gradient Boosting can be seen in Figure 4.

$$F_m(x) = F_{m-1}(x) + \gamma h_m(x) \quad (2)$$

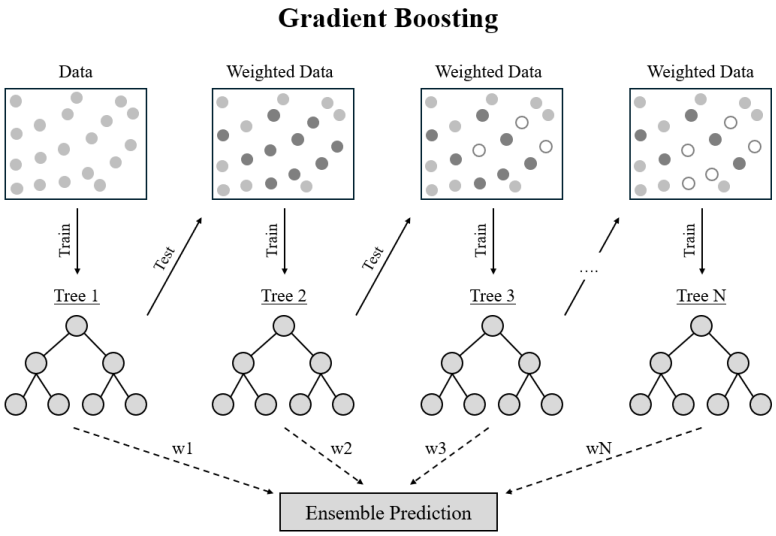

Figure 4. Gradient Boosting Block Diagram (adapted from [35])

XGBoost (Extreme Gradient Boosting) is an optimized and highly efficient implementation of the Gradient Boosting framework, designed for speed and performance [35]. It extends the core Gradient Boosting concept by incorporating several advanced features. Key enhancements include regularization techniques (L1 and L2) to prevent overfitting, which is crucial for handling complex, high-dimensional cybersecurity data. XGBoost also introduces a more sophisticated tree-pruning algorithm that prunes trees backward after they have been fully grown, leading to more generalized models. Furthermore, it supports parallel processing during tree construction, significantly speeding up computation, especially on large datasets. XGBoost handles missing values gracefully and includes built-in cross-validation at each iteration. These optimizations collectively make XGBoost not only faster but also more robust and accurate than traditional Gradient Boosting implementations, establishing it as a leading algorithm for various machine learning tasks, including complex classification problems like ransomware detection [40].

The last stage is Evaluation. A 70:30 split method is used to divide the data into training and testing sets. Classification performance is measured using a confusion matrix, which provides accuracy and F1-Score. Accuracy is chosen because it effectively represents the overall correctness of the model in classifying ransomware and goodware instances. Meanwhile, F1-Score is a crucial metric for ransomware detection because it balances precision and recall, ensuring that the model effectively minimizes both false positives and false negatives. Precision is essential for reducing false positives, where goodware is mistakenly classified as ransomware, preventing unnecessary security alerts. Recall, on the other hand, is vital for minimizing false negatives, ensuring that actual ransomware threats are not overlooked. Since the consequences of false negatives and false positives in cybersecurity can be severe, F1-Score provides a more comprehensive evaluation than accuracy alone, making it particularly suitable for detecting sophisticated ransomware threats, including zero-day attacks [20].

3. Results and Discussion

This section presents the experimental results and provides an in-depth interpretation of the performance of the classification algorithms applied for ransomware detection. We compare the performance of Gradient Boosting and AdaBoost, and evaluate their relative performance against other algorithms from previous studies.

For data processing, modeling, and evaluation, we utilized Python programming language along with Google Colab, a cloud-based environment for executing code and performing model analysis.

The dataset employed in this study is the RISS Ransomware Dataset, comprising 1,525 instances with a slightly imbalanced, yet tolerable, class distribution (582 ransomware and 942 goodware). The data preparation process involved removing two irrelevant features (ID and Ransomware Family) and designating the third feature as the target variable, resulting in an initial 30,967 features.

Our comprehensive comparative analysis of various ensemble machine learning algorithms for ransomware detection revealed a clear hierarchy in performance. As presented in Figure 5, XGBoost consistently demonstrated the most superior results across key metrics when compared against Random Forest, AdaBoost, and Gradient Boosting.

Specifically, XGBoost achieved an impressive accuracy and F1-Score of 97.60%. This strong performance highlights its robustness and effectiveness in identifying sophisticated ransomware threats within the dataset.

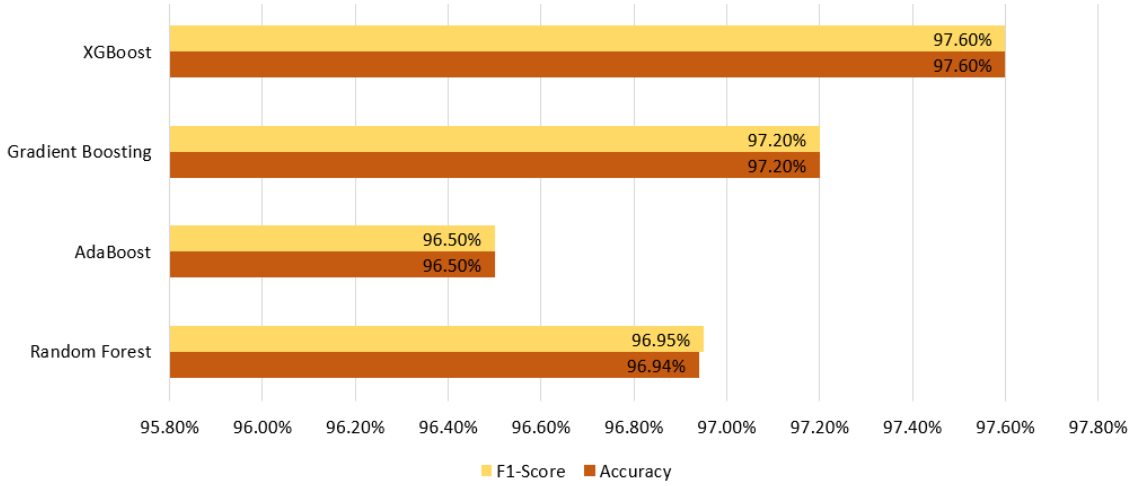

Figure 5. Performance Comparison of Four Ensemble Algorithms on Ransomware Dataset

The other ensemble methods, while still performing commendably, did not quite match XGBoost's peak. Gradient Boosting secured the second-highest performance, with both accuracy and F1-Score at 97.20%. Following closely was Random Forest, which delivered solid results at 96.94% accuracy and 96.95% F1-Score. Lastly, AdaBoost showed a respectable performance with 96.50% for both metrics. These results underscore the general efficacy of ensemble learning in cybersecurity applications, particularly for complex classification tasks like ransomware detection, where combining multiple weak learners often yields more stable and accurate predictions than single models.

The observed superiority of XGBoost can be attributed to its advanced design and inherent optimizations within the gradient boosting framework. Unlike standard Gradient Boosting, XGBoost incorporates sophisticated regularization techniques (L1 and L2) which are crucial for preventing overfitting, a common challenge in high-dimensional cybersecurity datasets. Furthermore, its ability to gracefully handle missing values and leverage parallel processing support significantly enhances its efficiency and scalability. These technical refinements allow XGBoost to more effectively capture intricate attack patterns and adapt to the evolving complexity of ransomware data, making it an exceptionally powerful tool for proactive threat detection.

With XGBoost firmly established as the leading ensemble algorithm in this initial internal evaluation, the research naturally progresses to a more stringent benchmark: comparing its performance against existing state-of-the-art (SOTA) methods in ransomware detection. This next phase is crucial for validating XGBoost's potential beyond its ensemble peers and assessing its standing within the broader landscape of cutting-edge cybersecurity solutions. The aim is not merely to identify the best among the tested ensemble methods but to determine whether XGBoost can truly represent the forefront of ransomware detection technology.

For this critical comparison, we will specifically pit XGBoost against two highly-regarded SOTA approaches: Support Vector Machine (SVM) and a deep learning approach, SA-CNN-IS. SVM, as reported in [30], achieved a commendable accuracy of 97.1% on the RISS Ransomware Dataset, showcasing its strong discriminative power. Similarly, the SA-CNN-IS deep learning method, detailed in [31], demonstrated robust performance with 96.3% accuracy, highlighting the potential of neural networks in analyzing complex Portable Executable (PE) header features for early detection. The results of this comprehensive SOTA comparison are presented in Table 2.

Our detailed analysis, further supported by the Confusion Matrix for XGBoost (Figure 6), revealed compelling insights into its detection capabilities. The matrix shows True Positives (TP) of 172 and True Negatives (TN) of 275, indicating a high rate of correctly identified ransomware and goodware samples, respectively. Critically, XGBoost achieved remarkably low misclassification rates with only 6 False Positives (FP) and 5 False Negatives (FN). These low FP and FN counts are vital; minimizing false positives prevents legitimate files from being mistakenly quarantined, while a low false negative rate ensures that malicious ransomware samples are not overlooked.

Table 2. Comparison of XGBoost with state-of-the-arts

No.	Model	Accuracy	F1-Score
1	XGBoost (Proposed)	97.60%	97.60%
2	SVM [30]	97.1%	UNK
3	SA-CNN-IS [31]	96.3%	UNK

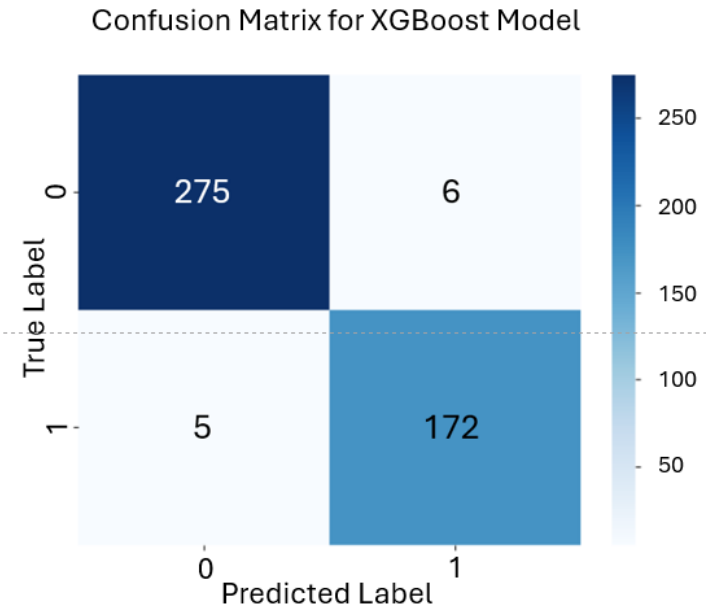

Figure 6. Confusion Matrix of XGBoost Model for Ransomware Detection

This impressive performance, particularly the minimal false classifications, demonstrates XGBoost's exceptional precision and recall in a real-world context. While the overall metrics in Table 3 indicate its competitive or superior standing against SVM and SA-CNN-IS, the breakdown in the confusion matrix specifically highlights its ability to make accurate distinctions across both classes. This superiority can be attributed to XGBoost's inherent capability to handle feature interactions and noise through regularization, tree pruning, and optimal split selection, which collectively enhance generalization and reduce overfitting—limitations often encountered in both SVM and deep learning models when applied to relatively small or imbalanced datasets. This makes XGBoost not only highly accurate, but also a reliable and trustworthy model for deployment in critical cybersecurity infrastructure, where the costs of misclassification are high.

These findings carry significant practical implications for real-world cybersecurity systems, particularly in enterprise and governmental environments, where early and accurate ransomware detection is critical. The low false positive and false negative rates achieved by XGBoost reduce operational disruptions caused by misclassification, enabling more reliable quarantine decisions and faster incident response. Moreover, its computational efficiency and scalability make it suitable for integration into existing endpoint protection platforms and network-based intrusion detection systems, offering a robust layer of defense against both known and emerging ransomware threats.

Ultimately, this final comparative step was vital to conclusively validate XGBoost's position as the most effective and adaptive ransomware detection solution. The results confirm its strong potential to surpass or at least match these current SOTA techniques, reinforcing the importance of advanced gradient-based machine learning in the dynamic field of cybersecurity.

4. Conclusion

This study unequivocally demonstrates the significant effectiveness of machine learning, particularly advanced ensemble methods, in addressing the pervasive threat of ransomware. Our rigorous comparative analysis of four prominent ensemble algorithms revealed that XGBoost emerged as the most optimal algorithm for ransomware detection, achieving the highest accuracy and F1-Score of 97.60%. This superior performance, consistently outperforming Gradient Boosting (97.20%), Random Forest (96.94%), and AdaBoost (96.50%), indicates its exceptional ability to minimize both false positives and false negatives, which is crucial in real-world cybersecurity scenarios.

Furthermore, extending our evaluation, this research solidifies XGBoost's position as a leading solution by demonstrating its superior performance even against established state-of-the-art (SOTA) methods such as Support Vector Machine (SVM) [30] and the SA-CNN-IS deep learning approach [31]. While SVM and SA-CNN-IS showed strong capabilities, XGBoost's enhanced optimizations for gradient boosting—including advanced regularization, efficient handling of missing values, and parallel processing—enable it to more effectively learn and adapt to the complex, evolving patterns of ransomware.

These findings strongly emphasize the inherent limitations of traditional signature-based antivirus systems, which are perpetually challenged by polymorphic ransomware and zero-day attacks. Our results reinforce the critical need for more adaptive and intelligent detection techniques. By leveraging sophisticated machine learning models like XGBoost, cybersecurity defenses can be significantly bolstered, offering a robust shield against rapidly evolving and increasingly stealthy ransomware variants.

For future research, efforts should focus on continually refining feature selection techniques and optimizing hyperparameters to further enhance detection capabilities and model generalization. Expanding datasets to include a wider variety of recent ransomware strains and attack vectors would also be beneficial. Additionally, investigating the robustness of these models against adversarial attacks and exploring hybrid approaches that combine the strengths of ensemble methods with deep learning could contribute to even more resilient and future-proof security solutions. Ultimately, this study reinforces the indispensable role of data-driven, adaptive machine learning approaches in strengthening our collective cybersecurity posture against emerging threats.

Acknowledgement

We extend our sincere gratitude to the Institute for Research and Community Services (LPPM) at Universitas Dian Nuswantoro for the internal grant that supported this research. We hope this study makes a significant contribution to the advancement of technology in Indonesia.

References

[1] F. A. Rafrastara, C. Supriyanto, C. Paramita, Y. P. Astuti, and F. Ahmed, “Performance Improvement of Random Forest Algorithm for Malware Detection on Imbalanced Dataset using Random Under-Sampling Method,” *JPIT*, vol. 8, no. 2, pp. 113–118, 2023. https://doi.org/10.30591/jpit.v8i2.5207

[2] F. A. Rafrastara, C. Supriyanto, C. Paramita, and Y. P. Astuti, “Deteksi Malware menggunakan Metode Stacking berbasis Ensemble,” *JPIT*, vol. 8, no. 1, pp. 11–16, 2023. https://doi.org/10.30591/jpit.v8i1.4606

[3] S. Singh, T. Khanna, and D. K. Verma, “Enhanced Ransomware Classification with a Hybrid RF-SVM Framework Using PCA and RFE,” 2025. https://doi.org/10.1109/ICPCT64145.2025.10940253

[4] U. Urooj, B. A. S. Al-rimy, A. Zainal, F. A. Ghaleb, and M. A. Rassam, “Ransomware Detection Using the Dynamic Analysis and Machine Learning: A Survey and Research Directions,” *Applied Sciences*, vol. 12, no. 1, p. 172, Dec. 2021. http://doi.org/10.3390/app12010172

[5] Md. A. Hossain, T. Hasan, F. Ahmed, S. H. Cheragee, M. H. Kanchan, and M. A. Haque, “Towards superior android ransomware detection: An ensemble machine learning perspective,” *Cyber Security and Applications*, vol. 3, p. 100076, Dec. 2025. http://doi.org/10.1016/j.csa.2024.100076

[6] Er. Kritika, “A comprehensive literature review on ransomware detection using deep learning,” *Cyber Security and Applications*, vol. 3, p. 100078, Dec. 2025. http://doi.org/10.1016/j.csa.2024.100078

[7] V. Anand, S. K. G, S. K. K, and S. C, “Enhancing Ransomware Detection - A Comparative Review of XGBoost, Random Forest, and Neural Network Approaches,” in *2025 International Conference on Emerging Systems and Intelligent Computing (ESIC)*, Bhubaneswar, India: IEEE, Feb. 2025, pp. 710–715. http://doi.org/10.1109/ESIC64052.2025.10962609

[8] A. Kapoor, A. Gupta, R. Gupta, S. Tanwar, G. Sharma, and I. E. Davidson, “Ransomware Detection, Avoidance, and Mitigation Scheme: A Review and Future Directions,” *Sustainability*, vol. 14, no. 1, p. 8, Dec. 2021. http://doi.org/10.3390/su14010008

[9] U. Urooj, B. A. S. Al-Rimy, A. B. Zainal, F. Saeed, A. Abdelmaboud, and W. Nagmeldin, “Addressing Behavioral Drift in Ransomware Early Detection Through Weighted Generative Adversarial Networks,” *IEEE Access*, vol. 12, pp. 3910–3925, 2024. http://doi.org/10.1109/ACCESS.2023.3348451

[10] G. Murray, M. Falkeling, and S. Gao, “Trends and challenges in research into the human aspects of ransomware: a systematic mapping study,” *ICS*, Jul. 2024. http://doi.org/10.1108/ICS-12-2022-0195

[11] G. Munoz Cornejo, J. Lee, and B. A. Russell, “A thematic analysis of ransomware incidents among United States hospitals, 2016–2022,” *Health Technol.*, vol. 14, no. 6, pp. 1059–1070, Nov. 2024. https://doi.org/10.1007/s12553-024-00890-3

[12] M. Robles-Carrillo and P. García-Teodoro, “Ransomware: An Interdisciplinary Technical and Legal Approach,” *Security and Communication Networks*, vol. 2022, pp. 1–17, Aug. 2022. http://doi.org/10.1155/2022/2806605

[13] G. Kirubavathi, W. Regis Anne, and U. K. Sridevi, “A recent review of ransomware attacks on healthcare industries,” *Int J Syst Assur Eng Manag*, vol. 15, no. 11, pp. 5078–5096, Nov. 2024. http://doi.org/10.1007/s13198-024-02496-4

[14] H. Oz, A. Aris, A. Levi, and A. S. Uluagac, “A Survey on Ransomware: Evolution, Taxonomy, and Defense Solutions,” *ACM Comput. Surv.*, vol. 54, no. 11s, pp. 1–37, Jan. 2022. http://doi.org/10.1145/3514229

[15] S. Jawad and H. M. Ahmed, “Machine Learning Approaches to Ransomware Detection: A Comprehensive Review,” *IJSSE*, vol. 14, no. 6, pp. 1963–1973, Dec. 2024. https://doi.org/10.18280/ijsse.140630

[16] M. S. Abbasi, H. Al-Sahaf, and I. Welch, “Particle Swarm Optimization: A Wrapper-Based Feature Selection Method for Ransomware Detection and Classification,” in *Applications of Evolutionary Computation*, vol. 12104, P. A. Castillo, J. L. Jiménez Laredo, and F. Fernández De Vega, Eds., in Lecture Notes in Computer Science, vol. 12104. , Cham: Springer International Publishing, 2020, pp. 181–196. http://doi.org/10.1007/978-3-030-43722-0_12

[17] B. A. S. Al-rimy, M. A. Maarof, and S. Z. M. Shaid, “Ransomware threat success factors, taxonomy, and countermeasures: A survey and research directions,” *Computers & Security*, vol. 74, pp. 144–166, May 2018. http://doi.org/10.1016/j.cose.2018.01.001

[18] M. S. Balamurugan, V. Rajendran, and S. C. Mary, "A Review on Cognitive Based Ransomware Detection using Machine Learning and Deep Learning Techniques," *JATIT*, vol. 102, no. 10, pp. 4572–4581, May 2024.
[19] A. Alqahtani and F. T. Sheldon, "A Survey of Crypto Ransomware Attack Detection Methodologies: An Evolving Outlook," *Sensors*, vol. 22, no. 5, p. 1837, Feb. 2022. http://doi.org/10.3390/s22051837
[20] F. A. Rafrastara *et al.*, "Integrating Information Gain and Chi-Square for Enhanced Malware Detection Performance," *JICT*, vol. 24, no. 1, pp. 80–104, Jan. 2025. http://doi.org/10.32890/jict2025.24.1.4
[21] R. R. Sani, F. A. Rafrastara, and W. Ghozi, "Integrating Ensemble Learning and Information Gain for Malware Detection based on Static and Dynamic Features," *KINETIK*, Jan. 2025. http://doi.org/10.22219/kinetik.v10i1.2051
[22] A. Hussain, A. Saadia, M. Alhussein, A. Gul, and K. Aurangzeb, "Enhancing ransomware defense: deep learning-based detection and family-wise classification of evolving threats," *PeerJ Computer Science*, vol. 10, p. e2546, Nov. 2024. http://doi.org/10.7717/peerj-cs.2546
[23] Y. A. Ahmed, B. Koçer, S. Huda, B. A. Saleh Al-rimy, and M. M. Hassan, "A system call refinement-based enhanced Minimum Redundancy Maximum Relevance method for ransomware early detection," *Journal of Network and Computer Applications*, vol. 167, p. 102753, Oct. 2020. http://doi.org/10.1016/j.jnca.2020.102753
[24] M. Al-Hawawreh, M. Alazab, M. A. Ferrag, and M. S. Hossain, "Securing the Industrial Internet of Things against ransomware attacks: A comprehensive analysis of the emerging threat landscape and detection mechanisms," *Journal of Network and Computer Applications*, vol. 223, p. 103809, Mar. 2024. http://doi.org/10.1016/j.jnca.2023.103809
[25] J. Hernandez-Castro, A. Cartwright, and E. Cartwright, "An economic analysis of ransomware and its welfare consequences," *R. Soc. open sci.*, vol. 7, no. 3, p. 190023, Mar. 2020. http://doi.org/10.1098/rsos.190023
[26] M. Hansel and J. Silomon, "Ransomware as a threat to peace and security: understanding and avoiding political worst-case scenarios," *Journal of Cyber Policy*, vol. 9, no. 2, pp. 159–178, May 2024. http://doi.org/10.1080/23738871.2024.2357092
[27] C. F. Azubuike, O. I. Akinwumi, and E. O. Ezeamu, "Assessing the Global Economic Impact of Ransomeware Attacks and Strategic Global Response," *Nnamdi Azikiwe Journal of Political Science (NAJOPS)*, vol. 9, no. 4, pp. 1–17, 2024.
[28] M. S. Abbasi, "Automating Behavior-based Ransomware Analysis, Detection, and Classification Using Machine Learning," Open Access Te Herenga Waka-Victoria University of Wellington, 2023. http://doi.org/10.26686/wgtn.22180858
[29] A. Alraizza and A. Algarni, "Ransomware Detection Using Machine Learning: A Survey," *BDCC*, vol. 7, no. 3, p. 143, Aug. 2023. http://doi.org/10.3390/bdcc7030143
[30] W. Z. A. Zakaria, M. F. Abdollah, O. Mohd, S. M. W. M. S. M. M. Yassin, and A. Ariffin, "RENTAKA: A Novel Machine Learning Framework for Crypto-Ransomware Pre-encryption Detection," *IJACSA*, vol. 13, no. 5, 2022. http://doi.org/10.14569/IJACSA.2022.0130545
[31] M. Cen, X. Deng, F. Jiang, and R. Doss, "Zero-Ran Sniff: A zero-day ransomware early detection method based on zero-shot learning," *Computers & Security*, vol. 142, p. 103849, Jul. 2024. http://doi.org/10.1016/j.cose.2024.103849
[32] Imperial College London, "RISS: Resilient Information Systems Security," Ransomware Dataset.
[33] D. Sgandurra, L. Muñoz-González, R. Mohsen, and E. C. Lupu, "Automated Dynamic Analysis of Ransomware: Benefits, Limitations and use for Detection," Sep. 10, 2016, *arXiv*: arXiv:1609.03020. Accessed: May 14, 2023. https://doi.org/10.48550/arXiv.1609.03020
[34] R. R. Sani, F. A. Rafrastara, and W. Ghozi, "Integrating Ensemble Learning and Information Gain for Malware Detection based on Static and Dynamic Features," *KINETIK*, Jan. 2025. https://doi.org/10.22219/kinetik.v10i1.2051
[35] M. Ibadullah, S. A. Amalina, W. Ghozi, and F. A. Rafrastara, "Machine Learning-based Malware Detection on Android Operating System using AdaBoost Algorithm and ReliefF Feature Selection Method," in *2024 International Seminar on Application for Technology of Information and Communication (iSemantic)*, Semarang, Indonesia: IEEE, Sep. 2024, pp. 359–364. http://doi.org/10.1109/iSemantic63362.2024.10762096
[36] W. Shan, D. Li, S. Liu, M. Song, S. Xiao, and H. Zhang, "A random feature mapping method based on the AdaBoost algorithm and results fusion for enhancing classification performance," *Expert Systems with Applications*, vol. 256, p. 124902, Dec. 2024. http://doi.org/10.1016/j.eswa.2024.124902
[37] A. Sharma, H. Babbar, and A. K. Vats, "Enhanced Ransomware Detection Using Gradient Boosting Algorithms: A Cybersecurity Dataset Approach," in *2024 5th IEEE Global Conference for Advancement in Technology (GCAT)*, Bangalore, India: IEEE, Oct. 2024, pp. 1–5. http://doi.org/10.1109/GCAT62922.2024.10923841
[38] A. Sharma and H. Babbar, "Implementing Gradient Boosting Techniques for Real-Time Attack Detection in Vehicular Networks," in *Proc. - Int. Conf. Technol. Adv. Comput. Sci., ICTACS*, Chaudhary N., Ed., Institute of Electrical and Electronics Engineers Inc., 2024, pp. 213–218. http://doi.org/10.1109/ICTACS62700.2024.10840804
[39] J. Wu and C. Li, "Illustrating the nonlinear effects of urban form factors on transportation carbon emissions based on gradient boosting decision trees," *Science of The Total Environment*, vol. 929, p. 172547, Jun. 2024. http://doi.org/10.1016/j.scitotenv.2024.172547
[40] A. Ramadhani, F. A. Rafrastara, S. Rosyada, W. Ghozi, and W. M. Osman, "Improving Malware Detection using Information Gain and Ensemble Machine Learning," *J. Tek. Inform. (JUTIF)*, vol. 5, no. 6, pp. 1673–1686, Dec. 2024. http://doi.org/10.52436/1.jutif.2024.5.6.3903