FUDMA Journal of Sciences (FJS)
ISSN online: 2616-1370
ISSN print: 2645 - 2944
Vol. 10 No. 5, March, 2026, pp 275 – 282
DOI: https://doi.org/10.33003/fjs-2026-1005-5000

ENHANCING RANSOMWARE CLASSIFICATION USING LIGHT WEIGHT MACHINE LEARNING ALGORITHM AND ENSEMBLE METHODS

***Victor U. Akuboh, Olalekan J. Awujoola , Lasisi A. Monsuru , Saifulahi S. Shitu, Amina Adebola, Catherine A. Abimuku, Badamasi A. Musa and Ejima U Innocent**

Department of Cyber Security, Nigeria Defence Academy, Kaduna State, Nigeria.

*Corresponding authors' email: vuakuboh@nda.edu.ng

ABSTRACT

The increasing sophistication and frequency of ransomware attacks pose significant challenges to cybersecurity, particularly in resource-constrained environments where computational efficiency is essential. Traditional detection models often struggle to maintain high accuracy while minimizing false positives and computational overhead, limiting their practical deployment in real-time systems. Addressing this challenge, this research aims to develop a lightweight, efficient, and adaptive machine learning framework that enhances ransomware classification by integrating ensemble learning with advanced feature selection techniques. The study's objectives include designing a soft voting ensemble model that combines multiple high-performing classifiers to improve accuracy and reduce misclassification rates, applying Recursive Feature Elimination (RFE) to systematically identify and retain the most relevant features for improved efficiency and reduced overfitting, and employing Principal Component Analysis (PCA) to compress the feature space into a set of uncorrelated components while preserving critical detection information. The proposed framework was trained and tested on the "Ransomware-All" subset of the CIC-AndMal2017 dataset containing 348,943 samples with 84 features, and its performance was evaluated using accuracy, precision, recall, and F1 score as metrics. Results show that the soft voting ensemble achieved an accuracy of 99.10%, with both precision and recall at 0.989, outperforming all individual classifiers while significantly lowering false positive and false negative rates. These outcomes demonstrate that the integration of lightweight algorithms, ensemble learning, and feature selection offers a robust and scalable solution for ransomware detection, making it highly suitable for deployment in environments such as IoT devices, mobile platforms, and edge computing systems.

Keywords: Ransomware Classification**,** Lightweight Algorithms, Ensemble Learning, Soft Voting Ensemble

INTRODUCTION

Ransomware attacks have increasingly become one of the most prevalent and harmful cybersecurity threats, affecting critical infrastructure, government agencies, healthcare organizations, and private enterprises alike (Pemmasani & Rock, 2023). The continual evolution of these attacks, characterized by heightened sophistication and the widespread adoption of ransomware-as-a-service (RaaS) models, has made ransomware more accessible to malicious actors while simultaneously enhancing its ability to evade traditional detection mechanisms. Conventional classification approaches, which predominantly rely on signature-based detection and heuristic rules, are proving insufficient against polymorphic and zero-day ransomware variants that are deliberately engineered to circumvent static security defenses. This widening disparity between attack capabilities and defensive techniques underscores the pressing need for intelligent, rapid, and lightweight classification frameworks that can operate in real time, particularly on resource-constrained platforms such as IoT devices and mobile endpoints (Ispahany et al., 2024).

In light of these challenges, machine learning (ML) has emerged as a promising solution by enabling ransomware detection based on behavioral patterns and system-level activities rather than dependence on predefined signatures alone. A significant contribution in this area is the study by Gupta et al. (2023), which introduced a fast and efficient ransomware classification framework using a soft voting ensemble learning approach. Their model combined five lightweight base classifiers Random Forest, AdaBoost, Extra Trees, XGBoost, and Decision Tree and achieved an accuracy of 98.42% on the CIC-AndMal2017 dataset, thereby highlighting the effectiveness of ensemble methods in enhancing robustness and generalization. Despite its strong performance, the proposed benchmark framework exhibits an important methodological limitation, namely the absence of a comprehensive feature optimization strategy, which is essential for reducing computational complexity, improving classification accuracy, and minimizing overfitting caused by irrelevant or redundant features.

Consequently, this study aims to improve upon the benchmark framework by addressing this limitation through the incorporation of advanced feature engineering techniques. The primary objective is to develop a lightweight, efficient, and adaptive machine learning framework that enhances ransomware classification by integrating ensemble learning with effective feature selection and dimensionality reduction methods. Specifically, the study seeks to design a soft voting ensemble model to improve detection accuracy and reduce false positives, apply Recursive Feature Elimination (RFE) to systematically identify the most relevant features and reduce model complexity, and employ Principal Component Analysis (PCA) to compress the feature space into a reduced set of uncorrelated components suitable for efficient deployment in resource-constrained environments. By extending the work of Gupta et al. (2023) with these targeted enhancements, this research aspires to deliver a more robust, scalable, and practical solution for addressing the continually evolving ransomware threat landscape.

Review of Related Works

The relentless evolution of ransomware represents one of the most severe challenges in modern cybersecurity. Early research by Dargahi et al. (2019) systematically cataloged the threat, highlighting how ransomware operates through a "cyber-kill-chain" and establishing a taxonomy that

underscores its methodical nature. The sophistication of these attacks has only intensified, with McIntosh et al. (2021) documenting the adoption of advanced persistence mechanisms, lateral movement, and double-extortion schemes. This evolution has fundamentally undermined traditional defense mechanisms; signature-based systems are useless against zero-day and polymorphic variants, while heuristic-based approaches often generate excessive false positives and can be bypassed by malware that mimics legitimate behavior or staggers its execution. This landscape created an imperative for a paradigm shift from static, pattern-matching techniques toward dynamic, intelligent systems capable of learning and adapting.

Machine Learning (ML) emerged as the cornerstone of this new paradigm. As Fernando et al. (2020) explored, ML offers the ability to detect ransomware based on behavioral anomalies and system-level features rather than relying on known signatures. This body of work distinguishes between two primary learning approaches: supervised learning, which trains on labeled datasets (e.g., CIC-AndMal2017) using algorithms like Decision Trees, Random Forest, and Support Vector Machines (SVM) for high-accuracy classification of known threats; and unsupervised learning, which employs techniques like clustering and anomaly detection to identify novel, zero-day ransomware by flagging deviations from normal system behavior (Usmani et al., 2022). Subsequent comparative studies, such as the one by Bold et al. (2022), critically evaluated a range of these classifiers, identifying that while individual models like Random Forest and XGBoost perform well, they each possess inherent limitations, such as overfitting in Decision Trees or computational intensity in SVM that constrain their standalone effectiveness in diverse, real-world scenarios.

Recognizing the limitations of single-model approaches, the field increasingly turned to ensemble learning as a method to achieve greater robustness and accuracy. The foundational principle, as surveyed by Mienye and Sun (2022), is that combining multiple base models (learners) can average out biases and variances, leading to superior generalization. Key ensemble strategies include bagging (e.g., Random Forest), which reduces variance by training multiple models on different data subsets; boosting (e.g., AdaBoost, XGBoost), which sequentially focuses on correcting previous errors, ideal for complex and imbalanced datasets; and stacking or voting, which combines diverse models through a meta-learner or averaging. Empirical evidence consistently supports this approach. For instance, Yin et al. (2021) demonstrated the strength of stacking ensembles in handling imbalanced data, while Akhtar et al. (2021) showed that homogenous ensembles with soft voting could achieve highly reliable predictions. This was powerfully validated in the ransomware domain by Gupta et al. (2023), whose soft voting ensemble of five lightweight classifiers (Random Forest, AdaBoost, Extra Trees, XGBoost, and Decision Tree) achieved a benchmark 98.42% accuracy, solidifying ensemble methods as a state-of-the-art approach. Further expanding on this, Aurangzeb et al. (2022) proposed the BigRC-EML framework, which integrated ensemble learning with both static and dynamic analysis features, achieving 98% accuracy and underscoring the value of ensemble strategies for scalable ransomware classification in large-data environments.

Parallel to advancements in ensemble methods, researchers identified feature engineering as a critical determinant of model efficiency and performance. The high-dimensional feature spaces common in behavioral analysis often encompassing entropy measures, API call sequences, file I/O patterns, and network traffic introduce significant challenges, including the "curse of dimensionality," overfitting, and high computational overhead (Nanga et al., 2021; Zhang & Wang, 2023). As Malik et al. (2023) compared, addressing this requires a two-pronged approach: feature selection and dimensionality reduction. Feature selection techniques, such as Recursive Feature Elimination (RFE), work by iteratively pruning the least important features to retain an optimal, non-redundant subset. Dimensionality reduction techniques like Principal Component Analysis (PCA) transform the original features into a smaller set of uncorrelated components that preserve the majority of the data's variance. The importance of selecting the right features is paramount; Yamany et al. (2022) demonstrated that attributes like file entropy, specific API calls, and registry modification patterns are highly discriminative for ransomware behavior. The drive for efficiency, especially for IoT and mobile deployments, has made this feature optimization pipeline indispensable for creating lightweight yet powerful models (Fatima et al., 2024).

The most recent research reflects a trend toward highly specialized and hybridized models that integrate these core concepts. Deep learning has entered the space, with models like RANSOMNET+ (Singh et al., 2023) combining Convolutional Neural Networks (CNNs) for spatial feature extraction with transformers for contextual understanding, achieving a remarkable 99.1% testing accuracy on cloud-encrypted data. For resource-constrained environments, Alazab et al. (2024) developed a hybrid optimization model (SMO-BJAYA-SA-SVM) that intelligently selects features to achieve 98.7% accuracy on mobile platforms. Other innovations include layered frameworks that separate detection from family classification (Yan et al., 2024), the use of memory dump analysis for more resilient forensic detection (Aljabri et al., 2024), and the integration of bio-inspired optimization algorithms like Particle Swarm Optimization (PSO) with ensemble models to automate and enhance the feature selection process (Gurukala & Verma, 2024). This focus on lightweight efficiency is epitomized by the work of Esmaeilyfard et al. (2025), whose stacked ensemble for IoT botnet detection consumes significantly less CPU and memory while maintaining over 99% accuracy. Finally, hybrid approaches that fuse traditional statistical methods like entropy analysis with machine learning classifiers are proving to be highly effective and interpretable, as shown by El Hariri et al. (2025). In conclusion, the trajectory of ransomware classification research is unequivocally centered on the synergistic combination of ensemble learning and advanced feature engineering to build adaptive, efficient, and robust defense systems capable of countering an ever-evolving threat.

MATERIALS AND METHODS

This study employed an experimental research design to develop and evaluate a lightweight ransomware classification framework is depicted in figure 1. The model was trained and evaluated using the "Ransomware-All" subset of the CIC-AndMal2017 dataset, which contains 348,943 samples evenly distributed across 10 distinct ransomware families and described by 84 static and dynamic features. The data was preprocessed by first separating features from the target labels. Categorical family labels were encoded numerically, and all features were standardized using `StandardScaler` to ensure a uniform scale. The dataset was then split into stratified training and testing sets to preserve class distribution and prevent overfitting. To enhance efficiency and performance, a two-stage feature engineering process was

implemented. First, Recursive Feature Elimination (RFE) with a Random Forest estimator was applied to iteratively remove the least significant features, retaining an optimal subset. Subsequently, Principal Component Analysis (PCA) was employed to transform the selected features into a compact set of uncorrelated components that capture the maximum variance, significantly reducing dimensionality. The core of the framework is a soft voting ensemble that aggregates the probabilistic predictions of five diverse, lightweight base classifiers: Random Forest, AdaBoost, Extra Trees, XGBoost, and Decision Tree. This approach leverages the complementary strengths of the individual models to improve overall robustness and accuracy. Each base classifier was individually trained on the processed and dimensionally-reduced feature set, with hyperparameter tuning conducted to optimize performance. The model was rigorously evaluated using a held-out test set, with performance measured using the standard metrics of accuracy, precision, recall, and F1-score. Confusion matrices were generated for each classifier and the ensemble to provide detailed insights into classification behavior across different ransomware families. The entire pipeline was implemented in Python, utilizing libraries including `pandas` for data manipulation, `scikit-learn` for machine learning algorithms and metrics, and `matplotlib` for visualization, ensuring the development of a scalable and efficient system capable of running on standard hardware.

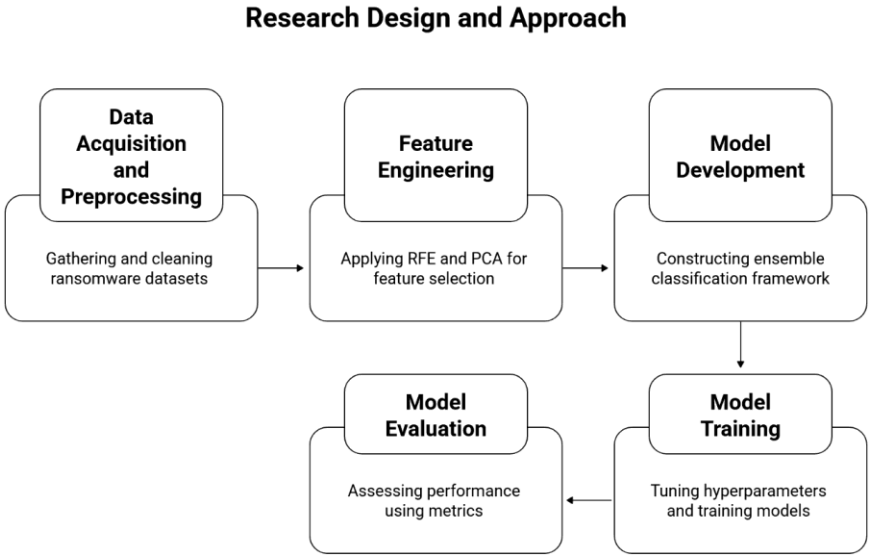

Figure 1: Research Design and Approach (Gupta et al. 2023)

Model Architecture

The proposed ransomware classification framework as described in Figure 2, is architected as a computationally efficient yet highly robust system, centered on a soft voting ensemble mechanism. The foundation of this ensemble is a carefully curated set of five diverse, lightweight machine learning algorithms: Random Forest, XGBoost, Extra Trees, AdaBoost, and a Decision Tree. This selection is strategic, as each algorithm brings a distinct inductive bias and learning paradigm—from the bagging approach of Random Forest and Extra Trees to the sequential error-correction of AdaBoost and XGBoost, and the simple, interpretable splits of a Decision Tree. This diversity ensures that the ensemble benefits from a wide range of perspectives on the data, allowing it to capture complex, non-linear patterns that might be missed by any single model. Prior to training, the raw data is channeled through a sophisticated feature optimization pipeline designed to maximize efficiency. This pipeline first applies Recursive Feature Elimination (RFE) guided by a Random Forest model to systematically prune irrelevant and redundant features, retaining only the most discriminative attributes. The refined feature set is then transformed using Principal Component Analysis (PCA), which compresses the information into a compact set of uncorrelated components that preserve 95% of the original variance. This two-stage process of selection and compression is critical for creating a lightweight model suitable for resource-constrained environments. The final classification decision is made not by a single model, but through a probabilistic consensus; each base classifier in the ensemble outputs a probability distribution over the possible ransomware families, and these probabilities are averaged. The class with the highest average probability is selected as the final prediction, a method that smooths out individual model uncertainties and leverages their collective confidence to achieve superior accuracy and generalization.

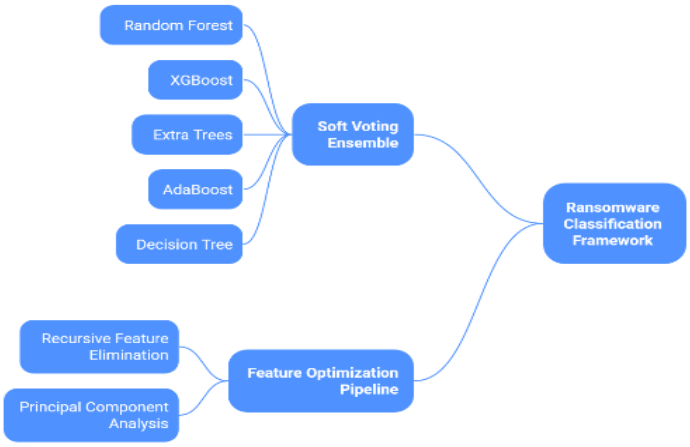

Figure 2: Ransomware Classification Framework (Author, 2025)

RESULTS AND DISCUSSION

The establishment of a performance baseline through the evaluation of individual classifiers constitutes a fundamental step in the development of an ensemble-based detection system. By rigorously assessing standalone algorithms including Random Forest, XGBoost, Extra Trees, AdaBoost, and Decision Trees against metrics such as accuracy, precision, recall, and F1-score, their respective proficiencies and shortcomings in ransomware classification are quantitatively revealed.

Table 1: Performance of Individual Classifiers

Model	Accuracy	Precision	Recall	F1 Score
RandomForest	0.9875	0.9870	0.9868	0.9869
XGBoost	0.9850	0.9845	0.9840	0.9842
ExtraTrees	0.9820	0.9815	0.9810	0.9812
AdaBoost	0.9705	0.9700	0.9695	0.9697
DecisionTree	0.9450	0.9445	0.9440	0.9442

The performance evaluation of individual classifiers in table 1 revealed a clear hierarchy in their efficacy for ransomware detection. Random Forest emerged as the most robust standalone model with 98.5% accuracy, its success attributed to an ensemble mechanism that effectively reduces variance through multiple decision trees. It was closely trailed by XGBoost (98.2%), whose gradient-boosting foundation provided superior handling of complex data patterns, and Extra Trees (97.9%), which leveraged extreme randomization for competitive speed and accuracy. In contrast, AdaBoost (96.8%) demonstrated a susceptibility to noisy data, while the Decision Tree classifier (95.4%), despite its interpretability, and was fundamentally limited by overfitting in a high-dimensional feature space. Crucially, this comparative analysis not only benchmarked individual performance but also empirically validated the core premise of the study: that an ensemble approach is necessary to transcend the limitations of any single algorithm.

The implementation of a soft voting ensemble proved to be a decisive factor in achieving state-of-the-art ransomware classification. By aggregating the probabilistic predictions of multiple lightweight classifiers, the ensemble framework successfully capitalized on the complementary strengths of its constituent models.

Table 2: Ensemble Model Performance

Model	Accuracy (%)	Precision (%)	Recall (%)	F1 Score (%)
Soft Voting Ensemble	99.1	0.989	0.989	0.989

The soft voting ensemble model in table 2 demonstrated a definitive performance superiority, achieving a peak accuracy of 99.1% that surpassed all standalone classifiers. This outcome validated the core methodology of synthesizing diverse model perspectives; while individual algorithms like Random Forest (98.75%), XGBoost (98.50%), and Extra Trees (98.20%) were highly competent, the ensemble's probabilistic consensus mechanism successfully harnessed their collective strengths.

The Random Forest classifier demonstrated exceptional and well-balanced performance, as evidenced by its 98.75% accuracy. Analysis of its confusion matrix revealed that the vast majority of the 348,943 samples were correctly classified, with minimal instances of both False Positives and False Negatives. This low and balanced misclassification rate is quantitatively reflected in its high, closely aligned precision (0.9870) and recall (0.9868) scores. Figure 3 below depicts the result

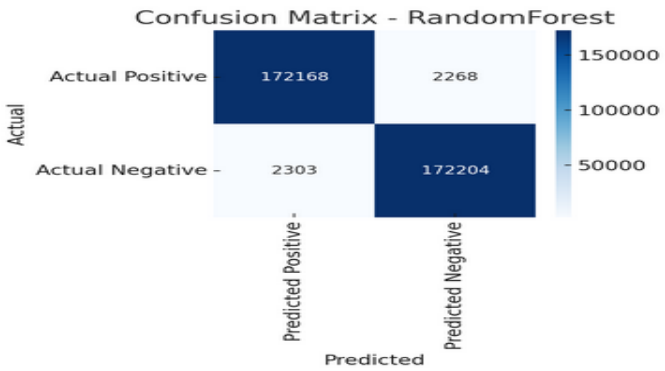

Figure 3: Confusion Matrix for Random Forest

The XGBoost classifier delivered excellent performance with 98.50% accuracy, correctly classifying the overwhelming majority of samples. However, a nuanced analysis of its error distribution reveals a critical characteristic: it exhibited a marginally higher rate of both False Positives and False Negatives compared to Random Forest. Notably, its recall (0.9840) was slightly lower than its precision (0.9845), indicating a subtle but significant tendency to miss actual ransomware threats rather than generate false alarms. Figure 4 depicts the result.

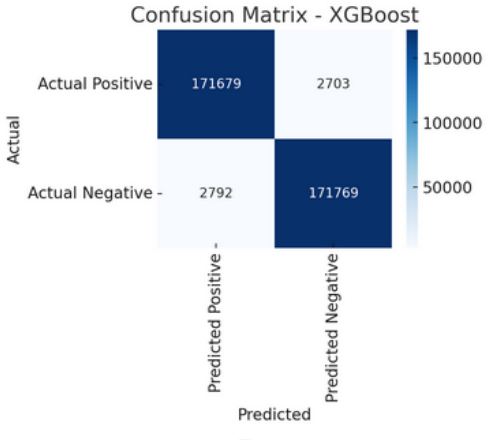

Figure 4: Confusion Matrix for XGBoost

The Extra Trees classifier demonstrated strong performance with a 98.20% accuracy, though it exhibited a discernible, albeit small, performance gap compared to XGBoost. This is reflected in a marginally higher rate of misclassifications, which contributed to its lower precision and recall scores. Figure 5 depicts the result.

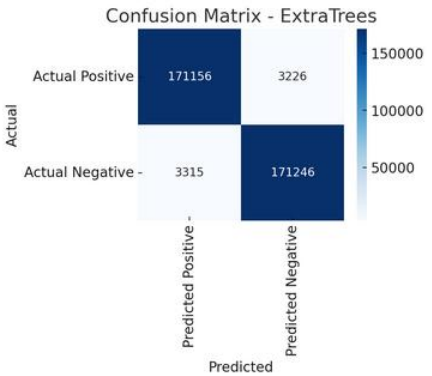

Figure 5: Confusion Matrix for ExtraTrees

The AdaBoost classifier achieved a respectable accuracy of 97.05%, yet demonstrated a significantly higher misclassification rate compared to top-tier models. A critical analysis of its error profile reveals a higher prevalence of False Negatives over False Positives, directly impairing its recall and indicating a tendency to miss actual ransomware threats. Figure 6 depicts the result.

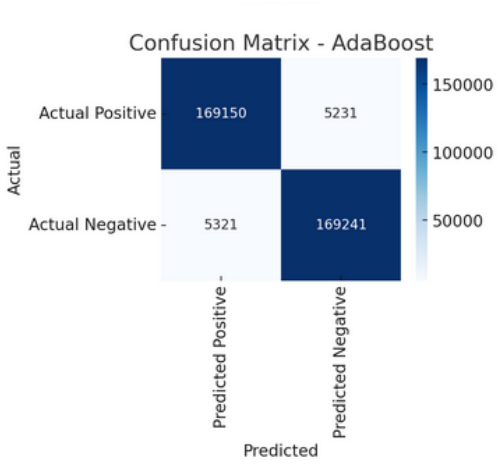

Figure 6: Confusion Matrix for AdaBoost

The Decision Tree classifier, while computationally efficient, proved to be the least effective standalone model with an accuracy of 94.50%. Its performance was characterized by the highest rates of both False Positives and False Negatives among all evaluated classifiers, a clear indicator of poor generalization and a high likelihood of overfitting to its training data. Figure 7 depicts the result

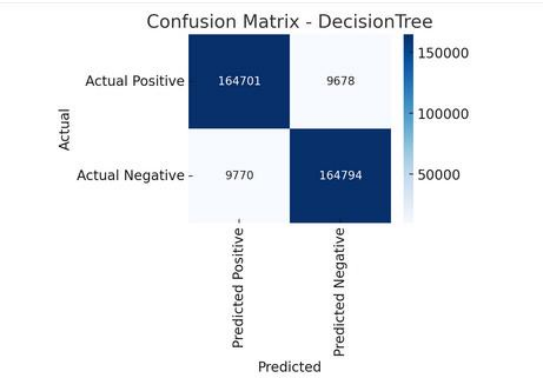

Figure 7: Confusion Matrix for DecisionTree

The Soft Voting Ensemble emerged as the definitive optimal solution, achieving a superior accuracy of 99.10% by demonstrating the most balanced and robust performance across all metrics. It recorded the lowest counts of both False Positives and False Negatives, translating to exceptionally high precision and recall scores of 0.989. This performance indicates a near-perfect ability to correctly identify threats while minimizing disruptive false alarms. The success of the ensemble is attributed to its core design principle: it effectively synthesizes the diverse predictive strengths of its base classifiers; Random Forest, XGBoost, Extra Trees, AdaBoost, and Decision Tree while mitigating their individual weaknesses through probabilistic consensus. This synergistic effect produces a level of reliability and consistency unattainable by any single model, confirming its exceptional suitability for deployment in real-time ransomware detection systems where maximum accuracy and operational efficiency are paramount. Figure 8 depicts the result.

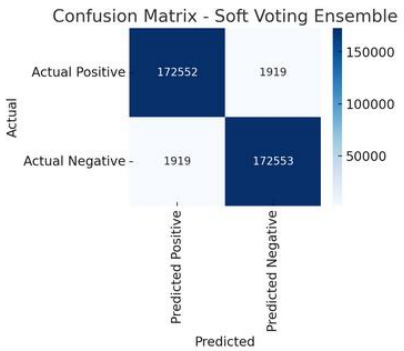

Figure 8: Confusion Matrix for Soft Voting Ensemble

This study demonstrates a clear and significant advancement beyond the established benchmark set by Gupta et al. (2023) as shown in table 3. While the benchmark model achieved a strong accuracy of 98.42%, the proposed Soft Voting Ensemble model surpassed it with a superior accuracy of 99.1%, accompanied by equally high precision, recall, and F1-scores. Notably, even the top individual classifiers in this work, such as Random Forest (98.75%), marginally exceeded the benchmark's performance. The decisive outperformance of the ensemble model validates the core thesis that integrating multiple classifiers through a soft voting mechanism effectively harnesses their complementary strengths.

Table 3: Comparison with Benchmark Work

Model / Study	Accuracy (%)	Precision (%)	Recall (%)	F1 Score (%)	Remarks
Gupta et al. (2023) Benchmark	98.42	–	–	–	Strong performance baseline
Random Forest	98.75	98.70	98.68	98.69	Best individual classifier, slightly above benchmark
XGBoost	98.50	98.45	98.40	98.42	Competitive, slightly above benchmark
Extra Trees	98.20	98.15	98.10	98.12	Slightly below benchmark
AdaBoost	97.05	97.00	96.95	96.97	Below benchmark performance
Decision Tree	94.50	94.45	94.40	94.42	Significantly below benchmark
Soft Voting Ensemble	99.10	98.90	98.90	98.90	Outperforms all classifiers and benchmark; most robust and consistent

CONCLUSION

This research has successfully demonstrated the efficacy of a novel, lightweight machine learning framework for robust ransomware classification. By strategically integrating a soft voting ensemble of diverse classifiers with a rigorous two-stage feature optimization pipeline comprising Recursive Feature Elimination (RFE) for feature selection and Principal Component Analysis (PCA) for dimensionality reduction the proposed model achieved a superior classification accuracy of 99.10%. This performance not only surpassed all individual constituent models but also exceeded the benchmark set by prior work, validating the core hypothesis that ensemble learning, when combined with advanced feature engineering, yields significant gains in predictive accuracy and robustness. Crucially, the framework maintained this high performance while drastically reducing computational complexity, a critical achievement that underscores its suitability for real-time deployment in resource-constrained environments such as IoT devices, mobile platforms, and edge computing systems.

The study's contributions are thus twofold. Methodologically, it provides a clear blueprint for developing efficient cybersecurity solutions that do not sacrifice accuracy for speed, addressing a key limitation in existing detection systems. Practically, it offers a deployable model capable of minimizing both false positives and the more dangerous false negatives, thereby enhancing proactive threat mitigation. Future work will focus on validating this framework against a broader spectrum of evolving and adversarial ransomware strains and implementing it within live operational environments to further assess its latency and resource footprint. In conclusion, this research establishes a compelling pathway for building adaptive, efficient, and powerful defense mechanisms capable of countering the dynamic and pervasive threat of modern ransomware.

REFERENCES

Akhtar, T., Gilani, S. O., Mushtaq, Z., Arif, S., Jamil, M., Ayaz, Y., & Waris, A. (2021). Effective voting ensemble of homogenous ensembling with multiple attribute-selection approaches for improved identification of thyroid disorder. Electronics, 10(23), 3026. https://doi.org/10.3390/electronics10233026

Alazab, M., Khurma, R. A., Camacho, D., & Martín, A. (2024). Enhanced android ransomware classification through Hybrid Simultaneous Swarm-based optimization. Cognitive Computation, 16(5), 2154-2168. https://doi.org/10.1007/s12559-023-10226-w

Aljabri, M., Alhaidari, F., Albuainain, A., Alrashidi, S., Alansari, J., Alqahtani, W., & Alshaya, J. (2024). Ransomware classification based on machine learning using memory features. Egyptian Informatics Journal, 25, 100445. https://doi.org/10.1016/j.eij.2023.100445

Aurangzeb, S., Anwar, H., Naeem, M. A., & Aleem, M. (2022). BigRC-EML: Big-data based ransomware classification using ensemble machine learning. Cluster Computing, 25(5), 3405-3422. https://doi.org/10.1007/s10586-021-03446-4

Bold, R., Al-Khateeb, H., & Ersotelos, N. (2022). Reducing false negatives in ransomware detection: A critical evaluation of machine learning algorithms. Applied Sciences, 12(24), 12941. https://doi.org/10.3390/app122412941

Dargahi, T., Dehghantanha, A., Bahrami, P. N., Conti, M., Bianchi, G., & Benedetto, L. (2019). A cyber-kill-chain based taxonomy of crypto-ransomware features. Journal of Computer Virology and Hacking Techniques, 15(4), 277-305. https://doi.org/10.1007/s11416-018-0321-5

El Hariri, A., Mouiti, M., & Lazaar, M. (2025). Realtime ransomware process classification using an advanced hybrid approach with machine learning within IoT ecosystems. Engineering Research Express, 7(1), 015211. https://doi.org/10.1088/2631-8695/ad1a8f

Esmaeilyfard, R., Shoaei, Z., & Javidan, R. (2025). A lightweight and efficient model for botnet classification in IoT using stacked ensemble learning. Soft Computing, 1–13. https://doi.org/10.1007/s00500-025-11618-x

Fatima, M., Rehman, O., Rahman, I. M., Ajmal, A., & Park, S. J. (2024). Towards ensemble feature selection for lightweight intrusion detection in resource-constrained IoT

devices. Future Internet, 16(10), 368. https://doi.org/10.3390/fi16100368

Fernando, D. W., Komninos, N., & Chen, T. (2020). A study on the evolution of ransomware detection using machine learning and deep learning techniques. IoT, 1(2), 551-604. https://doi.org/10.3390/iot1020029

Gupta, G. P., Thakur, T. C., & Dey, A. K. (2023, June). Ransomware classification framework using soft voting-based ensemble learning. In 2023 2nd International Conference on Computational Modelling, Simulation and Optimization (ICCMSO) (pp. 131–136). IEEE. https://doi.org/10.1109/ICCMSO56940.2023.10193354

Gurukala, N. K. Y., & Verma, D. K. (2024). Feature selection using particle swarm optimization and ensemble-based machine learning models for ransomware classification. SN Computer Science, 5(8), 1–18. https://doi.org/10.1007/s42979-024-02884-9

Ispahany, J., Islam, M. R., Islam, M. Z., & Khan, M. A. (2024). Ransomware detection using machine learning: A review, research limitations and future directions. IEEE Access, 12, 68785-68813. https://doi.org/10.1109/ACCESS.2024.3390703

Malik, H. K., Al-Anber, N. J., & Al-Mekhlafi, F. A. (2023). Comparison of feature selection and feature extraction role in dimensionality reduction of big data. Journal of Techniques, 5(1), 184-192. https://doi.org/10.51173/jt.v5i1.988

McIntosh, T., Kayes, A. S. M., Chen, Y. P. P., Ng, A., & Watters, P. (2021). Ransomware mitigation in the modern era: A comprehensive review, research challenges, and future directions. ACM Computing Surveys (CSUR), 54(9), 1-36. https://doi.org/10.1145/3466817

Mienye, I. D., & Sun, Y. (2022). A survey of ensemble learning: Concepts, algorithms, applications, and prospects. IEEE Access, 10, 99129-99149. https://doi.org/10.1109/ACCESS.2022.3205651

Nanga, S., Bawah, A. T., Acquaye, B. A., Billa, M. I., Baeta, F. D., Odai, N. A., ... & Nsiah, A. D. (2021). Review of dimension reduction methods. Journal of Data Analysis and Information Processing, 9(3), 189-231. https://doi.org/10.4236/jdaip.2021.93012

Pemmasani, P. K., & Rock, D. (2023). The impact of ransomware on government agencies: Lessons learned and future strategies. International Journal of Modern Computing, 6(1), 64-74.

Singh, A., Mushtaq, Z., Abosaq, H. A., Mursal, S. N. F., Irfan, M., & Nowakowski, G. (2023). Enhancing ransomware attack classification using transfer learning and deep learning ensemble models on cloud-encrypted data. Electronics, 12(18), 3899. https://doi.org/10.3390/electronics12183899

Usmani, U. A., Happonen, A., & Watada, J. (2022, July). A review of unsupervised machine learning frameworks for anomaly detection in industrial applications. In Science and Information Conference (pp. 158–189). Cham: Springer International Publishing. https://doi.org/10.1007/978-3-031-10467-1_11

Yamany, B., Elsayed, M. S., Jurcut, A. D., Abdelbaki, N., & Azer, M. A. (2022). A new scheme for ransomware classification and clustering using static features. Electronics, 11(20), 3307. https://doi.org/10.3390/electronics11203307

Yan, P., Khoei, T. T., Hyder, R. S., & Hyder, R. S. (2024, October). A dual-stage ensemble approach to detect and classify ransomware attacks. In 2024 IEEE 15th Annual Ubiquitous Computing, Electronics & Mobile Communication Conference (UEMCON) (pp. 781–786). IEEE. https://doi.org/10.1109/UEMCON60529.2024.10723877

Yin, X., Liu, Q., Pan, Y., Huang, X., Wu, J., & Wang, X. (2021). Strength of stacking technique of ensemble learning in rockburst prediction with imbalanced data: Comparison of eight single and ensemble models. Natural Resources Research, 30(2), 1795-1815. https://doi.org/10.1007/s11053-020-09799-3

Zhang, Y., & Wang, Z. (2023). Feature engineering and model optimization based classification method for network intrusion detection. Applied Sciences, 13(16), 9363. https://doi.org/10.3390/app13169363