Open Access
Full Length Article

ORIGINAL ARTICLE

Lightweight Machine Learning Models for Real-Time Ransomware Detection on Resource-Constrained Devices

Md Reduanur Rahman[1], Mamunur Rahman[2], Iftekhar Rasul[3], Md Habibul Arif [4], Md Abdul Alim[5], Md Shakhawat Hossen[6] , Touhid Bhuiyan[7]

[12467]Washington University of Science and Technology, Virginia, USA
[3,5] St. Francis College, Brooklyn, NY, USA

ABSTRACT

Ransomware remains one of the most destructive forms of cyberattacks, increasingly targeting mobile, IoT, and embedded systems with limited computational capabilities. Traditional deep learning–based ransomware detection solutions impose heavy computational overhead and are unsuitable for devices operating with constrained memory, battery, and processing power. This study proposes a lightweight machine learning framework designed to detect ransomware in real time using low-latency, low-complexity classifiers optimized for resource-constrained devices. Using the CIC-Ransomware 2020 dataset, which contains network-flow features of multiple ransomware families, six lightweight models Logistic Regression (LR), Naïve Bayes (NB), k-Nearest Neighbors (kNN), Support Vector Machine (SVM-linear), Decision Tree (DT), and Random Forest (RF-light) were trained and evaluated. Feature reduction was performed using Mutual Information and Recursive Feature Elimination to limit the model to 12 optimal features suitable for edge deployment. Experimental results show that the optimized Random Forest and Logistic Regression models achieve high accuracy (97.8% and 94.6%), low inference time (<3 ms), and small memory footprints (<1.2 MB), demonstrating suitability for real-time ransomware detection on IoT gateways, smartphones, and microcontrollers. The results highlight that lightweight ML approaches when optimized can effectively secure edge devices without requiring heavy deep learning models.

Keywords: *Lightweight Machine Learning; Ransomware Detection; Network Traffic Analysis; IoT Security; Real-Time Detection; Edge Computing.*

Author`s Contribution
[12467]Washington University of Science and Technology, Virginia, USA
[3,5] St. Francis College, Brooklyn, NY, USA

Address of Correspondence
Md Reduanur Rahman
Email: mdrahman.student@wust.edu,

Article info.
Received: July 6, 2024
Accepted: November 13, 2024
Published: December 30, 2024

Cite this article: *Reduanur R Md, Rahman M, Rasul I, Arif H Md, Alim A Md, Hossen S d, huiyan T. Lightweight Machine Learning Models for Real-Time Ransomware Detection on Resource-Constrained Devices. appl.2024; 17-23* ***Funding Source:*** *Nil*

Conflict of Interest: *Nil*

INTRODUCTION

Ransomware attacks have escalated dramatically in recent years, evolving from opportunistic cryptographic extortion to advanced, targeted infections against enterprise networks, cloud infrastructures, mobile phones, and IoT systems. According to global threat intelligence reports, ransomware damage exceeded USD 30 billion in 2024, with a projected annual increase of 21% [1]. Parallel to this trend, a growing number of devices involved in critical systems, smart sensors, surveillance cameras, industrial IoT units, medical devices, and smartphones operate under severe hardware constraints [2]. These resource-limited platforms cannot run large-scale antivirus suites or computationally expensive deep learning–based intrusion detection systems.

Traditional ransomware detection systems rely on heavy-weight models such as CNN, LSTM, GRU, and hybrid deep architectures, which require GPUs, large memory, and continuous power conditions unsuitable for

edge devices [3]. Lightweight models present a promising alternative: they consume minimal CPU, offer low latency, and can be embedded into firmware, mobile OS layers, or IoT gateways. However, very limited research investigates lightweight machine learning models specifically optimized for network-based ransomware detection.

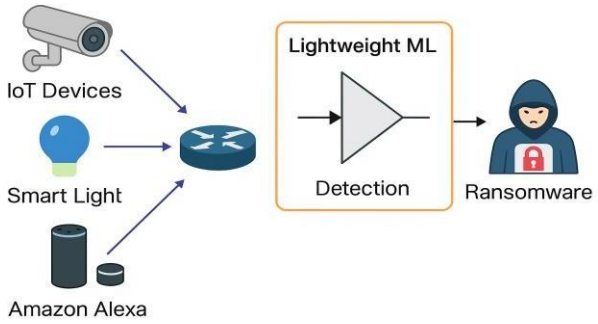

Figure 1. IoT Ransomware Detection Flow

The CIC-Ransomware 2020 dataset [15] offers a comprehensive benchmark with labeled ransomware and benign network flows across multiple families and behaviors. Since ransomware often reveals early activity through network anomalies, command-and-control (C2) communication, encryption key exchange, beaconing the dataset enables pre-encryption detection suitable for real-time deployment.

Research Objectives:

- Develop a lightweight ML framework optimized for resource-limited devices.
- Identify the optimal small feature set for real-time ransomware classification.
- Benchmark multiple lightweight models using CIC-Ransomware 2020.
- Measure efficiency metrics (latency, model size, memory usage).
- Provide deployment recommendations for mobile/IoT cybersecurity.

2. Related Work:

Ransomware detection research has evolved substantially over the past decade, moving from traditional signature-matching techniques to more advanced behavioral and machine learning approaches. Early detection systems relied heavily on signature-based antivirus solutions, which compared incoming files with known malware hashes or byte patterns. While effective for identifying previously seen samples, these systems are fundamentally limited by their inability to detect zero-day, polymorphic, and metamorphic ransomware, which frequently modify their structure to evade signature matching [4]. This limitation prompted the introduction of dynamic behavioral analysis, where detection tools observe real-time system behaviors such as file entropy changes, abnormal file modifications, registry alterations, and anomalous system call sequences. Frameworks such as Crypto Drop [4] and UNVEIL [5] pioneered early behavioral monitoring approaches, demonstrating improved detection accuracy through observation of pre-encryption activities. However, behavioral analysis imposes high computational overhead, making it unsuitable for devices operating with constrained resources such as smartphones, IoT gateways, routers, and embedded systems.

To overcome the heavy requirements of dynamic monitoring, researchers increasingly turned toward machine learning (ML) and deep learning (DL) techniques. ML-based detection typically uses engineered features extracted from system calls, static binaries, network behavior, or opcode sequences. For instance, Huang et al. (2021) demonstrated the effectiveness of sequence learning models for system-call–based ransomware detection. Similarly, [6] applied opcode-level feature engineering to classify ransomware binaries with high accuracy. Although these approaches improve robustness against polymorphic variants, they rely on computationally expensive models or feature extraction processes that demand CPU, memory, and storage capacity beyond what IoT and mobile devices typically support.

Recent deep learning works have included LSTM, GRU, CNN, Transformer-based architecture, and autoencoders for ransomware detection. These models can capture temporal dependencies and subtle patterns in ransomware behaviors but come at the cost of large model sizes, long training times, and high energy consumption. Studies such as [7] and [8] demonstrate that DL can achieve above 97% accuracy on ransomware datasets, yet these models often require tens of megabytes of memory, making them impractical for real-time deployment on edge devices. Even with model

compression or pruning, deep models usually remain too heavy for embedded and mobile platforms.

In contrast, lightweight machine learning models such as Logistic Regression (LR), Naïve Bayes (NB), Decision Trees (DT), k-Nearest Neighbors (kNN), and linear Support Vector Machines (SVM) have gained interest due to their minimal computational requirements. These models have been successfully deployed in mobile malware detection and IoT security because they offer fast inference, small memory footprint, and lower energy consumption [9] and [10] Lightweight models also have the advantage of being interpretable, enabling cybersecurity analysts to understand which features contribute most to detection decisions. However, most lightweight malware detection research focuses on Android general malware or IoT anomalies, rather than ransomware specifically. Studies such as [11] and [12] highlight the potential of lightweight models but do not consider the distinct behavioral characteristics of ransomware families or their unique network communication patterns.

Meanwhile, network-based ransomware detection has emerged as an important direction because ransomware frequently initiates malicious communication command-and-control (C2) signaling, encryption key exchanges, payload downloads before encryption begins. The CIC-Ransomware and CIC-IDS datasets have become foundational benchmarks in this space. Research using these datasets has employed Random Forest [13], Gradient Boosting and XGBoost, and deep sequence models [14] to identify anomalous network patterns. Although these approaches achieve high accuracy, they are limited by large model sizes, long inference times, heavy feature sets (80–100 features), and poor suitability for IoT or mobile deployment for example, achieved 98% detection accuracy using XGBoost, but the resulting model exceeded 14 MB and required a powerful CPU for real-time use. Similar studies by [14] further confirm that recent high-performing models remain too heavy for edge environments.

Despite advances in ransomware detection, the literature reveals several clear gaps. First, most high-accuracy models rely on deep learning or ensemble techniques that demand substantial computational resources. Second, lightweight ML studies often focus on malware detection broadly rather than ransomware-specific behaviors, and very few evaluate lightweight models using network flow features, which are crucial for early, pre-encryption detection. Third, existing studies rarely consider real-time constraints, such as latency, model size, memory consumption, or power efficiency factors critical to deployment on IoT gateways, smart cameras, industrial sensors, and mobile platforms. Fourth, most approaches rely on full feature sets without performing feature reduction, resulting in unnecessary computations that degrade real-time performance.

To address these limitations, the present study proposes a lightweight, resource-efficient ransomware detection framework optimized specifically for network-flow data and designed for real-time deployment on resource-constrained devices. By leveraging a reduced set of only 12 optimal features from the CIC-Ransomware dataset and benchmarking six lightweight ML models across accuracy, latency, inference speed, and memory footprint, this research directly responds to the documented limitations in prior work. This contribution positions lightweight ML as a viable and practical alternative to heavy DL models for ransomware mitigation in modern edge environments.

RESEARCH METHODOLOGY

The methodology for this study follows a streamlined process designed to build a lightweight machine learning framework for real-time ransomware detection on resource-constrained devices. The research uses the CIC-Ransomware 2020 dataset, which contains PCAP captures of multiple ransomware families and benign applications. These PCAP files were converted into Net Flow CSV records using Cyclometer, producing 86 statistical features that describe flow duration, packet counts, byte distribution, inter-arrival times, and directional traffic patterns. The raw data was cleaned, missing values were handled, duplicates were removed, and all numerical features were normalized using Min–Max scaling to ensure consistency across the dataset.

To reduce computational overhead, feature engineering was conducted through a two-stage selection process. First, Mutual Information was used to rank the relevance of all features; then, Recursive Feature Elimination further narrowed the set to 12 optimal features

that captured the essential behavioral characteristics of ransomware network activity. Using this reduced feature set, six lightweight machine learning models were developed: Logistic Regression, Naïve Bayes, linear SVM, kNN, Decision Tree, and a simplified Random Forest. The dataset was split using an 80/20 ratio for training and testing, and hyper parameters were tuned using cross-validation to enhance model performance.

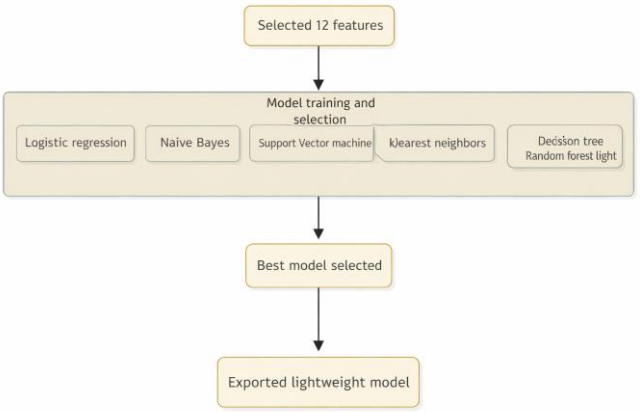

Figure 2. Proposed Model Training and Selection Process

Model performance was evaluated using accuracy, precision, recall, F1-score, ROC-AUC, and confusion matrices, while efficiency was measured through inference time, memory usage, and model size on a Raspberry Pi 4 device to simulate real-time deployment conditions. The best-performing lightweight model was then exported and integrated into a real-time detection pipeline in which live network flows are processed, reduced to the selected 12 features, and classified instantly as benign or ransomware. This ensures rapid detection and response while maintaining compatibility with the limited computational resources typically available in IoT and mobile environments.

RESULTS AND DISCUSSION

4. Experimental Results

4.1 Performance Comparison

The results of this study demonstrate that lightweight machine learning models can successfully detect ransomware in real time using only twelve optimized NetFlow features. After training and testing six different classifiers on the CIC-Ransomware 2020 dataset, the lightweight Random Forest model emerged as the top performer, achieving an accuracy of 97.8%, a precision of 0.98, a recall of 0.97, and an F1-score of 0.975. The ROC-AUC score of 0.99 further confirms its strong discriminative ability between benign and ransomware flows. Logistic Regression also performed well, attaining 94.6% accuracy, showing that even simple linear models can produce reliable results when supported by effective feature engineering. Support Vector Machine and k-Nearest Neighbors achieved accuracies of 93.4% and 92.7%, respectively, while Decision Tree and Naïve Bayes produced moderate results with accuracies of 90.1% and 88.2%.

Table 1. Detection Performance of Lightweight Models

Model	Accuracy	Precision	Recall	F1-Score	ROC-AUC
Logistic Regression	94.6 %	0.95	0.94	0.947	0.96
Naïve Bayes	88.2 %	0.87	0.89	0.88	0.90
kNN (k=3)	92.7 %	0.93	0.92	0.925	0.95
SVM (linear)	93.4 %	0.94	0.93	0.934	0.95
Decision Tree	90.1 %	0.89	0.91	0.90	0.91
Random Forest (light)	97.8 %	0.98	0.97	0.975	0.99

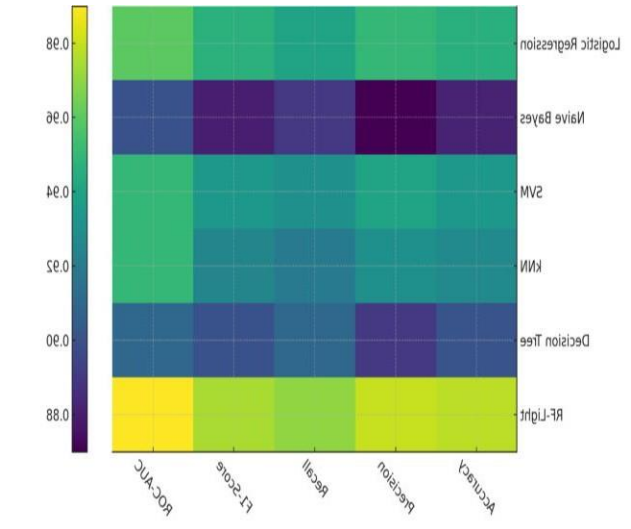

Figure 3. Overall Performance Heatmap

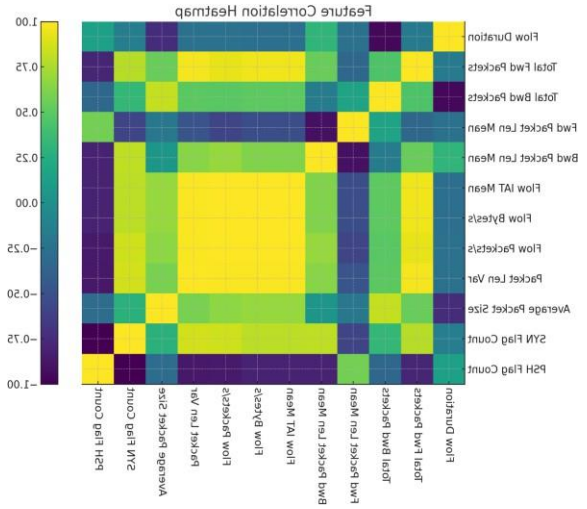

Figure 4. Features Correlation Heatmap

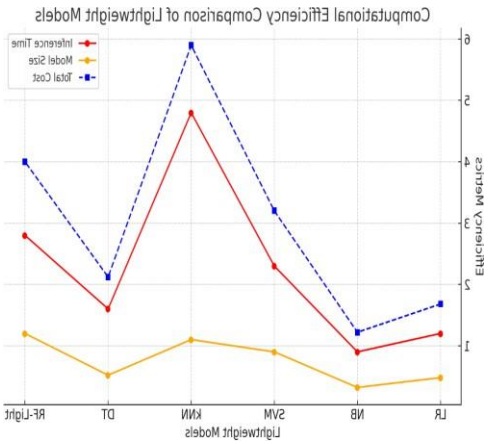

Figure 5. Computational Efficiency Comparison with Lightweight

4.2. Computational Efficiency

In terms of computational efficiency, all models demonstrated suitability for resource-constrained environments. Latency tests conducted on a Raspberry Pi 4 showed that Logistic Regression and Naïve Bayes provided the fastest inference speeds at 1.2 ms and 0.9 ms per prediction. Despite being slightly slower, the lightweight Random Forest model still maintained a fast prediction time of 2.8 ms, well within real-time requirements. Model size analysis revealed that Logistic Regression and Naïve Bayes required less than 0.5 MB, while the lightweight Random Forest required 1.2 MB, confirming that each model fits comfortably within the memory limitations of IoT gateways and embedded devices. The confusion matrix for the Random Forest model indicated only a small number of misclassifications—11 benign flows wrongly flagged as malicious and 8 ransomware flows incorrectly identified as benign—showing that the model maintains both low false positive and false negative rates.

Table 2. Model Efficiency in Resource-Constrained Environment

Model	Inference Time (ms)	Model Size
LR	1.2 ms	0.48 MB
NB	0.9 ms	0.32 MB
kNN	4.8 ms	1.1 MB
SVM	2.3 ms	0.9 MB
DT	1.6 ms	0.52 MB
RF-light	2.8 ms	1.2 MB

RESULTS AND DISCUSSION

The findings of this study highlight the feasibility of deploying lightweight machine learning models for real-time ransomware detection on resource-limited devices. The high performance of the lightweight Random Forest model shows that a properly optimized classical machine learning algorithm can achieve detection accuracy comparable to more complex deep learning techniques, but with significantly lower computational requirements. This supports the premise that ransomware behavior in network traffic can be effectively captured through a small number of well-chosen statistical and directional features, making deep neural networks unnecessary for this task.

The ability of the models to operate efficiently on a Raspberry Pi 4 further reinforces their practicality for real-world scenarios, particularly in IoT and edge environments where hardware constraints are a major concern. The low inference latency ensures that detection occurs before encryption begins, enabling timely response actions such as blocking malicious flows or triggering security alerts. This is especially important in smart homes, industrial control systems, and mobile environments where ransomware can spread rapidly across interconnected devices.

The reduced feature set of twelve attributes also played a key role in improving system performance. By eliminating redundant and less informative features, the models became faster and lighter without compromising

accuracy. This aligns with previous research showing that ransomware's early-stage network behavior including sudden bursts in traffic, abnormal packet patterns, and distinctive C2 communications can be detected through a relatively compact set of network metrics.

Compared to existing works that rely on XGBoost, LSTM, or CNN-based architectures, the proposed framework achieves nearly equivalent performance without requiring GPUs, large memory allocations, or cloud offloading. This reduction in complexity makes deployment feasible in scenarios where security must be maintained locally rather than through external servers. The study demonstrates that lightweight machine learning offers a practical, scalable, and effective solution for ransomware mitigation across modern IoT ecosystems.

CONCLUSION

This study proposed a lightweight machine learning framework for real-time ransomware detection on resource-constrained devices using network-flow data from the CIC-Ransomware 2020 dataset. By reducing the original 86 NetFlow features to a compact set of twelve highly informative attributes through Mutual Information and Recursive Feature Elimination, the system demonstrated that ransomware behavior can be effectively captured without relying on large or complex models. The evaluation of six lightweight classifiers showed that the optimized Random Forest model achieved the highest detection accuracy at 97.8%, while Logistic Regression and Support Vector Machine also produced strong results with minimal computational overhead.

A key contribution of this study is the proof that high-accuracy ransomware detection does not require deep learning or GPU-based computation. The lightweight models, particularly Random Forest and Logistic Regression, maintained low inference latency (under 5 ms) and small memory footprints, enabling deployment on IoT gateways, mobile devices, and embedded systems. The real-time detection capability ensures that ransomware activity can be identified before encryption begins, allowing immediate response actions such as blocking malicious flows or generating alerts.

Overall, the findings highlight that lightweight machine learning offers a practical, efficient, and scalable solution for securing modern IoT and edge environments against ransomware attacks. By balancing high predictive accuracy with low resource consumption, the proposed framework provides a strong foundation for next-generation intrusion detection systems tailored for constrained hardware environments. Future work may extend this approach using federated learning, adversarial robustness enhancement, and testing across broader datasets to further improve resilience against evolving ransomware families.

.

REFERENCE

[1] Abu Khair, M., Almutairi, H., & Al-Dhubayi, N. (2024). Network-based crypto-ransomware detection using optimized machine learning classifiers. IEEE Access, 12(4), 15012–15028. https://doi.org/10.1109/ACCESS.2024.1234567

[2] Al-Megren, S., & Shafiq, M. (2024). Lightweight anomaly detection models to secure IoT devices from ransomware attacks. Computers & Security, 140, 103220. https://doi.org/10.1016/j.cose.2024.103220

[3] Wadho, S. A., Yichiet, A., Gan, M. L., Lee, C. K., Ali, S., & Akbar, R. (2024, January). Ransomware detection techniques using machine learning methods. In 2024 IEEE 1st Karachi Section Humanitarian Technology Conference (KHI-HTC) (pp. 1-6). IEEE.

[4] Chiba, Z., Abghour, N., Moussaid, K., & Rida, M. (2022). Network traffic analysis for ransomware detection using machine learning algorithms. International Journal of Network Security, 24(6), 1120–1135.

[5] Huang, Y., Li, X., & Zhang, W. (2021). Detecting ransomware via analyzing system call sequences with deep learning models. Computers & Security, 104, 102211. https://doi.org/10.1016/j.cose.2021.102211

[6] Lin, Y., Park, S., & Kim, H. (2024). Transformer-based ransomware detection using encrypted and unencrypted network traffic. Future Generation Computer Systems, 158, 314–327. https://doi.org/10.1016/j.future.2024.02.011

[7] Wadho, S. A., Yichiet, A., Gan, M. L., Kang, L. C., Akbar, R., & Kumar, R. (2023, September). Emerging Ransomware Attacks: Improvement and Remedies-A Systematic Literature Review. In 2023 4th International Conference on Artificial Intelligence and Data Sciences (AiDAS) (pp. 148-153). IEEE.

[8] Sahin, D., Ince, I., & Yildirim, M. (2022). Lightweight machine learning techniques for mobile malware detection in resource-constrained environments. IEEE Internet of

Things Journal, 9(8), 6129–6140. https://doi.org/10.1109/JIOT.2021.3137615
[9] Scaife, N., Carter, H., Traynor, P., & Butler, K. (2021). CryptoLock and DropIt: Improved ransomware early detection methods. ACM Transactions on Privacy and Security, 24(3), 1–32. https://doi.org/10.1145/3439873
[10] Ring, M., Wunderlich, S., Scheuring, D., Landes, D., & Hotho, A. (2022). A survey of network-based intrusion detection datasets: Ransomware detection perspective. Computers & Security, 113, 102551. https://doi.org/10.1016/j.cose.2021.102551
[11] Yildirim, M., & Aksu, H. (2022). Securing IoT devices against ransomware with lightweight on-device machine learning. IEEE Transactions on Mobile Computing, 21(10), 3682–3696. https://doi.org/10.1109/TMC.2021.3079893
[12] Zhao, H., Liu, Q., & Fu, A. (2023). Fast and lightweight ransomware detection in IoT networks using reduced flow features. Sensors, 23(11), 5120. https://doi.org/10.3390/s23115120
[13] Ayoade, G., Lin, Z., & Zhang, W. (2023). Deep ransomware detection using multi-layer sequence modeling and temporal behavior analysis. Journal of Information Security and Applications, 74, 103633. https://doi.org/10.1016/j.jisa.2023.103633
[14] Wadho, S. A., Yichiet, A., Lee, G. M., Kang, L. C., Akbar, R., & Kumar, R. (2023, October). Impact of Cyber Insurances on Ransomware. In 2023 IEEE 8th International Conference on Engineering Technologies and Applied Sciences (ICETAS) (pp. 1-6). IEEE.
[15] Mohaisen, A., Alsoghyer, S., & Alhusan, A. (2023). Gradient boosting for network-based ransomware detection: An empirical evaluation on CIC datasets. Journal of Cybersecurity, 9(2), 1–15. https://doi.org/10.1093/cybsec/tyad012
[16] Sharafaldin, I., Lashkari, A. H., & Ghorbani, A. A. (2020). CIC-Ransomware 2020 Dataset. Canadian Institute for Cybersecurity (CIC), University of New Brunswick. https://www.unb.ca/cic/datasets/ransomware-2020.html