Ransomware Detection in Healthcare: Enhancing Cybersecurity with Processor and Disk Usage Data

Rongali Abhiram, Kunal and J. Shobana
Department of Data Science and Business System, School of Computing, SRM Institute of Science and Technology, Kattankulathur, Tamil Nadu, India

Keywords: Ransomware Detection, Healthcare Cybersecurity, Processor-Disk Monitoring, Machine Learning.

Abstract: Security of sensitive medical information and critical patient care is threatened with the increasing digitization of healthcare infrastructure, making hospitals prime targets for ransomware attacks. Besides having a significant computational burden, conventional detection methods such as signature-based and behavioural analysis are often lagging the rapidly evolving ransomware variants. This paper presents a new approach to detecting ransomware through host machine-level processor and disk I/O event monitoring and is tailored for use in healthcare environments. With the use of an RF classifier and machine learning-based method, the solution can detect threats in real time without affecting system performance. The framework is rigorously tested under simulated hospital workload scenarios and 22 ransomware variants, demonstrating its robustness, effectiveness, and adaptability to safeguard healthcare infrastructures from modern ransomware attacks.

1 INTRODUCTION

Ransomware is now one of the most dangerous forms of cyberattacks, affecting individuals, organizations, and critical infrastructure worldwide. Ransomware malware encrypts data or locks devices and demands a ransom to be paid in exchange for unlocking them. With the progress of digitalization and cloud storage, cyberattacks have taken advantage of vulnerabilities to carry out sophisticated ransomware attacks, which are likely to cause permanent harm. Ransomware first surfaced on the threat horizon in the late 1980s, some of the earliest well-documented attacks being the PC Cyborg Trojan, which was used to hide directories and encrypt files. Ransomware started off as a single hacker tool and developed into an extremely profitable cybercrime-as-a-service business, with cybercriminals selling ransomware kits to others for a price. Advanced campaigns such as WannaCry and Covid Lock have proven the havoc that ransomware can wreak in healthcare, government, and financial institutions, disrupting vital services and stealing sensitive information. Prolific remote work and cloud infrastructure have fed the threat, with attackers exploiting phishing campaigns, malicious links, and software vulnerabilities to spread ransomware. Signature and behaviour-based ransomware detection technologies are not able to identify new and emerging types of ransomware attacks. Hence, machine learning, artificial intelligence, and situational awareness technologies are being created to enhance ransomware detection and prevention. This research work is targeted to explore current ransomware attack mechanisms, evaluate the prevailing detection frameworks, and outline a real-time detection framework using processor and disk activity information. Integrating machine learning techniques, this endeavour is intended to propose an effective mechanism to suppress ransomware attacks with low system overhead and ensure strong security to critical infrastructure.

2 LITERATURE SURVEY

1. "Ransomware Detection Using Machine Learning: A Review, Research Limitations and Future Directions" (2024): This paper summarizes machine learning solutions for ransomware detection, reports trends in current research, detection strategies, and design strategies. It recognizes some significant limitations, including outdated ransomware trends, a lack of real-world practical solutions, and a lack of consideration of early and real-time detection. The

Abhiram, R., Kunal, and Shobana, J.
Ransomware Detection in Healthcare: Enhancing Cybersecurity with Processor and Disk Usage Data.
DOI: 10.5220/0013910000004919

In *Proceedings of the 1st International Conference on Research and Development in Information, Communication, and Computing Technologies (ICRDICCT'25 2025) - Volume 4*, pages 188-193
ISBN: 978-989-758-777-1

study also suggests directions for future studies to enhance ransomware detection systems and make them more effective against cyber-attacks.

2. "AI-Based Ransomware Detection: A Comprehensive Review." (2024): The paper is an end-to-end review of AI-based ransomware detection methods, comparing different machine learning and deep learning models on the basis of their efficiency. The paper discusses future directions in ransomware, data acquisition problems, feature engineering methods, and performance metrics. The paper also proposes a systematic evaluation framework to enhance ransomware detection and formulates research gaps, proposing future research areas such as hybrid and transfer learning mechanisms for enhancing detection.

3. "Detection and Decryption of Ransomware." (2023): The article discusses detection and decryption of ransomware by identifying changed file extensions of malware. It reports on the growing danger of a ransomware attack, overviews current detection techniques such as antivirus programs and intrusion detection systems, and describes a detection and decryption tool for ransomware-affected files. The research shows the need for proactive security and proposes stronger ransomware defence mechanisms.

4. "Ransomware Classification and Detection with Machine Learning Algorithms." (2023): This article talks about various methods and techniques used for detection and prevention against ransomware. The authors give a description of various types of ransomwares and various methods of detection and prevention used by each type.

5. "Detection of Ransomware Attacks Using Processor and Disk Usage Data." (2023): It suggests a machine learning technique based on CPU and disk I/O event monitoring for ransomware attack detection. In order to reduce overhead and stop ransomware manipulation, it does away with direct process monitoring. The random forest classifier has proved to be most effective with 98% detection accuracy and a response time of 400 milliseconds. The study points out why it is even more crucial to train models for varied workloads to develop resilience. The process provides a guaranteed and efficient method to detect ransomware in real time, and it is especially well suited for cloud environments in the form of virtual computers.

6. “Enhanced Ransomware Detection Techniques using Machine Learning Algorithms.” (2021): The paper talks about different machine learning methods for ransomware detection. It compares a variety of algorithms, such as KNN, Naïve Bayes, Random Forest, and Decision Trees, to test their performance on ransomware data. The research gathers behavioural information like API calls, target files, registry accesses, signatures, and network accesses to distinguish between ransomware and normal samples. The paper discovers Random Forest to possess the highest accuracy rate of 96%, thus being a strong candidate in ransomware detection.

3 PROBLEM STATEMENT

Hospitals are increasingly the target of ransomware attacks due to the growing reliance of healthcare on digital infrastructure. These assaults can lead to system outages, patient data loss, and compromised patient care. Healthcare networks are susceptible to attack since existing detection methods, such as signature-based detection and endpoint security solutions, cannot handle recently developed ransomware strains. Current methods of behavioural analysis incur heavy computational overhead and degrade the performance of hospital-critical infrastructure. Moreover, sophisticated ransomware uses stealthy techniques to remain undetected, further complicating real-time detection. To overcome these issues, efficient, low-overhead, and precise ransomware detection is required for ensuring the protection of healthcare settings. This paper explores the application of processor and disk I/O event monitoring and machine learning algorithms for devising a strong detection system that provides effective real-time threat detection with negligible system performance impact.

4 EXISTING SYSTEM

Modern methods for detecting ransomware are based on more conventional approaches such as behaviour analysis, signature-based detection, and endpoint protection technologies. By comparing the malware signatures that already exist, signature-based detection can detect ransomware that has already been identified. It cannot detect recently discovered and spreading ransomware strains, and it is helpless against zero-day attacks. Behavioural analysis traces system activity for anomalous patterns, but this incurs heavy system monitoring with resultant high computational overhead and performance degradation, which is especially undesirable in healthcare IT infrastructure. Furthermore, endpoint security products can be evaded by sophisticated ransomware operating in stealth mode or by misusing

legitimate system processes to stay under the radar. Such constraints indicate that there is a requirement for more efficient real-time ransomware detection with less system performance overhead but better detection of sophisticated ransomware threats. Sandbox procedures also execute suspect executables within a sandboxed environment to observe their behaviour before allowing them to run on a real system. Though useful against known ransomware signatures, sandbox is performance-intensive and can be evaded.

5 PROPOSED SYSTEM

To overcome the shortcomings of current ransomware detection techniques, we introduce a new approach that gathers processor and disk I/O event information at the host machine level. Utilizing machine learning methods, our model provides fast ransomware detection with low system overhead. Contrary to conventional approaches that monitor all processes, our system uses low-impact monitoring, targeting hardware performance counters (HPCs) and patterns of disk I/O, diminishing performance slowdowns in mission-critical hospital settings. The model is further made configurable to accommodate hospital workload variability for the realization of real-world applicability across various medical systems. Our method utilizes a random forest classifier to attain high detection rates with minimal compromises in overall efficiency of hospital IT infrastructure.

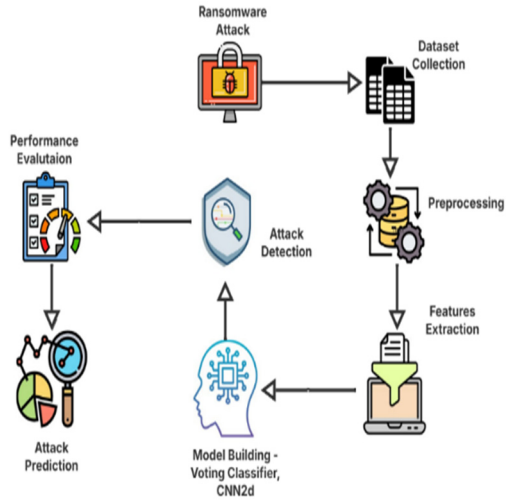

Figure 1: Architecture diagram.

The major benefits of this system are early ransomware detection, which avoids data encryption and system lockout, and protection of electronic health records (EHRs), medical imaging files, and other sensitive healthcare information from ransomware attacks. Figure 1 shows the Architecture Diagram.

6 MODEL TRAINING AND EVALUATION

6.1 Data Collection

The information is derived from Harvard Dataverse, an open data repository.

The chosen HPC and I/O dataset is system-level behaviour change that happens when a system is under ransomware infection. The dataset has 6,000 samples with high-volume capture of major hardware performance counters (HPCs) and disk I/O values. The dataset has features such as instructions retired, cache misses, branch load misses, disk read/write requests, and flush operations, which provide insights on how ransomware modifies system behaviour. Using a heatmap, we analyse the correlation of attributes and discover key parameters that diverge greatly in the presence of ransomware.

6.2 Feature Selection

Important system metrics such as instructions executed, cache misses, disk read/write requests, and flush operations are extracted to analyse ransomware impact.

6.3 Data Preprocessing

The dataset is cleaned and normalized to remove missing values and scale numerical features for better model performance.

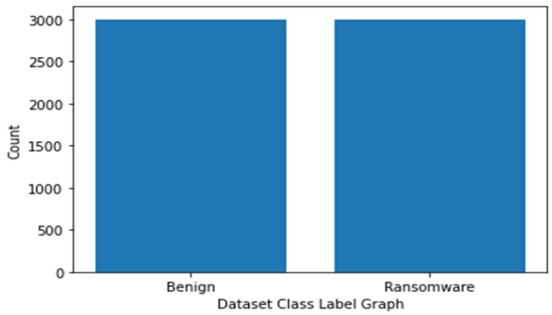

Figure 2: Dataset class label graph.

The above Figure 2 presents a summary of observed system behaviour changes, highlighting significant alterations in processor activity and disk operations when a ransomware attack occurs. This dataset forms the foundation for training our machine learning-based ransomware detection model, ensuring high accuracy and real-world applicability in securing critical systems.

In the dataset, the 'label' column represents the target or dependent variable, where:

- 0 → Indicates legitimate (benign) samples
- 1 → Indicates ransomware-infected samples

6.4 Data Splitting

The dataset is divided into 80% training and 20% testing data, ensuring a balanced model evaluation.

- Training Size (80%): 4800
- Testing Size (20%): 1200

6.5 Model Training

Machine Learning Models (XG Boost, Random Forest, SVM, KNN, Decision Trees) and Deep Learning Models (CNN, LSTM, DNN) are trained using Scikit-learn and TensorFlow/Keras libraries.
The Voting Classifier integrates predictions from various algorithms here we integrated AdaBoost and Random Forest. This ensemble approach improves the ransomware detection system's robustness and accuracy by considering diverse perspectives from different models.

7 MODEL EVALUATION

Table 1: Possible classification outcomes.

Classification	Description
True Positive (TP)	Test passes on benign file
True Negative (TN)	Test fails on ransomware file
False Positive (FP)	Test passes on ransomware file
False Negative (FN)	Test fails on benign file

Performance is evaluated using Accuracy, Precision, Recall, and F1-score to determine how well the models detect ransomware. Table 1 show Possible classification outcomes.

Accuracy: Accuracy is defined as the ratio of correctly classified instances to the total number of instances. Figure 3 illustrate Accuracy Graph.

$$Accuracy = \frac{TP+TN}{TP+FP+TN+FN} \quad (1)$$

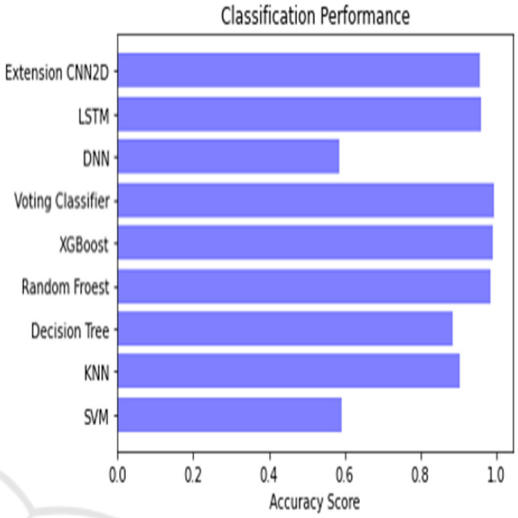

Figure 3: Accuracy graph.

Precision: Mathematically, Precision is defined as the ratio of true positives to the sum of true positives and false positives (FP). Figure 4 illustrate Precision Graph.

$$Precision = \frac{True\ Positives}{True\ Positives + False\ Positives} \quad (2)$$

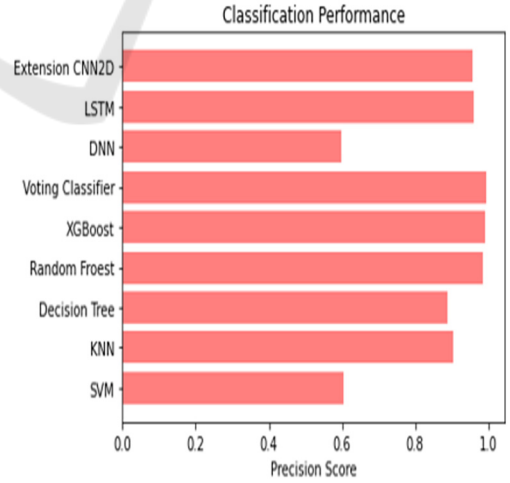

Figure 4: Precision graph.

Recall: Mathematically, recall is defined as the ratio of true positives (TP) to the sum of true positives and false negatives (FN). Figure 5 illustrate Recall Graph.

$$Recall = \frac{True\ Positives}{True\ Positives + False\ Negatives} \quad (3)$$

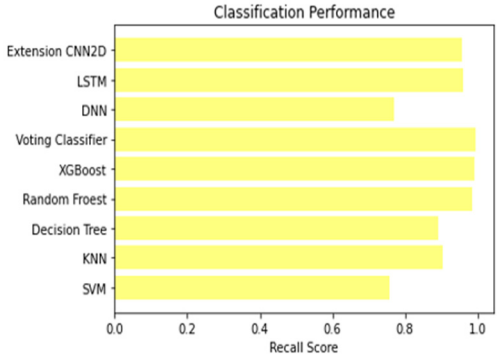

Figure 5: Recall graph.

F1 Score: F1 score is the harmonic mean of precision and recall, and provides a single metric that balances both measures. Figure 6 illustrate F1 Score Graph.

$$F1 = 2 * \frac{Precision * Recall}{Precision + Recall} \quad (4)$$

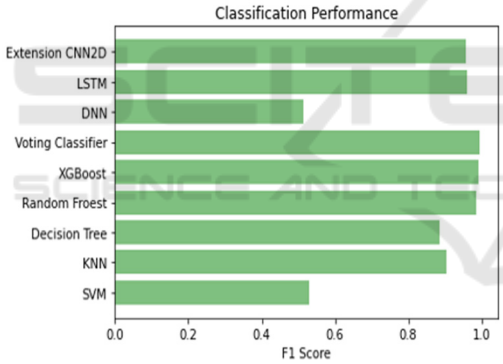

Figure 6: F1 score graph.

8 RESULT

The performance of different machine learning and deep learning models was tested on accuracy, precision, recall, and F1-score. Table 2 shows Possible classification outcomes. The results show that Voting Classifier and XG Boost performed better than all others in accuracy, followed by Random Forest and CNN.

- Voting Classifier was the top one, with an accuracy rate of 99.5%, and the most consistent method of detecting ransomware.
- Random Forest and XG Boost also did not disappoint, posting 99.2% and 98.5% respectively.
- CNN and LSTM performed very well, with CNN attaining 95.6% accuracy and LSTM 96.0%, which indicates their performance in pattern recognition.
- SVM and DNN were the worst on accuracy, which suggests that they may not be the best to use for classifying ransomware.

Table 2: Possible classification outcomes.

ML Model	Accuracy	Precision	Recall	F1-score
SVM	0.593	0.605	0.758	0.529
KNN	0.904	0.905	0.905	0.905
Decision Tree	0.885	0.887	0.891	0.885
Random Forest	0.985	0.985	0.985	0.985
XGBoost	0.992	0.991	0.992	0.992
Extension Voting Classifier	0.995	0.995	0.995	0.995
DNN	0.585	0.587	0.513	0.513
LSTM	0.96	0.961	0.961	0.961
Extension CNN2D	0.956	0.957	0.957	0.956

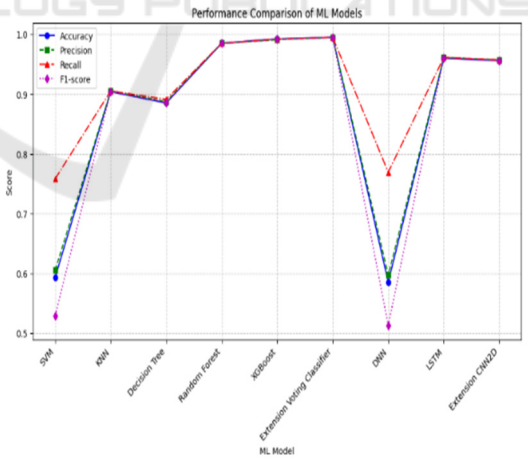

Figure 7: Comparison graph.

9 CONCLUSIONS

Ransomware poses a serious risk to healthcare settings. It can throw a wrench in operations, put

patient data at risk, and jeopardize vital medical services. Unfortunately, traditional detection methods frequently struggle to keep up with the fast-paced changes in ransomware tactics, leaving systems open to further attacks. This study seeks to enhance healthcare security through the proposal of a machine learning method for ransomware detection via tracking processor and disk I/O behaviour. Our Random Forest and XG Boost classifier model were highly accurate at low system performance cost while providing real-time protection without interrupting hospital-critical functions. Our findings indicate that hardware-based monitoring is an efficient and effective way to detect the presence of ransomware. The proposed solution was evaluated against various ransomware variants and varied hospital workloads, indicating its flexibility and reliability. Looking ahead, the combination of deep learning models and adversarial machine learning methods would continue to enhance ransomware detection through enhanced resilience against sophisticated attacks. In addition, growing dataset heterogeneity and incorporating real-time dynamic analysis will further improve performance. With the use of proactive, AI-driven cybersecurity technologies, healthcare organizations can significantly reduce the risk of ransomware attacks, ensuring the security, integrity, and availability of critical patient data and medical systems.

10 FUTURE SCOPE

Our detection features could be improved, even though our proposed approach is yielding some genuinely encouraging results. To improve our feature extraction and pattern recognition skills and stay ahead of the consistently changing threat of new ransomware types, we might think about deploying deep learning algorithms like LSTM and CNNs. Plus, incorporating real-time threat intelligence and cloud-based monitoring could speed up our response times, reducing the potential damage from attacks. As we move forward, it'll be essential to diversify our datasets and bring in adversarial machine learning techniques to make our detection models tougher against clever evasion tactics. Future research might also delve into using blockchain-based security frameworks to add an extra layer of data integrity and protection. By embracing AI-driven cybersecurity solutions, healthcare institutions can take proactive steps to fend off ransomware threats, ensuring that medical services run smoothly, protecting patient data, and bolstering overall system security.

REFERENCES

A. Vehabovic, N. Ghani, E. Bou-Harb, J. Crichigno, and A. Yayimli, "Ransomware detection and classification strategies," 2023.

B. Marais, T. Quertier, and S. Morucci, "Ai-based malware and ransomware detection models," 2022.

Ferdous, J., Islam, R., Mahboubi, A., & Islam, M. Z. (2024). "AI-Based Ransomware Detection: A Comprehensive Review." IEEE Access, Digital Object Identifier: 10.1109/ACCESS.2024.3461965. Received: 12 August 2024, Accepted: 11 September 2024, Published: 16 September 2024, Current Version: 30 September 2024

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, Digital Object Identifier: 10.1109/ACCESS.2024.3397921. Received: 11 February 2024, Accepted: 3 May 2024, Published: 7 May 2024, Current Version: 22 May 2024.

Jayanthi, M., Prakash, S. G., Ajay, A. J., & Vijayakumar, K. (2023). "Detection and Decryption of Ransomware." Proceedings of the Second International Conference on Applied Artificial Intelligence and Computing (ICAAIC 2023), IEEE Xplore, Part Number: CFP23BC3-ART, ISBN: 978-1-6654-5630-2.

K. Lee, J. Lee, S. Lee, and K. Yim, "Effective ransomware detection using entropy estimation of files for cloud services," Sensors, vol. 23, p. 3023, 03 2023.

Kunku, K., Zaman, A., & Roy, K. (2023). "Ransomware Detection and Classification using Machine Learning." Department of Physics & Computer Science, Wilfrid Laurier University, Waterloo, ON, Canada; Department of Computer Science, North Carolina A&T State University, Greensboro, NC, USA.

M. Masum, M. J. H. Faruk, H. Shahriar, K. Qian, D. Lo, and M. I. Adnan, "Ransomware classification and detection with machine learning algorithms," in 2022 IEEE 12th Annual Computing and Communication Workshop and Conference (CCWC). IEEE, Jan 2022.

R. A. Mowri, M. Siddula, and K. Roy, "Application of explainable machine learning in detecting and classifying ransomware families based on api call analysis," 2022.

S. Razaulla, C. Fachkha, C. Markarian, A. Gawanmeh, W. Mansoor, B. C. M. Fung, and C. Assi, "The age of ransomware: A survey on the evolution, taxonomy, and research directions," IEEE Access, vol. 11, pp. 40 698–40 723, 2023.

Usha, G., Madhavan, P., Cruz, M. V., Vinoth, N. A. S., Veena, & Nancy, M. (2023). "Enhanced Ransomware Detection Techniques using Machine Learning Algorithms." Department of Computing Technology, SRM Institute of Science and Technology (SRM IST), Kattankulathur, India.