ISSN:0975-9646

Proactive Ransomware Detection with Machine Learning: Comprehensive Techniques and Future Directions

S R Surya[#1], **T Sivakumar**[#2]
[#]*Department of Computer Science, Pondicherry University*
Ponndicherry, India
1suryaunnithan0603@gmail.com, 2tsivakumar72@gmail.com

Abstract — **Ransomware attacks have become increasingly advanced and frequent, posing a significant danger to digital infrastructures. This paper gives a comprehensive analysis that examines the critical role of machine learning in the prevention and detection of ransomware attacks through file attributes and network traffic analysis. Techniques of tree ensembles, neural networks and decision trees are used to identify ransomware activities. Handling unbalanced data is crucial for improving model robustness and techniques such as SMOTE, undersampling and oversampling play a key role in achieving this. Additionally, analyzing file entropy using Shannon entropy, helps in identifying encrypted data, which is often associated with ransomware attacks. Utilizing dynamic tools such as CRSTATIC can provide a comprehensive approach to cybersecurity. Integrating machine learning with memory forensics, utilizing tools such as Volatility, is essential in detecting file-less malware. Innovations in feature extraction, classification models, and hybrid machine learning techniques significantly strengthen ransomware detection efficiency. Despite advances, current methods face limitations, particularly in handling different ransomware types, outdated signature-based detections and adapting to evolving ransomware strategies. Addressing these problems with proactive, adaptive and comprehensive solutions can improve the effectiveness of ransomware detection and protection. This research demonstrates the potential of machine learning to transform cybersecurity by offering strong protections against one of the biggest threats of the digital era.**

Keywords — Ransomware, Machine Learning, Cybersecurity, Network traffic analysis

I. INTRODUCTION

Ransomware attacks are one of the biggest cybersecurity risks that businesses are currently facing. The use of ransomware by cybercriminals to extort money has increased recently. The malware encrypts the data of its victims and demands payment to unlock it. These attacks have had far-reaching consequences on a variety of sectors, including government, healthcare, banking and education [1]. The late 1970s witnessed the emergence of ransomware. Typically, it gains access to reputable resources to disrupt regular activities. This kind of virus encrypts user data without authorization, preventing users from accessing their files. Ransomware attacks are distinct from other forms of malware because of their irreversible nature. The decryption key is the only means of decrypting data once it has been encrypted. To access the data, attackers typically demand payment in untraceable currencies like Bitcoin. These attacks target organizations as well as people to extract more money. Attackers take advantage of ransomware attack's irrevocable effects and the untraceable nature of their payments. Threats to victims include data loss or misuse, as well as the disclosure of private information like search histories [2].

Machine Learning (ML) is a technique for predicting computer behaviour and trends by analysing and learning from current data using regression models. At its foundation, ML involves developing powerful algorithms capable of identifying and learning patterns in data. Its modern applications include computer vision, speech processing, natural language processing, autonomous vehicle control, and many more fields of science and technology. Furthermore, ML is used in real-world scenarios, such as AI security systems that detect harmful objects at airports, ports and trains, as well as intelligent Closed Circuit Television (CCTV) that recognize anomalous behaviour [3].

The application of ML, which tackles complex problems in a variety of fields, to the information security domain has been actively pursued in recent years. In the past, malicious attacks exhibited comparatively simple and predictable patterns, which enabled an Intrusion Detection System (IDS) and attack analysis via pattern recognition. However, identifying and stopping cyberattacks has been harder due to the spread of networked devices such as smartphones, automobiles, houses, factories, and grids, as well as the creation of sophisticated attack methods that can get past conventional security mechanisms. Attackers take advantage of the irreversible effects of ransomware attacks as well as the untraceable nature of their payments. The loss or improper use of their data, as well as the disclosure of private information like search histories, pose threats to victims. Virus detection and network attack analysis are two examples of current technologies that have demonstrated limitations in protecting systems as malware and cyberattacks get more sophisticated, employing Artificial Intelligence (AI) to act intelligently. To be ready for ever-changing cyber threats, developing systems that can learn from complex and sophisticated cyberattacks is essential, as well as employ ML and AI to make accurate predictions in the information security space [4].

This paper is organized into four sections. Section I provide an overview to the Ransomware introduction. Section II analysis of ML techniques that have been researched to determine if they are capable of identifying and preventing ransomware attacks. Section III reviews related studies and outline the limitations faced in current ransomware detection and prevention strategies. This paper concludes with Section IV, summarize the key findings and implications of the research. Each section provides a comprehensive examination of the topic, increasing understanding of challenges and developments in ransomware security on a larger scale.

II. Machine Learning Techniques For Ransomware Detection And Prevention

Ransomware detection and prevention have evolved to leverage advanced ML techniques and hybrid approaches. Key methods include analysing file-sharing traffic to detect ransomware, employing decision trees, neural networks, and tree ensembles to evaluate network traffic performance. For efficient malware classification, Support Vector Machine (SVM), Random Forest, and Decision Trees are used together with data balancing techniques like Synthetic Minority Oversampling Technique (SMOTE) and undersampling, while Logistic Regression models predict Uniform Resource Locator (URL) security [3]. Detecting encrypted files involves analysing file entropy and computing Shannon entropy to gauge encryption levels. File-less malware is identified through memory forensics and machine learning, utilizing classification methods and the Volatility tool to analyse characteristics. For ransomware detection through static analysis, the CRSTATIC tool is employed, while Android malware identification leverages the GreatEatlon program, which uses code flow analysis [4].

Packet inspection, traffic analysis, anomaly detection, signature-based detection, and behavior-based detection are other detection techniques. Static feature analysis for ransomware identification often relies on Import Address Table (IAT) analysis and Strings, with automated calculation of the Jaccard Index through feature extraction[5]. Evasion strategies like functional separation, process splitting, and imitation are addressed, alongside innovative attacks distributing the virus workload across multiple collaborating processes. Malware analysis on virtual machines using the Cuckoo agent combines network traffic analysis, dynamic API call analysis, and supervised ML techniques, including decision trees, K-Nearest Neighbors (KNN), SVM, and logistic regression[6]. A hybrid machine learning approach evaluates attack intensity by employing semantic similarity algorithms hence enhances ransomware detection [7].

Proposals for ML-based malware classification involve converting malware binary files into images for categorization with Androguard used for static analysis of Android Application Package (APK) files [8]. Supervised ML models, such as Random Forest for regression and classification, are developed and evaluated across various ransomware datasets. Integrating behavior-based detection with machine learning algorithms, the "snowballing" search technique, and six internet databases aids in identifying ransomware [9]. Comparative studies of ML based ransomware detection frameworks often use decision trees in Sequential Feature Selection (SFS) to identify relevant features and evaluate model performance and feature importance through classifiers [10]. Feature extraction through static analysis, balanced datasets using hybrid, SMOTE, and undersampling techniques, and detection models built using KNN, SVM, and Iterative Dichotomiser 3 (ID3) classifiers further enhance detection capabilities. Binary particle swarm optimization is utilized for feature selection and hyperparameter tuning, with classification performed using SMOTE and SVM on datasets compiled from various APK files. The Cuckoo Sandbox supports dynamic analysis, while ML methods like gradient-boosted trees, random forests, and neural networks are employed [11]. Comprehensive threat classification, cryptographic module analysis, and understanding of ransomware spreading techniques support early detection strategies, combining signature matching, behavior analysis, machine learning, regular updates and backups to guard against all ransomware strains [12].

III. Review Of Related Studies

The field of ransomware detection is characterized by numerous challenges and limitations that necessitate ongoing research and innovation. Existing studies predominantly focus on crypto-ransomware, with limited attention given to other ransomware descendants. This narrow scope restricts the applicability of these detection methods to a broader range of ransomware variants, thereby necessitating the development of more comprehensive solutions.

One significant limitation of current detection methodologies is their inapplicability to mobile operating systems due to the primary focus on file-sharing scenarios typical of desktop environments [13]. Consequently, mobile devices remain vulnerable to ransomware attacks, highlighting the need for research tailored to mobile operating systems. Additionally, file-less ransomware cases are often excluded from evaluations, further limiting the effectiveness of current detection strategies. File-less ransomware, which operates in memory without leaving a footprint on the file system, poses a unique challenge that is not adequately addressed by existing detection mechanisms.

Data deletion resulting from undersampling techniques in machine learning model training can lead to the loss of critical information, which is detrimental to the accuracy of these models [14]. Furthermore, high False Positive Rates (FPR) are a persistent issue, with benign applications frequently misidentified as malware. This is not only undermining the reliability of the detection systems but also causes unnecessary disruptions for users. The packed samples present another challenge, as they hinder accurate static feature similarity checks [5]. These samples, often compressed or concealed, require dynamic analysis to uncover their true nature. However, the development of dynamic analyzers to enhance classification accuracy is still in its initial stages, necessitating further research and development. Decryption challenges posed by sophisticated cryptographic algorithms, large key sizes and significant computational power requirements complicate data recovery efforts. When ransomware overwrites and encrypts data, recovering the original information becomes exceedingly difficult. The scarcity of high-quality cybersecurity datasets further hinders the training of machine learning models, limiting their effectiveness [15]. The difficulty in obtaining file-less malware samples, often due to inactive servers, worsens this problem. In addition to these technical challenges, there are methodological limitations in current research. The need for diverse dataset ages to ensure accurate machine learning model training is often overlooked. The impact of split points on accuracy evaluation is another critical factor left for future investigation. Single ML techniques have shown poor accuracy, ranging from 11% to 51% underscoring the need for more effective detection methods [7]. In comparison, hybrid machine learning algorithms have greatly enhanced accuracy, achieving levels of up to 91%. While decision trees are widely used, they are susceptible to overfitting and can be affected by outliers. Moreover, previous ransomware detection works have often overlooked the importance of false negatives, which are critical in understanding the true performance of detection models. The alteration of file extensions by ransomware also poses usability issues, necessitating automated recovery mechanisms to restore original file extensions [12].

Network-based deployment of detection tools is an emerging area of focus, aiming to provide enhanced protection across interconnected systems. However, there is a lack of differentiation in research concerning desktop versus mobile-based environments and insufficient coverage of IoT-based research in surveys. This gap highlights the need for more comprehensive studies that include various platforms and devices. Current literature also lacks a consolidated framework for ransomware detection, avoidance and mitigation. Without a unified approach, efforts remain fragmented and less effective. Additionally, the absence of specific mention of external funding for research raises concerns about the sustainability and scope of ongoing projects. Data quality and quantity are recurrent challenges in training machine learning models for ransomware detection. Collecting and preprocessing data is often arduous, limiting the development of robust models. EldeRan, a known ransomware detection tool, fails to detect silent ransomware within sandbox environments, primarily due to the lack of running applications, which reduces detection accuracy [16]. Moreover, the limited availability of original ransomware and goodware data samples further affects detection efficiency. EldeRan also detects ransomware only after the infection has occurred, underscoring the need for more proactive detection strategies.

Ransomware attacks frequently target predefined features to evade detection, necessitating feature overlap to ensure comprehensive coverage. However, the large number of features increases model complexity, requiring fine-grained modeling to maintain accuracy. Modeling errors and uncertainties can impact the effectiveness of imitation techniques, while the lack of detailed feature descriptions in studies makes replication challenging. Feature datasets for ransomware detection experiments are often limited, restricting the scope of research. Previous methods have used a limited subset of file types, focusing on overall entropy rather than specific file characteristics. Dataset issues require extensive preprocessing techniques to resolve inconsistencies and enhance model performance [3]. Initial model predictions are typically inaccurate, necessitating hyperparameter tuning to improve results. Categorical attributes often need encoding into numerical values for model training, while dataset imbalance is addressed using techniques such as undersampling, oversampling, Balanced Bagging and SMOTE. However, SMOTE does not consider majority classes during oversampling and may not address strong correlations between minority and majority classes.

The scarcity of cybersecurity datasets continues to hinder machine learning model training. Obtaining file-less malware samples remains difficult due to inactive servers, while decryption challenges posed by complex cryptographic algorithms and key sizes complicate data recovery efforts [15]. Packed samples hinder accurate static feature similarity checks, necessitating the development of dynamic analyzers to enhance classification accuracy. Data deletion during undersampling may result in the loss of important information, while high FPR for benign applications further complicates detection efforts.

In summary, the current state of ransomware detection research is marked by significant challenges and limitations. Future work must focus on developing comprehensive solutions that address these issues, including expanding the scope of detection to various ransomware descendants, enhancing mobile operating system protection, improving data collection and preprocessing techniques, and developing more robust machine learning models. Additionally, a consolidated framework for ransomware detection, avoidance, and mitigation, along with standardized metrics for evaluating anti-ransomware solutions is essential for advancing the field.

Table 3.1 describes the application of machine learning to analyse file-sharing traffic for ransomware detection focusing on assessing the performance of decision trees, neural networks, and tree ensembles in network traffic [13]. This analysis only looks at crypto-ransomware; file-less ransomware cases and other ransomware varieties are not included. Additionally, because of the file-sharing environment, it is not suitable for mobile operating systems.

A study proposed data balancing strategies including SMOTE and undersampling were applied to SVM, Random Forest, Decision Trees, and Logistic Regression models to improve malware classification and URL security forecasts [3]. Initial dataset inaccuracies were fixed via preprocessing and hyperparameter adjustments. Numerical values were encoded for categorical attributes, and SMOTE, balanced bagging, undersampling, and oversampling were used to effectively handle dataset imbalance.

A study utilized file entropy analysis and Shannon entropy to assess encryption levels, introducing a method for detecting encrypted files. Unlike previous approaches that focused on specific file types, this method considers overall entropy across all file types, providing a more comprehensive evaluation of encryption. By expanding the scope beyond particular file formats, this approach enhances the accuracy and effectiveness of encrypted file detection. It allows for a more complete assessment by analyzing a wider range of file types, making it possible to identify encrypted data more reliably. This broader perspective significantly improves detection capabilities, offering a more robust solution for identifying encrypted files across various formats, thus enhancing overall detection accuracy and performance [18].

Table 3.1- Ransomware Detection: Merging Machine Learning and Traffic Analysis

Techniques Used	Results	Contributions
Evaluate decision trees and neural networks on file-sharing traffic.	Ransomware tool detects 99% of ransomware.	Compare encryption-based detection tools with other methods.
Apply SMOTE, undersampling with SVM and trees for malware detection.	Random Forest outperformed AdaBoost, Gradient Boosting, Decision Trees.	Uses XGBoost, Random Forest, Decision Tree for malware detection.
Analyzes encryption levels using Shannon entropy to detect data.	The entropy values effectively differentiate ransomware encryption from compression.	This model identifies encrypted files and high-entropy data.
Detects file-less malware using machine learning and Volatility.	Logistic Regression: 75% True Positive Rate (TPR), 86.7% accuracy; Random Forest: 93.33% accuracy, 87.5% TPR.	

Table 3.1 focuses on memory forensics and machine learning are used in file-less malware detection, together with classification techniques and the Volatility tool for characteristic analysis [15]. Effective model training is, however, hindered by the shortage of cybersecurity datasets and the difficulty of getting file-less malware samples from inactive servers. This strategy seeks to improve malware detection capabilities despite these challenges.

Table 3.2 - Evaluating the Effectiveness of Ransomware Detection Tools: GreatEatlon and CRSTATIC

Techniques Used	Results	Contributions
Use GreatEatlon for Android malware and CRSTATIC for ransomware.	Data loss, technical expertise, or paying ransom are outcomes.	DAM architecture detects, prevents and mitigates ransomware; Continuum aids prevention.
Employ packet inspection, traffic analysis, and anomaly detection.	Achieved 99.9% detection rate with no false positives.	Ransomware's effects on vital infrastructure during the pandemic.

In Table 3.2, a study has been discussed that proposed ransomware detection using static analysis with the CRSTATIC tool and malware identification on Android via the GreatEatlon program, which uses code flow analysis [4]. However, there is no consolidated framework for ransomware detection, avoidance, and mitigation. Additionally, the research does not specify any external funding sources. Ransomware detection employs packet inspection, traffic analysis, anomaly detection, signature-based, and behavior-based methods [19]. Challenges include data quality and quantity issues for machine learning models, alongside difficulties in collecting and preprocessing data. These factors impact the effectiveness and accuracy of ransomware detection systems. Methods are proposed to improve classification accuracy, a dynamic analyser will be developed, as packed samples prevent accurate static feature similarity checks [5].

In Table 3.3, discusses static feature analysis that employs strings and IAT analysis for ransomware identification. An automated method for determining the Jaccard Index via feature extraction is also used. The goal of this combination strategy is to increase ransomware detection accuracy. To avoid detection, research showed attacks target predefined features, which requires feature overlap. Having a lot of features make problems more complicated and calls for more detailed modelling [20]. The efficacy of imitation can be impacted by modelling errors and uncertainty. The difficulties of precise identification are made more difficult by evasion techniques such as functional separation, process splitting, imitating, and creative attacks that divide the virus workload among cooperating systems. A discussion on network traffic analysis, dynamic Application Programming Interface (API) call analysis, and supervised machine learning techniques such as decision trees, logistic regression, SVM and KNN are all integrated into malware analysis on virtual machines using the Cuckoo agent [6]. Overfitting and outliers are common problems with decision trees. In earlier ransomware detection studies, the importance of false negatives was often underestimated. This oversight led to gaps in understanding the true effectiveness of the detection method.

Transforming malware binary files into graphics to classify them [21]. Although this strategy improves categorization, it has drawbacks, such as the requirement for more effective malware detection techniques in Table 3.4. Enhancing these techniques is crucial to overcoming existing limitations. By doing so, improved detection capabilities have been achieved, ensuring more effective and reliable ransomware detection methods.

Androguard is utilized for static analysis of APK files, with supervised machine learning models, including Random Forest for both regression and classification, developed and evaluated on various ransomware datasets [8]. Future work will address the need for diverse dataset ages to improve ML model training and evaluate the impact of different data split points on accuracy.Using the "snowballing" search method and six online databases, behavior-based detection combined with machine learning algorithms improves malware detection [9]. To prevent inaccuracies dynamic analysis demands a thorough investigation, and behavior-based detection could result in false positives if possibilities are not addressed. Enhancing detection accuracy and resolving possible false positive concerns are the goals of this integrated method. This study analysed many ML-based detection frameworks and investigated machine learning techniques for ransomware identification. However, EldeRan's accuracy is reduced when it comes to silent ransomware detection in sandbox situations because there aren't any active apps [22].

Table 3.3 - Analysis of Ransomware Detection Techniques: Combining Static Analysis, Instance Evasion, and Machine Learning

Techniques Used	Results	Contributions
Use IAT analysis, strings, and the Jaccard Index for ransomware.	Salsa20 decryptor and IAT indexing enhance ransomware classification.	Reviewed ransomware detection methods and proposed automated feature extraction.
Instance evasion techniques include process splitting and imitation.	Distributed malware tasks can drop detection accuracy to 0%.	Developed Cerberus to demonstrate strategies for evading ransomware detection.
Malware analysis on VMs uses network traffic and ML methods.	Shallow decision tree and Tanh activation achieved 98.65% accuracy.	Assessed ML methods and API calls for ransomware detection.

In Table 3.4, EldeRan identifies ransomware post-infection detection performance is impacted by the small sample sizes of original ransomware and goodware. In contrast, SFS makes use of classifiers to evaluate feature relevance and model performance in addition to decision trees to find pertinent features [10]. Although the limitations of this approach are not specifically addressed in the research, decision trees are useful in SFS because they help determine the relative relevance of various characteristics, which enhances the efficacy and accuracy of the model. Though they aren't specifically addressed, this method may have inherent drawbacks including sensitivity to frequent variations in the data or the possibility of overfitting. Despite these potential drawbacks, SFS decision trees are nevertheless useful feature selection tools because they provide insights into how particular attributes affect model behavior and overall performance, which can help develop more effective ransomware detection techniques. Despite specific shortcomings like overfitting, SFS evaluates feature relevance and model performance using classifiers and decision trees to help with ransomware detection. Decision trees identify significant components that improve model accuracy, while their limits are rarely explicitly addressed.

A study discussed detection of ransomware was much improved by an integrated machine learning technique that used semantic similarity algorithms to assess the intensity of the attack. Accuracy for single machine learning techniques ranged widely, from 11% to 51%. Hybrid machine learning techniques, on the other hand, significantly improved accuracy, achieving 91%.

Table 3.4 – Hybrid approach to Ransomware detection using ML and Semantic similarity analysis.

Techniques Used	Results	Contributions
Static analysis extracts features; SMOTE, undersampling balance datasets for detection.	KNN with SMOTE achieved 98.69% accuracy and 99.49% virus detection.	SMOTE and undersampling balanced datasets with effective feature ranking.
SMOTE and SVM classify APK files, with particle swarm optimization for tuning.	SMOTE-tBPSO-SVM achieved 97.5% g-mean.	Utilizes evolutionary machine learning for malware detection on unbalanced datasets.

Table 3.5 describes static analysis is used to extract features, while hybrid, SMOTE, and undersampling approaches are used to balance the datasets. To create a detection model, classifiers KNN, SVM, and ID3 are used [23].

Table 3.5 - Detection Using Static Analysis and Machine Learning

Techniques Used	Results	Contributions
Semantic similarity techniques enhance ransomware detection and attack evaluation.	Identify ransomware type, classify, assess criticality and suggest countermeasures.	Hybrid ML techniques boost ransomware detection using semantic similarity.
Proposed classifying malware binaries by converting them to images.	Proposed using Inception-based CNN to classify malware images.	Proposed machine learning approach and examined classification techniques.
Androguard analyses APKs statically; Random Forest models are tested.	The approach achieves 97.48% accuracy, surpassing modern techniques.	Shows dataset age affects detection accuracy and proposes improvement.
Uses machine learning and behavior-based detection with "snowballing" and databases.	SVM excelled over DT and N-grams; behavior-based methods addressed limitations.	
Compares ML techniques and frameworks for ransomware detection.	Compares ransomware detection methods, highlighting DRTHIS with CNN and LSTM.	Analyses ransomware detection frameworks and machine learning results.
Sequential Feature Selection with decision trees identifies and evaluates features.	Random Forest achieved 99.56% accuracy; MALWD&C ensured fast classification.	MALWDC classifies PE malware using key performance criteria.

However, there is a significant false positive rate, meaning that occasionally benign apps get incorrectly classified as malware, and data deletion during undersampling may result in the loss of crucial information. For feature selection and hyperparameter tuning, binary particle swarm optimization is employed, SMOTE and SVM are used for classification on a dataset of APK files from many sources [24]. SMOTE, however, may not be able to adequately handle strong correlations between minority and majority classes because it does not take majority class considerations into account while oversampling.

Table 3.6 - Comparative Analysis of Ransomware Detection Methods: Cuckoo Sandbox and Others

Techniques Used	Results	Contributions
Cuckoo Sandbox uses GBT, random forests.	Ransomware detection achieved high accuracy with 50 features.	Machine learning and dynamic features improve ransomware detection.
Analysis of threat classification and ransomware propagation methods.	Identifies ransomware strategies, cryptography, and forensic methods.	Focuses on memory and network malware detection.
Combine signatures, behavior analysis and backups.	This method protects files from Crypto-ransomware.	Multi-layered Defense blocks ransomware. attacks.

A study mainly focuses on dynamic analysis using Cuckoo Sandbox given in Table 3.6, ML techniques including random forests, neural networks and gradient boosted trees are used for detection. Comprehensive evaluation and comparison are restricted by the limited availability of feature datasets for ransomware detection tests and the frequent lack of broad feature descriptions in research, which complicate replication [11].

Ransomware spreading techniques cryptography modules, and threat classification are all covered in the analysis [25]. Cryptographic techniques, key sizes, and processing power present decryption issues. Furthermore, ransomware complicates data recovery by encrypting and overwriting files, making it challenging to recover infected data. Innovations in data recovery techniques and decryption techniques are needed to solve these problems.

In Table 3.6, several methods are proposed for ransomware protection, combining signature matching, behavior analysis, machine learning, frequent updates, and backups to defend against all variants and enable early detection [26]. However, when ransomware changes file extensions, it reduces file usability, highlighting the need for automatic recovery of original extensions. A key objective is to develop a network-based solution that enhances both usability and protection. The integration of these techniques can provide a more comprehensive defense by detecting ransomware early and minimizing damage through the automatic restoration of files and constant updates. Frequent backups also play a critical role in safeguarding data. The goal is to design a system that is simple for users to use that enhances security, enabling quick ransomware detection and improved data protection.

IV. Conclusion

The dynamic nature of ransomware demands sophisticated ML Techniques for efficient detection and prevention. Although they have made significant progress, current approaches such as feature extraction techniques, various classification models and static and dynamic assessments have certain drawbacks. Managing imbalanced datasets, adjusting to novel ransomware strains and resolving the high false-positive rates connected to traditional detection techniques are some of the main obstacles. To stay up with more complex ransomware tactics, future research must concentrate on expanding dataset diversity, upgrading proactive detection techniques and incorporating modern technologies like deep learning and quantum computing. By addressing these complicated concerns, we can build a more powerful and efficient ransomware protection solution that provides excellent defense against both existing and future threats.

References

[1] A. Alraizza and A. Algarni, "Ransomware Detection Using Machine Learning: A Survey," *Big Data and Cognitive Computing*, vol. 7, no. 3. Multidisciplinary Digital Publishing Institute (MDPI), Sep. 01, 2023. doi: 10.3390/bdcc7030143.

[2] 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," *Appl. Sci.*, vol. 12, no. 1, Jan. 2022, doi: 10.3390/app12010172.

[3] A. Kamboj, P. Kumar, A. K. Bairwa, and S. Joshi, "Detection of malware in downloaded files using various machine learning models," *Egypt. Informatics J.*, vol. 24, no. 1, pp. 81–94, Mar. 2023, doi: 10.1016/j.eij.2022.12.002.

[4] 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 (Switzerland)*, vol. 14, no. 1. MDPI, Jan. 01, 2022. doi: 10.3390/su14010008.

[5] B. Yamany, M. S. Elsayed, A. D. Jurcut, N. Abdelbaki, and M. A. Azer, "A New Scheme for Ransomware Classification and Clustering Using Static Features," *Electron.*, vol. 11, no. 20, Oct. 2022, doi: 10.3390/electronics11203307.

[6] R. Bold, H. Al-Khateeb, and N. Ersotelos, "Reducing False Negatives in Ransomware Detection: A Critical Evaluation of Machine Learning Algorithms," *Appl. Sci.*, vol. 12, no. 24, Dec. 2022, doi: 10.3390/app122412941.

[7] B. N. Chaithanya and S. H. Brahmananda, *Revolutionizing ransomware detection and criticality assessment : Multiclass hybrid machine learning and semantic similarity - based end2end solution*, vol. 83, no. 13. Springer US, 2024. doi: 10.1007/s11042-023-16946-x.

[8] Q. M. Yaseen, “The Effect of the Ransomware Dataset Age on the Detection Accuracy of Machine Learning Models,” pp. 1–23, 2023.

[9] N. Z. Gorment, A. Selamat, L. K. Cheng, and O. Krejcar, “Machine Learning Algorithm for Malware Detection : Taxonomy , Current Challenges , and Future Directions,” *IEEE Access*, vol. 11, no. December, pp. 141045–141089, 2023, doi: 10.1109/ACCESS.2023.3256979.

[10] A. Buriro, A. B. Buriro, T. Ahmad, S. Buriro, and S. Ullah, “MalwD&C: A Quick and Accurate Machine Learning-Based Approach for Malware Detection and Categorization,” *Appl. Sci.*, vol. 13, no. 4, Feb. 2023, doi: 10.3390/app13042508.

[11] J. A. Herrera-Silva and M. Hernández-Álvarez, “Dynamic Feature Dataset for Ransomware Detection Using Machine Learning Algorithms,” *Sensors*, vol. 23, no. 3, Feb. 2023, doi: 10.3390/s23031053.

[12] M. M. Khan, M. F. Hyder, S. M. Khan, J. Arshad, and M. M. Khan, “Ransomware prevention using moving target defense based approach,” *Concurr. Comput. Pract. Exp.*, vol. 35, no. 7, Mar. 2023, doi: 10.1002/cpe.7592.

[13] E. Berrueta, D. Morato, E. Magaña, and M. Izal, “Crypto-ransomware detection using machine learning models in file-sharing network scenarios with encrypted traffic,” *Expert Syst. Appl.*, vol. 209, Dec. 2022, doi: 10.1016/j.eswa.2022.118299.

[14] D. Tehrany and D. Abbas, “A new machine learning-based method for android malware detection on imbalanced dataset,” *Multimed. Tools Appl.*, pp. 24533–24554, 2021, doi: 10.1007/s11042-021-10647-z.

[15] O. Khalid *et al.*, “An Insight into the Machine-Learning-Based Fileless Malware Detection,” *Sensors*, vol. 23, no. 2, Jan. 2023, doi: 10.3390/s23020612.

[16] D. Smith, S. Khorsandroo, and K. Roy, “Machine Learning Algorithms and Frameworks in Ransomware Detection,” *IEEE Access*, vol. 10, no. October, pp. 117597–117610, 2022, doi: 10.1109/ACCESS.2022.3218779.

[17] A. Kamboj, P. Kumar, A. K. Bairwa, and S. Joshi, “Detection of malware in downloaded files using various machine learning models,” *Egypt. Informatics J.*, vol. 24, no. 1, pp. 81–94, 2023, doi: 10.1016/j.eij.2022.12.002.

[18] S. R. Davies, R. Macfarlane, and W. J. Buchanan, “Differential area analysis for ransomware attack detection within mixed file datasets,” *Comput. Secur.*, vol. 108, Sep. 2021, doi: 10.1016/j.cose.2021.102377.

[19] A. Alraizza, “Ransomware Detection Using Machine Learning : A Survey,” pp. 1–24, 2023.

[20] F. De Gaspari, D. Hitaj, G. Pagnotta, L. De Carli, and L. V. Mancini, “Evading behavioral classifiers: a comprehensive analysis on evading ransomware detection techniques,” *Neural Comput. Appl.*, vol. 34, no. 14, pp. 12077–12096, Jul. 2022, doi: 10.1007/s00521-022-07096-6.

[21] J. Moon, S. Kim, J. Song, and K. Kim, “Study on Machine Learning Techniques for Malware Classification and Detection,” *KSII Trans. Internet Inf. Syst.*, vol. 15, no. 12, pp. 4308–4325, Dec. 2021, doi: 10.3837/TIIS.2021.12.003.

[22] D. Smith, S. Khorsandroo, and K. Roy, “Machine Learning Algorithms and Frameworks in Ransomware Detection,” *IEEE Access*, vol. 10, pp. 117597–117610, 2022, doi: 10.1109/ACCESS.2022.3218779.

[23] D. T. Dehkordy and A. Rasoolzadegan, “A new machine learning-based method for android malware detection on imbalanced dataset,” *Multimed. Tools Appl.*, vol. 80, no. 16, pp. 24533–24554, Jul. 2021, doi: 10.1007/s11042-021-10647-z.

[24] I. Almomani *et al.*, “Android Ransomware Detection Based on a Hybrid Evolutionary Approach in the Context of Highly Imbalanced Data,” *IEEE Access*, vol. 9, pp. 57674–57691, 2021, doi: 10.1109/ACCESS.2021.3071450.

[25] N. Mangilal, C. Ankur, and S. Rijwan, “A review on spreading and Forensics Analysis of Windows- Based ransomware,” 2022.

[26] M. Mubashir, K. Muhammad, F. Hyder, and M. M. Khan, “Ransomware prevention using moving target defense based approach,” no. December 2022, pp. 1–15, 2023, doi: 10.1002/cpe.7592.

[27] N. M. Chayal, A. Saxena, and R. Khan, “A review on spreading and Forensics Analysis of Windows-Based ransomware,” *Annals of Data Science*. Springer Science and Business Media Deutschland GmbH, 2022. doi: 10.1007/s40745-022-00417-5.