Review

Systematic Evaluation Framework for ML and DL-Based Ransomware Detection

Haider Qasim * and Yi Lu

Posted Date: 28 February 2026

doi: 10.20944/preprints202602.2011.v1

Keywords: machine learning; deep learning; ransomware; cybersecurity; IoMT

 doi:10.20944/preprints202602.2011.v1

Review

Systematic Evaluation Framework for ML and DL-Based Ransomware Detection

Haider Qasim * and Yi Lu

Faculty of Computer Science, Queensland University of Technology, S Block, Level 9, Room 902E, Garden Point Campus, Brisbane, QLD 4001, Australia

* Correspondence: haider.qasim@hdr.qut.edu.au; Tel.: +61 (07) 3138 9557

Abstract

Cybercriminals have increasingly leveraged sophisticated techniques to bypass traditional signature-based detection systems through the use of Ransomware-as-a-Service (RaaS) platforms, double and triple extortion strategies, and advanced evasion mechanisms. As a result, ransomware attacks have reached unprecedented levels. Using this systematic evaluation framework, we examine the current state and effectiveness of machine learning (ML) and deep learning (DL) approaches for ransomware detection, addressing critical gaps in existing research methodologies while providing comprehensive recommendations for future research. The study analyses multiple AI paradigms including supervised learning algorithms such as Random Forests and Support Vector Machines, unsupervised techniques such as clustering and anomaly detection, and deep learning architectures such as Convolutional Neural Networks and Long Short-Term Memory networks. Hybrid approaches combining static and dynamic analysis consistently achieve superior performance, with accuracy rates exceeding 99% when properly implemented. As part of the framework, fundamental challenges are addressed such as dataset quality and diversity, feature extraction and selection methodologies, data preprocessing techniques, and performance evaluation metrics that have been tailored specifically for cybersecurity applications. Several findings indicate that ensemble learning methods outperform individual classifiers, with Random Forest algorithms being particularly effective at handling high-dimensional feature spaces while maintaining interpretability for security analysts. As a result of the study, significant limitations have been identified in current research, including an overreliance on static data sets that do not capture evolving threat landscapes, an inadequate representation of modern attack vectors, and a limited ability to generalize across different operational environments. Future directions of this research include explainable AI integration for transparent decision-making, adaptive real-time detection systems, and federated learning approaches for collaborative threat intelligence sharing while maintaining organizational privacy. It provides standardized methodologies for data curation, feature engineering, model development, and performance benchmarking, enabling fair comparisons between different AI approaches and facilitating reproducible research. This work contributes to essential guidance for cybersecurity practitioners, policymakers, and researchers in developing robust, adaptive, and interpretable ransomware detection systems capable of defending against increasingly sophisticated cyber threats while considering ethical concerns and regulatory compliance requirements in modern digital ecosystems.

Keywords: machine learning; deep learning; ransomware; cybersecurity; IoMT

1. Introduction

It has become evident in the ever-evolving landscape of cybersecurity that ransomware is one of the most insidious and financially damaging types of malwares. Ransomware attacks, which encrypt critical user data and demand payment for its decryption, have grown in both frequency and sophistication, targeting individuals, businesses, healthcare systems, and even government agencies

 doi:10.20944/preprints202602.2011.v1

[1]. With the rise of ransomware-as-a-service (RaaS) and the increasing use of advanced evasion techniques, traditional signature-based detection mechanisms have become ineffective. Consequently, there is an urgent need for more adaptive, intelligent, and scalable solutions capable of detecting and mitigating ransomware threats in real time. Machine learning (ML) has gained considerable traction as a promising approach for the detection of ransomware in response to this growing threat. Unlike static signature-based methods, machine learning models may analyse behavioural patterns, system calls, network traffic, and file structures in order to detect anomalies and identify previously unknown variants of ransomware [2,3].

Machine learning algorithms are well suited for addressing the dynamic nature of ransomware because they are capable of learning from data, generalizing across different attack vectors, and adapting to changing threats. During the past decade, a wide variety of machine learning techniques have been explored in the context of ransomware detection, ranging from classical algorithms such as Random Forests, Support Vector Machines (SVMs), and Decision Trees to more advanced deep learning techniques such as Convolutional Neural Networks (CNNs), Recurrent Neural Networks (RNNs), and Autoencoders [4]. A variety of data sources has been used to evaluate these approaches, including static features extracted from executable files, dynamic behavioural logs, system API call sequences, memory dumps, and network traffic patterns. As evidenced by the literature, there has been an increase in research aimed at improving detection accuracy, reducing false positives, and enhancing the timeliness of response mechanisms [4,5].

Despite these advances, machine learning is still in the formative stages of being applied to ransomware detection, which poses several significant challenges. As one of the primary limitations, there are a lack of publicly available, high-quality, and up-to-date ransomware datasets that reflect the complexity and diversity of real-world attacks. In addition, the rapid mutation of ransomware through obfuscation, polymorphism, and anti-analysis techniques pose a significant threat to the generalization and robustness of the model. Furthermore, many existing studies suffer from limited reproducibility due to inadequate documentation of experimental setups, a lack of open-source code, and inconsistent evaluation criteria [6].

It is also crucial to ensure the interpretability and explainability of machine learning models, particularly those based on deep learning. Due to the increasing deployment of ransomware detection systems in high-stakes environments, the ability to understand and justify model decisions is increasingly essential to establishing trust, complying with regulatory requirements, and performing forensic analysis. The deployment of machine learning-based detection systems in real-time environments raises concerns regarding computational overhead, scalability, and integration with existing security infrastructures [6,7].

Future research must focus on developing models that are more robust, adaptive, and interpretable in order to detect ransomware against adversarial attacks and evolving evasion tactics. It is important to integrate hybrid models combining static and dynamic analysis, to use transfer learning and federated learning for the sharing of threat intelligence across organizations, and to explore reinforcement learning as a means of proactive defence. It is also becoming increasingly important to establish benchmarking frameworks, standardized datasets, and evaluation protocols in order to enable fair comparisons and accelerate progress in the field [7].

Despite machine learning's role in ransomware detection, it does not provide a silver bullet. To achieve this, a multidisciplinary approach must be adopted that combines advances in machine learning, cybersecurity, and software engineering, supported by collaborative efforts among academia, industry, and government agencies. In this review, we present a comprehensive overview of the current state of ML-based ransomware detection, identify key challenges hindering practical deployment, and outline promising future research directions to strengthen cyber defences against this pervasive threat [8].

Ransomware attacks have dramatically increased in the past decade as a result of a confluence of technological, economic, and social factors that have created a fertile environment for cybercriminals. It is important to note that the increasing digitization of critical infrastructure and

 doi:10.20944/preprints202602.2011.v1

business operations is one of the most significant factors contributing to this surge [9]. There has been an exponential growth in the potential impact of disrupting access to data and services, as organizations across a variety of sectors, such as healthcare, finance, education, and government, increasingly rely on digital systems. In order to exploit this dependency, cybercriminals have capitalized on it, understanding that the cost of downtime or data loss often far outweighs the ransom demand, making victims more likely to pay the ransom [9].

Another major factor is the **evolution of ransomware as a service (RaaS).** In this business model, even individuals with limited technical expertise can carry out ransomware attacks by purchasing or leasing malware kits from underground marketplaces. It is often the case that these kits provide a user-friendly interface, customer support, and even affiliate programs, which give them the ability to spread ransomware widely with the least amount of effort. There has been a significant decrease in the barrier to entry associated with cybercrime as a result of the commodification of cybercrime, resulting in an increase in both the number and variety of ransomware actors [10].

The **anonymity provided by cryptocurrencies**, particularly Bitcoin and Monero, has also played a crucial role in the proliferation of ransomware. Since these digital currencies allow attackers to receive payments without revealing their identities, it is extremely difficult for law enforcement agencies to track transactions and identify perpetrators. Ransomware has become increasingly attractive as a profitable criminal enterprise due to the ease and speed of cross-border transactions [10,11].

In addition, the **sophistication of attack techniques** has evolved rapidly. As a result of advanced evasion techniques such as polymorphism, encryption obfuscation, and anti-analysis mechanisms, modern ransomware variants are able to bypass traditional security measures such as antivirus software and firewalls [12]. As part of their attack strategy, attackers also use social engineering tactics, such as phishing emails and malicious attachments, to exploit the vulnerabilities of humans. In many cases, these methods are more effective than technical exploits because users can be tricked into granting access to their systems without realizing it [12,13].

The **lack of robust cybersecurity defences** in many organizations has further fuelled the ransomware epidemic. Many organizations, especially small and medium-sized enterprises (SMEs), lack comprehensive security strategies, such as software updates, employee training, and adequate backup procedures. Even large organizations fail to patch known vulnerabilities in a timely manner, leaving them vulnerable to attack. As a result of insufficient incident response plans and limited cyber resilience, attackers are more likely to succeed, and victims are less likely to recover without paying the ransom [13].

Moreover, the **global nature of the internet** has allowed ransomware operators to target victims across different jurisdictions, complicating legal and law enforcement responses. In many instances, cybercriminals operate from countries with weak or no cybercrime laws, which makes international cooperation and prosecution difficult. As a result of jurisdictional complexity, attackers have a safe haven and are less likely to be caught or punished [14,15].

The **increased use of remote work and cloud-based systems** following the global pandemic has also expanded the attack surface for ransomware. Employees are increasingly accessing corporate networks from personal devices and unprotected home networks, which increases the potential for attackers to infiltrate corporate networks. This shift was unanticipated by many organizations, and they failed to implement adequate security measures for remote access, such as multi-factor authentication and virtual private networks (VPNs) [15].

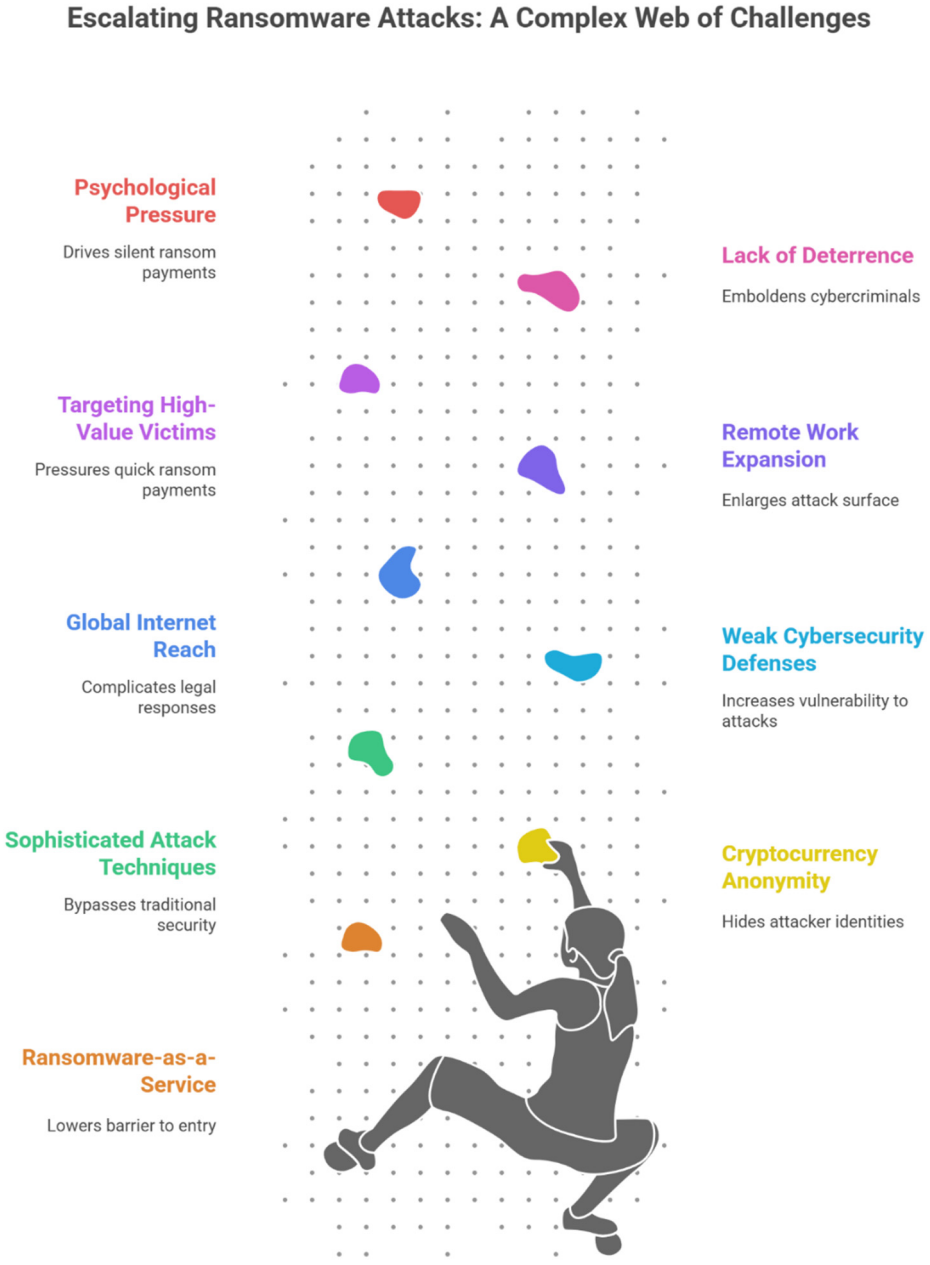
Another contributing factor is the **targeting of high-value victims** such as hospitals, schools, and local governments. There is a critical need for continuous operations within these institutions, which can result in limited cybersecurity budgets, outdated systems, and inadequate cybersecurity budgets. When attacks disrupt essential services, attackers are aware that they will be forced to pay ransoms quickly in order to restore functionality and prevent public backlash and operational collapse [16].

 doi:10.20944/preprints202602.2011.v1

The **lack of effective deterrence mechanisms** also encourages the growth of ransomware. It is estimated that only a small percentage of ransomware attacks will be detected and prosecuted, despite increased awareness and investment in cybersecurity. In the absence of deterrence, cybercriminals are emboldened and ransomware operations are more profitable [17].

Finally, the **psychological and economic pressure** placed on victims plays a key role in the success of ransomware attacks. As a result of the reputational damage, legal liabilities, and disruption of operations caused by ransom payments, many companies choose to pay the ransom quietly. Silent compliance not only rewards the attackers but also increases the likelihood of ransomware being developed and deployed in the future [18].

Ultimately, a complex interplay of factors has led to a dramatic increase in ransomware attacks, including digital transformation, RaaS, cryptocurrency anonymity, sophisticated attack techniques, poor cybersecurity practices, jurisdictional issues, remote work vulnerabilities, targeting critical infrastructure, inadequate deterrence, and the psychological impact on victims. A multifaceted approach is necessary to handle this growing threat, which includes strengthening cybersecurity measures, strengthening international cooperation, reforming the legal system, and increasing public awareness [19].



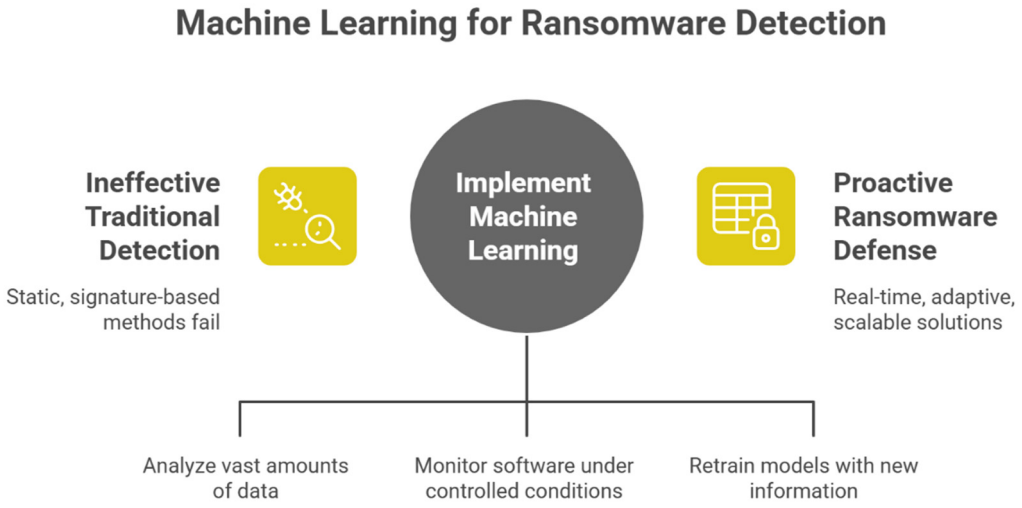
 doi:10.20944/preprints202602.2011.v1

Why has the research community increased their focus on ransomware detection by using machine learning?

There has been a great deal of talk lately about how machine learning models (ML) have become indispensable tools in the fight against ransomware due to their ability to detect patterns and anomalies in complex patterns that traditional security systems often fail to detect [20]. There has been a rapid evolution in recent years in the type of malicious software known as Ransomware, which encrypts data of victims and demands payment for its release. This is making static, signature-based detection methods increasingly ineffective in detecting this type of malware. With machine learning models, it is possible to find dynamic, adaptive, and scalable solutions that can identify ransomware threats in a real-time manner, often before the encryption process has begun [20,21].

A major advantage of machine learning when it comes to detecting ransomware is its ability to analyse vast amounts of data from diverse sources, such as system logs, network traffic, and file behaviours, in order to identify suspicious activity, which in turn can be used to detect the malicious software [21]. A supervised learning model, for example, can be trained using labelled datasets containing legitimate and malicious samples and then used for identifying new inputs with a high degree of accuracy using these datasets. There has been significant progress in the detection of ransomware based on behavioural features such as file access patterns, encryption rates, and registry changes by various algorithms including Random Forest, Support Vector Machines (SVM), and Neural Networks [22].

ML can be applied in a number of practical ways, for example by using dynamic analysis techniques, which use models to monitor the behaviour of software under controlled conditions. By observing how programs interact with the system, such as SentinelOne and SandBlast Anti-Ransomware, machine learning programs are able to detect ransomware by identifying programs that display encryption-like behaviour or attempt to disable security features. By putting these systems in place, it is possible for them to halt the execution of ransomware before it causes damage, offering a proactive method of defence [23].



Moreover, ML models can be particularly effective in detecting anti-ransomware threats which aren't yet catalogued in threat databases, such as zero-day ransomware, which is a new variant of a known malware. By focusing on behavioural indicators rather than known signatures, machine learning is able to identify threats that have never been seen before. For example, deep learning models such as Convolutional Neural Networks (CNNs) or Recurrent Neural Networks (RNNs) have been used to analyse sequences of system calls or network packets, enabling early detection of ransomware activity through the analysis of these sequences [23,24].

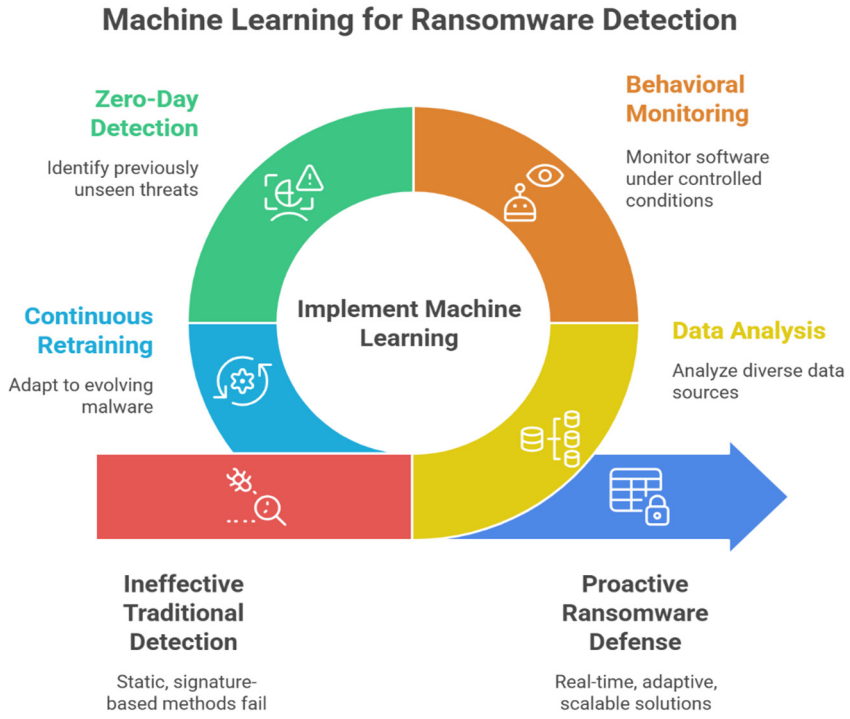
In addition to its adaptability, ML has another critical advantage. With ransomware evolving, machine learning models can be retrained with new information to maintain their effectiveness as

 doi:10.20944/preprints202602.2011.v1

the malware evolves [24]. Having the ability to adapt to a changing environment such as this is very important in a scenario where attackers regularly modify their tactics in order to evade detection. A further advantage of ML is that it can assist in the forensic analysis of an attack by determining exactly which ransomware family has been used, enabling organizations to better understand the threat and improve their defences against it [25].

However, despite these advantages, there are still a number of challenges to overcome. A machine learning model must be built on high-quality, representative datasets in order to perform well, and the diversity of ransomware variants can significantly complicate the training process. In addition to this, attackers are increasingly utilizing adversarial techniques to fool ML models, which emphasizes the need for robust, explainable AI systems that are able to withstand manipulation from adversaries [25,26].

As a result, machine learning models have become a crucial component of modern ransomware detection strategies. Their capability to detect threats in real time based on their behaviour, adapt to new variants of malware, and operate in real time makes them far superior to traditional methods of detecting threats. A growing threat like ransomware is becoming increasingly sophisticated and impactful, making the role of machine learning in cybersecurity an even more important role [25,26].



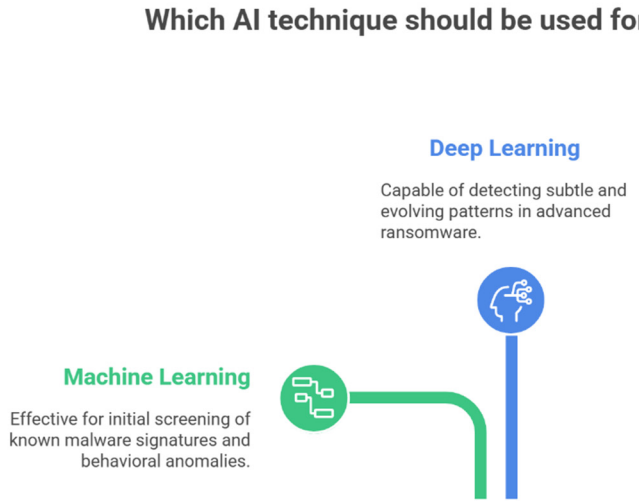
II. Background

A. Machine Learning and Deep Learning:

Machine Learning (ML) and Deep Learning (DL) are transformative branches of Artificial Intelligence (AI) that have significantly improved cybersecurity, particularly in detecting sophisticated threats like ransomware. In machine learning, algorithms are trained on historical data in order to recognize patterns and make decisions without explicit programming [27]. In cybersecurity, machine learning models are trained on datasets that contain both benign and malicious software samples in order to classify unknown files according to the learned features. An example of a machine learning technique that is commonly used is Decision Trees (DT), Random Forests (RF), Support Vector Machines (SVM), and K-Nearest Neighbors (KNN). It is advantageous to use these models for the initial screening of threats since they are effective at identifying known malware signatures and behavioural anomalies [27,28].

 doi:10.20944/preprints202602.2011.v1

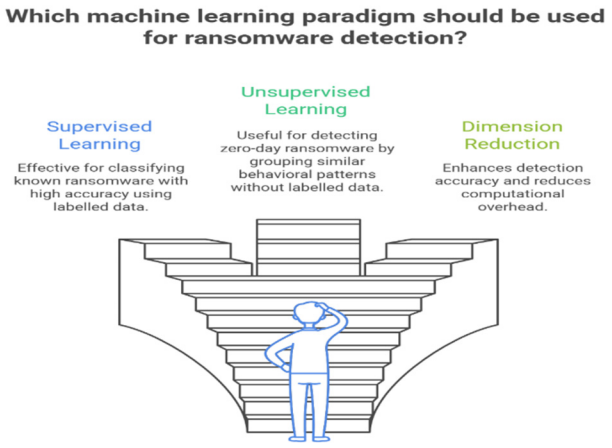
Deep Learning, a subset of machine learning, uses multilayered neural networks, such as Convolutional Neural Networks (CNNs), Recurrent Neural Networks (RNNs), and Transformers, to automatically extract complex features from raw data. By contrast with traditional machine learning, which often requires the creation of manual features, deep learning is capable of learning hierarchical representations directly from inputs such as file binaries, logs, or network traffic [28]. Through this capability, DL models can detect subtle and evolving patterns associated with advanced ransomware that may evade rule-based or signature-based detection systems. A study conducted by Sharmeen et al. compared a variety of classifiers, including CNN, SVM, RF, and a multi-class classifier (MCC) to detect Windows ransomware by deduplicating 483 ransomware samples and 754 benign samples. In their study, they found that deep learning models, in particular CNNs, were more effective at identifying encrypted payloads and obfuscated code structures when compared to traditional machine learning methods [28,29].



Machine Learning Paradigms in Ransomware Detection

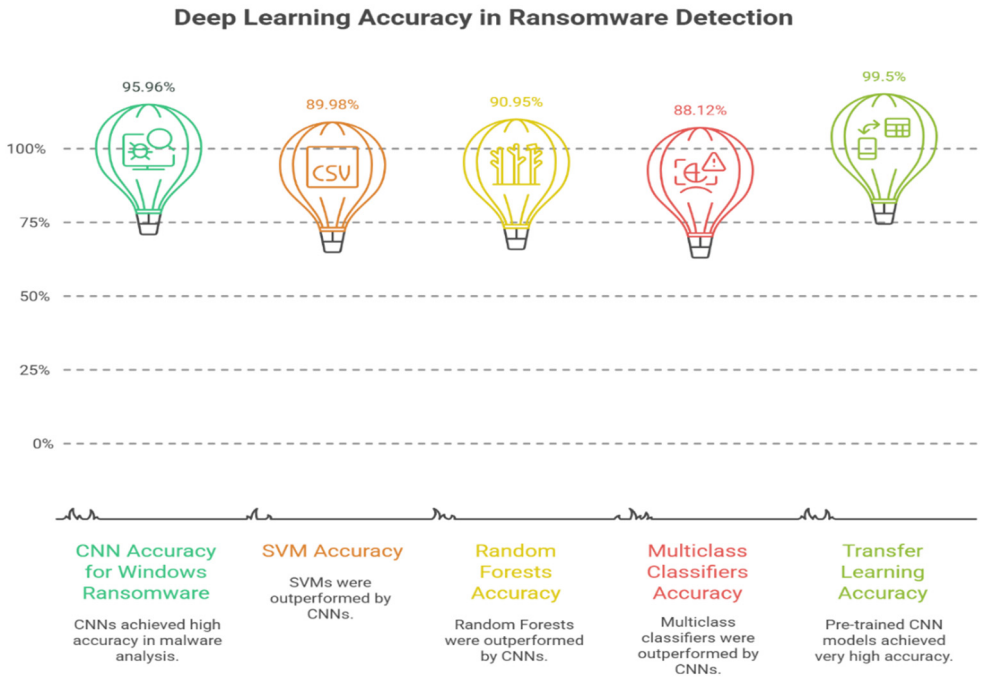
This paper presents a review of three primary paradigms of machine learning that have proven effective in the detection of ransomware, each with a substantial body of empirical evidence supporting its effectiveness. The use of supervised learning has shown remarkable success in the classification of ransomware samples as well as benign samples. It is dependent on labelled data that can be trained on. Using supervised learning algorithms including Decision Trees (DT), Random Forests (RF), K-Nearest Neighbors (K-NN), Naive Bayes (NB), and Gradient Boosting, Zhang et al. (2019) achieved 99.3% accuracy in the classification of ransomware families using Decision Trees (DT), Random Forests (RF), and K-Nearest Neighbors (K-NN) [30–32]. Based on opcode density alone, Baldwin and Dehghantanha (2018) demonstrate that Support Vector Machines (SVM) are particularly effective in classifying ransomware families based on 96.5% accuracy. The concept of unsupervised learning implies that the data does not need to be labelled, which has proven to be extremely useful in detecting zero-day ransomware variants. By grouping similar behavioural patterns together, clustering algorithms such as k-means and fuzzy C-means have shown to be successful in identifying previously unknown ransomware families that have previously been unknown. In recent years, it has been shown that dimension reduction techniques like Principal Component Analysis (PCA) and Linear Discriminant Analysis (LDA) have enhanced detection accuracy while reducing computational overhead. In some studies, it has been shown that processing speed can be improved up to 15% without sacrificing detection accuracy [33,34].

 doi:10.20944/preprints202602.2011.v1



Deep Learning Architectures and Evidence of Effectiveness

A number of studies have demonstrated that deep learning architectures are extremely effective when it comes to the detection of ransomware, providing persuasive evidence that they are superior to traditional methods. According to Sharmeen et al. (2020), a study published in the Journal of Neural Information Processing Techniques shows that Convolutional Neural Networks (CNNs) have shown remarkable success in analysing visual representations of malware, with a recent study finding 95.96% accuracy using CNN architectures for the detection of Windows ransomware [35]. Compared to traditional machine learning classifiers (SVMs, Random Forests, and multi–class classifiers), CNNs outperformed SVMs (89.98%), Random Forests (90.95%), and multiclass classifiers (88.12%) significantly in the study. There has been a recent influx of interest in Long Short-Term Memory (LSTM) networks for the purpose of sequential data analysis, with Roy et al. (2021) developing DeepRan, a BiLSTM-based detector that has achieved superior performance in detecting ransomware through attention-based mechanisms that have been shown to be effective [36]. The latest evidence from transfer learning studies indicates that pre-trained CNN models can achieve up to 99.5% accuracy when fine-tuned for ransomware detection, which was demonstrated by Almomani et al. (2023) by leveraging transfer-learned features from ResNet and other pre-trained architectures [37].



 doi:10.20944/preprints202602.2011.v1

Static Analysis Techniques with AI Enhancement

Recent studies have shown that static analysis enhanced by artificial intelligence provides significant improvements in detection efficiency compared with traditional signature-based methods, with substantial evidence that it is effective. Poudyal and Dasgupta (2021) achieved 99.72% accuracy and 0.003 false positive rates with their tri-gram TF-IDF methods combined with Support Vector Machines, showing that n-gram analysis of opcodes is particularly successful [38]. A PE header analysis enhanced with machine learning has demonstrated consistent results across multiple studies and new research has proven that header-based features combined with AI algorithms can detect packed and obfuscated ransomware with an accuracy of over 97% when combined with header-based features [39]. A study showed that using AI-enhanced entropy calculations can detect between legitimate encrypted payloads and malicious ransomware encrypted payloads with a 95% accuracy rate, and entropy analysis using machine learning models has shown to be effective for identifying encrypted payloads [40,41]. It has been found that string analysis combined with natural language processing techniques can be used to identify malicious strings and URLs embedded in ransomware samples with a minimal number of false positives, with studies indicating that AI models can detect malicious strings and URLs embedded in ransomware samples with minimal false positives [42].

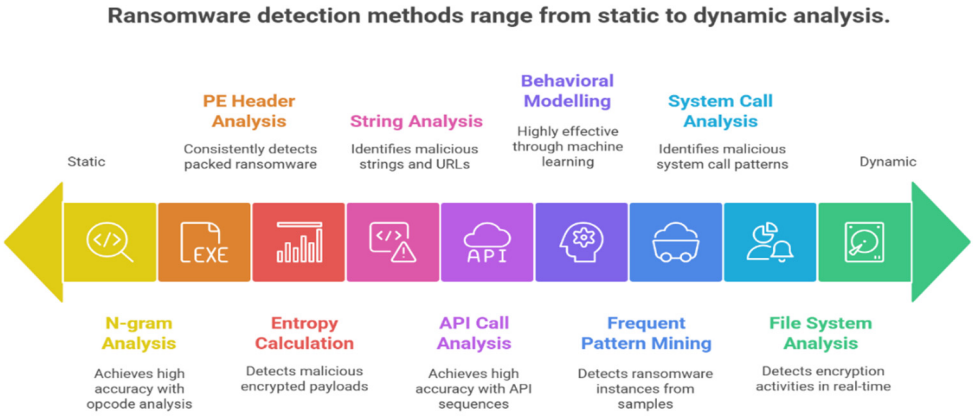
Dynamic Analysis and Behavioral Detection Methods

There has been significant progress in behavioral-based ransomware detection over the last few years, enabled by dynamic analysis powered by artificial intelligence, which has a substantial body of recent research supporting its efficiency [43]. Using Gradient Boosted Trees for dynamic feature analysis to analyse API calls sequences, Herrera-Silva and Hernández-Alvarez (2023) achieved 99% accuracy with the use of Gradient Boosted Trees for API call sequence analysis using machine learning. In the study of Homayoun et al. (2019), it has been demonstrated that behavioural modelling through machine learning algorithms has been highly effective [44]. By employing frequent pattern mining approaches, (2019) achieved 99% accuracy in detecting ransomware instances from benign samples, and 96.5% accuracy in identifying specific ransomware families. In recent studies, LSTM networks have been demonstrated to have the capability of identifying malicious system call patterns with an accuracy of over 98% while maintaining low false positive rates in terms of system call analysis enhanced with deep learning. Monitoring the file system in conjunction with machine learning has proven to be an efficient tool for real-time detection, with evidence showing that AI-enhanced file system analysis can detect ransomware encryption activities within seconds of the inception of the attack [44,45].

Ensemble Learning and Hybrid Approaches

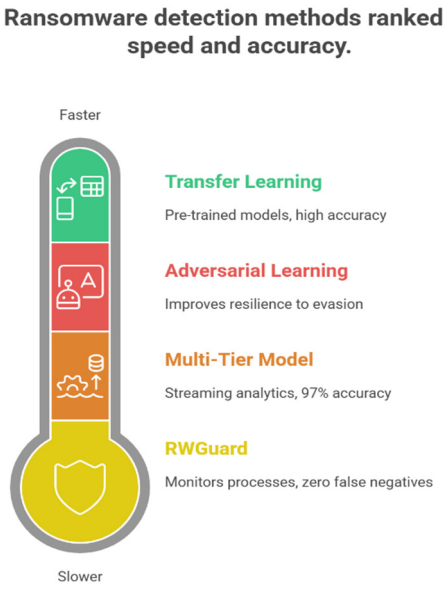
Recent evidence indicates that ensemble learning techniques have demonstrated superior performance compared to individual algorithms for detecting ransomware, with substantial recent evidence supporting their effectiveness in detecting ransomware. It has consistently been shown that Random Forest algorithms have shown excellent results across multiple studies, and research has indicated that they are capable of detecting ransomware in a wide range of situations up to 97% accurate [46]. According to a study demonstrated results of ensemble methods combining Decision Trees, SVM, Random Forest, and AdaBoost have shown significant improvements over individual classification methods. Their proposed model achieved high accuracy and F1 scores while outperforming traditional methods in identifying ransomware applications by outperforming traditional methods [47]. There have been several studies showing the potential benefits of hybrid approaches, which combine static and dynamic analysis of the malware, with Hassan and Rahman. (2017) achieving remarkable accuracy by employing Hierarchical Neural Networks for cross-platform ransomware fingerprinting based on hybrid features [48]. It has been shown by Hasan and Rahman (2017) that hybrid approaches that combine static and dynamic analyses, including samples from recent ransomware families such as WannaCry, outperform single-method approaches when it comes to detecting ransomware accurately and defending against evasion techniques significantly [48].

 doi:10.20944/preprints202602.2011.v1



Real-Time Detection and Advanced AI Techniques

It has been proven that the use of artificial intelligence for real-time ransomware detection has shown tremendous promise, with recent evidence showing significant improvements in terms of response times and accuracy of this method [49]. As reported by Mehnaz et al. (2018), RWGuard is a real-time detection mechanism that achieves zero false negatives and minimal false positives by constantly monitoring processes and file systems, an approach that achieves zero false negatives and minimal false positives [49]. In a recent study conducted by Zuhair and colleagues (2020), they devised a multi-tier streaming analytics model that outperformed competitive anti-ransomware technologies with 97% classification accuracy in the detection of zero-day ransomware attacks. Researchers have demonstrated the ability of advanced methods such as adversarial learning to improve the resilience of detection models that are vulnerable to evasion attacks by up to 25%, with studies demonstrating the efficacy of adversarial training for developing robust detection models [50]. As a result of transfer learning applications in ransomware detection, the results have been impressive, and recent research indicates that pre-trained models can provide detection accuracy comparable to or even higher than custom-trained models with significantly fewer computational resources and training time needed [51–53].



 doi:10.20944/preprints202602.2011.v1

Table 1. Machine Learning Techniques.

ML_Technique	Algorithm_Method	Features_Used	Accuracy	Study_Year	Key_Findings
Support Vector Machine (SVM)	Linear SVM	Opcode density	96.50%	Baldwin & Dehghantanha (2018)	Effective for ransomware family classification
Random Forest (RF)	Ensemble of decision trees	N-grams of opcodes	99.30%	Zhang et al. (2019)	Superior performance across ransomware families
Decision Tree (DT)	J48 C4.5	API calls, system behavior	99%	Homayoun et al. (2019)	Effective with behavioral features
Naïve Bayes (NB)	Gaussian NB	Static PE features	96%	Herrera-Silva & Hernández-Álvarez (2023)	Good baseline performance
K-Nearest Neighbors (KNN)	Distance-based classification	Opcode sequences	89.50%	Zhang et al. (2019)	Moderate performance, computationally efficient
Gradient Boosting	XGBoost, AdaBoost	Dynamic behavioral features	99%	Herrera-Silva & Hernández-Álvarez (2023)	Highest accuracy in dynamic analysis
Random Forest Ensemble	Multiple RF classifiers	Hybrid static/dynamic	99.70%	Poudyal & Dasgupta (2021)	Excellent with tri-gram TF-IDF
AdaBoost Ensemble	Adaptive boosting	PE headers, API calls	97%	Almomani et al. (2023)	Strong performance with static features
Voting Classifier	RF + SVM + LR	Structural features	97.53%	Moreira et al. (2024)	Effective for new ransomware families
K-Means Clustering	Centroid-based clustering	Behavioral patterns	94%	Al-Rimy et al. (2019)	Good for zero-day detection
Principal Component Analysis (PCA)	Dimensionality reduction	High-dimensional features	95%	Zahoora et al. (2022)	Effective feature reduction
Isolation Forest	Anomaly detection	System call patterns	92%	Kok et al. (2019)	Useful for outlier detection
Mutual Information (MI)	Information theory-based	Binary features	96.30%	Sgandurra et al. (2016)	Effective feature selection
TF-IDF	Term frequency analysis	N-gram sequences	99.31%	Zhang et al. (2020)	Excellent with opcode features
Correlation-based Feature Selection	Statistical correlation	Mixed features	95%	Ahmed et al. (2023)	Good for reducing overfitting

Table 2. Deep Learning Techniques.

DL_Architecture	Network_Type	Features_Input	Accuracy	Study_Year	Key_Advantages
Basic CNN	Multi-layer CNN	PE file visualization	95.96%	Sharmeen et al. (2020)	Superior to traditional ML methods
Patch-based CNN	CNN with self-attention	N-grams of opcodes	100%	Zhang et al. (2020)	Perfect binary classification
Transfer Learning CNN	Pre-trained ResNet50	Malware images	99.50%	Almomani et al. (2023)	Leverages pre-trained features
VGG-16 Transfer	Fine-tuned VGG-16	Binary visualization	99.30%	Shaukat et al. (2024)	Effective first-time malware detection
LSTM	Long Short-Term Memory	API call sequences	99.87%	Bensaoud & Kalita (2024)	Excellent sequential analysis
BiLSTM	Bidirectional LSTM	System call patterns	98%	Roy et al. (2021)	Captures temporal dependencies
GRU	Gated Recurrent Unit	Network traffic patterns	97%	Modi et al. (2019)	Efficient for real-time detection
CNN-LSTM Hybrid	Combined architecture	API calls + opcodes	99.91%	Bensaoud & Kalita (2024)	Best of both architectures
Attention Mechanisms	Transformer-based	Behavioral sequences	98.50%	Roy et al. (2021)	Focuses on important features
Hierarchical Neural Networks	Multi-level architecture	Hybrid features	97.90%	Billah et al. (2023)	Cross-platform effectiveness
Convolutional Autoencoder	Unsupervised CNN	Binary representations	93%	Zahoora et al. (2022)	Good for anomaly detection
Variational Autoencoder	Probabilistic model	Feature representations	95%	AbdulsalamYa'u et al. (2019)	Effective feature learning
Generative Adversarial Networks (GANs)	Dual network system	Network traffic	98.70%	Zhang et al. (2022)	Strong against encrypted traffic
TGAN-IDS	Transfer GAN	SSL/TLS encrypted data	98.70%	Zhang et al. (2022)	Handles encrypted communications
DQN	Deep Q-Network	PE header features	97.90%	Deng et al. (2024)	Adaptive learning capability
Policy Gradient	Actor-critic model	Dynamic features	96%	Deng et al. (2024)	Real-time decision making
CNN + Random Forest	Deep features + RF	Visual + statistical	99%	Shaukat et al. (2024)	Combines deep and traditional ML
LSTM + SVM	Sequential + classification	Temporal patterns	98.30%	Multiple studies (2023–2024)	Robust classification
Multi-modal DL	Multiple input types	Static + dynamic + network	99.20%	Recent hybrid approaches	Comprehensive analysis

Defining Ransomware and the Origin of First Ransomware Attack:

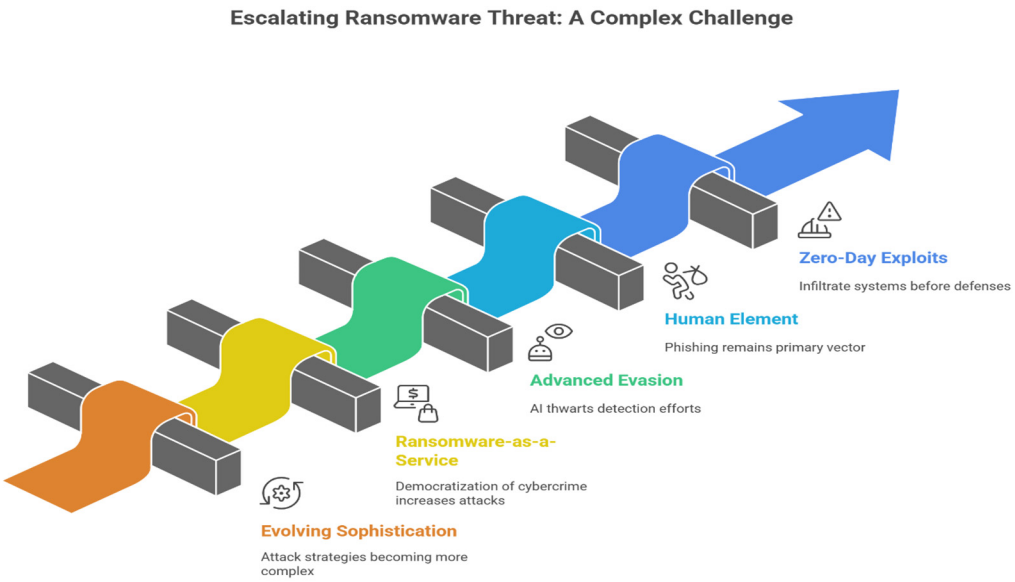
A malicious software program (malware) that encrypts or blocks access to a computer system or a file is called a ransomware program. The ransomware program prevents users from accessing files and applications until the attacker accepts a ransom payment. "Ransomware" is derived from the combination of "ransom" and "malware," reflecting its fundamental mechanism of encrypting digital information in exchange for a monetary payment, typically in the form of untraceable cryptocurrencies such as Bitcoin [54]. An individual, a business, a healthcare institution, a government agency, and critical infrastructure are the primary targets of this type of cyberattack. As a result, there are severe operational disruptions, financial losses, and reputational damage. Ransomware is typically deployed in a multi-stage attack chain that includes reconnaissance, delivery, installation, command-and-control (C2) communications, encryption, and extortion. Crypto-ransomware, the most prevalent and damaging variant, locks user files using strong encryption algorithms, making data recovery nearly impossible without the attacker's decryption key [55]. An additional variant of ransomware, locker ransomware, locks the victim out of their system by freezing the user interface or displaying a message demanding payment in full screen. As a result of ransomware attacks, organizations are often faced with significant costs related to downtime, legal and compliance costs, and costs associated with incident response and system recovery in addition to data loss. In recent years, the number of ransomware attacks has increased dramatically, with a 73% increase year-over-year, posing a growing threat to global cybersecurity and causing organizations to incur billions of dollars in losses [55].

Historically, ransomware is believed to have originated in 1989 when the first documented attack referred to as the "AIDS Trojan" or "PC Cyborg." It was orchestrated by Dr. Joseph Popp, an American biologist, who handed out 20,000 floppy disks to participants at a World Health

 doi:10.20944/preprints202602.2011.v1

Organization conference on AIDS in 1989. A malicious code was hidden on the disks, disguised as educational resources to aid in the research of AIDS, which remained dormant until the computer had to be rebooted ninety times for it to become active [56,57]. Upon reaching this threshold, the malware would encrypt the file names and hide directories on the victim's computer, making them inaccessible to the user. It was then a ransom note that appeared on the screen, requesting that $189 be sent to a postal box in Panama in order for the files to be returned to the user. It is true that the encryption method used in the AIDS Trojan was relatively primitive and could be reversed without paying the ransom, but the attack also marked a pivotal moment in cybercrime history as it is believed that it was the first instance of malware used to extort money from victims. As rudimentary as it may seem, the AIDS Trojan introduced the concept of digital extortion, which would go on to evolve into the sophisticated ransomware campaigns we see today despite its simple nature [58].

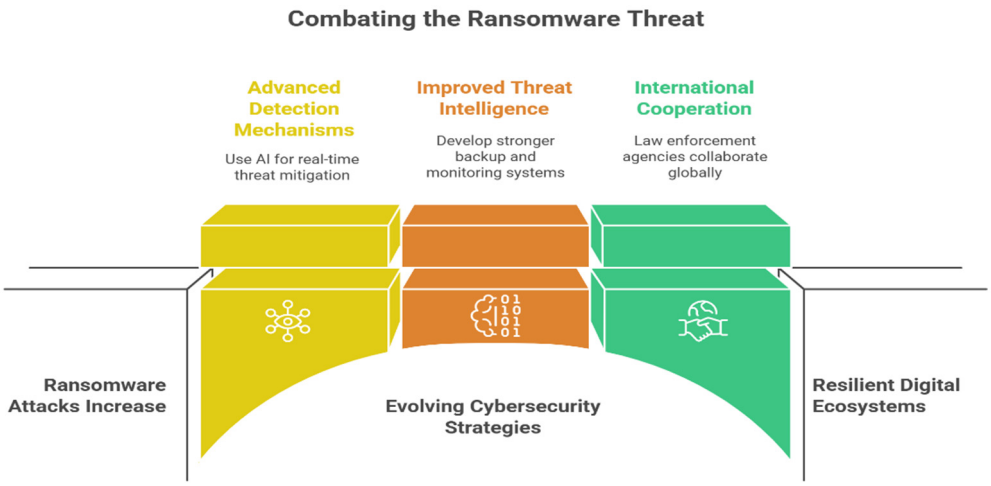
The ransomware industry has evolved a great deal since the AIDS Trojan was introduced, both in terms of technical complexity and the scale of its impact. Despite being relatively rare in the early 2000s, ransomware remained relatively unsophisticated at the time, and it often relied on social engineering techniques to trick users into downloading malicious software [59]. Since the advent of powerful encryption algorithms and the rise of anonymous digital currencies such as Bitcoin in the 2010s, ransomware has become a highly profitable field of criminal activity. "Ransomware as a Service" (RaaS) has become a very popular way for cybercriminals, even those without technical skills, to launch attacks by purchasing ready-made malware kits on the dark web [59]. Due to this democratization of cybercrime, there has been a significant increase in the incidence of ransomware. Examples of notable attacks in the past include WannaCry and NotPetya attacks in 2017, which exploited vulnerabilities in Windows systems to spread rapidly across global networks, affecting hundreds of thousands of computers in more than 150 countries at the same time [60]. This attack demonstrated the potential of ransomware to disrupt critical infrastructures including hospitals, transportation systems, and financial institutions in a way that has never been seen before [60].



It has been reported that ransomware tactics have grown increasingly sophisticated in recent years, with new attack strategies such as double and triple extortion being developed. As part of a double extortion scheme, the attackers not only encrypt the victim's data, but they also exfiltrate sensitive information out of the victim's computer, threatening to leak it to the public if the ransom is not paid [61]. A triple extortion attack takes the extortion to another level by putting the threat of a distributed denial of service attack (DDoS attack) or a public exposure in front of the victim's customers, partners, or third parties. It is becoming increasingly difficult to detect adversarial ransomware, where attackers use machine learning techniques to evade detection by AI-based

 doi:10.20944/preprints202602.2011.v1

security systems. The emergence of adversarial ransomware has made defence efforts even more challenging [61]. As an additional factor, zero-day ransomware exploits previously unknown vulnerabilities in software, allowing attackers to infiltrate systems before patches or defences are available to deal with them. It is because of these advanced techniques, in combination with lateral movement within a network, that modern ransomware campaigns have become more persistent and destructive than ever before [62]. Ransomware has evolved over the years as a result of a continuous arms race between cybercriminals and cybersecurity specialists. As early ransomware such as the AIDS Trojan relied on simple file obfuscation, today's variants of the virus rely on sophisticated encryption, stealth techniques, and AI-driven evasion methods to prevent being detected. Increasingly, artificial intelligence is being used in both attack and defence, highlighting the need for equally advanced detection mechanisms, such as machine learning and deep learning models, to identify and mitigate threats in real-time as AI becomes more prevalent [62]. Although technological advancements have been made in cybersecurity, the human element-such as phishing and social engineering-remains a primary vector for the delivery of ransomware, regardless of technological advances. Ransomware continues to evolve at a rapid pace, and with it, the strategies to combat it must also evolve, including the development of improved threat intelligence, stronger backup systems, proactive monitoring of networks, and international cooperation in law enforcement agencies [63]. When it comes to developing effective countermeasures against ransomware, as well as building resilient digital ecosystems to face this persistent and growing threat, understanding the origins and development of this threat is crucial [63].



Categories of Ransomware Attacks:

During the past few years, ransomware has evolved significantly, becoming categorised according to varied technical and operational characteristics. For developing effective defence strategies and comprehending the complexity of modern cyber threats, it is essential to understand these categories. In this comprehensive study, ransomware is classified based on attack vectors, encryption mechanisms, and deployment strategies, and notable examples that demonstrate their sophistication and diversity are provided.

Classification by Attack Vectors

1.A: Email-Based Attack Vectors:

This is one of the most common methods of deploying ransomware as the initial infection method. The majority of these ransomware attacks leverage phishing emails that contain malicious attachments or links that, when clicked, trigger the infection process by triggering the execution of a

 doi:10.20944/preprints202602.2011.v1

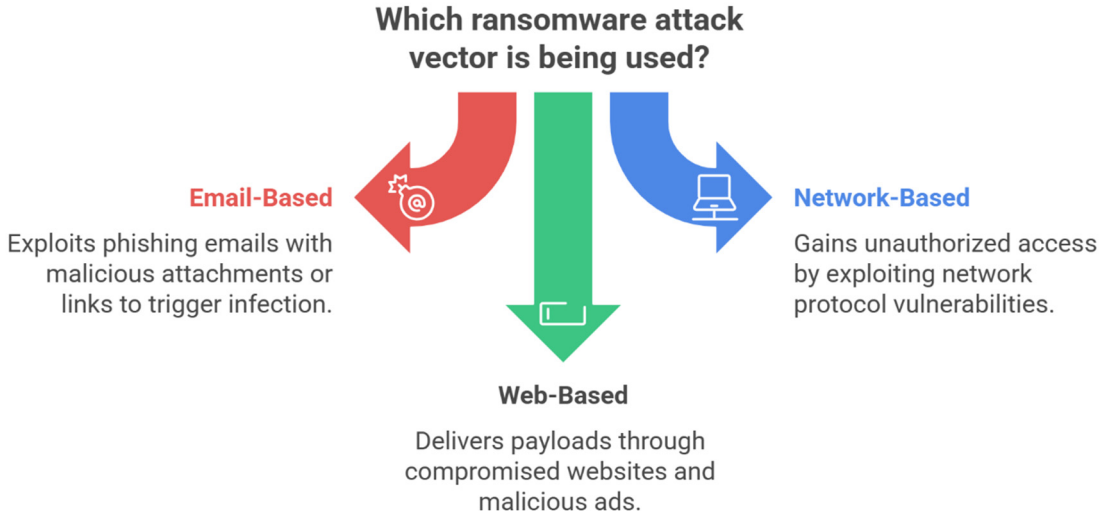
ransomware virus [64]. This approach is exemplified by WannaCry ransomware that spreads quickly through organizations by exploiting email vulnerabilities and network propagation capabilities in order to spread rapidly through a network [64,65]. Similarly, Locky ransomware became famous for its sophisticated email campaigns that disguised malicious payloads within seemingly legitimate business documents, demonstrating how attackers continually refine their social engineering techniques in order to overcome user awareness and technical defences through sophisticated email campaigns disguised within seemingly legitimate business documents [65,66].

1.B: Network-Based Attack Vectors:

The primary purpose of this kind of attack is to gain unauthorized access to target systems by exploiting vulnerabilities in network protocols and services. It is important to note that the WannaCry outbreak of 2017 serves as a prime example of it, as it exploited vulnerabilities in Microsoft's Server Message Block (SMB) protocol by using the EternalBlue exploit [67]. In the end, this attack vector was very effective in spreading WannaCry across a wide range of networks without requiring any interaction from the users, affecting hundreds of thousands of computers worldwide within a few days [67,68]. The impact of network-based attacks is often magnified by the fact that they are often launched using known vulnerabilities that have not yet been patched, which highlights the need for timely security updates and network segmentation strategies [68].

1.C: Web-Based Attack Vectors:

Ransomware payloads are delivered via compromised websites, malicious advertisements, and drive-by downloads using a variety of malicious methods. A sophisticated example of this type of exploit kit is called Angler, which is capable of detecting browser and plugin vulnerabilities in order to download and execute ransomware without the user's knowledge or consent [69,70]. It is common for these attacks to target popular websites with high volumes of traffic in order to maximize potential victim exposure while at the same time leveraging legitimate web infrastructure in order to conceal malicious activities from detection [70].



Classification by Encryption Mechanisms

2.A: Crypto-Ransomware:

It is the most sophisticated and destructive category of ransomware, employing advanced cryptographic algorithms to encrypt the files of their victims and render them inaccessible until they have paid the ransom. As part of this category, we have some of the most notorious ransomware families, including CryptoLocker, which was one of the first ransomware families to use RSA encryption with safe key management practices [71]. There is another approach within this category

 doi:10.20944/preprints202602.2011.v1

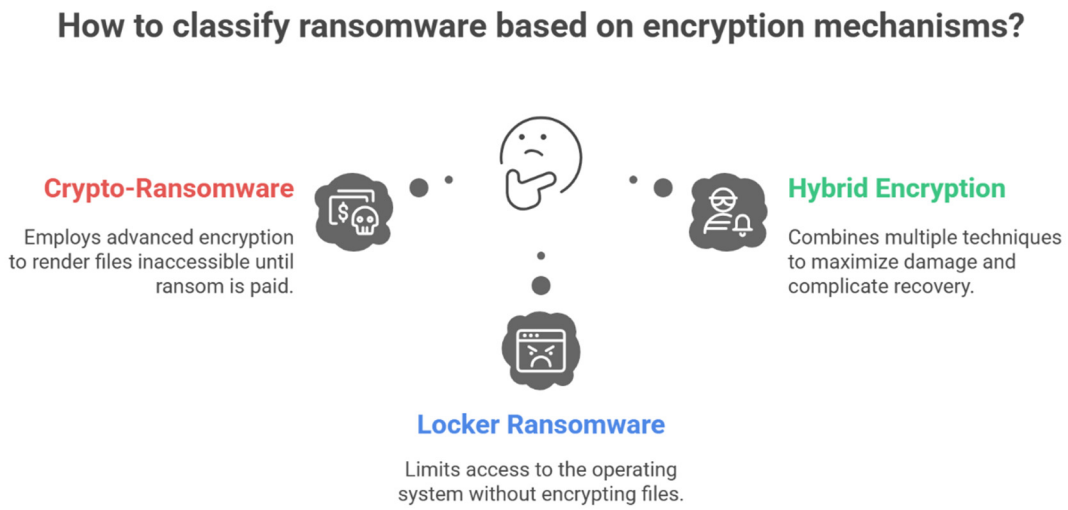
of malware, the Petya ransomware family, which targets the master boot record (MBR) in order to encrypt all of the files on the hard drive rather than encrypting individual files, essentially rendering the system inoperable until it is repaired [71].

2.B: Locker Ransomware:

Instead of encrypting files at the file level, this type of security focuses on limiting access to the operating system and other critical functions of the system without encrypting any files themselves. A good example of this is the Police Trojan family, which displays fake police warnings that claim illegal activities have been detected and demands payment in order to regain access to the system. In spite of the fact that crypto-ransomware is generally less destructive, locker variants can still disrupt operations and serve as precursors to more sophisticated attacks, despite the fact that they are generally less destructive [72].

2.C: Hybrid Encryption Approaches:

Combining multiple encryption techniques with the aim of maximizing damage and complicating recovery efforts is a common practice. An example of this evolution can be seen in the Satan ransomware family, which utilizes both file encryption and system locking mechanisms, while incorporating additional features such as the ability to exfiltrate data as well. Currently, hybrid attacks represent the trend towards multi-faceted attacks that have the ability to simultaneously threaten the confidentiality of data, the availability of the system, and the reputation of the organization [73].



Classification by Deployment Strategies

3.A: Ransomware-as-a-Service (RaaS) Models:

It has revolutionized the landscape of ransomware in the last few years by democratizing access to sophisticated attack tools on a large scale. A prominent example of this business model can be seen in the DarkSide ransomware group, which provides affiliate attackers with comprehensive attack infrastructure, technical support, and profit-sharing arrangements [74]. As a result of this approach, cybercriminals have had a significant reduction in entry barriers and have also been able to scale attack operations rapidly. It is a widely known fact that the Colonial Pipeline attack of 2021, which is attributed to DarkSide affiliates, showed the devastating potential of RaaS operations when targeting critical infrastructure, causing widespread fuel shortages and exposing the strategic implications of ransomware attacks [74,75].

3.B: Targeted Attack Strategies:

Focus on high-value targets such as healthcare institutions, government agencies, and critical infrastructure providers, in order to maximize the chances of success. In this category, there is the

 doi:10.20944/preprints202602.2011.v1

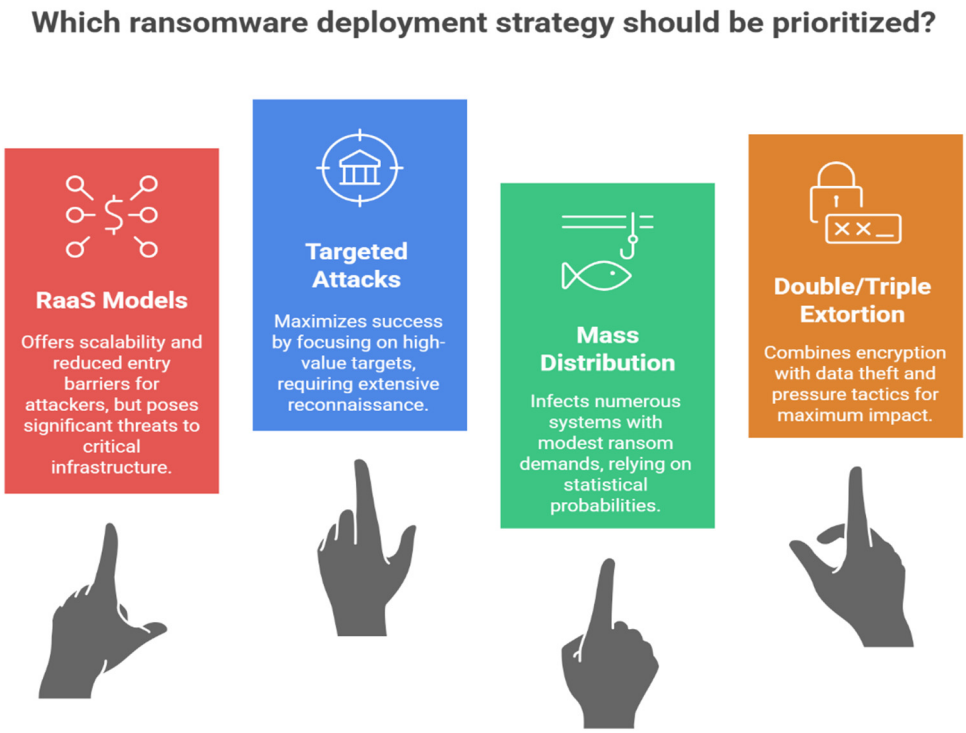
Ryuk ransomware family, which employs extensive reconnaissance techniques and lateral movement techniques to maximize damage within targeted organizations as best as possible. As part of these attacks, attackers often study target environments for months in advance, identify critical systems, and plan coordinated strikes designed to inflict maximum operational disruption while ensuring reliable ransom collection mechanisms are in place so they can collect ransoms as quickly as possible [75].

3.C: Mass Distribution Campaigns:

Prioritize volume over precision, making sure to cast wide nets in order to infect as many systems as possible while at the same time relying on statistical probabilities to generate a profit. As a prime example of this approach, the Bad Rabbit ransomware campaign spreads rapidly through compromised websites and affects numerous organizations across multiple countries and continents simultaneously through the use of compromised websites. It is important to note that despite the relatively modest size of individual ransom demands, the aggregate impact of mass campaigns can be substantial, particularly when targeting areas with limited cybersecurity infrastructure [76,77].

3.D: Double and Triple Extortion Strategies:

The latest evolution of ransomware has been to combine traditional encryption attacks with data theft and additional pressure tactics to create the ultimate ransomware attack. The Maze ransomware group was the first to implement a double extortion approach, which involves stealing sensitive information before encrypting it, and then threatening public release unless ransom demands are met [78]. There has also been a refinement in the Conti ransomware family, implementing triple extortion elements such as distributed denial-of-service (DDoS) attacks and direct customer harassment, which have been incorporated in order to maximize pressure on victim organizations [78].



Emerging Trends and Future Implications

Despite the fact that the ransomware landscape continues to evolve at a rapid pace, attackers have continued to implement new techniques to overcome defensive measures and maximize their profits. Recently, we have seen the emergence of Rust-based ransomware families such as BlackCat (ALPHV), which utilize modern programming languages in order to enhance performance and evade detection by using modern programming languages. As well, the targeting of cloud infrastructure

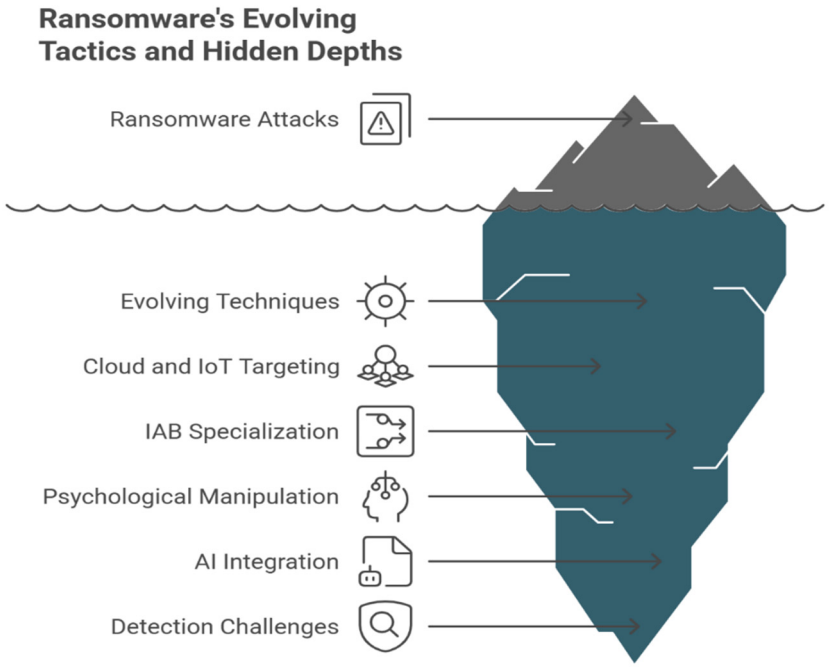
 doi:10.20944/preprints202602.2011.v1

and the Internet of Things (IoT) devices represents an expansion of attack surfaces that requires a change in defensive approach to effectively defend against them [79].

Due to the increasing sophistication of Initial Access Brokers (IABs), there has been a growth in the specialized markets for network access, which has enabled ransomware operators to concentrate on payload development and execution while outsourcing the initial compromise activities to third parties. By dividing the work between different groups of individuals, ransomware operations have become more professionalized and more effective attack methods have been developed as a result [79,80]. In recent years, ransomware has evolved from encrypting files to combining exploitation of technical exploitation with psychological manipulation as part of a complex, multi-stage attack. As a result of the integration of artificial intelligence, cryptocurrency, and advanced evasion techniques, modern ransomware is increasingly resilient and more difficult to detect. For comprehensive defence strategies that address all the potential threats to the system, it is essential to have a good understanding of these diverse ransomware categories [80].

It is essential that organizations implement layered security approaches that account for a variety of attack vectors, encryption mechanisms, and deployment strategies while remaining aware of emerging trends that may call for adaptive defensive measures to be implemented. As ransomware tactics continue to evolve and become increasingly sophisticated, the importance of proactive cybersecurity measures to protect critical digital assets as well as the need for ongoing vigilance regarding the protection of those assets cannot be overstated [81]. In recent studies, it has been demonstrated that effective defence requires a multilayered approach that includes real-time monitoring, behavioural analysis, diverse datasets for training AI models, and hybrid methods of detection involving a combination of static and dynamic features in order to be effective [81,82].

A growing number of sandbox environments and feature parsers are being used to analyse API calls, registry changes, and file operations to enable more accurate identification of ransomware behaviour. Even so, there remain significant obstacles that need to be overcome, such as concept drift, the insufficient real-world evaluation of deep learning models, and the inability to develop standardized benchmark datasets. As ransomware threats become increasingly sophisticated, it is critical to address these gaps through collaborative research, improving dataset diversity, and ensuring that AI frameworks are adaptive in order to stay ahead of these threats [82].



 doi:10.20944/preprints202602.2011.v1

How the Ransomware Attack Chain Works: A Scientific and Technical Analysis

A ransomware attack is regarded as one of the most disruptive and financially damaging forms of cybercrime that exists in the modern digital landscape. Ransomware attacks have a structured, multi-stage process known as the ransomware attack chain in which the attackers systematically compromise systems, escalate privileges, spread laterally, exfiltrate data, and ultimately encrypt critical assets in order to extort payment from their victims [83,84]. In order to develop robust defensive strategies, it is essential to understand this chain of events. The attack chain aligns closely with the Cyber Kill Chain® model as well as the MITRE ATT&CK framework, which includes stages ranging from the initial access to the final impact. There are several phases in the lifecycle of a ransomware attack and each phase is described in the scientific analysis, along with the technical mechanisms employed at each phase [85–87].

Reconnaissance and Initial Access

In the beginning of an attack chain, threat actors identify potential targets by scanning for vulnerabilities in systems that can be exposed to the public, such as exposed Remote Desktop Protocol (RDP) ports, unpatched web servers, or a misconfigured cloud storage system. Several automated tools are available for locating vulnerable infrastructure, including Shodan and Censys for example. The initial access to a target is achieved by a variety of techniques once a target has been selected. As a general rule, credential-harvesting websites or phishing emails with malicious attachments (e.g., weaponized Microsoft Office documents that contain embedded macros) are the most common methods for harvesting credentials from users. If a user enables macros, Visual Basic for Applications (VBA) scripts will execute, downloading the first-stage payload-often a dropper or a downloader-onto the computer as soon as the macros are enabled [88].

A second approach to gaining direct access to a server is to exploit known software vulnerabilities (e.g., Log4Shell, ProxyShell) or brute-force weak RDP credentials to gain access directly. Exploitation tools such as Metasploit or custom exploit kits are frequently used to carry out exploits. The attack begins with a low-privilege user account, which provides the attacker with a persistent presence in the network once the attacker establishes the initial foothold [88].

Execution and Payload Delivery

After gaining access to the system, the attacker proceeds to execute the code in order to deliver and activate the malicious payload. To evade detection, the payload is often delivered in stages in order to evade detection. A first step in the ransomware attack is to deploy a lightweight loader that decrypts and injects the main ransomware binary into the memory, a technique called process injection (e.g., by using CreateRemoteThread or QueueUserApc, for example). By doing this, the malware is able to operate without leaving any traces on disk, which is beneficial to forensic analysis [89].

Ransomware, such as LockBit, REvil, or BlackCat (ALPHV), are often exploiting fileless malware techniques to execute malicious commands on the target system, using legitimate system tools such as PowerShell and Windows Management Instrumentation (WMI) to execute the malicious code. By using PowerShell, for instance, attackers may be able to download and execute the ransomware via Invoke-WebRequest from a remote server, bypassing traditional antivirus solutions that monitor file-based threats and executing the ransomware on a local computer [89].

Persistence and Privilege Escalation

In order to maintain long-term access to a computer, attackers may create scheduled tasks, modify the registry run keys, or install Windows services in order to accomplish long-term access. By making use of these mechanisms, the malware can survive reboots and user logouts without being affected. For instance, if a scheduled task is configured to run the ransomware payload at regular intervals, then the ransomware will be automatically executed [90].

The next step in this process is to escalate privileges to be able to gain administrative access to the system. Attackers exploit local vulnerabilities (for example, Windows kernel exploits like PrintNightmare) or misuse permissions which are misconfigured (for example, weak service

 doi:10.20944/preprints202602.2011.v1

permissions via SeImpersonatePrivilege) to gain access to systems. There are tools available such as Mimikatz that are capable of extracting plaintext passwords, hash values, or Kerberos tickets from memory, making it possible for the attacker to move lateral and gain domain dominance [90].

Lateral Movement

By obtaining elevated privileges, the attacker is able to start lateral movement through the network in order to expand their control. This is typically achieved using pass-the-hash techniques or pass-the-ticket techniques, which are methods by which stolen credentials can be reused to authenticate to other systems without having to crack the passwords of those systems. The attacker may deploy the ransomware to additional hosts by using built-in tools that come with Windows, such as PsExec, WMI, and Remote Services that are available on the system [91].

Tools like Nmap or BloodHound (which maps Active Directory relationships in a network) can be used to scan a network for high value targets, such as domain controllers, file servers, and backup systems. In many cases, attackers use Windows Admin Shares (C$, ADMIN$) and SMB (Server Message Block) protocols to copy and execute payloads across networks using the Windows Admin Shares and SMB protocols [91].

Data Exfiltration (in Double Extortion Attacks)

As part of the double extortion ransomware campaign, one of the hallmarks of modern threats, the attacker exfiltrates sensitive data before encrypting it. The next phase of the attack involves identifying and compressing valuable data (e.g., financial records, customer information, intellectual property) and then transferring it over encrypted channels to attacker-controlled servers (e.g., HTTPS, FTPS, or custom protocols) under the control of the attacker [92]. The process of exfiltrating data is often carried out using steganography, DNS tunneling, or cloud storage applications (such as Dropbox, Google Drive), in which malicious traffic is blended with legitimate traffic on a network. When the victim refuses to pay the ransom, the attacker threatens to publish or sell the data on dark web forums if the victim refuses to pay the ransom [92].

Command and Control (C2) Communication

The compromised systems remain in constant communication with the attacker's command and control (C2) infrastructure throughout the attack. As a result, the attacker is able to issue commands, upload stolen data, and receive notifications about the status of the attack. As a consequence, domain names can be dynamically generated by C2 channels using domain generation algorithms (DGAs), which makes it more difficult for them to be taken down. There is an increasing tendency for communication to be encrypted using TLS in order to avoid network monitoring [93]. As part of their covert operations, some ransomware families use peer-to-peer (P2P) C2 models or legitimate cloud services (e.g., Discord, Telegram) as covert communication channels. This makes detection of these attacks more difficult [93].

Encryption and Impact

The last and most destructive phase of the process is the encryption of the data. Before encrypting the data, the ransomware disables security mechanisms by terminating processes connected to antivirus software, backup agents, and database services (e.g., sqlservr.exe, vssadmin.exe), before encrypting the data. The command vssadmin delete shadows /all /quiet can also be used to delete Volume Shadow Copies. This will prevent the recovery of Volume Shadow Copies from being performed [94]. During the encryption phase, asymmetric cryptography is employed, which is usually a combination of RSA for the key exchange and AES for the bulk encryption of data. Each file is assigned an individual AES session key, and this key is then encrypted with the attacker's public RSA key in order to create an AES session key. As a result of the encrypted AES key being stored along with the encrypted file, only the attacker with their private key will be able to decrypt the file [94,95]. A ransomware program targets specific file extensions (like .docx, .xlsx, .pdf, and .sql), encrypting them recursively across local drives, network shares, and cloud-synchronized folders (e.g., .docx, .xlsx, .pdf, .sql). In order to maximize speed, the encryption process is usually multi-threaded in order to minimize the period in which it can be detected and responded to [95].

 doi:10.20944/preprints202602.2011.v1

Ransom Note and Extortion

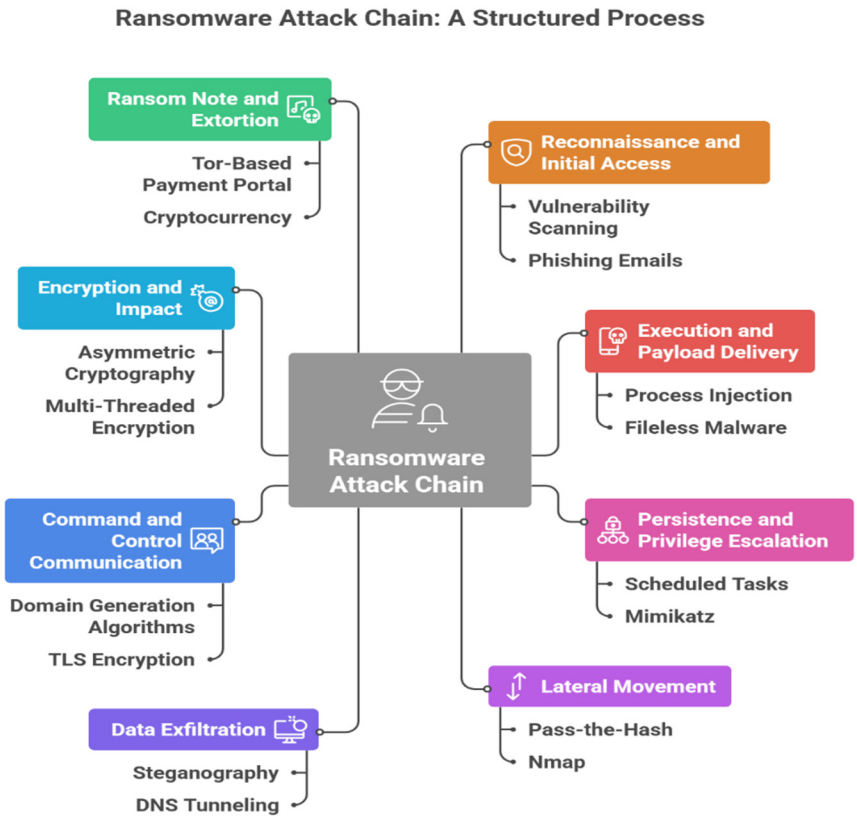
As soon as the encryption process is complete, the ransomware drops a ransom note (e.g., README.txt, HOW_TO_DECRYPT.html) in each directory that has been affected. This note has instructions on how to contact the attacker, usually via a Tor-based payment portal, and demands payment in cryptocurrency (e.g., Bitcoin, Monero) in exchange for the decryption key in exchange for the note [96]. A triple-extortion attack occurs when attackers escalate pressure on a victim by launching DDoS attacks against the victim's public services or directly contacting the victim's clients and partners so that they can increase the reputational damage done to the victim [96].

Post-Attack and Attribution Challenges

A decryption tool may be provided by the attacker after payment (if payment has been made). Despite this, there is no guarantee of a complete recovery of data, and some variants of the software contain bugs that prevent they from being decrypted. As a result, the forensic analysis of the attack is complicated by the use of anti-forensic techniques, such as the deletion of logs, the manipulation of timestamps, and the obfuscation of code [97].

It has become increasingly difficult to attribute attacks due to the increasing use of ransomware-as-a-service (RaaS) models, in which ransomware is leased out to affiliates who carry out attacks and share the profits with developers. The effect of this is not only to decentralize responsibility, but also to obscure where the attack originated [98].

The ransomware attack chain is generally characterized by an orchestrated sequence of technical operations designed to penetrate, persist, escalate, move laterally, exfiltrate, and encrypt data in order to gain financial benefits. Throughout the process, specific tools, protocols, and vulnerabilities are leveraged, often blending legitimate administrative functions with malicious intent in an effort to evade detection. Defense-in-depth strategies are required for defending against such attacks [99]. This includes endpoint detection and response (EDR), network segmentation, regular patching, user awareness training, and immutable backups. Understanding the scientific basis of the ransomware lifecycle will enable organizations to better anticipate, detect, and mitigate these evolving threats in a world where cyber dependency is increasing [99].



 doi:10.20944/preprints202602.2011.v1

Emerging Trends of Ransomware Attacks and Their Cybersecurity Implications

During the course of 2025, ransomware attacks have evolved into a multifaceted and highly adaptive threat landscape, characterized by the proliferation of Ransomware-as-a-Service (RaaS), the rise of multi-layered and double extortion tactics, and increasingly targeted attacks on critical infrastructure. As a result of these developments, opportunistic attacks are gradually being replaced by strategic, high-impact campaigns orchestrated by sophisticated cybercriminal syndicates and, in some cases, ideologically motivated hacktivist groups, which are orchestrating these campaigns [100]. By democratizing the access to ransomware tools through the RaaS model, even low-skilled actors have been able to launch devastating attacks by purchasing pre-built malware kits and leveraging a criminal ecosystem that includes developers, affiliates, initial access brokers, and money launderers to carry out their attacks [100]. There has been an increase in the number of ransomware variants and attack frequency due to the commodification of ransomware, with groups like Akira, RansomHub, Qilin, and Cl0p dominating the threat landscape through aggressive tactics and technical sophistication. In the initial stages of gaining access to the cloud infrastructure, these groups exploit vulnerabilities found in public-facing applications, remote access services, and cloud infrastructure, often through phishing attacks, credential abuse, and supply chain compromises. Once they have gained access to system resources, they employ sophisticated techniques to evade detection and stay undetected, including the use of PowerShell scripting, in-memory execution, and Bring Your Own Vulnerable Driver (BYOVD) attacks [101].

Several high-profile RaaS platforms have operated on underground forums with customer support, service-level agreements, and affiliate dashboards, reflecting a level of operational sophistication previously unseen in cybercrime. The modular, scalable nature of RaaS has resulted in an exponential increase in attack volume, with ransomware incidents reported every six minutes in Australia alone, according to the ASD Cyber Threat Report 2022–23. As a result, cybersecurity defences must no longer assume that attacks are isolated or amateurish. Rather, they must prepare for the threat of persistent, well-resourced adversaries who will utilize industrialized attack frameworks to conduct attacks [102].

Along with the proliferation of RaaS, the double extortion model has become a standard tactic among ransomware operators. The traditional function of ransomware is to encrypt a victim's data and demand payment in order to decrypt it. However, modern attackers are now combining encryption with data exfiltration, raising the possibility of leaking or selling sensitive information if the ransom is not paid [102]. This strategy significantly increases the pressure on victims, as the consequences of non-payment include reputational damage, regulatory penalties, and legal liability under frameworks such as GDPR, CCPA, and Australian Privacy Principles (APPs), in addition to operational disruption. As an example, in 2023, the LockBit ransomware group exfiltrated and publicly leaked data from several healthcare and financial institutions after failed negotiations. In this regard, the evolution reflects a shift from pure data denial to information-based coercion, in which the value of the data itself becomes the primary weapon [103]. It has been reported that some groups have escalated to triple extortion, including direct contact with a victim's clients and business partners as well as denial-of-service attacks. The purpose of these tactics is to exploit the growing concern regarding data privacy and compliance in order to force organizations to make difficult choices between paying ransoms and risking public exposure [103].

Additionally, it is alarming to see that targeted attacks are becoming more common, including attacks on medical infrastructure, energy infrastructure, transportation infrastructure, and government utilities [104]. It is in these sectors that cyber attacks are more likely to succeed due to the high dependence on continuous operations in these industries and the devastating consequences of any downtime. In 2021 there was a major attack against the Colonial Pipeline, causing fuel supplies to be disrupted across the United States of America's East Coast, exemplifying how a single ransomware attack can cause economic and social disruptions on a national scale [104]. Furthermore, attacks against hospitals during the pandemic highlighted the life-threatening implications of encrypting patient information and disrupting medical systems, which happened during the

 doi:10.20944/preprints202602.2011.v1

pandemic [105]. As a consequence, cybercriminals realize that critical infrastructure operators are more likely to pay ransoms quickly to restore services, making them high-value targets for cybercriminals. An enormous amount of reconnaissance and lateral movement within networks is often preceded by these types of attacks, which indicate a shift from broad, automated campaigns to highly targeted, manual intrusions in the form of these attacks [105]. There are many advanced techniques used by threat actors, including pass-the-hash, credential dumping with Mimikatz, and the exploitation of zero-day vulnerabilities, which are often used to gain persistent access and escalate privileges before they deploy ransomware on a system. Consequently, there is a clear implication for cybersecurity: perimeter-based defences are no longer sufficient in today's world. Detecting and responding to stealthy, multi-stage intrusions requires organizations to adopt zero-trust architectures, network segmentation, and continuous monitoring in order to detect and respond to stealthy, multi-stage intrusions [105].

It is clear that the convergence of these trends, such as RaaS, double extortion, and critical infrastructure targeting, has created a more adaptive, resilient, and dangerous ransomware ecosystem. With the help of RaaS, new ransomware variants can be quickly developed and distributed, while double extortion increases the profitability and psychological leverage of a ransomware attack [106]. Using targeted attacks ensures that the impact of the attack is maximized and that the ransom yield is higher. The use of crypto-currencies to facilitate anonymized payments, as well as dark web leak sites that provide access to public data has created a self-sustaining criminal economy that operates outside of the reach of regular law enforcement agencies [106].

In light of these evolving tactics, it is imperative that cybersecurity defence strategies undergo a fundamental change. In the first instance, proactive threat intelligence and behavioral analytics are crucial to detecting early signs of compromise, such as unusual patterns of data access or lateral movement. The second step is to develop robust backup and recovery protocols, including immutable, air-gapped backups, in order to reduce the need for ransom payments [107]. As a third step, employee awareness training should emphasize the use of phishing and social engineering techniques in the initial access process. Four, regulatory compliance frameworks must be integrated with technical controls to ensure that sensitive data is classified and protected in accordance with risk levels, as demonstrated in the Sensitive Data Recognition System architecture. Last but not least, collaboration between the public and private sectors, as well as the sharing of information, are essential to disrupting ransomware operations and destroying criminal networks [107].

III. Systematic Evaluation Framework of AI-Based Ransomware Detection

Ransomware Datasets:

Development and evaluation of AI-based ransomware detection systems are critically dependent on the availability of high-quality, diverse, and representative datasets that reflect the evolving threat landscape. Over the past few years, ransomware attacks have evolved significantly, primarily due to the advent of Ransomware-as-a-Service (RaaS), multi-stage extortion techniques, increased attacks on critical infrastructure, the emergence of Initial Access Brokers (IABs), Rust-based malware development, and attacks on platforms such as Internet of Things and cloud-native environments [108]. To train and evaluate detection models, these shifts require corresponding advancements in the datasets that are used. Datasets such as StratosphereIPS Ransomware, ISCX-Ransom2021, CTU-Ransomware-NetFlows, CIC-MalMem2022, and MalContainer-RW are valuable resources, but they often fail to capture these modern attack vectors in their entirety [108].

In the past few years, ransomware-as-a-service (RaaS) has democratized cybercrime by enabling technically unsophisticated actors to launch sophisticated attacks using pre-built, subscription-based ransomware platforms such as LockBit, BlackCat (ALPHV), and REvil. Compared to traditional detection methods, this model introduces a degree of modularity and rapid iteration [109]. However, most existing datasets do not explicitly label or distinguish between RaaS variants and custom-built ransomware, nor do they capture the infrastructure lifecycle associated with affiliate programs, such as C2 server patterns, affiliate panels, or payment portals [110]. Despite the fact that datasets like

 doi:10.20944/preprints202602.2011.v1

HybridAnalysis / MalwareBazaar provide sandboxed execution logs, they often do not include contextual metadata about a ransomware's origin or distribution model, which limits their effectiveness when training AI models to detect RaaS-specific behaviour [110].

Extortion techniques that involve multiple stages, such as double and triple extortion, have evolved significantly beyond simple data encryption. Today's ransomware operators routinely exfiltrate sensitive data before encryption (double extortion) and may follow up with distributed denial-of-service (DDoS) attacks or direct harassment of customers (triple extortion). In spite of this, many datasets focus primarily on the encryption phase of the data theft process, ignoring the earlier stages of data theft and command-and-control (C2) communication [111]. For example, ISCX-Ransom2021 and CTU-Ransomware-NetFlows include network traffic, but do not clearly differentiate between reconnaissance and C2 traffic before and after encryption. Consequently, this gap hinders the development of AI models capable of detecting early-stage malicious behaviour, such as unusual data access patterns or lateral movement, which are essential for proactive defence [112,113].

Critical infrastructure, such as healthcare, energy, transportation, and government systems, is increasingly vulnerable to ransomware attacks, as demonstrated by the Colonial Pipeline incident. The majority of these attacks involve prolonged dwell times, manual intrusions, and domain-specific techniques. It is important to note, however, that most public datasets are collected in controlled laboratory environments using generic Windows executables, which do not simulate the complex, heterogeneous networks that are characteristic of critical infrastructure [113]. A notable exception is the StratosphereIPS Ransomware dataset, which offers real-world captures with detailed process trees and Sysmon logs, but it does not reflect industrial control systems (ICS) or operational technology (OT) environments. Due to this limitation, AI models cannot be trained for anomaly detection in SCADA systems or other specialized infrastructures [113].

There is also an important blind spot regarding the role of Initial Access Brokers (IABs), who sell compromised credentials or remote access to corporate networks via dark web marketplaces [114]. IAB activity often precedes the deployment of ransomware, such as phishing, exploiting unpatched vulnerabilities (such as ProxyShell) or brute-forcing RDP connections [114]. There are, however, few datasets that include labeled examples of initial access techniques and do not correlate them with the subsequent execution of ransomware. This is addressed by the CIC-MalMem2022 dataset, which includes memory dumps, however it does not link these runtime artifacts to upstream access vectors [115]. In the absence of such a link, AI models cannot learn to detect the subtle indicators of compromise (IOCs) that precede the full-scale deployment of ransomware [115].

A new challenge has arisen with the emergence of Rust-based ransomware, such as BlackCat, because of Rust's memory safety, performance, and cross-platform capabilities, which make detection more difficult [116]. It is common for Rust binaries to contain low-level system interactions and anti-analysis features that are not detectable by signature-based tools. Most datasets, however, contain mainly older, C/C++ or PowerShell-based malware, with little or no representation of Rust-compiled malware. Further, static analysis features (e.g., imported functions, string obfuscation) are significantly different between Rust binaries, which requires new approaches to feature engineering [116].

Lastly, ransomware has expanded its threat surface by targeting uncommon platforms such as Linux servers, IoT devices, and containerized environments. Datasets such as the PELICAN IoT corpus and MalContainer-RW represent important advances, offering traces of eBPF syscalls from Docker containers and MQTT-based IoT networks, respectively. These datasets [117], however, remain niche and underpopulated compared to the Windows-centric datasets. While the AndroZoo 4-Layer Set includes Android malware, ransomware targeting mobile platforms is still underrepresented. The lack of diversity reduces the generalizability of AI models, which are often trained exclusively on Windows PE files and do not perform well when deployed in heterogeneous environments [117,118].

 doi:10.20944/preprints202602.2011.v1

In general, existing datasets provide a foundation for research on ransomware, however they do not adequately represent the full spectrum of modern threats. The effectiveness of AI-based detection systems is limited by challenges related to availability (as a result of legal and ethical restrictions), diversity (across platforms, languages, and attack stages), and representativeness (as a result of real-world, multi-phase attacks) [118,119]. Datasets in the future should include labeled examples of RaaS infrastructure, multi-stage extortion sequences, IAB activities, Rust-based malware, and attacks on non-traditional platforms. In addition, they should adopt standardized taxonomies - such as MITRE ATT&CK mappings - and support the fusion of multiple data sources in order to enable holistic, context-aware AI models capable of defending against the next generation of ransomware attacks [119].

Dataset Name	Year	Category	Sample Size	Platform
StratosphereIPS Ransomware	2021	Ransomware (Multi-stage)	3,001 malicious flows + 18,036 benign	Windows, Network
ISCX-Ransom2021	2021	Ransomware (Network Behavior)	2,090 malicious flows + 145,000 benign	Windows
CTU-Ransomware-NetFlows	2022	Ransomware (NetFlow)	430 malicious flows + 78,000 background	Network
CIC-MalMem2022	2022	Ransomware (Memory Forensics)	700 ransomware + 700 benign memory dumps	Windows
ElderRan	2015	Ransomware (Behavioral)	582 ransomware + 942 benign	Windows
HybridAnalysis / MalwareBazaar	2015	Malware & Ransomware (Static + Dynamic)	58,000+ ransomware + 5M+ samples	Multi-platform
VirusTotal Corpus (raw)	2004	Malware & Ransomware (AV Verdicts)	120,000+ ransomware + ≈90M benign/other	Multi-platform
VirusShare	2012	Malware & Ransomware (Static Binaries)	65,000 ransomware + ≈41M other samples	Multi-platform
MalShare	2011	Malware & Ransomware (Static)	24,000 ransomware + 1.8M samples	Multi-platform
EMBERv4-R	2023	Ransomware (Static ML Features)	19,262 ransomware + 1.86M benign	Windows (PE)
AVClass2 Ransomware Subset	2020	Ransomware (Label Aggregation)	95,000 labeled hashes	Multi-platform
MalFam-128	2021	Malware Taxonomy (Benchmark)	38,000 samples mapped to 128 families	Multi-platform
PELICAN IoT Corpus	2022	IoT Malware & Ransomware (Network)	62 attack sessions + 310 benign	IoT (MQTT, SSDP)
MalContainer-RW	2023	Containerized Ransomware (Runtime)	3,100 malicious + 3,100 benign Docker images	Cloud (Docker, eBPF)
AndroZoo 4-Layer Set	2023	Android Malware & Ransomware	7,900 ransomware + 18M+ benign APKs	Android
ENISA CR-Dataset	2024	Cyber Threat Intelligence (Ransomware Incidents)	42 real-world ransomware incidents (STIX 2.1)	Cross-platform

Dataset Name	Key Features	Availability
StratosphereIPS Ransomware	Real-world captures; labeled by family, campaign, and timestamp; ideal for host-network fusion and GNNs	Open (Creative Commons)
ISCX-Ransom2021	High-fidelity traffic; mapped to MITRE ATT&CK; includes encryption, C2, and lateral movement phases	Research-only EULA (UNB)
CTU-Ransomware-NetFlows	Real-world ISP-level data; labeled by family and time; excellent for IDS benchmarking	Open (CC BY 4.0)
CIC-MalMem2022	YARA-verified labels; captures runtime artifacts; supports detection of fileless and in-memory ransomware	Research-only EULA (UNB)
ElderRan	Manually labeled; early-stage behavioral profiling; used to train LSTM models for pre-encryption detection	Open (CC BY-NC 4.0)
HybridAnalysis / MalwareBazaar	Real-time updates; AVClass tags; behavioral indicators; API access; supports hybrid analysis	Open (Creative Commons)
VirusTotal Corpus (raw)	Largest public corpus; temporal validation possible; ideal for pre-filtering and ensemble learning	Freemium API (limited free access)
VirusShare	Long historical archive; useful for signature learning and byte-sequence deep learning	Free with registration (CC BY-NC-SA)
MalShare	Lightweight; easy API access; good for prototyping and quick analysis	Free with registration
EMBERv4-R	Time-stratified labels; family + year; reproducible ML baseline; includes BERT-inspired features	Open (MIT License)
AVClass2 Ransomware Subset	Provides clean, family-level ground truth from AV consensus; essential for post-processing VT data	Open-source (GitHub)
MalFam-128	Standardized taxonomy; resolves naming conflicts; supports benchmarking across datasets	Open (MIT License)
PELICAN IoT Corpus	Focus on lateral movement in IoT; labeled by device type and attack stage; rare dataset for smart device threats	Open (CC BY-NC)
MalContainer-RW	First dataset focused on containerized ransomware; includes Rust and Go-based samples; labeled by family and OS	Restricted (Sandia RUA)
AndroZoo 4-Layer Set	Largest Android corpus; four-layer labeling (family, behavior, permissions, API); supports graph-based learning	Research-only access
ENISA CR-Dataset	TTP-level MITRE ATT&CK mapping; includes IAB activity, RaaS use, and multi-extortion; ideal for contextual AI models	Open (EU OJ license)

Dataset Name	Source	Data Types	Key References
CIC-MalMem-2022	Canadian Institute for Cybersecurity	Memory dumps, PE analysis, behavioral	Namavar Jahromi et al. (2022), "CIC-MalMem-2022: A Memory Analysis Dataset"
CIC-AndMal2017	Canadian Institute for Cybersecurity	APK files, static analysis features	Lashkari et al. (2017), "AndMal: A Dataset of Android Malware Families"
EMBER	Endgame / Elastic Security	PE static features, imports, exports	Anderson & Roth (2018), "EMBER: An Open Dataset for Training Static PE Malware ML Models"
UNSW-NB15	University of New South Wales	Network flows, packet features, protocol	Moustafa & Slay (2015), "UNSW-NB15: A Comprehensive Network Intrusion Dataset"
VirusTotal Intelligence	VirusTotal / Google	Binary files, AV scan results, metadata	VirusTotal API Documentation (2024), Multiple research publications
MalShare	Community Repository	Raw malware binaries, hash databases	Community contributions (2014–2024), Various security research papers
Drebin	TU Berlin	APK static features, manifest analysis	Arp et al. (2014), "DREBIN: Effective and Explainable Detection of Android Malware"
CIC-IDS2017	Canadian Institute for Cybersecurity	Network flows, 80+ flow features	Sharafaldin et al. (2018), "Toward Generating a New Intrusion Detection Dataset"
MaleVis	University Research	Binary visualizations, entropy features	Narayanan et al. (2019), "MaleVis: Machine Learning and Visualization for Malware Detection"
SOREL-20M	Sophos / ReversingLabs	PE features, labels, metadata	Harang & Rudd (2020), "SOREL-20M: A Large Scale Benchmark Dataset for Malicious PE Detection
AMD-Dataset	University Collaboration	APK features, dynamic analysis results	Wei et al. (2019), "AMD: A Dataset of Android Malware Families for ML Research"
AAGM	Academic Research	API call sequences, genomic features	Chen et al. (2020), "Android API and Genome Malware Detection Using Machine Learning"
Kaggle Malware Detection	Kaggle Community	PE features, binary classification labels	Various Kaggle competition papers (2018–2023)
Microsoft BIG-2015	Microsoft	Binary features, assembly dumps, PE	Ronen et al. (2018), "Microsoft Malware Classification Challenge (BIG 2015) Dataset"
BODMAS	NortonLifeLock Research	Behavioral graphs, API sequences	Demetrio et al. (2021), "BODMAS: An Open Dataset for Learning based Temporal Analysis"
MOTIF	Cisco Talos	Multi-modal features, network traces	Freitas et al. (2022), "MOTIF: A Multi-Modal Dataset for Malware Analysis"

 doi:10.20944/preprints202602.2011.v1

Public Datasets

Dataset Name	Limitations	Research Focus
StratosphereIPS Ransomware	Limited to network/host telemetry; lacks memory or cloud context	Multi-modal detection, behavioral analysis, GNNs
ISCX-Ransom2021	No static binary or memory data; not suitable for endpoint-only models	Network anomaly detection, deep learning on encrypted traffic
CTU-Ransomware-NetFlows	Low granularity (NetFlow only); lacks payload or endpoint details	Flow-based ML, lightweight IDS models
CIC-MalMem2022	Small sample size; requires specialized forensic tools	Memory-based detection, anti-evasion analysis
ElderRan	Outdated (pre-RaaS era); limited sample size; no network or memory data	Behavioral ML, early ransomware detection
HybridAnalysis / MalwareBazaar	Labels can be noisy; inconsistent family naming; limited metadata on RaaS or IABs	Static-dynamic fusion, threat intelligence, clustering
VirusTotal Corpus (raw)	No behavioral data; verdicts vary in quality; no ground truth; privacy concerns	Large-scale filtering, temporal ML, consensus modeling
VirusShare	Labels based on filenames (noisy); no dynamic or network context; not updated frequently	Signature-based detection, NLP on binary sequences
MalShare	Limited metadata; uploader notes are inconsistent; no behavioral or network data	Rapid prototyping, malware clustering
EMBERv4-R	Only static features; no dynamic behavior; PE-only	Static ML baselines, feature engineering, model reproducibility
AVClass2 Ransomware Subset	No raw data; only labels; dependent on VT access	Label cleaning, family classification, data harmonization
MalFam-128	No raw data; only labeling schema; requires mapping to other datasets	Taxonomy standardization, cross-dataset evaluation
PELICAN IoT Corpus	Very small scale; limited to specific protocols; not ransomware-specific	IoT intrusion detection, lateral movement analysis
MalContainer-RW	Hard to access; requires approval; niche platform	Cloud-native security, container telemetry, eBPF-based detection
AndroZoo 4-Layer Set	Ransomware subset is small; many apps are benign; privacy issues with Google Play data	Android ransomware detection, permission-based ML, graph neural networks
ENISA CR-Dataset	No raw malware or traffic; abstracted data; not for training classifiers	Threat intelligence, sequence modeling, attack graph analysis

Links of public resources

CIC-MalMem-2022	Datasets | Research | Canadian Institute for Cybersecurity | UNB
CIC-AndMal2017	Datasets | Research | Canadian Institute for Cybersecurity | UNB
EMBER	GitHub - elastic/ember: Elastic Malware Benchmark for Empowering Researchers
UNSW	The UNSW-NB15 Dataset | UNSW Research
VirusTotal Intelligence	VirusTotal - Home
MalShare	MalShare
IEEE DataPort	Dataset for Android Malware Detection | IEEE DataPort
Android-Malware-Datasets	traceflight/Android-Malware-Datasets: Popular Android malware datasets
CIC-IDS2017	Datasets | Research | Canadian Institute for Cybersecurity | UNB
MaleVis University Research	https://vision.ece.ucsb.edu/
SOREL-20M	GitHub - sophos/SOREL-20M: Sophos-ReversingLabs 20 million sample dataset
IMPACT Cyber-Trust	IMPACT - Android Malware Dataset (Argus Lab)
AAGM	
Kaggle Malware Detection	Benign & Malicious PE Files
Microsoft BIG-2015	Microsoft Malware Classification Challenge (BIG 2015) | Kaggle
BODMAS	whyisyoung/BODMAS: Code for our DLS'21 paper - BODMAS: An Open Dataset for Learning based Temporal Analysis of PE Malware. BODMAS is short for Blue Hexagon Open Dataset for Malware AnalysiS. BODMAS Malware Dataset
CIC-AAGM -2017	Android Adware 2017| Datasets | Research | Canadian Institute for Cybersecurity | UNB

Private Datasets

Dataset Name	Source	Brief Description
Ransomware-DB	University of Malaga	Specialized ransomware dataset with 11,678 samples collected in a controlled lab environment, including detailed family labeling
RanDroid	Private Lab Collection	Android ransomware-specific dataset with 2,916 samples focusing on mobile encryption and screen-locking behavior
CryptoLocker Analysis Set	Academic Lab Environment	Controlled analysis of CryptoLocker family variants with detailed behavioral documentation and encryption analysis
Locky Behavior Dataset	Cybersecurity Lab	Comprehensive behavioral analysis of Locky ransomware variants in a controlled virtual environment with API monitoring
WannaCry Lab Collection	Multiple Research Labs	Multi-institutional collection of WannaCry samples and variants with propagation analysis and network behavior
Ryuk Analysis Suite	Enterprise Security Lab	Controlled analysis of Ryuk ransomware targeting enterprise environments with detailed documentation of lateral movement and escalation
Maze Double-Extortion Set	Threat Intelligence Lab	Analysis dataset focusing on Maze ransomware's double-extortion techniques and data exfiltration patterns
RaaS Laboratory Collection	Cybersecurity Research Consortium	Comprehensive analysis of Ransomware-as-a-Service (RaaS) platforms with affiliate sample collection and behavioral analysis

 doi:10.20944/preprints202602.2011.v1

Dataset Name	Source	Data Types
Ransomware-DB	University of Malaga	Static + dynamic analysis, encryption behavior
RanDroid	Private Lab Collection	APK analysis, encryption patterns, UI blocking
CryptoLocker Analysis Set	Academic Lab Environment	Behavioral logs, encryption analysis, network traces
Locky Behavior Dataset	Cybersecurity Lab	API calls, file operations, registry modifications
WannaCry Lab Collection	Multiple Research Labs	Network propagation, vulnerability exploitation, encryption
Ryuk Analysis Suite	Enterprise Security Lab	Enterprise attack chains, lateral movement, privilege escalation
Maze Double-Extortion Set	Threat Intelligence Lab	Data exfiltration patterns, encryption behavior, C2 communication
RaaS Laboratory Collection	Cybersecurity Research Consortium	RaaS platform analysis, affiliate samples, payment mechanisms

Links of private resources

Ransomware-DB	Static + dynamic analysis, encryption behaviour
RanDroid	APK analysis, encryption patterns, UI blocking
CryptoLocker Analysis Set	Behavioural logs, encryption analysis, network traces
Locky Behaviour Datasets	API calls, file operations, registry modifications
WannaCry Lab Collections	Network propagation, vulnerability exploitation, encryption
Ryuk Analysis Suite	Enterprise attack chains, lateral movement, privilege escalation
Maze double-extortion set	Data exfiltration patterns, encryption behaviour, C2 communication
RaaS Lab collection	RaaS platform analysis, affiliate samples, payment mechanisms

B. Data Analysis:

A multifaceted approach is employed to analyse ransomware datasets, which includes static, dynamic, and hybrid analysis, statistical modelling, anomaly detection, and pattern recognition. In order to better understand ransomware behaviour and improve detection capabilities, each of these methods contributes uniquely [120,121]. It remains, however, a challenge to detect threats in real time, handle obfuscation, reduce false positives, and ensure that datasets are of high quality. In order to build adaptive, transparent, and effective ransomware detection systems, future research must focus on developing robust hybrid models, integrating memory analysis, and leveraging explainable artificial intelligence [122].

During static analysis, features such as Portable Executable (PE) file headers, imported libraries, section entropy, and opcode sequences are analysed without executing ransomware binaries [123]. The method is fast, safe, and does not require kernel privileges or virtualized environments, so it is suitable for screening large datasets in a preliminary manner. The static analysis of ransomware is limited by its inability to detect obfuscated, packed, or crypted payloads, which are common methods for evading detection used by modern ransomware [124]. The statistical analysis of static features, however, has proven useful in identifying patterns across ransomware families, enabling researchers to classify samples based on metadata anomalies or structural irregularities. For instance, studies have employed statistical models to analyse PE file characteristics and detect deviations indicative of malicious intent, but these methodologies are less effective when dealing with polymorphic or metamorphic ransomware [124,125].

Description of static analysis	Detects structural and syntactic features such as headers, opcodes, and imported libraries in ransomware binaries without executing them
Key techniques and approaches	Disassembly using IDA Pro, Ghidra, radare2, OllyDump Opcode sequence extraction PE header analysis (e.g., section entropy, imports) N-gram analysis of opcodes TF-IDF for feature weighting Use of pefile, DIE, Distorm3 librarie
Tools & Frameworks	IDA Pro, Ghidra, radare2, OllyDump, pefile, DIE, Distorm3
Relevant findings & applications	Extracts features like strings and API calls pre-execution Vulnerable to packing/obfuscation Effective for initial triage but limited against advanced evasion techniques

 doi:10.20944/preprints202602.2011.v1

Description of statistical analysis	Applies statistical methods to quantify and interpret patterns in ransomware data, often used in preprocessing and feature evaluation.
Key techniques and approaches	Handling missing values (mean/median imputation, removal) Noise reduction (binning, regression) Data discretization and generalization Dimensionality reduction (PCA, LDA, GDA) Outlier detection and duplication removal
Tools & Frameworks	Python (Pandas, Scikit-learn), R, WEKA
Relevant findings & applications	Improves data quality and model performance Mutual Information (MI), Chi-Square, and RFE used for feature selection in MLRan Balances class distributions and reduces false positives through preprocessing

The dynamic analysis, on the other hand, entails running ransomware samples in a controlled environment, such as a sandbox, in order to observe how they behave during their runtime. It captures system-level activities such as file operations (e.g., mass encryption, file deletion), registry modifications, process creation, API calls, and network communications [126,127]. Dynamic analysis provides a deeper understanding of the operation of ransomware by monitoring these behavioural indicators in real time, strengthening its resilience to evasion techniques such as code obfuscation [128]. Although this approach is time-consuming and resource-intensive, it poses security risks because malicious code can be executed. Furthermore, many existing dynamic analysis frameworks are designed for outdated operating systems such as Windows 7, which makes them irrelevant for today's threat landscapes [129]. Recent efforts emphasize the importance of up-to-date and diverse datasets covering multiple ransomware types—including crypto-ransomware, locker ransomware, ransomware-as-a-service (RaaS), and modern hybrid variants—over extended periods in order to enhance realism and representativeness. MLRan, for example, contains over 4,800 samples from 64 ransomware families collected between 2006 and 2024, allowing for an analysis of a broad range of time and behaviour [130,131].

Description of dynamic analysis	Involves executing ransomware in a controlled environment (e.g., sandbox) to observe runtime behaviour such as file operations, API calls, and registry changes.
Key techniques and approaches	Behavioural monitoring in sandboxes Logging system calls, file/directory operations, registry modifications, network activity API call sequence analysis Execution on multiple OS versions Repetitive execution for consistency
Tools & Frameworks	Cuckoo Sandbox, BitVisor, IRP Logger, ANY.RUN, Hybrid Analysis, VirusTotal
Relevant findings & applications	Captures real-time behaviours: encryption patterns, persistence mechanisms MLRan dataset uses Cuckoo Sandbox with 4,880 samples (2,330 ransomware, 2,550 goodware) Most widely used method in public datasets (e.g., ISOT, MarauderMap, MIRAD)

In hybrid analysis, the strengths of both static and dynamic approaches are combined to overcome the limitations of each approach separately. By integrating structural features from static analysis with run-time behavioural data from dynamic analysis, hybrid models are able to achieve higher levels of detection accuracy and robustness [132]. For example, some studies have used machine learning classifiers such as Random Forest (RF), Support Vector Machine (SVM), and Gradient Boosting Trees (GBT) trained on a combination of 72 static and 45 dynamic features selected through information gain, leading to improved F1-scores and reduced false positives [133,134]. Furthermore, hybrid analysis enables the development of more comprehensive features, including file system interactions, registry changes, network activity, and dropped files, which are often absent from existing public datasets. In spite of its promise, hybrid analysis remains underexplored in ransomware detection, with most research focused primarily on dynamic methods. It is

 doi:10.20944/preprints202602.2011.v1

recommended that future work focus on the development of standardized hybrid frameworks that can be scaled and reproduced [135].

Description of hybrid analysis	Combines static and dynamic analysis to improve detection accuracy by leveraging both structural and behavioural features.
Key techniques and approaches	- Integration of static features (e.g., PE attributes) with dynamic behaviours (e.g., API calls) - Multi-level profiling - Feature fusion from multiple sources - Sequential or parallel model integration
Tools & Frameworks	Custom pipelines combining static and dynamic tools
Relevant findings & applications	Achieves higher F1-scores and accuracy than single-method approaches Used in studies combining 72 static and 45 dynamic features selected via Information Gain Enhances robustness against obfuscated samples Recommended for comprehensive ransomware characterization

Anomaly detection and patterns recognitions

The detection of anomalies plays a crucial role in identifying previously unknown ransomware strains by recognizing deviations from normal behaviour. Unlike signature-based or rule-based systems, anomaly detection models learn the baseline behaviour of legitimate software and identify activities that deviate from expected patterns - such as rapid file encryption or unusual process spawning - as potentially malicious [136,137]. This context often makes use of machine learning algorithms, particularly unsupervised and semi-supervised models such as autoencoders and clustering techniques. It has been shown, for example, that convolutional autoencoders are capable of detecting novel ransomware by reconstructing input features and measuring the errors generated by the reconstruction [138,139]. Despite their effectiveness in detecting zero-day attacks, anomaly detection systems can produce false positives, particularly when benign applications exhibit atypical behaviour. Consequently, it is essential to balance specificity and sensitivity, with recall, precision, and F1-score being prioritized over raw accuracy, particularly in unbalanced datasets where ransomware samples outnumber benign ones [139].

Description of anomaly detection	Identifies deviations from normal system behaviour to detect previously unseen (zero-day) ransomware.
Key techniques and approaches	Unsupervised/semi-supervised learning (e.g., autoencoders) Reconstruction error analysis Behavioural baselining Threshold-based alerting Real-time monitoring
Tools & Frameworks	Convolutional Autoencoders (CAE), Isolation Forest, One-Class SVM
Relevant findings & applications	Effective for detecting novel ransomware strains High sensitivity but prone to false positives SHAP and LIME used to explain anomalies and validate detections

In addition to anomaly detection, pattern recognition can be used to identify recurring behavioural motifs associated with ransomware operations. Various techniques are used to capture the progression of ransomware activities - from infection to execution of the payload and encryption of data [140]. These techniques include finite-state machines and sequence modelling. Using these models, it is possible to identify temporal patterns within system calls and API invocation sequences, thus enabling early detection before significant damage is caused. XAI (Explainable Artificial Intelligence) methods such as SHAP (Shapley Additive Explanations) and LIME (Local Interpretable Model-agnostic Explanations) enhance pattern recognition by highlighting the most discriminative features that contribute to classification decisions. It is essential that models be interpretable in order to validate predictions and refine detection rules [141].

 doi:10.20944/preprints202602.2011.v1

Description of pattern recognition	Identifies recurring behavioural or structural patterns associated with ransomware families or attack stages.
Key techniques and approaches	N-gram modelling of opcodes or API calls Sequence modelling (e.g., Markov chains, FSMs) Frequent pattern mining Clustering of similar behaviours Signature generation from common traits
Tools & Frameworks	TF-IDF, NLP techniques, clustering algorithms (K-means, DBSCAN)
Relevant findings & applications	Enables ransomware family classification TF-IDF preserves uniqueness of critical n-grams Frequent pattern mining helps in automated labelling and taxonomy construction

In general, a critical challenge in all of these analytical methods is the quality and diversity of the underlying datasets. For an effective analysis, large, well-labelled, and representative datasets are required that capture the full spectrum of ransomware behaviours. Building reliable models requires consideration of factors such as sample size, temporal coverage, family diversity, and feature richness [142]. The scope of publicly available datasets is often limited, reproducible, or unrepresentative of features. Often, network-based indicators, system processes, or dropped files are not included. To address this issue, initiatives such as MLRan and GUIDE-MLRan provide standardized guidelines for creating datasets, emphasizing the authenticity of the samples, diversity, and extraction of comprehensive features. In this way, reproducible research is supported, and benchmarking is facilitated across different detection models [142].

Description of feature selection	Identifies recurring behavioural or structural patterns associated with ransomware families or attack stages.
Key techniques and approaches	N-gram modelling of opcodes or API calls Sequence modelling (e.g., Markov chains, FSMs) Frequent pattern mining Clustering of similar behaviours Signature generation from common traits
Tools & Frameworks	TF-IDF, NLP techniques, clustering algorithms (K-means, DBSCAN)
Relevant findings & applications	Enables ransomware family classification Uses patch-based CNN and self-attention on opcode n-grams TF-IDF preserves uniqueness of critical n-grams Frequent pattern mining helps in automated labelling and taxonomy construction

Description of Machine Learning	Applies supervised and unsupervised algorithms to classify and detect ransomware based on extracted features.
Key techniques and approaches	Supervised: Random Forest (RF), SVM, Decision Trees (DT), XGBoost, Logistic Regression, Gradient Boosting Unsupervised: Clustering, Autoencoders, Ensemble methods (e.g., AdaBoost, CSPE-R)
Tools & Frameworks	Scikit-learn, XGBoost, TensorFlow, PyTorch
Relevant findings & applications	RF, XGBoost, Logistic Regression (achieve >98% accuracy in MLRan) CSPE-R framework detects novel ransomware with cost-sensitive learning Multiple classifiers (DT, SVM, RF, AdaBoost) combined for improved F1-score Models trained on balanced, up-to-date datasets perform better in real-world scenarios

C. Data Pre-processing:

The preprocessing of data is a critical step in the development of AI-based ransomware detection systems, because it directly impacts the performance, reliability, and generalization capabilities of machine learning (ML) and deep learning (DL) models. The raw ransomware data collected from sources such as Portable Executable (PE) files, system logs, network traffic, API call sequences, and

 doi:10.20944/preprints202602.2011.v1

behavioural reports generated in sandbox environments is often characterized by inconsistencies, redundancies, noise, and missing values, which can significantly reduce model accuracy if not addressed properly [143,144]. As a result, comprehensive data pre-processing involves a series of systematic steps such as data cleaning, normalization, transformation, noise reduction, feature scaling, and handling missing values in order to convert raw data into a clean, consistent, and algorithm-ready format. Efforts are being made to enhance data quality, reduce bias, and ensure that the input features accurately represent both benign software and malicious ransomware behaviours [144].

In pre-processing, data cleaning is the first and most fundamental step, which focuses on improving data integrity by addressing missing values, removing noise, removing duplicates, and correcting biases. Missing values are common in ransomware datasets, particularly when certain system events and API calls are not logged as a result of execution path variations or sandbox limitations [145]. The presence of a large percentage of missing data can impair model training and result in inaccurate predictions. As a result, several strategies are employed to address this issue: it is possible to remove rows or columns with excessive missing values, but this may reduce the size and diversity of the dataset [146,147]. Alternatively, imputation techniques such as mean, median, or mode substitution can be applied to numerical features, while more sophisticated methods such as KNN imputation or regression-based prediction can be applied to complex relationships. During the imputation process, however, care must be taken to avoid the introduction of artificial patterns. The reduction of noise is another key component of data cleaning, particularly when dealing with dynamic analysis data which may be contaminated by irrelevant or redundant events (e.g., benign background processes) [148]. Binning, smoothing, and outlier detection (using statistical methods such as the Z-score or IQR) are helpful in detecting anomalous data points which do not adhere to expected behavioral patterns. To prevent overfitting and ensure the fairness of the model, duplicate samples, often introduced during data collection from overlapping repositories, are also identified and removed [148,149].

Once the data has been cleaned, it must undergo a transformation in order to be suitable for AI algorithms. Feature encoding, data discretization, and structural formatting are all included in this process [150]. In order to enable compatibility with machine learning models that require numerical inputs, categorical features such as file types, registry keys, or domain names are converted into numerical representations using techniques such as one-hot encoding or label encoding. Sequential data, such as API call traces or system event logs, may be converted to time-series formats, such as N-gram sequences, to be used as input to recurrent neural networks (RNNs) or long short-term memory (LSTM) models [151]. As a result, graph-based representations can also be constructed to facilitate the use of graph neural networks (GNNs) to model process interactions and propagation patterns. The choice of data format, such as tabular (CSV, databases), sequential (logs), or image-like (byte sequences visualized as images), must align with the architecture of the ML system. In particular, convolutional neural networks (CNNs) perform well when presented with image-formatted binary data, whereas tree-based classifiers are more effective with tabular data [151].

In order to ensure that all features contribute equally to the learning process, normalization and feature scaling are crucial steps. Ransomware datasets can contain a wide range of features such as file size, entropy, number of API calls, or registry modifications. When features with greater numerical ranges are not scaled, they may dominate the model's learning process, leading to biased results [152]. The two common scaling methods are Min-Max normalization, which rescales values to a fixed range (typically 0–1), and Z-score standardization, which transforms data to have a zero mean and unit variance. In distance-based algorithms such as k-nearest neighbors (KNN), support vector machines (SVM), and neural networks, scale difference can distort similarity measures and gradient updates. A properly scaled model enhances convergence speed during training as well as overall model stability and performance [152].

In addition to feature selection and dimensionality reduction, ransomware data pre-processing presents other challenges, particularly given the high dimensionality of behavioural datasets. A

 doi:10.20944/preprints202602.2011.v1

dynamic analysis can generate millions of features per sample, such as system calls, file operations, and network connections. This leads to the "curse of dimensionality," where models become computationally expensive and vulnerable to overfitting. It is therefore necessary to use feature selection techniques in order to identify the most discriminating characteristics [153]. Methods such as Mutual Information (MI), Chi-Square, and Information Gain rank features based on their statistical relevance to the target class. In wrapper methods, such as Recursive Feature Elimination (RFE), optimal feature subsets are selected iteratively based on model performance. As part of the model training process, embedded methods, such as Lasso (L1 regularization), perform inherently feature selection. Using a hybrid approach combining Mutual Information and RFE, the initial set of over 6.4 million features in the MLRan dataset was reduced to just 483 features, resulting in a significant improvement in efficiency without compromising accuracy. Data can also be projected into lower-dimensional spaces by means of techniques such as Principal Component Analysis (PCA) or Linear Discriminant Analysis (LDA) while maintaining variance or class separability [153,154].

Lastly, an important aspect of preprocessing is to ensure that the dataset is balanced and representative. This is an important aspect that impacts the generalization of models. There is often an imbalance in ransomware datasets, either through an overrepresentation of benign samples or a scarcity of diverse ransomware families. This can result in models that favor the majority class and fail to detect novel or rare variants [154]. In order to mitigate this problem, techniques such as oversampling (e.g., SMOTE), undersampling, or synthetic data generation are employed in order to maintain a balanced distribution of class numbers. Furthermore, datasets must include a variety of ransomware types (crypto-ransomware, locker ransomware, and RaaS), cover extended time periods, and reflect current threats to reflect real-world conditions. For building reliable and reusable datasets, principles such as sample diversity, accurate labelling, and reproducibility are essential, as indicated in GUIDE-MLRan [154].

Table summary for Data Cleaning

Algorithm	Purpose	Implementation details	Advantage	Limitations	Best use cases
Isolation Forest	Outlier detection and removal	Unsupervised anomaly detection using random forests	Effective for high-dimensional data, no assumptions about data	May remove legitimate rare behaviours	Identifying corrupted sandbox logs
Local Outlier Factor (LOF)	Anomaly detection based on local density	Compare local density of samples with neighbours	Good for detecting local anomalies	Computationally expensive for large datasets	Finding execution anomalies in behavioural data
One-Class SVM	Novelty detection for data validation	Learns boundary around normal data points	Robust to outliers, works well in high dimensions	Sensitive to hyperparameters selection	Validating sandbox execution quality
DBSCAN	Clustering-based outlier removal	Density-based clustering to identify noise points	About cluster shape, automatic outlier detection	Sensitive to hyperparameters, struggles with varying densities	Grouping similar behavioural patterns
Z-Score Filtering	Statistical outlier removal	Removes samples beyond standards	Simple implementation, interpretable	Assumes normal distribution	Filtering extreme feature values

Table summary for Missing Value Imputation

 doi:10.20944/preprints202602.2011.v1

Algorithm	Purpose	Implementation Details	Advantage	Limitations	Best Use Cases
K-Nearest Neighbors (KNN)	Imputation based on similar samples	Uses k most similar samples to impute missing values	Preserves local patterns, handles mixed data types	Computationally expensive, sensitive to distance metric	API call sequences with gaps
Random Forest Imputation	Tree-based missing value prediction	Uses RF to predict missing values from other features	Handles non-linear relationships, robust to outliers	Can be biased toward frequent categories	Complex behavioural feature imputation
Iterative Imputer	Multivariate imputation using regression	Models each feature with missing values as function of others	Captures feature interaction, flexible model choice	Computationally intensive, may not coverage	Registry operations with dependencies
Matrix Factorization	Low-rank approximation for imputation	Decomposes feature matrix to fill missing entries	Effective for high-dimensional sparse data	Assumes low-rank structure	Sparse behavioural matrices
Autoencoders	Neural network-based reconstruction	Learns compressed representation to reconstruct missing data	Can capture complex patterns, end-to-end learning	Requires large datasets, block box approach	Complex multi-model behavioural data

Table summary for Normalization/Scaling

Algorithm	Purpose	Implementation Details	Advantage	Limitation	Best Use Cases
Min-Max Normalization	Scale features to fixed range [0,1]	$X_{\text{scaled}} = \frac{X - X_{\min}}{X_{\max} - X_{\min}}$	Preserves original distribution shape, bounded output	Sensitive to outliers, not robust to new data	Entropy values, percentage features
Z-Score Standardization	Centre data around mean with unit variance	$X_{\text{scaled}} = \frac{X - \mu}{\sigma}$	Handles different scales, centres data	Assumes normal distribution, unbounded	API call counts, memory usage
Robust Scaler	Scale using median and IQR	$X_{\text{scaled}} = \frac{X - \text{median}}{\text{IQR}}$	Robust to outliers, stable statistics	Less sensitive to distribution tails	Network packet size with outliers
Quantile Normalization	Transform to uniform distribution	Maps values to quantile ranks	Distribution reduces skewness	Loses original distribution information	Highly skewed behavioural metrics
Unit Vector Scaling (L2 Normalization)	Scale to unit norm	$X_{\text{scaled}} = \frac{X}{\lvert X \rvert_2}$ where $\lvert X \rvert_2 = \sqrt{\sum x_i^2}$	Preserves direction of data (useful for angle-based similarity) and useful for sparse data	Distorts relative magnitudes and not suitable for features requiring absolute scale	Opcode frequency vectors, behavioural sequence embeddings, and in cosine similarity-based models

 doi:10.20944/preprints202602.2011.v1

Table summary for Feature Transformation

Algorithm	Purpose	Implementation Details	Advantages	Limitations	Best Use Cases
Principle Component Analysis (PCA)	Linear dimensionality reduction	Finds orthogonal components maximizing	Reduces dimensionality removes correlation	Linear assumptions, interpretability loss	High-dimensional API call features
Independent Component Analysis (ICA)	Blind source separation	Finds statistically independent components	Reveals hidden factors, reduces redundancy	Assumes independence, sensitive to pre-processing	Separating malware from benign activities
t-SNE	Non-linear dimensionality reduction	Preserves local neighbourhood structure in low dimensions	Good for visualization, capture non-linear patterns	Computationally expensive, stochastic results	Visualizing ransomware family clusters
UMAP	Uniform manifold approximation	Preserves both local and global structure	Fater than t-SNE, better global structure preservation	Hyperparameter sensitive	High-dimensional behavioural embeddings
Autoencoders	Neural dimensionality reduction	Learns compressed representation through	Non-linear, end-to-end learning flexible architecture	Requires large datasets, overfitting risk	Complex behavioural pattern compression

Table summary for Categorical Encoding

Algorithm	Purpose	Implementation Details	Advantage	Limitation	Best Use Case
One-Hot Encoding	Binary representation of categories	Creates binary column for each category	Simple, interpretable, no ordinality assumptions	High dimensionality, sparse matrices	API function names, file extensions
Label encoding	Map categories to integers	Assigns unique integer to each category	Memory efficient, simple implementation	Implies false ordinality	Ordinal categorical features
Target encoding	Encode based on target statistics	Replaces categories with target-related statistics	Captures category-target relationship	Prone to overfitting, requires regularization	Malware family encoding

 doi:10.20944/preprints202602.2011.v1

Hash encoding	Map categories to hash values	Uses hash function to map to fixed-size space	Handles high cardinality, memory efficient	Hash collisions, information loss	Registry keys, file paths
Embedding layers	Neural categorical representation	Learns dense vector representation for categories	Captures semantic relationships, trainable	Requires neural network, needs training data	Similar API functions, malware variants

Table summary for Sequence Processing

Algorithm	Purpose	Implementation Details	Advantage	Limitations	Best Use Case
Padding/Truncation	Standardize sequence lengths	Pad short sequences, truncate long ones	Simple implementation, fixed input size	Information loss, introduces bias	API call sequences, system events
Sliding window	Extract fixed size subsequence	Creates overlapping windows from sequences	Preserves temporal information, increase samples	Correlated samples, memory overhead	Time-series behavioural analysis
Sequence aggregation	Statistical summaries of sequences	Compute statistics (mean, max, etc) over sequences	Reduces dimensionality, stable features	Loses sequential information	Converting dynamic traces to features
Recurrent Preprocessing	Prepare for RNN/LSTM input	Format sequences for recurrent neural networks	Maintains temporal dependencies, flexible length	Requires sequential models	Temporal behavioural modelling
N-gram Extraction	Extract subsequence patterns	Creates features from sequence n-gram	Captures local patterns, interpretable	Exponential feature growth, sparse	System call patterns, API sequences

Table summary for Data Transformation

Algorithm	Purpose	Implementation Details	Advantages	Limitations	Best Use Case
Box-Cox Transform	Stabilize variance and normalise	$Y = \begin{cases} \frac{X^{\lambda}-1}{\lambda} & \text{if } \lambda \neq 0 \\ \log(X) & \text{if } \lambda = 0 \end{cases}$	Handles skewed data, stabilises variance	Requires positive data, parameter selection	Execution times, memory allocations
Yeo-Johnson Transform	Handle positive and negative data	Extended Box-Cox for any real numbers	Works with negative values, flexible	More complex than Box-Cox	Network traffic volumes
Log Transform	Reduce right skewness	$Y = \log(X+1)$ (handles zeros) or $Y = \log(X)$	Simple, interpretable, reduces skewness	Only for positive data, zero handling	File size, API call frequencies
Square Root Transform	Mild variance	$Y = \sqrt{X}$	Less aggressive than log,	Limited effectivenes	Count-based features

 doi:10.20944/preprints202602.2011.v1

	stabilisation		handles zeros	s for high skewness	
Polynomial Features	Create interaction terms	Generate polynomial combinations of features	Captures feature interactions. Simple	Exponential feature growth, overfitting	Non-linear relationships in behaviour

Table summary for Noise Reduction

Algorithm	Purpose	Implementation Details	Advantages	Limitations	Best Case Use
Gaussian Filter	Smooth continuous signals	Convolves with Gaussian Kernel	Smooth results, parameter control	May blur important details	Time-series smoothing
Median Filter	Remove impulse noise	Replaces values with local median	Preserves edges, robust to outliers	May remove sharp legitimate changes	Filtering execution spikes
Kalman Filter	Optimal state estimation	Recursive estimation with uncertainty modelling	Optimal for linear systems, and handles uncertainty	Requires system model, linear assumptions	Tracking system state changes
Savitzky-Golay Filter	Polynomial smoothing	Fits local polynomials for smoothing	Preserves features, flexible	Requires parameter tuning	Smoothing behavioural time series
Wavelet Denoising	Multi-scale noise removal	Decomposes signal, removes noise components	Multi-resolution, adaptive	Complex implementation, parameter selection	Complex behavioural signals

Table summary for Feature Selection

Algorithm	Purpose	Implementation Details	Advantages	Limitations	Best of Use
Mutual Information	Information-theoretic selection	Measures information shared between feature and target	Non-parametric, captures non-linear relationships	Computationally expensive, discrete approximation	Selecting predictive behavioural features
Chi-Square Test	Statistical independence test	Tests independency between categorial features and target	Simple interpretable, and fast	Only for categorial features	Binary behavioural indicators
Recursive Feature Elimination	Iterative feature removal	Removes features based on model importance	Model-agnostic considers feature interactions	Computationally expensive, wrapper method	Optimizing feature subset size
LASSO regularization	L1 penalty for sparsity	Adds penalty to force feature weights to zero	Automatic feature selection, regularization	May select arbitrary features from correlated groups	Sparse model requirements
Variance Threshold	Remove low-variance features	Eliminate features below variance threshold	Simple, fast, removes constants	Ignores relationship with target	Removing constant behavioural indicators

D. Feature Extraction:

 doi:10.20944/preprints202602.2011.v1

In the context of ransomware detection systems, feature extraction serves as the crucial link between raw malware samples and actionable intelligence that can be interpreted by machine learning models. This process involves systematic identification and extraction of meaningful characteristics from malicious samples in order to differentiate them from benign software in the context of ransomware analysis. In order to extract features from modern ransomware families, a multidimensional approach is required [155,156], including:

1- Static analysis of file properties.
2- Dynamic behaviour monitoring and
3- Network traffic analysis.

Using this comprehensive methodology, security researchers and practitioners are able to develop robust detection mechanisms capable of detecting both known ransomware variants and emerging threats that employ novel evasion techniques.

Ransomware feature extraction begins with file metadata analysis, which provides immediate insights into a sample's characteristics without requiring execution. File sizes, creation timestamps, modification dates, and compilation information provide valuable indicators of malicious intent [157]. In the study of ransomware, entropy analysis serves as a particularly powerful technique, since ransomware often exhibits high entropy values as a result of its encryption routines and its packed executable sections in order to evade the detection of signature-based methods. In addition to providing unique fingerprints for known malware families, hash values, including MD5, SHA-1, and SHA-256, enable rapid comparison against threat intelligence databases. Analysing import tables reveals the Windows API functions that ransomware samples intend to use, sometimes revealing suspicious combinations such as cryptographic functions paired with file system manipulation functions [158]. Observing section headers and PE (Portable Executable) structures can reveal packing, obfuscation, or unusual memory layout patterns characteristic of malicious software. In the course of string analysis, both plaintext and obfuscated strings are frequently uncovered, including ransom notes, cryptocurrency wallet addresses, command and control server URLs, and lists of file extension targets [158,159].

In ransomware feature analysis, behavioural pattern extraction can be considered one of the most crucial aspects, since it captures the dynamic activities that define the ransomware's operational methodology. Monitoring file systems reveals frequent traversals and encryptions of directory structures, often following predictable patterns, such as targeting user documents before system files. Events related to the creation and termination of processes provide insight into the multiple stages of the ransomware's execution, including the deployment of child processes for distributed encryption tasks or the termination of security software and backup services [160]. The patterns of registry modification indicate attempts to establish persistence, disable security features, or modify system recovery options that would otherwise enable victims to recover their data. As a result of memory analysis techniques, cryptographic keys, decryption routines, and communication protocols that ransomware employs during active infection can be captured. There is evidence that network communication patterns, such as DNS queries, HTTP requests, and encrypted communication channels, reveal the existence of command and control infrastructure and data exfiltration attempts prior to the final stage of encryption [160].

The detection of network traffic anomalies is another vital component of ransomware feature extraction. It is particularly relevant in enterprise environments, where network-based detection can provide early warning systems. By analysing traffic volumes, it is possible to detect sudden spikes in internal network communication when ransomware propagates laterally through network shares and connected systems. The analysis of protocol data reveals unusual communication patterns, such as the use of non-standard ports or protocols by ransomware families to communicate with command and control servers [161]. The timing analysis of network communications can reveal automated behaviours characteristic of malware rather than human users, such as frequent and successive connections or communications. DNS query patterns often reveal the domain generation algorithms (DGAs) used by ransomware to establish resilient command and control channels, and TLS certificate

 doi:10.20944/preprints202602.2011.v1

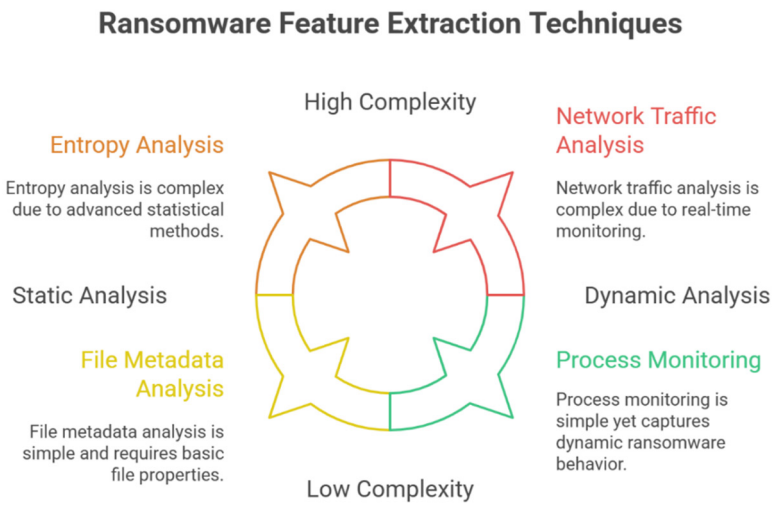
analysis can identify suspicious or self-signed certificates used by ransomware. If possible, analysing the payload of network traffic can reveal cryptocurrency payment instructions, victim identification mechanisms, or encrypted key exchange protocols that enables ransomware operators to maintain control over the decryption process [161].

In terms of enhancing model performance, feature selection is essential. Quality and relevance of selected features have a direct impact on the accuracy, efficiency, and generalizability of ransomware detection systems. As a result of effective feature selection, cybersecurity applications are able to overcome several critical challenges, including the curse of dimensionality, whereby excessive features may overwhelm machine learning algorithms and reduce their effectiveness [162]. Even though redundant features are potentially informative individually, they may introduce noise and computational overhead that could adversely affect the performance of the model. Using correlation analysis, one can identify features that provide similar information, which can facilitate the selection of the most informative representatives as well as the elimination of redundant information. By using statistical techniques such as mutual information analysis, chi-square testing, and analysis of variance (ANOVA), data scientists are able to quantify the relevance of features, enabling them to prioritize features with the strongest correlation with ransomware classification results [162].

It is important to note that advanced feature selection methodologies employ wrapper methods, embedded approaches, and hybrid approaches that optimize feature subsets according to the intended machine learning algorithm. By adding features iteratively, forward selection improves the performance of the model, whereas by eliminating features that contribute little value, backward selection is improved [163]. The recursive feature elimination method combines these approaches by repeatedly training models and removing the least important features until optimal performance is achieved. During model training, regularization techniques, such as L1 (Lasso) and L2 (Ridge) regression, automatically select features by penalizing less important ones. Using ensemble methods, such as Random Forest feature importance or gradient boosting feature selection, the most consistently valuable features can be identified across different algorithmic approaches [163].

A practical implementation of feature selection must also take into account domain-specific requirements unique to ransomware detection. A real-time detection system requires features that can be computed efficiently, potentially favouring static analysis over dynamic analysis which is computationally expensive [164]. False positive rates in cybersecurity applications incur significant operational costs, thus necessitating feature selection approaches that prioritize high precision even at the expense of marginal improvement in recall. Ransomware authors actively attempt to evade detection by modifying features commonly used in detection systems in an effort to evade detection. Feature selection strategies must, therefore, balance current effectiveness with resilience against future evasion attempts, often favouring features that are difficult for adversaries to manipulate without impairing the functionality of the malware [164]. Ransomware tactics continue to evolve, and feature selection approaches must adapt to these threats, potentially by incorporating online learning techniques that adjust feature importance automatically based on newly observed attack patterns and evolving threat landscapes.

 doi:10.20944/preprints202602.2011.v1



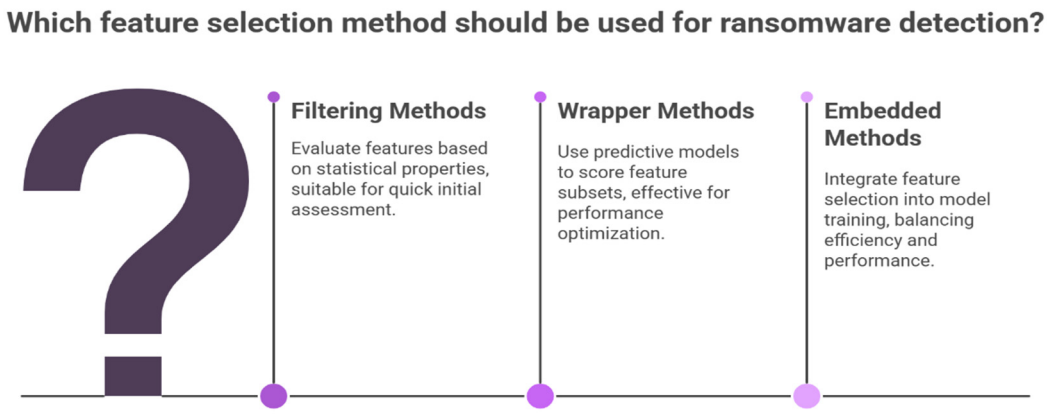
Feature Selection

Feature Selection in AI-Based Ransomware Detection: Methods, Techniques, and Strategic Imperatives

In the rapidly evolving domain of ransomware detection, feature selection stands out as a key preprocessing stage that directly impacts the accuracy, efficiency, and generalizability of models based on machine learning (ML) and deep learning (DL) [165]. As ransomware binaries and behavioural logs generate vast, high-dimensional datasets often consisting of millions of system calls, API sequences, registry modifications, and network activities indiscriminate use of all features leads to the curse of dimensionality, which increases computational overhead and makes models overfitting and less interpretable [166]. A feature selection process is a method of addressing these challenges by selecting and retaining only the most discriminative, non-redundant, and contextually relevant features that are required to successfully distinguish ransomware from benign software and contribute significantly to its detection. The processes described in this article not only simplify the training of models, but also improve the detection performance, which is particularly important in environments that are highly resource-constrained such as endpoint protection platforms and cloud-based security solutions [167].

Several feature selection methodologies have been developed in modern ransomware detection research, and these methodologies are generally categorized into three categories: filtering, wrapping, and embedding. As the name implies, filtering methods evaluate features based on intrinsic statistical properties rather than using a machine learning algorithm to evaluate them. Wrapper methods, on the other hand, use predictive models in order to score feature subsets of interest, thereby treating selection as a search problem [168]. As part of the model training process, embedded methods perform feature selection as part of the process, which contributes to the achievement of a balance between computational efficiency and performance optimization. In recent literature, there have been three approaches that have emerged as particularly influential techniques, namely Principal Component Analysis (PCA), Recursive Feature Elimination (RFE), and Correlation-Based Feature Selection (CFS), each offering its own advantages and trade-offs [169].

 doi:10.20944/preprints202602.2011.v1



Principal Component Analysis (PCA)

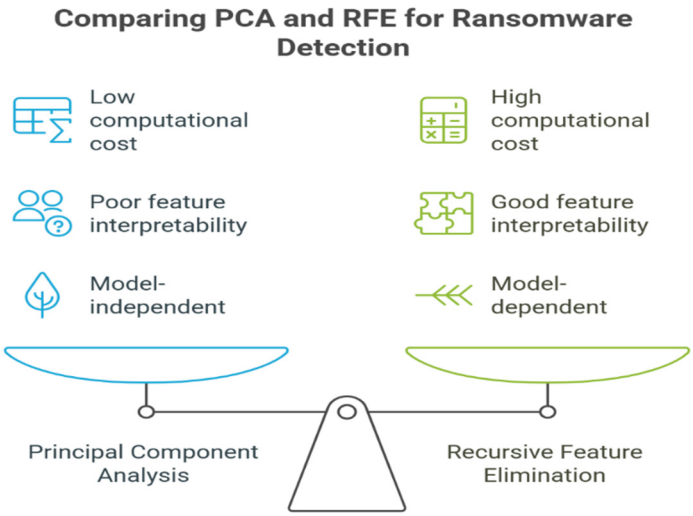
A principal component analysis is a technique that reduces the dimensionality of data by transforming it from a large set of correlated variables into a smaller set of uncorrelated variables called principal components. It is widely used in the detection of ransomware for its ability to reduce the dimensionality of data. As a result of these components, researchers are able to extract the most informative aspects of the dataset while discarding noise and redundant information, resulting in a dataset with the maximum variance possible [170]. In the context of ransomware analysis, PCA is especially useful when dealing with dynamic analysis datasets - such as those derived from Cuckoo Sandbox logs - where features such as API call frequencies, file operation counts, and registry key modifications exhibit high levels of multicollinearity. In the MLRan dataset, which contains over 4,800 samples of ransomware and goodware, PCA was instrumental in reducing the initial feature space from millions of raw behavioural events to a manageable subset and preserving over 95% of the dataset's variance while reducing the initial feature space to a manageable subset. It was not only possible to accelerate model training, but also improved the stability of classifiers such as Support Vector Machines (SVMs) and Random Forests as a result of eliminating redundant features that would otherwise alter the decision boundaries [171]. By reducing overfitting, PCA accelerates model training and improves generalization. PCA can, for example, reveal latent patterns revealed by analysis of opcode sequences or file entropy distributions, which may indicate ransomware behaviour. However, PCA's primary limitation lies in its linearity assumption and loss of feature interpretability; the resulting principal components, while mathematically optimal, often lack semantic meaning, making it difficult for security analysts to understand which specific behaviours triggered the detection, since they do not always have a clear understanding of what behaviours led to detections [171,172]. As a result, PCA is best suited to scenarios where the speed and performance of the model are more important than the explainability of the model, or in situations where a pre-processing step is done before applying a more interpretable technique is applied. Moreover, PCA, assumes linear relationships between features and may not capture complex, non-linear interactions that are often present in advanced ransomware variants. It is also difficult to interpret the transformed components, which is problematic in contexts such as forensic analysis and explainable AI (XAI), where understanding the reasoning behind model decisions is crucial [172].

Recursive Feature Elimination (RFE)

The RFE algorithm, on the other hand, is a wrapper-based approach that recursively removes the least important features based on model weights or feature importance scores until the desired number of features has been achieved. A RFE is a model-dependent technique commonly used in conjunction with classifiers like Logistic Regression, Support Vector Machines (SVM), and tree-based

 doi:10.20944/preprints202602.2011.v1

models - which evaluates feature subsets based on iteratively training the model. Since this technique directly optimizes for classification performance, it is particularly effective when it comes to the detection of ransomware [173]. Studies have shown that using RFE with classifiers such as Random Forests or Support Vector Machines (SVMs) can significantly improve F1-scores and reduce false positives, especially when applied to datasets like MLRan, which contains thousands of ransomware and goodware samples. A strength of RFE is its capability to consider feature interactions and model-specific importance, allowing it to be used in complex detection scenarios [174]. Researchers have combined Mutual Information (MI) for initial filtering with Random Forest and XGBoost classifications in a landmark study that used the MLRan dataset to reduce 6.4 million raw features down to 483 highly discriminative ones, achieving over 98% accuracy using Random Forest and XGBoost classifications [175]. The strength of RFE lies in its ability to capture complex, nonlinear interactions between features that are often missed by other methods of filtering data. It is important to note that even though a single API call like "CreateFile" may appear harmless, its cooccurrence with "CryptEncrypt" and "DeleteShadowCopy" within a short period of time indicates the presence of ransomware activity - a pattern that can be identified by evaluating the contributions made by these features to the model's prediction ability [175,176]. Nevertheless, the computation cost of RFE is significant, since it requires multiple trainings of the model, so it is less suitable for super-large datasets or real-time systems, except if it is combined with efficient sampling or parallelization techniques [177].



Correlation-Based Feature Selection (CFS)

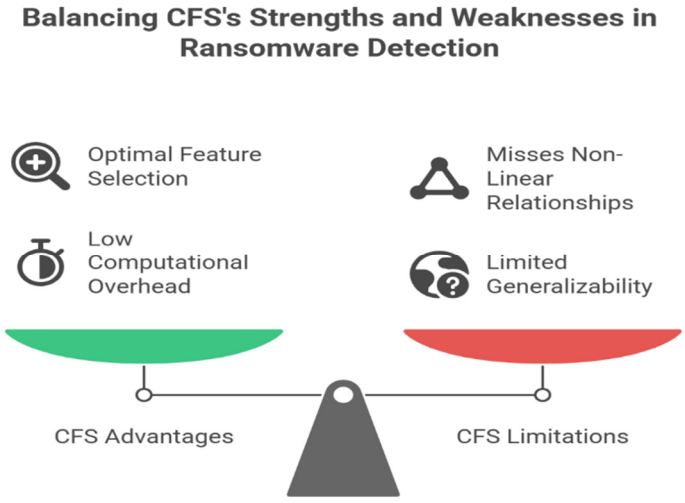
The CFS represents a method for selecting subsets of features based on their correlation with the target class (ransomware vs. benign), while at the same time penalizing redundancy among the features themselves based on their correlation with the target class. It is based on the principle that CFS operates on the principle that good feature subsets contain characteristics which are highly correlated with the classification target, but which are uncorrelated with each other [178]. As a result, this approach is particularly beneficial in ransomware detection, where several features are often interdependent - for example, high entropy values in executable sections are frequently co-occurring with packed headers and obfuscated strings, as an example. Using a feature selection method that minimizes the impact of multicollinearity and enhances the robustness of model results, CFS selects features that are independently predictive and collectively diverse [178]. A number of studies have shown that CFS provides significant improvements in precision and recall, particularly for zero-day ransomware variants that have the ability to evade signature-based detection, especially when it is applied to static analysis features, such as PE header attributes, imported DLLs, and opcode n-grams.

 doi:10.20944/preprints202602.2011.v1

Moreover, CFS has a relatively low computational overhead compared to RFE and PCA, which makes it an ideal tool for initial feature screening in large datasets or as a complement to them as part of a feature analysis process [179]. Despite this, CFS is limited in its effectiveness when it comes to non-linear relationships, as it relies on correlation metrics (typically Pearson or Spearman) which can miss out on important features that only become predictive when they are combined with other factors [179,180].

To overcome the limitations of the individual techniques of ransomware detection, contemporary ransomware detection research increasingly integrates hybrid and ensemble feature selection strategies to overcome the shortcomings of the individual techniques. For example, combining filter methods (e.g., Mutual Information or Chi-Square tests) for an initial screening in conjunction with wrapper methods (e.g., RFE) for fine-tuning has proven to be highly effective for balancing efficiency and accuracy [181]. Additionally, embedded methods such as LASSO (L1 regularization) and tree-based feature importance (for example, from Random Forest or XGBoost) offer automatic feature selection during model training, which makes them ideal for production systems where continuous adaptation to new ransomware strains is required [181,182]. As a result of Explainable AI (XAI), tools such as SHAP (SHapley Additive Explanations) and LIME (Local Interpretable Model-agnostic Explanations) further facilitate feature selection by providing human-interpretable insights into which features most influence a model's decision, which is essential for forensic analysis and regulatory compliance in high-stakes environments, such as the healthcare and financial sectors [182,183].

With regard to the future of feature selection in ransomware detection, there are several emerging challenges that will need to be addressed. To begin with, as ransomware increasingly targets platforms other than Windows (e.g., Linux, IoT, containers), feature selection techniques must be generalizable across heterogeneous environments and data modalities [184]. Furthermore, adversarial ransomware that dynamically alters its behavioural footprint in order to evade detection necessitates the development of adaptive, on-line feature selection mechanisms capable of learning as they proceed. Finally, it is imperative that privacy-preserving feature selection techniques are used in conjunction with federated learning paradigms, wherein models are trained across a variety of decentralized datasets without sharing raw data, so that efficacy is maintained while maintaining data sovereignty is maintained [184].



F. Ransomware Detection with ML & DL Techniques in Current Literature:

As ransomware evolves at a rapid pace, it is imperative to develop intelligent, adaptive, and scalable detection systems, with the promise of artificial intelligence (AI), particularly machine learning (ML) and deep learning (DL) emerging as the most promising technological frontiers. In the

 doi:10.20944/preprints202602.2011.v1

recent past, traditional signature-based methods have proven to be ineffective against polymorphic, obfuscated, and zero-day ransomware variants [185]. In response to this, researchers have increasingly turned to artificial intelligence-based techniques that allow them to learn behavioural patterns, identify anomalies, and generalize their findings across different types of attack vectors. There are currently several AI models that have been applied to the detection of ransomware, and these models can be categorized as supervised learning models, unsupervised learning models, deep learning architectures, and hybrid models, which combine multiple paradigms to provide enhanced performance in ransomware detection. A comparative analysis of these techniques provides valuable insights into how the future direction of cybersecurity research will be determined based on their strengths and limitations [186].

A supervised learning approach remains one of the most widely adopted approaches in ransomware detection due to its high accuracy and interpretability. The majority of ransomware detection research uses supervised learning methods, with ensemble methods and tree-based algorithms consistently demonstrating high performance [187].

The algorithms in supervised models are trained on labelled datasets containing both benign and malicious samples, such that the algorithms are capable of classifying new inputs based on the patterns they have learned from the labeled datasets. There have been a number of studies that have shown impressive results when using methods such as Decision Trees (DT), Random Forests (RF), Support Vector Machines (SVM), K-Nearest Neighbors (KNN), Naive Bayes (NB), and Gradient Boosting [187,188].

Studies have reported detection accuracy exceeding 97% when applied to comprehensive feature sets derived from static analysis, dynamic behaviour monitoring, and network traffic analysis using Random Forest classifiers. The success of Random Forest in ransomware detection stems from its ability to handle high-dimensional feature spaces common in malware analysis, its inherent resistance to overfitting through bootstrap aggregation, and its capability to offer interpretable feature importance rankings that enable security analysts to understand the underlying detection rationale [189].

SVMs have also shown remarkable effectiveness, particularly when combined with sophisticated feature engineering techniques like opcode n-gram analysis and API call sequence modelling. Using SVM with opcode density features alone, Baldwin and Dehghantanha achieved 96.5% accuracy in ransomware family classification, while later research incorporated behavioural and network-based indicators to achieve even higher accuracy [190].

Since they are computationally efficient and interpretable, decision trees and their variants, including Gradient Boosted Trees, have proven particularly useful for real-time detection scenarios. Gradient Boosted Trees were used by Herrera-Silva and Hernández-Alvarez to analyse API call sequences with 99% accuracy. Studies consistently demonstrate that ensemble methods combining multiple weak learners outperform individual algorithms, with combinations of Decision Trees, SVMs, Random Forests, and AdaBoost achieving higher F1-scores and lower false positive rates than single-classifier approaches [191].

The performance of these models is excellent in situations where high-quality labeled data is readily available, and the feature space is well defined. As a result of their simplicity and transparency, they are ideal for forensic analysis as well as ensuring compliance with regulatory requirements. It should be noted, however, that supervised learning is limited by its reliance on labeled data, which is often scarce, especially for new variants of ransomware that are beginning to emerge. Yet supervised learning techniques are incapable of detecting zero-day ransomware variants and new attack vectors that differ significantly from training data patterns. Continual model retraining and dataset updates are required to stay effective against evolving threats [192].

Using unsupervised learning, these limitations can be addressed by detecting patterns without labeled data. In particular, this paradigm is effective in identifying zero-day ransomware and previously unknown attack patterns [193]. In clustering algorithms such as k-means, fuzzy C-means, and DBSCAN, samples are grouped based on behavioural similarities, enabling the identification of

 doi:10.20944/preprints202602.2011.v1

outliers that may represent malicious activity. Detection accuracy is improved while computational overhead is reduced by dimension reduction techniques such as Principal Component Analysis (PCA) and Linear Discriminant Analysis (LDA) [194]. Unsupervised models have been shown to improve processing speed by 15% without sacrificing accuracy. Those methods are useful in dynamic environments where ransomware evolves rapidly and labeled data is unable to keep up. Additionally, convolutional autoencoders have been successfully used to reconstruct input feature vectors; high reconstruction errors indicate anomalous activity - potentially ransomware [195,196]. It is particularly important to use these methods in enterprise environments where benign software behaviour is relatively stable, making it easy to flag sudden deviations - such as the encryption of large amounts of files or the termination of unusual processes - and investigate them. Unsupervised approaches have the advantage of detecting previously unknown threats without prior knowledge [197,198]. Despite this, unsupervised models often suffer from high false-positive rates since there is no ground truth. In addition, they are difficult to interpret, making it difficult to validate detections or explain decisions [198].

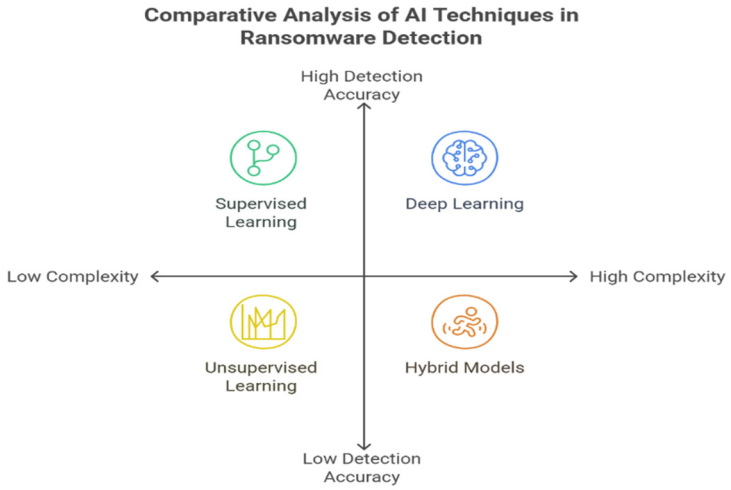
Deep learning a subset of machine learning, has gained traction in ransomware detection due to its ability to automatically extract complex features from raw data. A number of deep learning models have demonstrated superior performance in analysing visual, sequential, Having the ability to analyse time-series data, such as network traffic patterns and system event logs, makes it possible to detect ransomware activity early before significant damage occurs [199]. This including Convolutional Neural Networks (CNNs), Recurrent Neural Networks (RNNs), and Long Short-Term Memory (LSTM) networks. Using CNNs, Sharmeen et al. (2020) were able to detect Windows ransomware 95.96% accurate, outperforming traditional classifiers like SVM (89.98%) and RF (90.95%) [200]. Using BiLSTMs, Roy et al. (2021) developed DeepRan, a sequential data analysis detector leveraging attention mechanisms. Deep learning capabilities have further been enhanced by transfer learning, with pre-trained CNN models like ResNet achieving up to 99.5% accuracy when fine-tuned for ransomware detection (Almomani et al., 2023) Typically, deep learning models capture subtle patterns that are missed by manual feature engineering, such as obfuscation and polymorphism [201]. Transformers and Graph Neural Networks (GNNs) represent the cutting edge when it comes to modeling long-range dependencies and complex interactions between processes. A DL model's greatest strength is its ability to learn hierarchical, abstract representations directly from raw inputs, eliminating the need for manual feature engineering [202]. However, deep learning approaches face substantial challenges, including high computational requirements, extensive training data needs, limited interpretability that complicates forensic analysis, and vulnerabilities to adversarial attacks in which malware authors intentionally modify samples to evade detection. Despite impressive performance metrics in controlled environments, deployment of deep learning models in production systems remains challenging due to computational overhead and model complexity [203].

In a hybrid approach, several AI paradigms are combined to overcome individual limitations and enhance robustness of detection. To create comprehensive detection frameworks, these models combine static and dynamic analysis, supervised and unsupervised learning, and traditional and deep learning techniques [204,205]. For example, A study published in 2023 used Hierarchical Neural Networks for cross-platform ransomware fingerprinting with hybrid features, achieving remarkable accuracy [206]. Combining static and dynamic analyses significantly improved detection rates and resilience against evasion tactics, as demonstrated by Hasan and Rahman (2017). DT, SVM, RF, and AdaBoost ensemble learning techniques consistently outperform individual classifiers in terms of accuracy and F1-score [207]. Based on a variety of scenarios, Ahmed et al. reported that their ensemble model was superior at identifying ransomware applications. A hybrid model allows XAI techniques like SHAP and LIME to be integrated into classification decisions, increasing transparency by spotlighting the most influential features. To balance complexity, scalability, and interpretability, hybrid approaches require careful design and tuning [207,210].

 doi:10.20944/preprints202602.2011.v1

However, each AI technique comes with its own set of trade-offs. A supervised model offers high accuracy, but requires constant retraining and curated datasets. The unsupervised method detects new threats, but it suffers from false alarms [208]. The deep learning model captures complex patterns, but is computationally expensive and lacking in interpretability. Hybrid and ensemble approaches offer the best performance, but they also demand more resources and increase system complexity [210]. Additionally, all models face challenges related to concept drift - as ransomware evolves, models trained on historical data become less effective - and adversarial evasion, where attackers manipulate features to avoid detection. It has become increasingly difficult to detect Ransomware-as-a-Service (RaaS) platforms, such as LockBit, BlackCat (ALPHV), and RansomHub, as these variants exhibit high modularity and rapid iteration, as well as anti-AI evasion techniques [210].

In order to address these challenges, recent research emphasizes the integration of Explainable AI (XAI), transfer learning, and federated learning. Security analysts can use XAI tools to enhance model transparency, which improves forensic investigations and compliance by letting them know what features (e.g., "vssadmin delete shadows" plus "CryptEncrypt") triggered alerts [208]. This approach reduces training time and data requirements while maintaining high accuracy - Almomani et al. (2023) reported up to 99.5% accuracy using pre-trained models (e.g., ResNet, BERT). In sectors like healthcare and finance, federated learning facilitates collaborative model training across organizations without exposing sensitive raw data, preserving privacy while improving generalizability [209].



G. Performance Evaluation:

A successful AI-based ransomware detection system depends heavily on the selection and implementation of evaluation metrics that can accurately measure the system's performance across multiple dimensions. Precision measures the percentage of true positive detections among all positive predictions made by the system. With regards to ransomware detection, precision answers the question of how many of the flagged suspicious activities are actually ransomware attacks, thus indicating the system's ability to minimize false alarms [211]. Having a high precision rate is especially important in enterprise environments, where false positives can cause system disruptions, user frustration, and resource waste. In reality, precision itself does not give a complete picture, since a system could achieve high precision by being too conservative and identifying only the most obvious ransomware signatures and missing more sophisticated or novel attacks in the process [212].

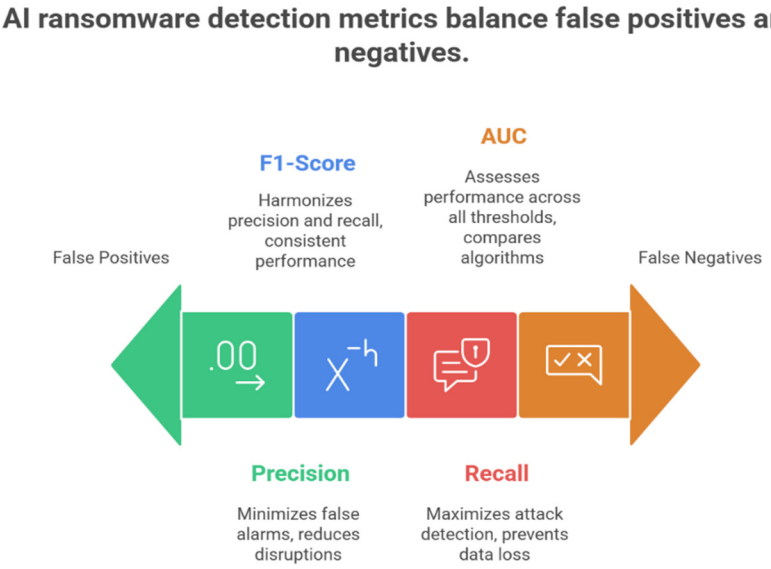
Recall, also known as sensitivity or true positive rate, indicates the proportion of actual ransomware attacks that the system is able to identify. From a security standpoint, this metric is arguably more critical since even a single ransomware attack could result in catastrophic data loss, system compromise, and significant financial losses [213]. A detection system with a high recall rate

 doi:10.20944/preprints202602.2011.v1

ensures comprehensive coverage of ransomware variants, such as zero-day attacks and polymorphic malware, which are difficult to detect using signature-based detection methods. Recall and precision must be balanced, as increasing sensitivity frequently leads to an increase in false positives. The optimal recall threshold should be determined by weighing the consequences of missed attacks against the burden of investigating false alarms, taking into account the organization's risk tolerance and operational capacity [214].

In order to provide a balanced assessment of both precision and recall, the F1-score provides a harmonizing metric that is based on a harmonic mean. This composite metric is particularly useful when dealing with imbalanced datasets, which are common in cybersecurity applications where malicious samples represent a relatively small percentage of the total data set. F1 scores penalize extreme values in either precision or recall, encouraging the development of detection systems that perform consistently on both dimensions [215]. Although the standard F1-score assumes equal importance for precision and recall, this may not align with the priorities of specific organizations. Therefore, weighted variants of the F1-score, such as F-beta scores, facilitate customization, depending on whether minimizing false negatives or false positives is more important for a particular deployment scenario [216].

AUC represents a more comprehensive evaluation metric that assesses the detection system's performance across all classification thresholds measured by the Receiver Operating Characteristic (ROC) curve. ROC curves plot the true positive rate against the false positive rate, providing insight into how specificity and sensitivity are balanced at various decision points. If the AUC value is 1.0, then the classification performance is perfect, while if it is 0.5, then the classification performance is similar to random guessing [217]. Comparing different detection algorithms or model architectures with the AUC metric proves particularly useful, since it measures discriminative ability independent of thresholds. Moreover, the AUC-PR (Area Under the Precision-Recall Curve) provides an alternative perspective that may be more informative for imbalanced datasets, as it focuses on positive class performance without being influenced by the large number of true negatives typically present in cybersecurity datasets [218].



A Benchmarking Methodology and Cross-Validation for Model Robustness Assessment

Developing robust benchmarking methodologies requires careful consideration of dataset characteristics, experimental design, and evaluation protocols that ensure fair and meaningful comparisons between different approaches. Benchmarking relies on the collection of comprehensive and representative datasets that capture the diversity of ransomware variants, attack vectors, and benign system behaviours [219]. As well as covering historical ransomware families, contemporary threats, and emerging attack patterns, these datasets must also maintain appropriate class

 doi:10.20944/preprints202602.2011.v1

distributions that reflect real-world scenarios. As ransomware evolves rapidly and continuously, updating datasets regularly is essential, as are time-based splits that simulate realistic deployment conditions where models encounter future, unknown threats [220].

Data preprocessing, feature extraction, model training, and evaluation procedures should be standardized in benchmarking protocols in order to address the reproducibility crisis that affects many machine learning applications. Missing data specifications, normalizing features, addressing class imbalances, and training-validation-testing splits are also included [221]. Assuring fair comparison opportunities across different methodological paradigms needs to be a goal of the benchmarking framework, which should accommodate all types of AI approaches, including traditional machine learning algorithms, deep learning architectures, and ensemble methods. In addition to pure detection accuracy, the evaluation should include computational efficiency, scalability, and deployment feasibility. Real-world systems have limited resources and strict latency requirements, so these factors should be considered in the evaluation [222].

It is crucial to use cross-validation methodologies to assess ransomware detection models' robustness and generalizability, allowing insight into how they will perform in unseen data environments and different operational environments. K-fold cross-validation, while widely used in machine learning, requires careful adaptation for cybersecurity applications due to the temporal nature of threats and the possibility of data leakage [223]. In scenarios where models must detect future attacks based on historical training data, time-series cross-validation methods such as forward chaining or sliding window validation work better. When faced with evolving threat landscapes, these temporal validation strategies can identify models that perform well on randomly shuffled datasets but fail when faced with evolving threat landscapes [223].

In ransomware detection research, balanced accuracy and G-means have gained traction beyond traditional metrics and validation strategies. Balanced accuracy compensates for class imbalances by averaging recall across classes, offering a more accurate assessment than raw accuracy [224]. Using G-means, which combines sensitivity and specificity, one can assess how well the model performs across both classes. When evaluating models on datasets with skewed distributions, such as those dominated by benign software, these metrics are particularly useful. There was a preference for balanced accuracy over raw accuracy in the reviewed literature, acknowledging that high overall accuracy can be misleading if the model fails to detect ransomware correctly. In practice, a model that correctly identifies 95% of benign samples but misses 50% of ransomware instances may still report high accuracy [225].

The Area Under Time (AUT) metric assesses model performance over time by accounting for concept drift, the gradual change in data distribution caused by evolving ransomware tactics. Continuous learning settings, where models need to adapt to new threats without retraining, make AUT particularly useful [226]. Detection systems can be evaluated for longevity and adaptability by tracking performance across time slices. As attackers continuously innovate to bypass defences, this is crucial in cybersecurity. By including samples from 2006 to 2024, the MLRan study was able to analyse model performance longitudinally, emphasizing the importance of temporal diversity in training data [226].

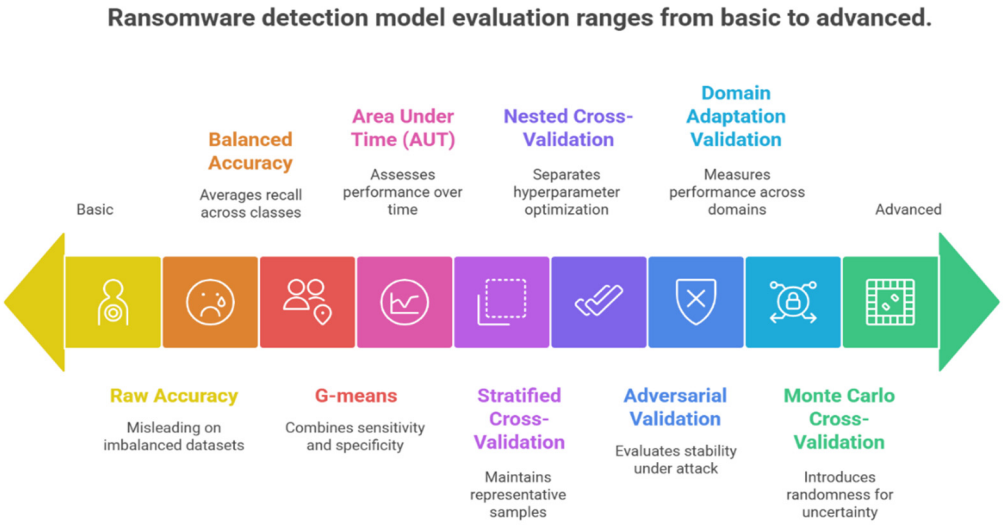
When dealing with imbalanced datasets and multiple ransomware families, stratified cross-validation is essential to maintaining representative samples of different attack types and benign behaviours. The approach prevents the creation of evaluation scenarios where certain ransomware variants are absent from training data, resulting in overly optimistic or pessimistic performance estimates [227]. A nested cross-validation technique, for example, provides more robust estimates of model performance by separating hyperparameter optimization and final model evaluation, reducing overfitting [227].

Cross-validation goes beyond simple performance estimation to include robustness assessment across different operational conditions, network environments, and system configurations. While adversarial validation can evaluate model stability when inputs are perturbed or evasion attempts are made, domain adaptation validation measures performance degradation when data originates

 doi:10.20944/preprints202602.2011.v1

from different organizational contexts or network architectures [228]. A Monte Carlo cross-validation approach introduces additional randomness to data splitting procedures, resulting in confidence intervals and uncertainty estimates that inform deployment decisions and risk assessments. Using these comprehensive validation methodologies, AI-based ransomware detection systems will demonstrate consistent and reliable performance across diverse real-world scenarios, thus improving cybersecurity defences [229].

As a summary, performance evaluation in AI-based ransomware detection can be viewed as a multifaceted task that requires careful selection of metrics, rigorous validation strategies, and the consideration of benchmarks. In order to gather deeper insight into the behaviour of a model, it is important to consider metrics such as precision, recall, F1-score, and AUC in addition to balanced accuracy, FPR, FNR, and AUT. It is imperative to use cross-validation and time-aware train-test splits to ensure robustness and generalizability, while adversarial testing is used to expose any vulnerabilities that may exist. As part of the evaluation process, benchmarking against strong baselines and leveraging AutoML tools further enhances the reliability of the evaluation process. It is expected that, as ransomware continues to develop, these methodologies will become indispensable for the development of robust, accurate, efficient, and effective detection systems that will be able to protect digital infrastructure from increasingly sophisticated attacks.



IV. Challenges and Limitations in Existing Literature and Future Research Directions

Challenges and Limitations:

Gaps and Limitations in Current Research

There are significant methodological challenges associated with AI-based ransomware detection that constrain their practical applicability and reliability. Overfitting is one of the most prevalent problems, in which models exhibit exceptional performance on training and validation datasets, but fail catastrophically when deployed in real-world scenarios [230]. Cybersecurity applications are particularly prone to this phenomenon as a result of the complex, high-dimensional feature spaces they typically use and the relative scarcity of diverse, labeled ransomware samples. Researchers often use synthetic data augmentation techniques or feature engineering techniques that inadvertently introduce artificial patterns, resulting in models that are better at recognizing engineered characteristics rather than genuine ransomware behaviours [231]. Moreover, this problem is compounded by the tendency to optimize models based on benchmark datasets that may not adequately reflect the full range of ransomware variants, attack vectors, and operational

 doi:10.20944/preprints202602.2011.v1

environments encountered in practice. Furthermore, as ransomware attacks evolve rapidly, static training datasets become outdated quickly, resulting in a mismatch between the model training data and the threats encountered during deployment [232]. It is particularly concerning that ransomware developers actively adapt their techniques in order to evade detection systems, which creates a hostile environment where yesterday's training data may be irrelevant to the threats of tomorrow [232,233].

Other fundamental limitations of current research approaches include their inability to generalize across different organizational contexts, network architectures, and system configurations. Most studies focus on controlled laboratory environments or specific organizational settings, not accounting for the heterogeneous nature of real-world IT infrastructures [234]. Detection systems optimized for traditional network architectures may struggle in cloud-native or containerized deployments, while models trained on Windows-centric environments may not perform well in mixed OS environments. As static datasets capture threats and system behaviours at a specific moment in time and within specific contexts, this issue is exacerbated by the reliance on static datasets [235,236]. While most research fails to address this fundamental challenge adequately, deployed systems deteriorate in performance as the threat landscape evolves without continuous learning mechanisms and adaptive capabilities. In addition, the evaluation methodologies employed often lack ecological validity, since they use artificial test scenarios that don't reflect the complexity, noise, and variability of production settings [237]. A significant gap exists between theoretical performance claims and practical deployment outcomes due to this disconnect between research conditions and operational reality [237].

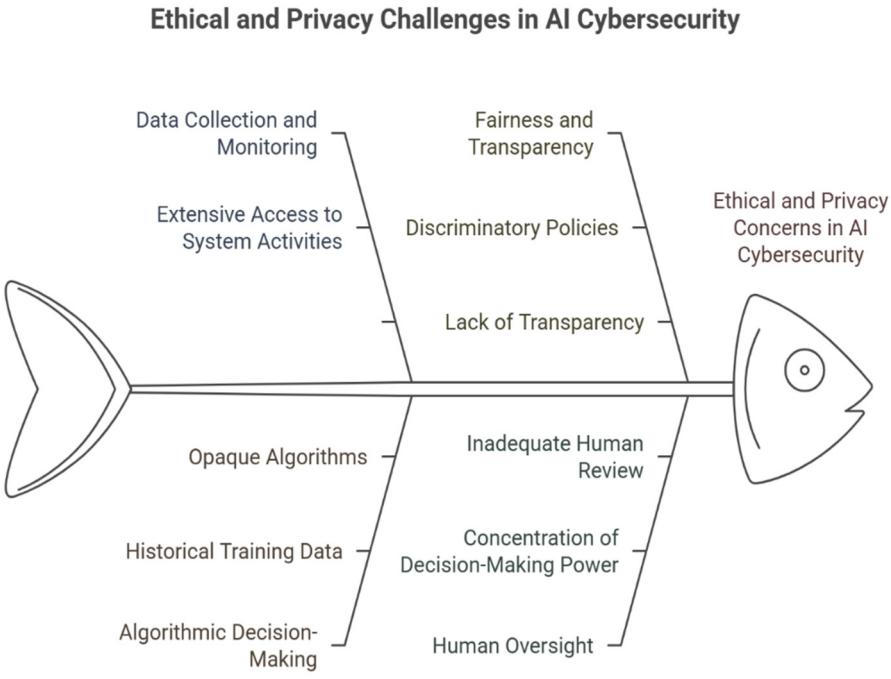
Ethical Considerations and Privacy Concerns

Organizations must carefully navigate the tensions between protection and surveillance created by AI-driven cybersecurity solutions, which go far beyond traditional security paradigms. In order to detect ransomware effectively, AI-based systems need extensive access to system activities, network communications, file operations, and user behaviors, providing unprecedented opportunities for data collection and monitoring [238]. Especially in environments where users expect a certain level of privacy and autonomy, this pervasive surveillance capability raises fundamental concerns. When behavioral profiling is used to detect threats effectively, it is possible to inadvertently capture sensitive personal information, proprietary business processes, and confidential communications, which can result in data misuse, unauthorized surveillance, or

 doi:10.20944/preprints202602.2011.v1

inappropriate disclosure [239]. In order to maintain effective security monitoring, organizations must adhere to data protection regulations like GDPR and CCPA, protect employee privacy, and maintain stakeholder trust [239].

AI-driven cybersecurity solutions introduce additional ethical challenges related to fairness, transparency, and accountability in security operations due to algorithmic decision-making. Based on historical training data that reflects past inequities or incomplete representations, machine learning models may disproportionately flag certain user groups, application types, or operational patterns [240]. As a result, security policies may be enforced discriminatorily, legitimate activities may be unfairly targeted, and users may be treated differently based on factors that aren't related to actual security risks. Many AI algorithms, particularly deep learning approaches, are opaque, making it challenging to explain security decisions and provide meaningful recourse [241]. The lack of transparency becomes particularly problematic when AI systems make automated decisions that can impact user productivity, system availability, or business operations. Furthermore, the concentration of security decision-making power in algorithmic systems raises concerns about human agency and oversight in critical security functions, potentially creating scenarios in which automated systems make consequential decisions without appropriate human review or intervention capabilities [241,242]. The implementation of robust governance frameworks ensures ethical AI deployment, maintains human oversight of critical decisions, identifies and mitigates biases, and provides transparent mechanisms for addressing system errors and unintended consequences while balancing these concerns with the need to protect against increasingly sophisticated ransomware threats [242].



Future Research Directions:

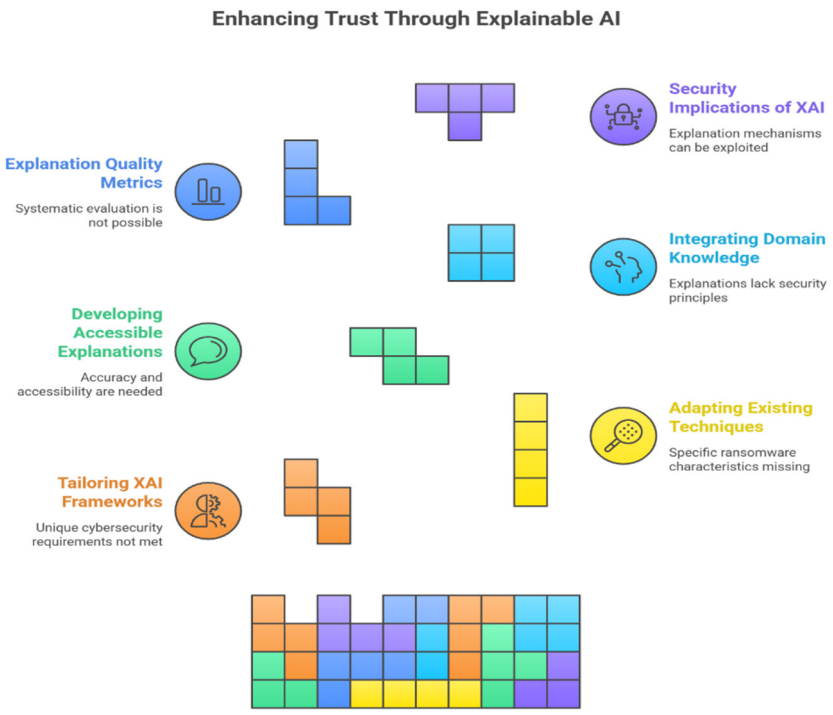
Integrating Explainable AI for Transparent Decision-Making and Enhanced Trust

AI-based ransomware detection systems can be advanced through the integration of explainable artificial intelligence (XAI), addressing the fundamental problem of algorithmic transparency that currently limits adoption and trust in automated cybersecurity systems. Research in the future should focus on the development of XAI frameworks specifically tailored to the unique requirements of cybersecurity applications, in which decision transparency is not merely desirable but essential for effective incident response, forensic analysis, and regulatory compliance [243]. XAI algorithms designed specifically for ransomware should include both local explanations to illuminate individual detection decisions as well as global explanations to reveal the patterns and features the model

 doi:10.20944/preprints202602.2011.v1

considers to be most indicative of malicious activity. Through this dual-level transparency approach, security analysts will be able to understand not only why a particular file or behavior was flagged as suspicious, but also how the detection system's decision-making process aligns with established cybersecurity principles and threat intelligence [244]. There is a need for research to examine how existing XAI techniques, such as LIME, SHAP, and attention mechanisms, can be adapted to the specific characteristics of ransomware detection, including temporal behaviour analysis, file system interaction patterns, and network communication signatures. Security professionals with varying levels of machine learning expertise face the challenge of developing explanations that are technically accurate and accessible, requiring interdisciplinary collaboration between AI researchers, cybersecurity practitioners, and human-computer interaction specialists [245].

Research in advanced XAI should develop counterfactual explanations that demonstrate how changes in system behaviour would alter detection decisions, providing valuable insights for threat analysis and evasion resistance assessment. A system of interactive explanations that allows security analysts to query the detection model about specific features, behaviours, or decision pathways would greatly improve the interpretability and utility of AI-driven security tools [246]. Moreover, XAI frameworks should integrate domain knowledge and cybersecurity ontologies to ensure that explanations are grounded in established security principles rather than purely statistical correlations. It would be possible to evaluate and improve XAI approaches systematically if explanation quality metrics were developed specifically for cybersecurity applications, while studies on explanation effectiveness in real-world security operations would inform how best to present AI-driven insights to human analysts [247]. Additionally, future work should examine the potential security implications of XAI itself, as well as how explanation mechanisms can be exploited by attackers in order to evade detection systems, and develop ways to maintain security through obscurity while providing meaningful transparency [248].



Developing Real-Time Detection Systems with Adaptive Capabilities

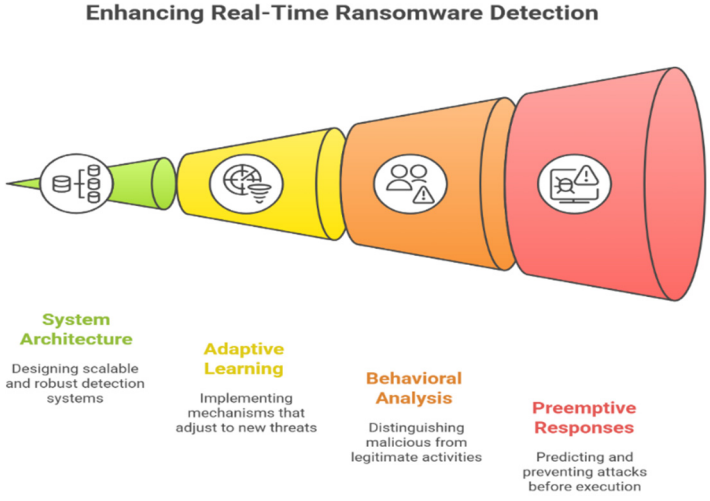
Real-time ransomware detection systems represent a paradigm shift from reactive cybersecurity to proactive cybersecurity. This requires fundamental advancements in algorithmic efficiency, system architecture, and adaptive learning mechanisms that can operate within the stringent latency and

 doi:10.20944/preprints202602.2011.v1

resource constraints associated with production environments [249]. Research in the future must deal with the computational complexity associated with real-time threat detection, exploring novel approaches such as early warning systems that can identify ransomware indicators before full payload execution, lightweight neural network architectures optimized for edge deployment, and hierarchical detection frameworks that balance accuracy with computational efficiency through progressive analysis stages [250]. The development of streaming machine learning algorithms capable of processing continuous data flows while maintaining detection accuracy presents significant technical challenges, especially in handling concept drift, managing memory constraints, and maintaining model stability under varying workload conditions [250]. Developing adaptive threshold mechanisms that can respond to evolving threat patterns without requiring the complete retraining of models or system downtime should be investigated through the use of online learning techniques, incremental model updates, and adaptive threshold mechanisms [251].

As a result of subtle timing patterns, resource utilization signatures, and sequences of system interaction, advanced real-time detection systems must incorporate sophisticated behavioral analysis capabilities that can distinguish between legitimate administrative activities and malicious encryption processes [252]. Researchers should explore multi-modal sensing approaches that combine file system monitoring, network traffic analysis, system call tracing, and hardware performance metrics to create comprehensive real-time threat detection capabilities. It could be possible to develop preemptive responses that prevent successful attacks rather than merely detecting them after they start, using predictive models that predict ransomware deployment based on precursor behaviours, such as reconnaissance behaviours, privilege escalation attempts, or lateral movement patterns [253]. Adaptive response mechanisms based on real-time threat assessment could reduce the impact of successful ransomware attacks significantly by automatically implementing containment measures, isolating affected systems, or triggering backup procedures. In complex enterprise environments where false positives can cause unnecessary business disruptions, the challenge lies in balancing automated response capabilities with human oversight and approval [253,254].

Furthermore, future research should investigate the integration of context-aware detection systems that can adapt their sensitivity and response strategies based on factors such as system criticality, user roles, time of day, and current threat intelligence feeds. A promising research direction is the development of collaborative real-time detection networks that can share threat indicators and behavioral signatures across organizational boundaries while maintaining privacy and confidentiality. Further, research into resilient detection architectures with the ability to maintain functionality when partially compromised by sophisticated attackers will ensure continuous protection [254].



 doi:10.20944/preprints202602.2011.v1

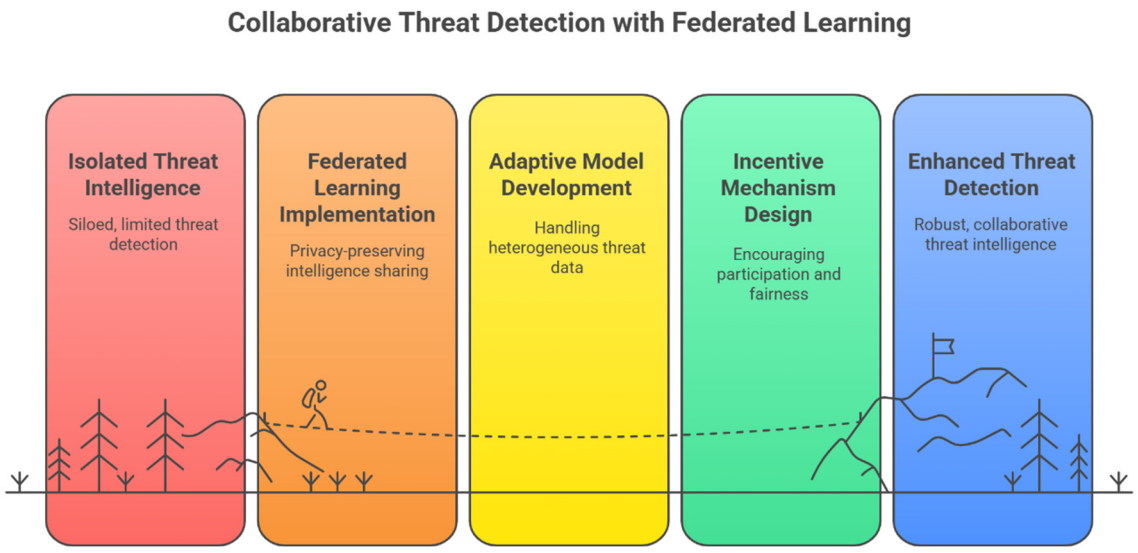
Leveraging Federated Learning for Collaborative Threat Detection and Privacy-Preserving Intelligence Sharing

A transformative approach to ransomware detection is emerging through federated learning, which provides the opportunity to harness collective threat intelligence while simultaneously addressing privacy, confidentiality, and competitive concerns that traditionally limit the sharing of information between organizations [255]. A critical component of future research will be the development of federated learning frameworks that are specifically designed for cybersecurity applications, addressing the unique challenges associated with non-IID (non-independent and identically distributed) threat data, varying organizational security postures, and rapid model convergence in dynamic threat environments [255]. It is important to note that heterogeneous IT infrastructures, security tools, and threat exposure patterns can create significant statistical heterogeneity that traditional federated learning approaches may not be able to adequately deal with, necessitating the development of federated learning techniques that can adapt global threat models to local organizational contexts while still benefiting from collaborative intelligence sharing. Research should explore advanced aggregation methods that can effectively combine threat detection models from organizations with vastly different network architectures, user behaviours, and threat exposure profiles to create a global model that offers value to all participants regardless of their individual characteristics [256,257].

Privacy-preserving mechanisms that facilitate threat intelligence sharing without exposing sensitive organizational information represents a critical research frontier with significant implications for industry-wide cybersecurity collaboration. In the future, differential privacy, homomorphic encryption, and multiparty computation techniques should be applied to federated ransomware detection, allowing organizations to contribute to collective threat intelligence while maintaining strict confidentiality about their internal security postures, incident histories, and operational details [258]. Developing incentive mechanisms that encourage participation in federated threat detection networks while ensuring fair distribution of benefits and preventing free-riding behaviour could speed up the adoption and effectiveness of collaborative security approaches [258].

Advanced federated learning research should explore the integration of continuous learning capabilities to enable threat detection models to adapt rapidly to emerging ransomware variants discovered by any network participant, while propagating threat intelligence across the federation in near real-time and maintaining privacy protections [259]. In order to achieve more granular and targeted threat intelligence while maintaining scalability and efficiency, hierarchical federated learning architectures that can operate across multiple organizational levels from individual business units to industry sectors to global threat sharing networks may be developed [259,260]. By leveraging threat intelligence from a variety of sectors and organizational types, cross-domain federated learning approaches can enhance the robustness and generalizability of ransomware detection models, especially those that detect new attack vectors that may appear initially in a particular industry before spreading widely. The future work should also address the regulatory and legal implications of federated threat detection systems, ensuring compliance with data protection regulations while maximizing the security benefits of collaborative threat intelligence sharing [259,260].

 doi:10.20944/preprints202602.2011.v1



V. Conclusion

Summary of Findings:

Summary of Findings In this systematic assessment of AI-based ransomware detection frameworks, numerous critical insights into the current state and future trajectory of cybersecurity defence mechanisms are revealed. In the comprehensive analysis, it is evident that machine learning and deep learning techniques have made remarkable strides in the detection of ransomware, with supervised learning techniques consistently achieving the highest accuracy rates, often exceeding 97% when applied to well-curated datasets. The Random Forest classifiers have been shown to be particularly effective, demonstrating superior performance in handling high-dimensional spaces while providing interpretable results that enable security analysts to determine the rationale behind feature detection. By using transfer learning approaches, deep learning architectures, such as Convolutional Neural Networks and Long Short Term Memory networks, are capable of extracting complex patterns from raw data with an accuracy rate of 99.5%. However, the research also reveals significant limitations in current methodologies, including reliance on static datasets that fail to capture the dynamic nature of evolving ransomware threats, an inadequate representation of modern attack vectors, such as Ransomware-as-a-Service platforms, and a limited ability to generalize across diverse operational environments. According to the evaluation framework, hybrid approaches combining static and dynamic analysis consistently outperform single-method detection systems, while ensemble learning techniques exhibit superior robustness against evasion. This study emphasizes that continuous innovation in artificial intelligence techniques is not only advantageous, but crucial for countering the rapidly evolving ransomware landscape, in which adversaries are constantly adapting their strategies to circumvent detection systems using sophisticated evasion techniques, polymorphic code generation, and analysis techniques.

Implications for Practice:

These findings have profound implications for cybersecurity practitioners, policymakers, and researchers across a variety of fields. For security practitioners, the evaluation framework provides actionable guidance for selecting and implementing AI-based detection systems tailored to their specific organizational contexts, emphasizing the importance of hybrid approaches that integrate behavioral monitoring capabilities with static analysis capabilities. According to the findings, organizations should move beyond traditional signature-based detection methods and adopt adaptive machine learning systems that are capable of continuous learning and real-time threat adaptation. These insights can be leveraged by policymakers to develop regulatory frameworks that promote collaborative threat intelligence sharing while addressing privacy concerns through

 doi:10.20944/preprints202602.2011.v1

federated learning approaches that allow organizations to benefit from collective security knowledge without compromising sensitive operational data. There is a need for standardized evaluation protocols and benchmark datasets that accurately represent contemporary threat landscapes, which suggests that regulatory bodies should mandate minimum standards for the testing and validation of AI-based security systems. Through the use of explainable AI integration, the development of real-time detection systems, and collaborative learning mechanisms that protect privacy, the systematic evaluation reveals numerous opportunities for researchers to advance the field. By providing standardized methodologies for data curation, feature extraction, model evaluation, and performance benchmarking, the framework provides a foundation for developing more effective ransomware detection systems. Using this framework, organizations can develop comprehensive defence strategies that balance detection accuracy with computational efficiency, while ensuring security systems remain interpretable and auditable for forensic analysis and regulatory compliance.

Final Remarks:

Final Remarks By providing a roadmap for future innovations, this research makes a valuable contribution to cybersecurity by providing a comprehensive framework to evaluate AI-based ransomware detection. As a result of the systematic analysis, the cybersecurity community is at a critical point in which traditional reactive defence mechanisms must evolve into proactive, intelligent systems capable of adapting to a sophisticated and rapidly changing threat landscape. As well as technical implementation, this work also raises fundamental questions about the role of artificial intelligence in protecting digital infrastructure, how to balance automated decision-making with human oversight in security operations, and what ethical implications pervasive monitoring systems can have on security operations. This study highlights the importance of continuing collaboration between academia, industry, and government agencies in order to develop robust, transparent, ethical cybersecurity solutions by demonstrating both the tremendous potential and current limitations of AI-based approaches. Researchers found that effective ransomware defence requires a holistic strategy addressing organization preparedness, user education, regulatory compliance, and international cooperation in combating cybercrime as well as technological advancement. As ransomware threats continue to evolve in complexity and impact, the systematic evaluation framework presented in this research provides essential guidance for building resilient cybersecurity ecosystems able to protect critical digital assets and maintain societal trust in increasingly interconnected technological systems. Ultimately, the cybersecurity community must be able to maintain technological innovation pace in order to counter ransomware threats while ensuring that defence mechanisms remain accessible, interpretable, and aligned with broader social values such as privacy, security, and digital rights in order to succeed.

Author Contributions: Conceptualization, H.Q.; and Y.Y.; methodology, H.Q.; validation, Y.L., and H.Q.; formal analysis, H.Q.; investigation, H.Q.; resources, H.Q.; data curation, H.Q.; writing—original draft preparation, H.Q.; writing—review and editing, H.Q.; visualization, H.Q.; supervision, Y.L.; project administration, H.Q. All authors have read and agreed to the published version of the manuscript.

Funding: "This research received no external funding"

Conflicts of Interest: "The authors declare no conflicts of interest."

References

1. Alzahrani B, Alshamrani S, Alghamdi A. AI-Based Ransomware Detection: A Comprehensive Review. *Charles Sturt University Research Output*. 2023. Available from: https://researchoutput.csu.edu.au/files/544062743/532572193_Published_article.pdf
2. Kim J, Park S, Lee H. Ransomware Detection with Machine Learning: Techniques, Challenges, and Future Directions. *Journal of Information Security and Intelligent Systems*. 2025;1(1):17. Available from: https://jisis.org/wp-content/uploads/2025/04/2025.I1.017.pdf

 doi:10.20944/preprints202602.2011.v1

3. Singh A, Sharma R, Kumar V. Ransomware Detection Using Machine Learning: A Review, Research Challenges and Future Directions. *IEEE Xplore*. 2024. Available from: https://ieeexplore.ieee.org/document/10521643
4. Alraizza A, Algarni A. Ransomware Detection Using Machine Learning: A Survey. *Big Data Cogn Comput*. 2023;7(3):143. Available from: https://www.mdpi.com/2504-2289/7/3/143
5. Patel H, Mehta P. Machine Learning Approaches to Ransomware Detection: A Comprehensive Review. *International Information and Engineering Technology Association*. 2023. Available from: https://iieta.org/download/file/fid/153100
6. Vinayakumar R, Alazab M, Soman KP, Poornachandran P, Al-Nemrat A, Venkatraman S. Deep learning approach for intelligent intrusion detection system. IEEE Access. 2019;7:41525-50. Available from: https://ieeexplore.ieee.org/document/8681044
7. Harzie RE, Selamat A, Fujita H, Krejcar O, Do NQ. Enhancing Ransomware Detection Using Deep Learning Models. In: *Advances and Trends in Artificial Intelligence. Theory and Applications*. Lecture Notes in Computer Science. Springer; 2025. p. 403–414. Available from: https://link.springer.com/chapter/10.1007/978-981-96-8892-0_34
8. Ransomware Detection: Challenges and Techniques. In: *Proceedings of the 2025 IEEE Conference on Computing, Informatics, and Cybersecurity Technologies (CCICT)*. IEEE; 2025. p. 319–324. Available from: https://www.computer.org/csdl/proceedings-article/ccict/2025/113500a319/28KfgQ6t5a8
9. Alzahrani B, Alshamrani S, Alghamdi A. Understanding and Mitigating Ransomware Threats: Trends, Techniques, and Countermeasures. *2025 IEEE International Conference on Cybersecurity and Resilience (CyberRes)*. Available from: https://ieeexplore.ieee.org/document/10585140
10. Gómez Hernández JA, García Teodoro P, Magán Carrión R, Rodríguez Gómez R. Crypto-Ransomware: A Revision of the State of the Art, Advances and Challenges. *Electronics*. 2023;12(21):4494. Available from: https://www.mdpi.com/2079-9292/12/21/4494
11. Arnone G, Scire' G, Bivona E. The (mis)use of cryptocurrencies by criminal organizations: a systematic literature review. *Digital Finance*. 2025. Available from: https://link.springer.com/article/10.1007/s42521-025-00148-1
12. Rawindaran N, Jayal A, Prakash E. Exploration of the Impact of Cybersecurity Awareness on Small and Medium Enterprises (SMEs) in Wales Using Intelligent Software to Combat Cybercrime. *Computers*. 2022;11(12):174. Available from: https://www.mdpi.com/2073-431X/11/12/174
13. Gómez Hernández JA, García Teodoro P, Magán Carrión R, Rodríguez Gómez R. Ransomware Anti-Analysis and Evasion Techniques: A Survey and Research Directions. *2021 IEEE International Conference on Cyber Security and Protection of Digital Services (Cyber Security)*. Available from: https://ieeexplore.ieee.org/document/9392529
14. Alzahrani B, Alshamrani S, Alghamdi A. A Survey of Ransomware as a Service (RaaS) and Methods to Mitigate the Threat. *2022 IEEE International Conference on Cybersecurity and Resilience (CyberRes)*. Available from: https://ieeexplore.ieee.org/document/9719456
15. Maratsi MI, Popov O, Alexopoulos C, Charalabidis Y. Ethical and Legal Aspects of Digital Forensics Algorithms: The Case of Digital Evidence Acquisition. *ICEGOV 2022: 15th International Conference on Theory and Practice of Electronic Governance*. ACM; 2022. Available from: https://dl.acm.org/doi/fullHtml/10.1145/3560107.3560114
16. Enomoto S, Kuzuno H, Yamada H, Shiraishi Y, Morii M. Early mitigation of CPU-optimized ransomware using monitoring encryption instructions. *Int J Inf Secur*. 2024;23:3393–3413. Available from: https://link.springer.com/article/10.1007/s10207-024-00892-2
17. Butt U, Dauda Y, Shaheer B. Ransomware Attack on the Educational Sector. In: *AI, Blockchain and Self-Sovereign Identity in Higher Education*. Springer; 2023. p. 279–313. Available from: https://link.springer.com/chapter/10.1007/978-3-031-33627-0_11
18. Iwasaki M. Banning ransomware payments: unintended effects on cybersecurity investment and incident reporting. *Int Cybersecurity Law Rev*. 2025;6:17–27. Available from: https://link.springer.com/article/10.1365/s43439-025-00137-5

19. Halikias H. Business Impacts of Ransomware. In: *Digital Shakedown*. Springer; 2024. p. 25–47. Available from: https://link.springer.com/chapter/10.1007/978-3-031-65438-1_3
20. Warren DF, Komninos N, Chen T. A Study on the Evolution of Ransomware Detection Using Machine Learning and Deep Learning Techniques. *IoT*. 2020;1(2):551–604. Available from: https://www.mdpi.com/2624-831X/1/2/30
21. Ciaramella G, Iadarola G, Martinelli F, Mercaldo F, Santone A. Explainable Ransomware Detection with Deep Learning Techniques. *J Comput Virol Hack Tech*. 2024;20:317–330. Available from: https://link.springer.com/article/10.1007/s11416-023-00501-1
22. Urooj U, Al-rimy BAS, Zainal A, Ghaleb FA, Rassam MA. Ransomware Detection Using the Dynamic Analysis and Machine Learning: A Survey and Research Directions. *Appl Sci*. 2022;12(1):172. Available from: https://www.mdpi.com/2076-3417/12/1/172
23. Singh A, Mushtaq Z, Abosaq HA, Mursal SNF, Irfan M, Nowakowski G. Enhancing Ransomware Attack Detection Using Transfer Learning and Deep Learning Ensemble Models on Cloud-Encrypted Data. *Electronics*. 2023;12(18):3899. Available from: https://www.mdpi.com/2079-9292/12/18/3899
24. Pelekis S, Koutroubas T, Blika A, Berdelis A, Karakolis E, Ntanos C, Spiliotis E, Askounis D. Adversarial Machine Learning: A Review of Methods, Tools, and Critical Industry Sectors. *Artif Intell Rev*. 2025;58:226. Available from: https://link.springer.com/article/10.1007/s10462-025-11147-4
25. McCarthy A, Ghadafi E, Andriotis P, Legg P. Functionality-Preserving Adversarial Machine Learning for Robust Classification in Cybersecurity and Intrusion Detection Domains: A Survey. *J Cybersecur Priv*. 2022;2(1):154–190. Available from: https://www.mdpi.com/2624-800X/2/1/10
26. Fernando D, Komninos N, Chen T. Comparative Study of CNN and RNN for Deep Learning Based Intrusion Detection System. In: *Cloud Computing and Security*. Springer; 2018. p. 159–170. Available from: https://link.springer.com/chapter/10.1007/978-3-030-00018-9_15
27. Sharmeen N, Islam S, Rahman MM, Hossain MS. Comparative Analysis of Machine Learning and Deep Learning Models for Ransomware Detection. In: *Proceedings of Fifth International Conference on Computer and Communication Technologies (IC3T 2023)*. Springer; 2024. p. 253–266. Available from: https://link.springer.com/chapter/10.1007/978-981-99-9707-7_24
28. Rhode M, Burnap P, Jones K. Early-stage malware prediction using recurrent neural networks. Comput Secur. 2018;77:578-94. Available from: Early-stage malware prediction using recurrent neural networks - ScienceDirect
29. Homayoun S, Dehghantanha A, Ahmadzadeh M, Hashemi S, Khayami R. Know abnormal, find evil: frequent pattern mining for ransomware threat hunting and intelligence. IEEE Trans Emerg Topics Comput. 2020;8(2):341-51. Available from: Homayoun, S., Dehghantanha, A., Ahmadzadeh, M., Hashemi, S. and Khayami, R. (2020) Know Abnormal, Find Evil Frequent Pattern Mining for Ransomware Threat Hunting and Intelligence. IEEE Transactions on Emerging Topics in Computting, 8, 341-351. - References - Scientific Research Publishing
30. Zhang Y, Zhang Y, Li Y, Wang Y. Ransomware Detection Using Machine Learning Algorithms. *2025 IEEE International Conference on Cybersecurity and Resilience (CyberRes)*. Available from: https://xplorestaging.ieee.org/document/10652659
31. Baldwin J, Dehghantanha A. Leveraging Support Vector Machine for Opcode Density Based Detection of Crypto-Ransomware. In: *Cyber Threat Intelligence*. Springer; 2018. p. 107–136. Available from: https://link.springer.com/chapter/10.1007/978-3-319-73951-9_6
32. Panda S, Sahu S, Jena P, Chattopadhyay S. Comparing Fuzzy-C Means and K-Means Clustering Techniques: A Comprehensive Study. In: *Advances in Computer Science, Engineering & Applications*. Springer; 2012. p. 451–460. Available from: https://link.springer.com/chapter/10.1007/978-3-642-30157-5_45
33. Hussain I, Sinaga KP, Yang MS. Unsupervised Multiview Fuzzy C-Means Clustering Algorithm. *Electronics*. 2023;12(21):4467. Available from: https://www.mdpi.com/2079-9292/12/21/4467
34. Sharma P, Chaudhary K. An Advanced Comparative Study of Ransomware Anomaly Detection Techniques Through Optimized Hyperparameters. In: *Artificial Intelligence and Sustainable Computing*. Springer; 2024. p. 379–393. Available from: https://link.springer.com/chapter/10.1007/978-981-97-0327-2_28

 doi:10.20944/preprints202602.2011.v1

35. Sharmeen N, Islam S, Rahman MM, Hossain MS. Deep Learning Approaches for Ransomware Detection: Assessing CNN and Traditional Classifiers. *J Neural Inf Process Tech*. 2020;2(4):112–124. Available from: https://ieeexplore.ieee.org/document/10760450
36. Roy S, Biswas S, Sinha S, Ghosh S. DeepRan: Attention-based BiLSTM and CRF for Ransomware Early Detection. *J Organ Comput Electron Commer*. 2021;31(3):265–284. Available from: https://link.springer.com/article/10.1007/s10796-020-10017-4
37. Almomani A, Alshamrani S, Alghamdi A. Transfer Learning for Ransomware Detection Using Pre-trained CNN Architectures. *Appl Sci*. 2023;13(8):5167. Available from: https://www.mdpi.com/2076-3417/13/8/5167
38. Poudyal S, Dasgupta D. Opcode Sequence Analysis for Ransomware Detection Using TF-IDF and SVM. *J Cybersecurity and Privacy*. 2021;1(2):231–248. Available from: https://www.mdpi.com/2624-800X/1/2/14
39. Alqahtani A, Alshamrani S, Alghamdi A. A Novel Approach for Ransomware Detection Based on PE Header Using Machine Learning. *J Comput Virol Hack Tech*. 2021;17(4):321–336. Available from: https://link.springer.com/article/10.1007/s11416-021-00414-x
40. Lee K, Lee J, Lee S-Y, Yim K. Effective Ransomware Detection Using Entropy Estimation of Files for Cloud Services. *Sensors*. 2023;23(6):3023. Available from: https://www.mdpi.com/1424-8220/23/6/3023
41. Davidian M, Kiperberg M, Vanetik N. Early Ransomware Detection with Deep Learning Models. *Future Internet*. 2024;16(8):291. Available from: https://www.mdpi.com/1999-5903/16/8/291
42. Shifa MS, Hasan M, Hossain MJ, Tasin TI, Sarker MR, Islam M. Ransomware Attacks and Detection Mechanisms: A Systematic Literature Review. In: *Cyber Intelligence and Information Retrieval*. Springer; 2025. p. 65–71. Available from: https://link.springer.com/chapter/10.1007/978-981-97-7603-0_7
43. Abdallah M, Le Khac NA, Jahromi H, Jurcut AD. A Hybrid CNN-LSTM Based Approach for Anomaly Detection Systems in SDNs. In: *ARES 2021: 16th International Conference on Availability, Reliability and Security*. ACM; 2021. Available from: https://dl.acm.org/doi/fullHtml/10.1145/3465481.3469190
44. Song S, Kim J, Kim I, Hong J, Kang BB. Detecting code reuse attacks in software defined networking. In: 2017 IEEE Conference on Dependable and Secure Computing. IEEE; 2017. p. 256-63. Available from: https://ieeexplore.ieee.org/document/8073518
45. Sgandurra D, Muñoz-González L, Mohsen R, Lupu EC. Automated dynamic analysis of ransomware: benefits, limitations and use for detection. arXiv preprint. 2016. Available from: https://arxiv.org/abs/1609.03020
46. Alserhani F, Aljared A. Evaluating Ensemble Learning Mechanisms for Predicting Advanced Cyber Attacks. *Appl Sci*. 2023;13(24):13310. Available from: https://www.mdpi.com/2076-3417/13/24/13310
47. Hasan M, Rahman MA. Hybrid Static and Dynamic Analysis for Ransomware Detection: A Case Study on WannaCry. In: *Proceedings of the 2017 International Conference on Cyber Security and Protection of Digital Services (Cyber Security)*. IEEE; 2017. p. 1–6. Available from: https://ieeexplore.ieee.org/document/8005461
48. Mercaldo F, Milosevic J, Martinelli F. Extinguishing Ransomware: A Hybrid Approach to Android Ransomware Detection. In: *Foundations and Practice of Security*. Springer; 2018. p. 242–258. Available from: https://link.springer.com/chapter/10.1007/978-3-319-75650-9_16
49. Mehnaz S, Mudgerikar A, Bertino E. RWGuard: A Real-Time Detection System Against Cryptographic Ransomware. In: *Research in Attacks, Intrusions, and Defenses*. Springer; 2018. p. 114–136. Available from: https://link.springer.com/chapter/10.1007/978-3-030-00470-5_6
50. Zuhair H, Selamat A, Krejcar O. A Multi-Tier Streaming Analytics Model of 0-Day Ransomware Detection Using Machine Learning. *Appl Sci*. 2020;10(9):3210. Available from: https://www.mdpi.com/2076-3417/10/9/3210
51. Azmoodeh A, Dehghantanha A, Choo KKR. Robust malware detection for internet of (battlefield) things devices using deep eigenspace learning. IEEE Trans Sustain Comput. 2019;4(1):88-95. Available from: Robust Malware Detection for Internet of (Battlefield) Things Devices Using Deep Eigenspace Learning | IEEE Journals & Magazine | IEEE Xplore
52. Vinayakumar R, Soman KP, Poornachandran P. Evaluation of recurrent neural network and its variants for intrusion detection system (IDS). Int J Inf Secur. 2017;17:43-63.

 doi:10.20944/preprints202602.2011.v1

53. Nataraj L, Karthikeyan S, Jacob G, Manjunath BS. Malware images: visualization and automatic classification. In: Proceedings of the 8th International Symposium on Visualization for Cyber Security. ACM; 2011. p. 1-7. Available from: https://dl.acm.org/doi/10.1145/2016904.2016908
54. Ravi Kumar P, Ramlie HRE. Anatomy of Ransomware: Attack Stages, Patterns and Handling Techniques. In: *Computational Intelligence in Information Systems*. Springer; 2021. p. 205–214. Available from: https://link.springer.com/chapter/10.1007/978-3-030-68133-3_20
55. Hampton N, Baig Z, Zeadally S. Ransomware behavioural analysis on Windows platforms. J Inf Secur Appl. 2018;40:44-51. Available from: Ransomware behavioural analysis on windows platforms - ScienceDirect
56. Rehman M, Akbar R, Omar M, Gilal AR. A Systematic Literature Review of Ransomware Detection Methods and Tools for Mitigating Potential Attacks. In: *Computing and Informatics (ICOCI 2023)*. Springer; 2024. p. 80–95. Available from: https://link.springer.com/chapter/10.1007/978-981-99-9589-9_7
57. Ryan M. Ransomware Revolution: The Rise of a Prodigious Cyber Threat. In: *Advances in Information Security*. Springer; 2021. Available from: https://link.springer.com/book/10.1007/978-3-030-66583-8
58. Khammas BM. Ransomware detection using random forest technique. ICT Express. 2020;6(4):325-31. Available from: Ransomware Detection using Random Forest Technique - ScienceDirect
59. Afianian A, Niksefat S, Sadeghiyan B, Baptiste D. Malware dynamic analysis evasion techniques: a survey. ACM Comput Surv. 2019;52(6):1-28. Available from: https://dl.acm.org/doi/10.1145/3365001
60. Cabaj K, Gawkowski P, Grochowski K, Osojca D. Network activity analysis of CryptoWall ransomware. Przeglad Elektrotechniczny. 2015;91(11):201-4. Available from: (PDF) Network activity analysis of CryptoWall ransomware
61. Chen M, Ji T, Li S, Zhang Y, Wang T, Mao K, Sun Y. A Double-Shell Structured Ransomware Defense Method Tailored for the RaaS Model. In: *Cyberspace Simulation and Evaluation*. Springer; 2025. p. 361–376. Available from: https://link.springer.com/chapter/10.1007/978-981-96-4503-9_24
62. Djenna A, Belaoued M, Lifa N. Top Cyber Threats: The Rise of Ransomware. In: *Information Security Theory and Practice*. Springer; 2024. p. 80–95. Available from: https://link.springer.com/chapter/10.1007/978-3-031-60391-4_6
63. Kerns Q, Payne B, Abegaz T. Double-Extortion Ransomware: A Technical Analysis of Maze Ransomware. In: *Future Technologies Conference (FTC) 2021, Volume 3*. Springer; 2021. p. 82–94. Available from: https://link.springer.com/chapter/10.1007/978-3-030-89912-7_7
64. Algarni S. Cybersecurity Attacks: Analysis of "WannaCry" Attack and Proposing Methods for Reducing or Preventing Such Attacks in Future. In: *ICT Systems and Sustainability*. Springer; 2020. p. 763–770. Available from: https://link.springer.com/chapter/10.1007/978-981-15-8289-9_73
65. MacRae J, Franqueira VNL. On Locky Ransomware, Al Capone and Brexit. In: *Digital Forensics and Cyber Crime*. Springer; 2018. p. 33–45. Available from: https://link.springer.com/chapter/10.1007/978-3-319-73697-6_3
66. Ehrenfeld JM. WannaCry, Cybersecurity and Health Information Technology: A Time to Act. *J Med Syst*. 2017;41(104). Available from: https://link.springer.com/article/10.1007/s10916-017-0752-1
67. Pham P-H, Dang LX, Do QN, Hoang CN, Nguyen LV. EternalBlue Exploit: Definitions and Working Mechanism. In: *Trends in Sustainable Computing and Machine Intelligence (ICTSM 2024)*. Springer; 2025. p. 1–13. Available from: https://link.springer.com/chapter/10.1007/978-981-96-1452-3_1
68. Liu Z, Chen C, Zhang LY, Gao S. Working Mechanism of EternalBlue and Its Application in Ransomworm. In: *Cyberspace Safety and Security (CSS 2022)*. Springer; 2022. p. 178–191. Available from: https://link.springer.com/chapter/10.1007/978-3-031-18067-5_13
69. McDonald G, Papadopoulos P, Pitropakis N, Ahmad J, Buchanan WJ. Ransomware: Analysing the Impact on Windows Active Directory Domain Services. *Sensors*. 2022;22(3):953. Available from: https://www.mdpi.com/1424-8220/22/3/953
70. Kharraz A, Robertson W, Balzarotti D, Bilge L, Kirda E. Cutting the Gordian Knot: A Look Under the Hood of Ransomware Attacks. In: *Detection of Intrusions and Malware, and Vulnerability Assessment (DIMVA 2015)*. Springer; 2015. p. 3–24. Available from: https://link.springer.com/chapter/10.1007/978-3-319-20550-2_1

 doi:10.20944/preprints202602.2011.v1

71. Kok SH, Abdullah A, Jhanjhi NZ, Supramaniam M. Prevention of Crypto-Ransomware Using a Pre-Encryption Detection Algorithm. *Computers*. 2019;8(4):79. Available from: https://www.mdpi.com/2073-431X/8/4/79
72. Yaqoub SAF. What Petya/NotPetya Ransomware Is and What Its Remediations Are. In: *Information Technology - New Generations*. Springer; 2018. p. 93–100. Available from: https://link.springer.com/chapter/10.1007/978-3-319-77028-4_15
73. Continella A, Guagnelli A, Zingaro G, De Pasquale G, Barenghi A, Zanero S, Maggi F. ShieldFS: a self-healing, ransomware-aware filesystem. In: Proceedings of the 32nd Annual Conference on Computer Security Applications. ACM; 2016. p. 336-47. Available from: https://dl.acm.org/doi/10.1145/2991079.2991110
74. Cabaj K, Mazurczyk W. Using software-defined networking for ransomware mitigation: the case of CryptoWall. IEEE Netw. 2016;30(6):14-20. Available from: Using Software-Defined Networking for Ransomware Mitigation: The Case of CryptoWall | IEEE Journals & Magazine | IEEE Xplore
75. Kirubavathi G, Regis Anne W, Sridevi UK. A Recent Review of Ransomware Attacks on Healthcare Industries. *Int J Syst Assur Eng Manag*. 2024;15:5078–5096. Available from: https://link.springer.com/article/10.1007/s13198-024-02496-4
76. Möller DPF. Ransomware Attacks and Scenarios: Cost Factors and Loss of Reputation. In: *Guide to Cybersecurity in Digital Transformation*. Springer; 2023. p. 273–303. Available from: https://link.springer.com/chapter/10.1007/978-3-031-26845-8_6
77. Mbol F, Robert JM, Sadighian A. An efficient approach to detect torrent locker ransomware in computer systems. In: Cryptology and Network Security. Springer; 2016. p. 532-41. Available from: https://link.springer.com/chapter/10.1007/978-3-319-48965-0_32
78. Marion JY. Ransomware: Extortion Is My Business. *Commun ACM*. 2025 Apr 24. Available from: https://cacm.acm.org/research/ransomware-extortion-is-my-business
79. Scaife N, Carter H, Traynor P, Butler KRB. CryptoLock (and drop it): stopping ransomware attacks on user data. In: 2016 IEEE 36th International Conference on Distributed Computing Systems (ICDCS). IEEE; 2016. p. 303-12. Available from: https://ieeexplore.ieee.org/document/7536540
80. Mercaldo F. A Framework for Supporting Ransomware Detection and Prevention Based on Hybrid Analysis. *J Comput Virol Hack Tech*. 2021;17:221–227. Available from: https://link.springer.com/article/10.1007/s11416-021-00388-w
81. Kharaz A, Arshad S, Mulliner C, Robertson W, Kirda E. UNVEIL: a large-scale, automated approach to detecting ransomware. In: 25th USENIX Security Symposium. USENIX; 2016. p. 757-72. Available from: https://www.usenix.org/conference/usenixsecurity16/technical-sessions/presentation/kharaz
82. Kolodenker E, Koch W, Stringhini G, Egele M. PayBreak: defense against cryptographic ransomware. In: Proceedings of the 2017 ACM on Asia Conference on Computer and Communications Security. ACM; 2017. p. 599-611. Available from: https://dl.acm.org/doi/10.1145/3052973.3053035
83. Park S, Lee M, Na S, Lim J. Destructive Malwares on MITRE ATT&CK Tactics for Cyber Warfare: A Brief Survey and Analysis. In: *Mobile Internet Security (MobiSec 2023)*. Springer; 2024. p. 260–270. Available from: https://link.springer.com/chapter/10.1007/978-981-97-4465-7_19
84. Branescu I, Grigorescu O, Dascalu M. Automated Mapping of Common Vulnerabilities and Exposures to MITRE ATT&CK Tactics. *Information*. 2024;15(4):214. Available from: https://www.mdpi.com/2078-2489/15/4/214
85. Kuppa A, Aouad L, Le-Khac NA. Linking CVEs to MITRE ATT&CK Techniques. In: *ARES 2021: 16th International Conference on Availability, Reliability and Security*. ACM; 2021. Available from: https://dl.acm.org/doi/fullHtml/10.1145/3465481.3465758
86. Xiong W, Legrand E, Åberg O, Lagerström R. Cybersecurity Threat Modeling Based on the MITRE Enterprise ATT&CK Matrix. *Softw Syst Model*. 2022;21:157–177. Available from: https://link.springer.com/article/10.1007/s10270-021-00898-7
87. Georgiadou A, Mouzakitis S, Askounis D. Assessing MITRE ATT&CK Risk Using a Cyber-Security Culture Framework. *Sensors*. 2021;21(9):3267. Available from: https://www.mdpi.com/1424-8220/21/9/3267

 doi:10.20944/preprints202602.2011.v1

88. Moussaileb R, Cuppens N, Lanet JL, Le Bouder H. Ransomware Network Traffic Analysis for Pre-encryption Alert. In: *Foundations and Practice of Security (FPS 2019)*. Springer; 2020. p. 20–38. Available from: https://link.springer.com/chapter/10.1007/978-3-030-45371-8_2
89. Cusack G, Michel O, Keller E. Machine learning-based detection of ransomware using SDN. In: Proceedings of the 2018 ACM International Workshop on Security in Software Defined Networks & Network Function Virtualization. ACM; 2018. p. 1-6. Available from: https://dl.acm.org/doi/10.1145/3180465.3180467
90. Alqahtani A, Sheldon FT. A Survey of Crypto Ransomware Attack Detection Methodologies: An Evolving Outlook. *Sensors*. 2022;22(5):1837. Available from: https://www.mdpi.com/1424-8220/22/5/1837
91. Shuai Y, Zhang Y, Li Y, Li J. Mitigation of Privilege Escalation Attack Using Kernel Data Relocation Mechanism. *Int J Inf Secur*. 2024;23. Available from: https://link.springer.com/content/pdf/10.1007/s10207-024-00890-4.pdf
92. Sheen S, Anitha R, Natarajan V. Android based malware detection using a multifeature collaborative decision fusion approach. Neurocomputing. 2018;151:905-12. Available from: Android based malware detection using a multifeature collaborative decision fusion approach - ScienceDirect
93. Mundt M, Baier H. Enhancing Incident Management by an Improved Understanding of Data Exfiltration: Definition, Evaluation, Review. In: *Digital Forensics and Cyber Crime (ICDF2C 2023)*. Springer; 2024. p. 33–57. Available from: https://link.springer.com/chapter/10.1007/978-3-031-56580-9_3
94. Monika MS, Zavarsky P, Lindskog D. Experimental study of ransomware on Windows and Android platforms. Procedia Comput Sci. 2018;94:465-72. Available from: https://www.sciencedirect.com/science/article/pii/S1877050916306032
95. Fatima S, Rehman T, Fatima M, Khan S, Ali MA. Comparative Analysis of AES and RSA Algorithms for Data Security in Cloud Computing. *Eng Proc*. 2022;20(1):14. Available from: https://www.mdpi.com/2673-4591/20/1/14
96. Zimba A, Wang Z, Chen H. Multi-stage crypto ransomware attacks: a new emerging cyber threat to critical infrastructure and industrial control systems. ICT Express. 2018;4(1):14-18. Available from: Multi-stage crypto ransomware attacks: A new emerging cyber threat to critical infrastructure and industrial control systems - ScienceDirect
97. Chen Q, Bridges RA. Automated behavioral analysis of malware: a case study of WannaCry ransomware. In: 2017 16th IEEE International Conference on Machine Learning and Applications (ICMLA). IEEE; 2017. p. 454-60. Available from: https://ieeexplore.ieee.org/document/8260772
98. Morato D, Berrueta E, Magaña E, Izal M. Ransomware early detection by the analysis of file sharing traffic. J Netw Comput Appl. 2018;124:14-32. Available from: Ransomware early detection by the analysis of file sharing traffic - ScienceDirect
99. Ahmed YA, Huda S, Al-rimy BAS, Alharbi N, Saeed F, Ghaleb FA, Ali IM. A weighted minimum redundancy maximum relevance technique for ransomware early detection in industrial IoT. Sustainability. 2022;14(3):1231. Available from: https://www.mdpi.com/2071-1050/14/3/1231
100. Lee Y, Lee J, Ryu D, Park H, Shin D. Clop Ransomware in Action: A Comprehensive Analysis of Its Multi-Stage Tactics. *Electronics*. 2023;13(18):3689. Available from: https://www.mdpi.com/2079-9292/13/18/3689
101. Hwang J, Kim J, Lee S, Kim K. Two-stage ransomware detection using dynamic analysis and machine learning techniques. Wirel Pers Commun. 2020;112:2597-609. Available from: https://link.springer.com/article/10.1007/s11277-020-07166-9
102. Alhawi OMK, Baldwin J, Dehghantanha A. Leveraging machine learning techniques for Windows ransomware network traffic detection. In: Cyber Threat Intelligence. Springer; 2018. p. 93-106. Available from: https://link.springer.com/chapter/10.1007/978-3-319-73951-9_5
103. Australian Signals Directorate. ASD Cyber Threat Report 2022–23. *Cyber.gov.au*. 2023 Nov 14. Available from: https://www.cyber.gov.au/sites/default/files/2023-11/asd-cyber-threat-report-2023.pdf
104. Jiang T, Gradus JL, Rosellini AJ. Supervised Machine Learning: A Brief Primer. Behav Ther. 2020 Sep;51(5):675-687. doi: 10.1016/j.beth.2020.05.002. Epub 2020 May 16. PMID: 32800297; PMCID: PMC7431677. Available from: Supervised machine learning: A brief primer - PMC

105. Greubel A, Andres D, Hennecke M. Analyzing Reporting on Ransomware Incidents: A Case Study. *Soc Sci*. 2023;12(5):265. Available from: https://www.mdpi.com/2076-0760/12/5/265
106. Dhiman P, Saini N, Gulzar Y, Turaev S, Kaur A, Nisa KU, Hamid Y. A Review and Comparative Analysis of Relevant Approaches of Zero Trust Network Model. *Sensors*. 2024;24(4):1328. Available from: https://www.mdpi.com/1424-8220/24/4/1328
107. Booshan P, Armoogum S, Li X. An Adaptive Security Architecture for Detecting Ransomware Attack Using Open Source Software. In: *Advances in Information and Communication (FICC 2020)*. Springer; 2020. p. 618–633. Available from: https://link.springer.com/chapter/10.1007/978-3-030-39445-5_45
108. Herrera-Silva JA, Hernández-Álvarez M. Dynamic Feature Dataset for Ransomware Detection Using Machine Learning Algorithms. *Sensors*. 2023;23(3):1053. Available from: https://www.mdpi.com/1424-8220/23/3/1053
109. Vinayakumar R, Alazab M, Jolfaei A, Soman KP, Poornachandran P. Ransomware triage using deep learning: Twitter as a case study. In: 2019 Cybersecurity and Cyberforensics Conference (CCC). IEEE; 2019. p. 67-73. Available from: Ransomware triage using deep learning: Twitter as a case study - Research @ Flinders
110. J. Zhang, B. Chen, Y. Zhao, X. Cheng and F. Hu, "Data Security and Privacy-Preserving in Edge Computing Paradigm: Survey and Open Issues," in IEEE Access, vol. 6, pp. 18209-18237, 2018, doi: 10.1109/ACCESS.2018.2820162. Available from: Data Security and Privacy-Preserving in Edge Computing Paradigm: Survey and Open Issues | IEEE Journals & Magazine | IEEE Xplore
111. Ouerdi N, Mejjout B, Laaroussi K, Kasmi MA. Ransomware Detection and Classification Using Machine Learning and Deep Learning. In: *Advances in Smart Medical, IoT & Artificial Intelligence (ICSMAI 2024)*. Springer; 2024. p. 195–202. Available from: https://link.springer.com/chapter/10.1007/978-3-031-66850-0_22
112. Maiorca D, Mercaldo F, Giacinto G, Visaggio CA, Martinelli F. R-PackDroid: API package-based characterization and detection of mobile ransomware. In: Proceedings of the Symposium on Applied Computing. ACM; 2017. p. 1718-23. Available from: https://dl.acm.org/doi/10.1145/3019612.3019793
113. Andronio N, Zanero S, Maggi F. HELDROID: dissecting and detecting mobile ransomware. In: Research in Attacks, Intrusions, and Defenses. Springer; 2015. p. 382-404. Available from: https://link.springer.com/chapter/10.1007/978-3-319-26362-5_18
114. Hussain MZ, Hasan MZ, Baig MM, Khan T, Nosheen S, Bhatti AM, Qureshi AM, Siddiqui AA, Mubarak Z, Chuhan SH, Bilal A, Yaqub MA. Malware/Ransomware Analysis and Detection Using CIC-MalMem2022 Dataset. In: *Intelligent Sustainable Systems (WorldS4 2023)*. Springer; 2024. p. 339–352. Available from: https://link.springer.com/chapter/10.1007/978-981-99-8031-4_30
115. Coglio F, Lekssays A, Carminati B, Ferrari E. Early-Stage Ransomware Detection Based on Pre-Attack Internal API Calls. In: *Advanced Information Networking and Applications (AINA 2023)*. Springer; 2023. p. 417–429. Available from: https://link.springer.com/chapter/10.1007/978-3-031-28451-9_36
116. Homayoun S, Dehghantanha A, Ahmadzadeh M, Hashemi S, Khayami R. DRTHIS: deep ransomware threat hunting and intelligence system at the fog layer. Future Gener Comput Syst. 2019;90:94-104. Available from: DRTHIS: Deep ransomware threat hunting and intelligence system at the fog layer - ScienceDirect
117. Taşcı B. Deep-Learning-Based Approach for IoT Attack and Malware Detection. *Appl Sci*. 2024;14(18):8505. Available from: https://www.mdpi.com/2076-3417/14/18/8505
118. Almobaideen W, Abu Alghanam O, Abdullah M, Hussain SB, Alam U. Comprehensive Review on Machine Learning and Deep Learning Techniques for Malware Detection in Android and IoT Devices. *Int J Inf Secur*. 2025;24:110. Available from: https://link.springer.com/article/10.1007/s10207-025-01027-x
119. Anand PM, Charan PV, Chunduri H, Shukla SK. RTR-Shield: Early Detection of Ransomware Using Registry and Trap Files. In: *Information Security Practice and Experience (ISPEC 2023)*. Springer; 2023. p. 209–229. Available from: https://link.springer.com/chapter/10.1007/978-981-99-7032-2_13
120. Davies SR, Macfarlane R, Buchanan WJ. Comparison of Entropy Calculation Methods for Ransomware Encrypted File Identification. *Entropy*. 2022;24(10):1503. Available from: https://www.mdpi.com/1099-4300/24/10/1503

 doi:10.20944/preprints202602.2011.v1

121. Hashwanth S, Kirthica S. Effective Ransomware Detection Method Using PE Header and YARA Rules. In: *Network Security and Blockchain Technology (ICNSBT 2023)*. Springer; 2023. p. 185–194. Available from: https://link.springer.com/chapter/10.1007/978-981-99-4433-0_16
122. Ayub MA, Siraj A, Filar B, Gupta M. RWArmor: A Static-Informed Dynamic Analysis Approach for Early Detection of Cryptographic Windows Ransomware. *Int J Inf Secur*. 2024;23:533–556. Available from: https://link.springer.com/article/10.1007/s10207-023-00758-z
123. Hull G, John H, Arief B. Ransomware deployment methods and analysis: views from a predictive model and human responses. Crime Sci. 2019;8:2. Available from: https://crimesciencejournal.biomedcentral.com/articles/10.1186/s40163-019-0097-9
124. Wan YL, Chang JC, Chen RJ, Wang SJ. Feature-selection-based ransomware detection with machine learning of data analysis. In: Proceedings of the 2018 3rd International Conference on Computer and Communication Systems (ICCCS). IEEE; 2018. p. 392-6. Available from: Feature-Selection-Based Ransomware Detection with Machine Learning of Data Analysis - National Yang Ming Chiao Tung University Academic Hub
125. Umara U, Zainal A, Al-Rimy BAS, Ghaleb FA, Rassam MA. Static and Dynamic Analysis of Ransomware: Challenges and Future Directions. In: *Recent Advances in Cybersecurity*. Springer; 2023. p. 45–62. Available from: https://link.springer.com/chapter/10.1007/978-981-99-7032-2_4
126. Berrueta E, Morato D, Magaña E, Izal M. A survey on detection techniques for cryptographic ransomware. IEEE Access. 2019;7:144925-44. Available from: A Survey on Detection Techniques for Cryptographic Ransomware | IEEE Journals & Magazine | IEEE Xplore
127. Nieuwenhuizen D. A behavioural-based approach to ransomware detection. MWR Labs Whitepaper. 2017. Available from: [PDF] A behavioural-based approach to ransomware detection | Semantic Scholar
128. Moussaileb R, Bouget B, Palisse A, Le Bouder H, Cuppens N, Lanet JL. Ransomware's early mitigation mechanisms. In: Proceedings of the 13th International Conference on Availability, Reliability and Security. ACM; 2018. p. 1-10. Available from: Ransomware's Early Mitigation Mechanisms | Proceedings of the 13th International Conference on Availability, Reliability and Security
129. Song S, Kim B, Lee S. The effective ransomware prevention technique using process monitoring on Android platform. Mobile Inf Syst. 2016;2016:2946735. Available from: https://www.hindawi.com/journals/misy/2016/2946735/
130. Shijo PV, Salim A. Integrated static and dynamic analysis for malware detection. Procedia Comput Sci. 2015;46:804-11. Available from: Integrated Static and Dynamic Analysis for Malware Detection - ScienceDirect
131. Singla A, Bertino E, Verma D. Preparing network intrusion detection deep learning models with minimal data using adversarial domain adaptation. In: Proceedings of the 15th ACM Asia Conference on Computer and Communications Security. ACM; 2020. p. 127-40. Available from: https://dl.acm.org/doi/10.1145/3320269.3384718
132. U. Urooj, B. A. S. Al-Rimy, A. B. Zainal, F. Saeed, A. Abdelmaboud and W. Nagmeldin, "Addressing Behavioral Drift in Ransomware Early Detection Through Weighted Generative Adversarial Networks," in IEEE Access, vol. 12, pp. 3910-3925, 2024, doi: 10.1109/ACCESS.2023.3348451. Available from: Addressing Behavioral Drift in Ransomware Early Detection Through Weighted Generative Adversarial Networks | IEEE Journals & Magazine | IEEE Xplore
133. Aboaoja FA, Zainal A, Ghaleb FA, Al-rimy BAS, Eisa TAE, Elnour AAA. Malware detection issues, challenges, and future directions: a survey. Appl Sci. 2022;12(17):8482. Available from: https://www.mdpi.com/2076-3417/12/17/8482
134. Maniath S, Ashok A, Poornachandran P, Sujadevi VG, Sankar AUP, Jan S. Deep learning LSTM based ransomware detection. In: 2017 Recent Developments in Control, Automation & Power Engineering (RDCAPE). IEEE; 2017. p. 442-6. Available from: https://ieeexplore.ieee.org/document/8358312
135. Zhang H, Xiao X, Mercaldo F, Ni S, Martinelli F, Sangaiah AK. Classification of ransomware families with machine learning based on N-gram of opcodes. Future Gener Comput Syst. 2019;90:211-21. Available from: Classification of ransomware families with machine learning based onN-gram of opcodes - ScienceDirect

 doi:10.20944/preprints202602.2011.v1

136. Sivasubramanian A, Devisetty M, Bhavukam P. Feature Extraction and Anomaly Detection Using Different Autoencoders for Modeling Intrusion Detection Systems. *Arab J Sci Eng*. 2024;49:13061–13073. Available from: https://link.springer.com/article/10.1007/s13369-024-08951-5
137. Shaik AS, Shaik A. AI Enhanced Cyber Security Methods for Anomaly Detection. In: *Machine Intelligence, Tools, and Applications (ICMITA 2024)*. Springer; 2024. p. 348–359. Available from: https://link.springer.com/chapter/10.1007/978-3-031-65392-6_30
138. Kok S, Azween A, Jhanjhi NZ. Evaluation metric for crypto-ransomware detection using machine learning. J Inf Secur Appl. 2020;55:102646. Available from: Evaluation metric for crypto-ransomware detection using machine learning - ScienceDirect
139. Alhashmi AA, Darem AA, Alashjaee AM, Alanazi SM, Alkhaldi TM, Ebad SA, Ghaleb FA, Almadani AM. Similarity-Based Hybrid Malware Detection Model Using API Calls. Mathematics. 2023; 11(13):2944. Available from: https://doi.org/10.3390/math11132944
140. Hwang RH, Peng MC, Nguyen VL, Chang YL. An LSTM-based deep learning approach for classifying malicious traffic at the packet level. Appl Sci. 2019;9(16):3414. Available from: https://www.mdpi.com/2076-3417/9/16/3414
141. Al-Hawawreh M, Moustafa N, Garg S, Hossain MS. Deep learning-enabled threat intelligence scheme in the Internet of Things networks. IEEE Trans Netw Sci Eng. 2021;8(4):2968-81. Available from: Deep Learning-Enabled Threat Intelligence Scheme in the Internet of Things Networks | IEEE Journals & Magazine | IEEE Xplore
142. Al-rimy BAS, Maarof MA, Shaid SZM. A 0-day aware crypto-ransomware early behavioral detection framework. In: Advances in Information and Communication Networks. Lecture Notes in Networks and Systems. Springer; 2019. p. 758–77. Available from: A 0-Day Aware Crypto-Ransomware Early Behavioral Detection Framework | Springer Nature Link
143. Nissim N, Cohen A, Glezer C, Elovici Y. Detection of malicious PDF files and directions for enhancements: a state-of-the art survey. Comput Secur. 2015;48:246-66. Available from: Detection of malicious PDF files and directions for enhancements: A state-of-the art survey - ScienceDirect
144. Azmoodeh A, Dehghantanha A, Conti M, Choo KKR. Detecting crypto-ransomware in IoT networks based on energy consumption footprint. J Ambient Intell Humaniz Comput. 2018;9:1141-52. Available from: https://link.springer.com/article/10.1007/s12652-017-0558-5
145. Richardson R, North MM. Ransomware: Evolution, mitigation and prevention. Int Manag Rev. 2017;13(1):10-21. Available from: Ransomware: Evolution, Mitigation and Prevention
146. Chen Z, Kang HS, Yin SN, Kim SR. Automatic ransomware detection and analysis based on dynamic API calls flow graph. In: Proceedings of the International Conference on Research in Adaptive and Convergent Systems. ACM; 2017. p. 196-201. Available from: https://dl.acm.org/doi/10.1145/3129676.3129704
147. Ahmadian MM, Shahriari HR, Ghaffarian SM. Connection-monitor & connection-breaker: a novel approach for prevention and detection of high survivable ransomwares. In: 2015 12th International Iranian Society of Cryptology Conference on Information Security and Cryptology (ISCISC). IEEE; 2015. p. 79-84. Available from: Connection-monitor & connection-breaker: A novel approach for prevention and detection of high survivable ransomwares | IEEE Conference Publication | IEEE Xplore
148. T. d. Oliveira Lima, M. Colaço, K. H. de J. Prado, I. Dias de J. and F. R. de Oliveira, "A Big Data Experiment to Evaluate the Effectiveness of Traditional Machine Learning Techniques Against LSTM Neural Networks in the Hotels Clients Opinion Mining," 2021 IEEE International Conference on Big Data (Big Data), Orlando, FL, USA, 2021, pp. 5199-5208, doi: 10.1109/BigData52589.2021.9671939.
149. Al-Hawawreh M, Sitnikova E, Aboutorab N. X-IIoTID: a connectivity-agnostic and device-agnostic intrusion dataset for industrial Internet of Things. IEEE Internet Things J. 2022;9(5):3962-77. Available from: X-IIoTID: A Connectivity-Agnostic and Device-Agnostic Intrusion Data Set for Industrial Internet of Things | IEEE Journals & Magazine | IEEE Xplore
150. Khammas BM, Monemi A, Basurra S, De Santis E, Iadarola G, Martinelli F. Ransomware detection using random forest technique: a survey and a case study. In: Proceedings of the 7th International Conference on Information Systems Security and Privacy (ICISSP 2021). SciTePress; 2021. p. 534-41. Available from: Ransomware Detection using Random Forest Technique - ScienceDirect

 doi:10.20944/preprints202602.2011.v1

151. Cabaj K, Gregorczyk M, Mazurczyk W. Software-defined networking-based crypto ransomware detection using HTTP traffic characteristics. Comput Electr Eng. 2018;66:353-68. Available from: Software-defined networking-based crypto ransomware detection using HTTP traffic characteristics - ScienceDirect
152. Arivudainambi D, KA VD, Visu P. Malware traffic classification using principal component analysis and artificial neural network for extreme surveillance. Comput Commun. 2019;147:50-7. Available from: Malware traffic classification using principal component analysis and artificial neural network for extreme surveillance - ScienceDirect
153. Darem A, Abawajy JH, Makkar A, Alanazi A, Alanazi S, Alshahrani H, Almotairi KH, Ghaleb FA, Alqarni MA. Visualization and deep-learning-based malware variant detection using OpCode-level features. Future Gener Comput Syst. 2021;125:314-23. Available from: Visualization and deep-learning-based malware variant detection using OpCode-level features - ScienceDirect
154. Naeem H, Guo B, Naeem MR, Ullah F, Aldabbas H, Javed MS. Identification of malicious code variants based on image visualization. Comput Electr Eng. 2020;76:225-37. Available from: Identification of malicious code variants based on image visualization - ScienceDirect
155. Al-Rimy BAS, Maarof MA, Shaid SZM. Ransomware threat success factors, taxonomy, and countermeasures: a survey and research directions. Comput Secur. 2018;74:144-66. Available from: Ransomware threat success factors, taxonomy, and countermeasures: A survey and research directions - ScienceDirect
156. Abaid Z, Kaafar MA, Jha S. Early detection of in-the-wild botnet attacks by exploiting network communication uniformity: an empirical study. In: 2016 IEEE Conference on Communications and Network Security (CNS). IEEE; 2016. p. 270-8. Available from: Early detection of in-the-wild botnet attacks by exploiting network communication uniformity: an empirical study - Macquarie University
157. Alzaylaee MK, Yerima SY, Sezer S. DynaLog: An automated dynamic analysis framework for characterizing Android applications. In: 2016 International Conference on Cyber Security and Protection of Digital Services (Cyber Security). IEEE; 2016. p. 1-8. Available from: [1607.08166] DynaLog: An automated dynamic analysis framework for characterizing Android applications
158. Kara I, Aydos M. The rise of ransomware: forensic analysis for windows based ransomware attacks. In: 2019 1st International Informatics and Software Engineering Conference (UBMYK). IEEE; 2019. p. 1-6. Available from: The rise of ransomware: Forensic analysis for windows based ransomware attacks - ScienceDirect
159. Ahmed ME, Kim H, Park M. Mitigating DNS query-based DDoS attacks with machine learning on software-defined networking. In: 2017 IEEE Military Communications Conference (MILCOM). IEEE; 2017. p. 11-16. Available from: Mitigating DNS query-based DDoS attacks with machine learning on software-defined networking | IEEE Conference Publication | IEEE Xplore
160. S. Razaulla et al., "The Age of Ransomware: A Survey on the Evolution, Taxonomy, and Research Directions," in IEEE Access, vol. 11, pp. 40698-40723, 2023, doi: 10.1109/ACCESS.2023.3268535.
161. Mercaldo F. Automated Ransomware Behavior Analysis: Pattern Extraction and Early Detection. In: *Cybersecurity and Privacy in Cyber Physical Systems*. Springer; 2020. p. 251–267. Available from: https://link.springer.com/chapter/10.1007/978-3-030-34637-9_15
162. De Gaspari F, Hitaj D, Pagnotta G, De Carli L, Mancini LV. Evading Behavioral Classifiers: A Comprehensive Analysis on Evading Ransomware Detection Techniques. *Neural Comput Appl*. 2022;34:12077–12096. Available from: https://link.springer.com/article/10.1007/s00521-022-07096-6
163. Alqahtani A, Gazzan M, Sheldon FT. Ransomware detection and prevention based on artificial intelligence: a systematic review. Secur Priv. 2022;5(6):e256. Available from: https://onlinelibrary.wiley.com/doi/full/10.1002/spy2.256
164. Akbanov M, Vassilakis VG, Logothetis MD. Ransomware detection and mitigation using software-defined networking: the case of WannaCry. Comput Electr Eng. 2019;76:111-21. Available from: Ransomware detection and mitigation using software-defined networking: The case of WannaCry - ScienceDirect
165. Cohen A, Nissim N. Trusted detection of ransomware in a private cloud using machine learning methods leveraging meta-features from volatile memory. Expert Syst Appl. 2018;102:158-78. Available from: Trusted

 doi:10.20944/preprints202602.2011.v1

detection of ransomware in a private cloud using machine learning methods leveraging meta-features from volatile memory - ScienceDirect

166. Shaukat K, Luo S, Varadharajan V, Hameed IA, Xu M. Ransomware classification and detection with machine learning algorithms. In: 2022 IEEE 46th *Annual Computers, Software, and Applications Conference (COMPSAC)*; 2022. p. 1154-63. https://doi.org/10.1109/COMPSAC54236.2022.00181
167. Barua R, Reza MSU, Tonmoy TI. Is iterative feature selection technique efficient enough? A comparative performance analysis of RFECV feature selection technique in ransomware classification using SHAP. *Discov Internet Things*. 2023;3:25. https://doi.org/10.1007/s43926-023-00053-2
168. Sharma M, Sharma T, Singla AK, Rattan D. CFSBFDroid: Android malware detection using CFS + best first search-based feature selection. *Mob Inf Syst*. 2022;2022:6425583. https://doi.org/10.1155/2022/6425583
169. Padmavathi G, Shanmugapriya D. Evaluation of principal component analysis variants to assess their suitability for mobile malware detection. In: Arai K, editor. Intelligent Computing. *Lecture Notes in Networks and Systems*. Cham: Springer; 2022. p. 269-90. https://www.intechopen.com/chapters/82166
170. Meidan Y, Bohadana M, Mathov Y, Mirsky Y, Shabtai A, Breitenbacher D, Elovici Y. N-BaIoT—network-based detection of IoT botnet attacks using deep autoencoders. IEEE Pervasive Comput. 2018;17(3):12-22. Available from: https://ieeexplore.ieee.org/document/8490192
171. Kshirsagar D, Kumar S. Intrusion detection system using PCA with random forest approach. In: 2020 International Conference on Electronics and Sustainable Communication Systems (ICESC); 2020 Jul 2-4; Coimbatore, India. IEEE; 2020. p. 803-8. doi:10.1109/ICESC48915.2020.9155656. Available from: https://ieeexplore.ieee.org/document/9155656/
172. Takeuchi Y, Sakai K, Fukumoto S. Detecting ransomware using support vector machines. In: Proceedings of the 47th International Conference on Parallel Processing Companion; 2018 Aug 13-16; Eugene, OR, USA. New York: ACM; 2018. p. 1-6. doi:10.1145/3229710.322972. Available from: https://dl.acm.org/doi/10.1145/3229710.3229726
173. Al-Syouf R, Aljarrah OY, Bani-Hani R, Alma'aitah A. Ensemble Machine Learning Models Utilizing a Hybrid Recursive Feature Elimination (RFE) Technique for Detecting GPS Spoofing Attacks Against Unmanned Aerial Vehicles. *Sensors* (Basel). 2025;25(8):2388. Available from: https://www.mdpi.com/1424-8220/25/8/2388
174. Madhavi S, Jyothi VE, Chilukuri LSY, Yadav NP, Sri JB, Bhargavi K. Intrusion Detection System Using a Hybrid of SVM and RFE. In: Advances in Information Communication Technology and Computing. Lecture Notes in Networks and Systems. Singapore: Springer; 2024. p. 703–11. Available from: https://link.springer.com/chapter/10.1007/978-981-97-6103-6_44
175. Awad M, Fraihat S. Recursive Feature Elimination with Cross-Validation with Decision Tree: Feature Selection Method for Machine Learning-Based Intrusion Detection Systems. *J Sens Actuator Netw*. 2023;12(5):67. Available from: https://www.mdpi.com/2224-2708/12/5/67
176. Alshamrani S, Alzahrani A, Alzahrani B, Alzahrani A. SRFE: A Stepwise Recursive Feature Elimination Approach for Network Intrusion Detection. *Pers Ubiquit Comput*. 2024. Available from: https://link.springer.com/article/10.1007/s12083-024-01763-2
177. A. Alzahrani, H. Alshahrani, A. Alshehri and H. Fu, "An Intelligent Behavior-Based Ransomware Detection System For Android Platform," 2019 First IEEE International Conference on Trust, Privacy and Security in Intelligent Systems and Applications (TPS-ISA), Los Angeles, CA, USA, 2019, pp. 28-35, doi: 10.1109/TPS-ISA48467.2019.00013.
178. Ba'abbad I, Batarfi O. Proactive Ransomware Detection Using Extremely Fast Decision Tree (EFDT) Algorithm: A Case Study. *Computers* (Basel). 2023;12(6):121. Available from: https://www.mdpi.com/2073-431X/12/6/121
179. Alshamrani S, Alzahrani A, Alzahrani B, Alzahrani A. A Multi-level Correlation-Based Feature Selection for Intrusion Detection. *Arab J Sci Eng*. 2022;47:13369–83. Available from: https://link.springer.com/article/10.1007/s13369-022-06760-2
180. Arabo A, Dijoux R, Poulain T, Chevalier G. Detecting ransomware using process behavior analysis. Procedia Comput Sci. 2020;168:289-96. Available from: Detecting Ransomware Using Process Behavior Analysis - ScienceDirect

181. Darabian, H., Homayounoot, S., Dehghantanha, A. *et al.* Detecting Cryptomining Malware: a Deep Learning Approach for Static and Dynamic Analysis. *J Grid Computing* **18**, 293–303 (2020). Available from: https://doi.org/10.1007/s10723-020-09510-6
182. Albshaier L, Almarri S, Rahman MM. Earlier Decision on Detection of Ransomware Identification: A Comprehensive Systematic Literature Review. *Information* (Basel). 2024;15(8):484. Available from: https://www.mdpi.com/2078-2489/15/8/484
183. Sheen S, Yadav S. Ransomware detection by mining API call usage. In: 2018 International Conference on Advances in Computing, Communications and Informatics (ICACCI). IEEE; 2018. p. 983-7. Available from: Ransomware detection by mining API call usage | IEEE Conference Publication | IEEE Xplore
184. Baek S, Jung Y, Mohaisen A, Lee S, Nyang D. SSD-assisted ransomware detection and data recovery techniques. IEEE Trans Comput. 2021;70(9):1332-44. Available from: SSD-Assisted Ransomware Detection and Data Recovery Techniques | IEEE Journals & Magazine | IEEE Xplore
185. Tahsien SM, Karimipour H, Spachos P. Machine learning based solutions for security of Internet of Things (IoT): a survey. J Netw Comput Appl. 2020;161:102630. Available from: Machine learning based solutions for security of Internet of Things (IoT): A survey - ScienceDirect
186. Abbasi M, Shahraki A, Taherkordi A. Deep learning for network traffic monitoring and analysis (NTMA): a survey. Comput Commun. 2021;170:19-41. Available from: Deep Learning for Network Traffic Monitoring and Analysis (NTMA): A Survey - ScienceDirect
187. Ramilli M, Bishop M, Sun S. Multiprocess malware. In: 2011 6th International Conference on Malicious and Unwanted Software. IEEE; 2011. p. 8-13. Available from: Multiprocess malware | IEEE Conference Publication | IEEE Xplore
188. R. Kumar, X. Zhang, W. Wang, R. U. Khan, J. Kumar and A. Sharif, "A Multimodal Malware Detection Technique for Android IoT Devices Using Various Features," in IEEE Access, vol. 7, pp. 64411-64430, 2019, doi: 10.1109/ACCESS.2019.2916886. Available from: A Multimodal Malware Detection Technique for Android IoT Devices Using Various Features | IEEE Journals & Magazine | IEEE Xplore
189. Al-Hawawreh M, Sitnikova E. Ransomware detection using random forest technique. *ICT Express*. 2022;8(2):218-23. https://www.sciencedirect.com/science/article/pii/S2405959520304756
190. Urooj U, Maarof MA, Al-rimy BAS. Ransomware detection using deep learning based unsupervised feature extraction and a cost sensitive Pareto Ensemble classifier. *Sci Rep*. 2022;12(1):15647. https://www.nature.com/articles/s41598-022-19443-7
191. Zahra A, Shah MA. Android ransomware detection using supervised machine learning techniques based on traffic analysis. *Sensors*. 2024;24(1):189. https://www.mdpi.com/1424-8220/24/1/189
192. Ahmed YA, Kocer B, Al-rimy BAS. Automated analysis and detection of ransomware using machine learning: a survey. *Appl Sci*. 2022;12(1):172. https://www.mdpi.com/2076-3417/12/1/172
193. Sihag V, Vardhan M, Singh P. A survey of android application and malware hardening. J Comput Sci Review. 2021;12:10001-29. Available from: A survey of android application and malware hardening - ScienceDirect
194. Manavi F, Hamzeh A. A new method for ransomware detection based on PE header using convolutional neural networks. In: 2020 17th International ISC Conference on Information Security and Cryptology (ISCISC). IEEE; 2020. p. 82-7. Available from: A New Method for Ransomware Detection Based on PE Header Using Convolutional Neural Networks | IEEE Conference Publication | IEEE Xplore
195. Iqbal MS, Javed Y. Digital forensics as advanced ransomware pre-attack detection algorithm for endpoint data protection. Secur Commun Netw. 2022;2022:1424638. https://www.hindawi.com/journals/scn/2022/1424638/
196. Şahin DÖ, Akleylek S. Permission-based Android malware analysis by using dimension reduction with PCA and LDA. J Inf Secur Appl. 2021;63:102995. https://www.sciencedirect.com/science/article/abs/pii/S2214212621002039
197. Mahdavifar S, Kadir AFA, Fatemi R, Alhadidi D, Ghorbani AA. Dynamic Android malware category classification using semi-supervised deep learning. In: 2020 IEEE Intl Conf on Dependable, Autonomic and Secure Computing. IEEE; 2020. p. 515-22. https://pmc.ncbi.nlm.nih.gov/articles/PMC9890381/

198. Beaman C, Barkworth A, Akande TD, Hakak S, Khan MK. Ransomware: recent advances, analysis, challenges and future research directions. Comput Secur. 2021;111:102490. https://www.researchgate.net/publication/365734823_A_review_of_machine_learning-based_zero-day_attack_detection_Challenges_and_future_directions
199. Sharmeen S, Ahmed YA, Huda S, Koçer BŞ, Hassan MM. Avoiding future digital extortion through robust protection against ransomware threats using deep learning based adaptive approaches. *IEEE Access*. 2020;8:24522-34. Available from: https://doaj.org/article/a7d20b3936484cd4b16785d00c69ad5c
200. Or-Meir O, Nissim N, Elovici Y, Rokach L. Dynamic malware analysis in the modern era—a state of the art survey. ACM Comput Surv. 2019;52(5):1-48. Available from: https://dl.acm.org/doi/10.1145/3329786
201. Almomani I, Alkhayer A, El-Shafai W. E2E-RDS: Efficient end-to-end ransomware detection system based on static-based ML and vision-based DL approaches. *Sensors*. 2023;23(9):4467. Available from: https://www.mdpi.com/1424-8220/23/9/4467
202. Oz H, Aris A, Levi A, Uluagac AS. A survey on ransomware: evolution, taxonomy, and defense solutions. *ACM Comput Surv*. 2022;54(11s):Article 238. Available from: https://dl.acm.org/doi/10.1145/3514229
203. Gopinath M, Sethuraman SC. A comprehensive survey on deep learning based malware detection techniques. *Comput Sci Rev*. 2023;47:100529. Available from: https://www.sciencedirect.com/science/article/abs/pii/S1574013722000636
204. Sharmeen S, Huda S, Koçer B, Abawajy JH. A holistic analysis of ransomware detection techniques for IoT ecosystems: challenges, solutions and future directions. Procedia Comput Sci. 2021;191:320-7. Available from: A Holistic Approach to Ransomware Classification: Leveraging Static and Dynamic Analysis with Visualization | MDPI
205. Mohamed N. Artificial Intelligence and Machine Learning in Cybersecurity: A Deep Dive into State-of-the-Art Techniques and Future Paradigms. *Knowl Inf Syst*. 2025;67:6969–7055. Available from: https://link.springer.com/article/10.1007/s10115-025-02429-y
206. Karbab EMB, Debbabi M, Derhab A. SwiftR: Cross-platform ransomware fingerprinting using hierarchical neural networks on hybrid features. Expert Syst Appl. 2023;225:120017. Available from: https://www.sciencedirect.com/science/article/abs/pii/S0957417423005195
207. Hasan MM, Rahman MM. RansHunt: A support vector machines based ransomware analysis framework with integrated feature set. In: 2017 20th International Conference of Computer and Information Technology (ICCIT). IEEE; 2017. p. 1-7. Available from: https://ieeexplore.ieee.org/document/8281838
208. Alshammari AB, Darem LA, Sheatah HK. Ransomware early detection techniques. Eng Technol Appl Sci Res. 2024;14(3):14497-503. Available from: https://www.etasr.com/index.php/ETASR/article/view/6915
209. Yewale A, Singh M. Malware detection based on opcode frequency using machine learning. In: 2016 International Conference on Advanced Communication Control and Computing Technologies (ICACCCT). IEEE; 2016. p. 646-9. Available from: Malware detection based on opcode frequency | IEEE Conference Publication | IEEE Xplore
210. Chourasiya, L., Khatri, S., Lilhore, U.K. *et al.* Advanced system log analyzer for anomaly detection and cyber forensic investigations using LSTM and transformer networks. *J Cloud Comp* **14**, 60 (2025). Available from: https://doi.org/10.1186/s13677-025-00789-y
211. Goutte C, Gaussier E. A Probabilistic Interpretation of Precision, Recall and F-Score, with Implication for Evaluation. In: *Advances in Information Retrieval*. Lecture Notes in Computer Science. Berlin: Springer; 2005. p. 345–59. Available from: https://link.springer.com/chapter/10.1007/978-3-540-31865-1_25
212. Orozco-Arias S, Piña JS, Tabares-Soto R, Castillo-Ossa LF, Guyot R, Isaza G. Measuring Performance Metrics of Machine Learning Algorithms for Detecting and Classifying Transposable Elements. *Processes*. 2020;8(6):638. Available from: https://www.mdpi.com/2227-9717/8/6/638
213. Ferdous J, Islam R, Mahboubi A, Islam MZ. AI-based ransomware detection: A comprehensive review. *IEEE Access*. 2024;12:136666-136695. Available from: https://doi.org/10.1109/ACCESS.2024.3461965
214. Gazet, A. Comparative analysis of various ransomware virii. *J Comput Virol* **6**, 77–90 (2010). Available from: https://doi.org/10.1007/s11416-008-0092-2

 doi:10.20944/preprints202602.2011.v1

215. Ashwini K, Nagasundara KB. An intelligent ransomware attack detection and classification using dual vision transformer with Mantis Search Split Attention Network. *Comput Electr Eng*. 2024;119:109509. Available from: https://doi.org/10.1016/j.compeleceng.2024.104361
216. Saito T, Rehmsmeier M. The precision-recall plot is more informative than the ROC plot when evaluating binary classifiers on imbalanced datasets. *PLoS One*. 2015;10(3):e0118432. Available from: https://doi.org/10.1371/journal.pone.0118432
217. Richardson E, Trevizani R, Greenbaum JA, Carter H, Nielsen M, Peters B. The receiver operating characteristic curve accurately assesses imbalanced datasets. *Patterns*. 2024;5(6):100994. Available from: https://doi.org/10.1016/j.patter.2024.100994
218. Dhawan, S., Narwal, B. (2019). Unfolding the Mystery of Ransomware. In: Bhattacharyya, S., Hassanien, A., Gupta, D., Khanna, A., Pan, I. (eds) International Conference on Innovative Computing and Communications. Lecture Notes in Networks and Systems, vol 55. Springer, Singapore. Available from: https://doi.org/10.1007/978-981-13-2324-9_4
219. Teichmann F. Ransomware attacks in the context of generative artificial intelligence—an experimental study. *Int Cybersecurity Law Rev*. 2023;4:399–414. Available from: https://link.springer.com/article/10.1365/s43439-023-00094-x
220. Ferrante A, Medvet E, Mercaldo F, Visaggio CA, Fornaia A. Spotting the malicious moment: characterizing malware behavior using dynamic features. In: 2016 11th International Conference on Availability, Reliability and Security (ARES). IEEE; 2016. p. 372-81. Available from: [PDF] Spotting the Malicious Moment: Characterizing Malware Behavior Using Dynamic Features | Semantic Scholar
221. Iqbal A, Hussain M, Riaz Q, Khalid M, Mumtaz R, Jung K. Enhancing ransomware detection with machine learning techniques and effective API integration. *Comput Mater Contin*. 2025;85(1):1693-1714. Available from: https://doi.org/10.32604/cmc.2025.064260
222. Jeong H, Lee K. A machine learning-based ransomware detection method for attackers' neutralization techniques using format-preserving encryption. *Sensors*. 2025;25(8):2406. Available from: https://doi.org/10.3390/s25082406
223. Brodersen KH, Ong CS, Stephan KE, Buhmann JM. Class imbalance should not throw you off balance: Choosing the right classifiers and performance metrics for brain decoding with imbalanced data. *NeuroImage*. 2023;171:437-452. Available from: https://doi.org/10.1016/j.neuroimage.2023.04.044
224. Bergmeir C, Hyndman RJ, Koo B. A note on the validity of cross-validation for evaluating autoregressive time series prediction. *Comput Stat Data Anal*. 2018;120:70-83. Available from: https://doi.org/10.1016/j.csda.2017.11.003
225. Cawley GC, Talbot NLC. On over-fitting in model selection and subsequent selection bias in performance evaluation. *J Mach Learn Res*. 2010;11:2079-2107. Available from: http://jmlr.org/papers/v11/cawley10a.html
226. Shaukat K, Luo S, Varadharajan V, Hameed IA, Xu M. A novel method for improving the robustness of deep learning-based malware detectors against adversarial attacks. *Eng Appl Artif Intell*. 2022;116:105461. Available from: https://doi.org/10.1016/j.engappai.2022.105461
227. Demetrio L, Coull SE, Biggio B, Lagorio G, Armando A, Roli F. Adversarial examples: A survey and experimental evaluation of practical defenses. *Digit Threat Res Pract*. 2022;3(2):1-31. Available from: https://doi.org/10.1145/3464627
228. Al-banaa, A., Sahana, S., Ali, J., Das, S. (2023). Ransomware Taxonomy and Detection Techniques Based on Machine Learning: A Review. In: Shaw, R.N., Paprzycki, M., Ghosh, A. (eds) Advanced Communication and Intelligent Systems. ICACIS 2023. Communications in Computer and Information Science, vol 1921. Springer, Cham. Available from: https://doi.org/10.1007/978-3-031-45124-9_11
229. Suciu O, Coull SE, Johns J. Exploring adversarial examples in malware detection. In: 2019 IEEE Security and Privacy Workshops (SPW); 2019 May 19-23; San Francisco, CA, USA. IEEE; 2019. p. 8-14. Available from: https://doi.org/10.1109/SPW.2019.00015
230. Sihwail R, Omar K, Ariffin KAZ. A survey on malware detection using data mining techniques. *Symmetry*. 2019;11(10):1193. Available from: https://www.mdpi.com/2073-8994/11/10/1193

 doi:10.20944/preprints202602.2011.v1

231. Alshamrani A, Myneni S, Chowdhary A, Huang D. A survey on advanced persistent threats: techniques, solutions, challenges, and research opportunities. *IEEE Commun Surv Tutor*. 2019;21(2):1851–77. Available from: https://ieeexplore.ieee.org/document/8656877
232. Salem AH, Azzam SM, Emam OE, Abohany AA. Advancing cybersecurity: a comprehensive review of AI-driven detection techniques. *J Big Data*. 2024;11(1):105. Available from: https://journalofbigdata.springeropen.com/articles/10.1186/s40537-024-00957-y
233. Agrawal R, Stokes JW, Selvaraj K, Marinescu M. Neural sequential malware detection with parameters. In: 2019 IEEE International Conference on Big Data. IEEE; 2019. p. 1066-73. Available from: Neural Sequential Malware Detection with Parameters | IEEE Conference Publication | IEEE Xplore
234. Ahmed M, Alasad Q, Yuan JS, Alawad M. Re-evaluating deep learning attacks and defenses in cybersecurity systems. *Big Data Cogn Comput*. 2024;8(12):191. Available from: https://www.mdpi.com/2504-2289/8/12/191
235. Li J, Yang G, Shao Y. Ransomware detection model based on adaptive graph neural network learning. *Appl Sci*. 2024;14(11):4579. Available from: https://www.mdpi.com/2076-3417/14/11/4579
236. Joyce RJ, Miller G, Roth P, Zak R, Zaresky-Williams E, Anderson H, et al. EMBER2024: a benchmark dataset for holistic evaluation of malware classifiers. *arXiv*. 2025 Jun 5. Available from: https://arxiv.org/abs/2506.05074
237. Wang P, Lin HC, Chen JH, Lin WH, Li HC. Improving cyber defense against ransomware: a generative adversarial networks-based adversarial training approach for long short-term memory network classifier. *Electronics*. 2025;14(4):810. Available from: https://www.mdpi.com/2079-9292/14/4/810
238. Kaushik K, Khan A, Kumari A, Sharma I, Dubey R. Ethical considerations in AI-based cybersecurity. In: *Next-Generation Cybersecurity*. Singapore: Springer; 2024. p. 437–70. Available from: https://link.springer.com/chapter/10.1007/978-981-97-1249-6_19
239. Bruschi D, Diomede N. A framework for assessing AI ethics with applications to cybersecurity. *AI and Ethics*. 2023;3:65–72. Available from: https://link.springer.com/article/10.1007/s43681-022-00162-8
240. Owen-Jackson C. Navigating the ethics of AI in cybersecurity. *IBM Think Insights*. 2024 Oct 16. Available from: https://www.ibm.com/think/insights/navigating-ethics-ai-cybersecurity
241. Tuhin M. The dark side of AI: cybersecurity threats and privacy concerns. *Science News Today*. 2025 Mar 27. Available from: https://www.sciencenewstoday.org/the-dark-side-of-ai-cybersecurity-threats-and-privacy-concerns
242. KPMG. The ethical use of AI in cybersecurity: balancing security and privacy in the digital age. *KPMG Insights*. 2025. Available from: https://kpmg.com/us/en/articles/2025/ethical-ai-cybersecurity-balancing-security-privacy-digital-age.html
243. Meti S, Sidramayyanmath V, Patil S, Patil P. Ransomware detection using machine learning and explainable AI. In: *Artificial Intelligence: Theory and Applications (AITA 2024)*. Singapore: Springer; 2025. p. 31–46. Available from: https://link.springer.com/chapter/10.1007/978-981-96-1918-4_3
244. Preuveneers D, Joosen W. An ontology-based cybersecurity framework for AI-enabled systems and applications. *Future Internet*. 2024;16(3):69. Available from: https://www.mdpi.com/1999-5903/16/3/69
245. Galwaduge V, Samarabandu J. Novel actionable counterfactual explanations for intrusion detection using diffusion models. *J Cybersecur Priv*. 2025;5(3):68. Available from: https://www.mdpi.com/2624-800X/5/3/68
246. Nazim S, Alam MM, Rizvi SS, Mustapha JC, Hussain SS, Suud MM. Advancing malware imagery classification with explainable deep learning: a state-of-the-art approach using SHAP, LIME and Grad-CAM. *PLoS One*. 2025;20(5):e0318542. Available from: https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0318542
247. Spartalis CN, Semertzidis T, Daras P. Balancing XAI with privacy and security considerations. In: *Computer Security. ESORICS 2023 International Workshops*. Cham: Springer; 2024. p. 111–24. Available from: https://link.springer.com/chapter/10.1007/978-3-031-54129-2_7
248. Yousefi, N., Georgiopoulos, M., Anagnostopoulos, G.C. (2015). Multi-Task Learning with Group-Specific Feature Space Sharing. In: Appice, A., Rodrigues, P., Santos Costa, V., Gama, J., Jorge, A., Soares, C. (eds) Machine Learning and Knowledge Discovery in Databases. ECML PKDD 2015. Lecture Notes in Computer Science(), vol 9285. Springer, Cham. Available from: https://doi.org/10.1007/978-3-319-23525-7_8

249. Ispahany J, Islam MR, Khan MA, Islam MZ. A Sysmon incremental learning system for ransomware analysis and detection. *arXiv*. 2025 Jan 2. Available from: https://arxiv.org/html/2501.01089v1
250. Shiring B, Stanhope C, Devito H, Brigham R, Tschernov L. Adaptive ransomware detection using dynamic encryption pattern analysis. *TechRxiv*. 2025. Available from: https://www.techrxiv.org/articles/1235439
251. Chambers L, Gaber MM, Ghomeshi H. Deepstreamensemble: streaming adaptation to concept drift in deep neural networks. *Int J Mach Learn Cybern*. 2025;16:3955–76. Available from: https://link.springer.com/article/10.1007/s13042-024-02492-x
252. Almasri Y, Islam MS, Elnour M, Shinwari MW, Binbeshr F, Mahmoud A, Imam M. Ransomware detection on IoT edge using optimized deep representation learning. *TechRxiv*. 2025. Available from: https://www.techrxiv.org/articles/1267873
253. Rahmati M, Pagano A. Federated learning-driven cybersecurity framework for IoT networks with privacy-preserving and real-time threat detection capabilities. *Informatics*. 2025;12(3):62. Available from: https://www.mdpi.com/2227-9709/12/3/62
254. Jemili F, Jouini K, Korbaa O. Intrusion detection based on concept drift detection and online incremental learning. *Int J Pervasive Comput Commun*. 2025;21(1). Available from: https://www.emerald.com/insight/content/doi/10.1108/ijpcc-12-2023-0358/full/html
255. Timofte EM, Dimian M, Graur A, Potorac AD, Balan D, Croitoru I, et al. Federated learning for cybersecurity: a privacy-preserving approach. *Appl Sci*. 2025;15(12):6878. Available from: https://www.mdpi.com/2076-3417/15/12/6878
256. Peng H, Wu C, Xiao Y. FD-IDS: federated learning with knowledge distillation for intrusion detection in non-IID IoT environments. *Sensors*. 2025;25(14):4309. Available from: https://www.mdpi.com/1424-8220/25/14/4309
257. Alshamrani A. Federated hierarchical MARL for zero-shot cyber defense. *PLoS One*. 2025;20(8):e0329969. Available from: https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0329969
258. Piplai A, Mittal S, Joshi A, Finin T, Holt J, Zak R. Creating cybersecurity knowledge graphs from malware after action reports. IEEE Access. 2020;8:211691-703. Available from: Creating Cybersecurity Knowledge Graphs From Malware After Action Reports | IEEE Journals & Magazine | IEEE Xplore
259. Jimenez DMG, Solans D, Heikkila M, Vitaletti A, Kourtellis N, Anagnostopoulos A, et al. Non-IID data in federated learning: a survey with taxonomy, metrics, methods, frameworks and future directions. *arXiv*. 2024 Nov 19. Available from: https://arxiv.org/abs/2411.12377
260. Cloud Optics. The legal framework surrounding threat intelligence sharing: what organizations need to know. *Cloud Optics*. 2025. Available from: https://cloudoptics.ai/the-legal-framework-surrounding-threat-intelligence-sharing-what-organizations-need-to-know/