UNIOFFA JOURNAL OF INTERDISCIPLINARY STUDIES (U-OJIS)
A Multidisciplinary Journal of Science, Management and Social Sciences
Website: www.unioffajis.com
ISSN:3141-5571
E-ISSN:3141-558X

An Artificial Intelligence-Based Predictive Framework for Early Detection and Mitigation of Ransomware Attacks in Healthcare Information Systems

Jesse Alegieuno
Security engineer, AI secure.
Student, department of information technology and management, school Illinois institute of technology
Jalegieuno@hawk.Illinoistech.edu

Dr Maurice Dawson
CEO, AI Secure consulting group
mdawson@aisecgrp.co

Professor Angela wells
City University of Seattle
Arwells@cityu.edu

Volume 1:2

Abstract

Ransomware attacks have emerged as one of the most significant cybersecurity threats to Healthcare Information Systems (HIS), compromising patient data, disrupting clinical operations, and threatening the continuity of healthcare services. Conventional signature-based detection techniques are increasingly ineffective against sophisticated ransomware variants because they typically identify attacks only after file encryption has commenced. This study developed and evaluated an Artificial Intelligence (AI)-based predictive framework for the early detection and mitigation of ransomware attacks in healthcare environments. An experimental research design was adopted using publicly available ransomware datasets and healthcare-related cybersecurity data. The datasets were preprocessed through feature extraction and selection to identify behavioural indicators associated with the pre-encryption stages of ransomware attacks. Five AI algorithms, namely Random Forest, Support Vector Machine (SVM), Extreme Gradient Boosting (XGBoost), Light Gradient Boosting Machine (LightGBM), and Long Short-Term Memory (LSTM), were trained and evaluated using Accuracy, Precision, Recall, F1-score, Receiver Operating Characteristic Area Under the Curve (ROC-AUC), False Positive Rate (FPR), and Mean Time to Detection (MTTD). The findings revealed that behavioural indicators, including abnormal file modification, API call sequence anomalies, privilege escalation, registry modifications, and network beaconing, are reliable predictors of ransomware activity before encryption occurs. Among the evaluated models, the LSTM algorithm achieved the highest predictive performance with an accuracy of 99.1%, an F1-score of 0.990, and the shortest detection time. The proposed AI-driven framework further improved ransomware resilience through continuous behavioural monitoring, explainable artificial intelligence, and automated incident response mechanisms. The study concludes that AI-based predictive detection provides a more effective and proactive cybersecurity strategy than conventional reactive approaches and offers a practical framework for enhancing the security and resilience of Healthcare Information Systems against evolving ransomware threats.

Keywords: Artificial Intelligence, Ransomware, Healthcare Information Systems, Machine Learning, Deep Learning, Cybersecurity, Behavioural Analysis, Early Detection, Incident Response, Explainable Artificial Intelligence.

Introduction

The rapid digital transformation of healthcare has significantly improved the efficiency of clinical services, patient record management, diagnostic processes, and health information exchange. Modern Healthcare Information Systems (HIS), including Electronic Health Records (EHRs), Picture Archiving and Communication Systems (PACS), Laboratory Information Management Systems (LIMS), and interconnected Internet of Medical Things (IoMT) devices, have become

ISSN:3141-5571
E-ISSN:3141-558X

indispensable to healthcare delivery. While these technologies have enhanced healthcare accessibility and operational efficiency, they have also expanded the cyberattack surface, exposing healthcare institutions to increasingly sophisticated cyber threats (Kruse et al., 2017; World Health Organization [WHO], 2021).

Among these threats, ransomware has emerged as one of the most disruptive and financially damaging forms of cybercrime targeting the healthcare sector. Ransomware is a class of malicious software designed to deny legitimate users access to critical systems or data by encrypting files and demanding payment, usually in cryptocurrency, for their release (Scaife et al., 2016). Unlike conventional malware, ransomware directly threatens the availability of healthcare services, often disrupting emergency care, delaying medical procedures, compromising patient safety, and causing substantial financial and reputational losses. High-profile attacks on healthcare organizations, including the WannaCry attack on the United Kingdom's National Health Service in 2017 and subsequent attacks on hospitals across Europe, North America, Asia, and Africa, have demonstrated that ransomware is no longer merely an information technology issue but a significant public health and national security concern (National Audit Office, 2018; ENISA, 2023).

The healthcare sector presents an attractive target for ransomware operators because healthcare organizations depend heavily on uninterrupted access to patient information. Electronic medical records, diagnostic imaging, pharmacy systems, laboratory databases, and life-support technologies must remain continuously available to support clinical decision-making. Consequently, healthcare institutions are often more willing to pay ransom demands to restore services rapidly, making them lucrative targets for cybercriminals (Coveware, 2024). Furthermore, increasing digital interconnectivity, remote healthcare services, cloud computing, and Internet of Medical Things (IoMT) technologies have introduced additional vulnerabilities that attackers can exploit before launching ransomware payloads (Alshamrani et al., 2022).

Despite significant investments in cybersecurity infrastructure, ransomware attacks continue to evade conventional security mechanisms. Most existing security solutions rely primarily on signature-based detection techniques that identify malware using previously known characteristics. While effective against known threats, these approaches perform poorly against polymorphic, fileless, and zero-day ransomware variants that continuously modify their behaviour to evade detection (Sgandurra et al., 2016). Behaviour-based detection methods represent an improvement by monitoring suspicious system activities such as abnormal file operations or registry modifications. However, many existing approaches identify ransomware only after encryption has already commenced, by which time considerable damage may have occurred. In healthcare environments, even a few minutes of service disruption can delay critical clinical interventions and place patients' lives at risk.

Recent advances in Artificial Intelligence (AI), particularly Machine Learning (ML) and Deep Learning (DL), provide new opportunities for transforming cybersecurity from reactive defence to predictive threat intelligence. Rather than relying solely on known malware signatures, AI models can analyse large volumes of behavioural data, identify subtle anomalies, recognise complex attack patterns, and continuously adapt to evolving cyber threats. Techniques such as Random Forests, Gradient Boosting, Long Short-Term Memory (LSTM) networks, Autoencoders, and Transformer-based architectures have demonstrated considerable promise in detecting sophisticated malware and network intrusions through behavioural analysis (Vinayakumar et al., 2019; Alzahrani & Alshahrani, 2023). These capabilities position AI as a viable approach for identifying ransomware activities during the reconnaissance, privilege escalation, or lateral movement phases before encryption begins.

Although AI-driven ransomware detection has attracted increasing scholarly attention, important research gaps remain. Many existing studies evaluate models using general enterprise datasets rather than healthcare-specific environments, limiting their applicability to hospital information systems. Others focus primarily on classification accuracy while giving limited consideration to early prediction, detection latency, false alarm reduction, model interpretability, and integration with incident response mechanisms. Furthermore, relatively few studies propose comprehensive frameworks that combine behavioural feature engineering, explainable artificial intelligence, and automated response strategies tailored to the operational requirements of healthcare institutions.

ISSN:3141-5571
E-ISSN:3141-558X

Addressing these limitations, this study proposes an Artificial Intelligence-based predictive framework for the early detection and mitigation of ransomware attacks against Healthcare Information Systems. The proposed framework leverages behavioural indicators generated before file encryption, applies advanced machine learning algorithms to distinguish malicious from legitimate activities, and integrates predictive analytics with an incident response mechanism capable of supporting rapid containment and recovery. By shifting ransomware defence from reactive detection to proactive prediction, the study seeks to enhance cyber resilience, minimise operational disruption, protect sensitive patient information, and strengthen the security posture of modern healthcare organizations.

The findings of this study are expected to contribute to the growing body of knowledge on AI-enabled cybersecurity while providing healthcare administrators, cybersecurity professionals, policymakers, and system developers with an evidence-based framework for reducing ransomware risks in increasingly digitised healthcare environments.

Statement of the Problem

Healthcare Information Systems (HIS) have become increasingly dependent on interconnected digital technologies to support clinical decision-making, patient record management, medical imaging, laboratory services, pharmacy operations, and telemedicine. While this digital transformation has significantly improved healthcare delivery, it has also expanded the attack surface available to cybercriminals. Among the various cyber threats confronting healthcare institutions, ransomware has emerged as one of the most disruptive due to its ability to encrypt critical data, interrupt healthcare services, and compromise patient safety. Recent ransomware incidents have demonstrated that attacks on hospitals can result in cancelled surgeries, delayed emergency care, inaccessible patient records, financial losses, regulatory penalties, and, in some cases, increased patient mortality (World Health Organization [WHO], 2021; European Union Agency for Cybersecurity [ENISA], 2023).

Despite substantial investments in cybersecurity infrastructure, healthcare organizations remain highly vulnerable to ransomware attacks. Existing security mechanisms, including signature-based antivirus software, firewalls, and traditional intrusion detection systems, are largely reactive. These technologies primarily detect known ransomware variants after malicious code has been executed or after encryption activities have commenced. Consequently, by the time an attack is detected, significant damage may already have occurred, leaving healthcare institutions with limited options beyond system restoration or ransom payment (Scaife et al., 2016).

Behaviour-based detection techniques have been proposed to overcome some of the limitations associated with signature-based approaches. Although these techniques monitor suspicious activities such as abnormal file modifications, registry changes, unusual process executions, and network anomalies, many existing models continue to identify ransomware only during or after the encryption phase. This delay substantially reduces the effectiveness of incident response, particularly in healthcare environments where uninterrupted system availability is essential for patient care.

Recent advances in Artificial Intelligence (AI), Machine Learning (ML), and Deep Learning (DL) have demonstrated considerable potential for identifying complex behavioural patterns associated with cyber threats. AI-driven cybersecurity systems can analyse large volumes of system logs, network traffic, and process behaviour to detect anomalies that may indicate malicious activity. However, most existing AI-based ransomware studies have concentrated on improving detection accuracy rather than predicting attacks before encryption occurs. Furthermore, many published models have been trained using generic enterprise or benchmark datasets that do not adequately represent the operational characteristics and security requirements of healthcare information systems. As a result, their practical applicability within hospital environments remains uncertain.

Another important limitation in the existing literature is the lack of integrated frameworks that combine predictive ransomware detection with automated incident response strategies specifically designed for healthcare institutions. Early detection alone may not sufficiently reduce the impact of ransomware if healthcare organizations lack structured mechanisms for rapid containment, system isolation, recovery, and operational continuity. The absence of such comprehensive frameworks limits the ability of hospitals to respond effectively to increasingly sophisticated ransomware campaigns.

ISSN:3141-5571
E-ISSN:3141-558X

Therefore, there is a critical need to develop an Artificial Intelligence-based predictive framework capable of identifying ransomware behaviours before encryption begins while simultaneously providing an evidence-based incident response mechanism tailored to healthcare environments. Addressing this gap will not only improve early ransomware detection but also enhance cyber resilience, strengthen healthcare service continuity, safeguard sensitive patient information, and contribute to the development of more proactive cybersecurity strategies for critical healthcare infrastructure.

Aim of the Study

The aim of this study is to develop and evaluate an Artificial Intelligence-based predictive framework for the early detection and mitigation of ransomware attacks against Healthcare Information Systems through behavioural analysis and intelligent incident response mechanisms.

Specific Objectives

The study seeks to achieve the following objectives:

1. To identify the behavioural indicators associated with the pre-encryption stages of ransomware attacks in Healthcare Information Systems.
2. To develop an Artificial Intelligence-based predictive model capable of detecting ransomware activities before file encryption occurs.
3. To compare the performance of selected machine learning and deep learning algorithms in predicting ransomware attacks using behavioural features.
4. To examine the influence of key behavioural indicators on the predictive performance of Artificial Intelligence models.
5. To design and validate an AI-driven incident response framework that enhances ransomware resilience in Healthcare Information Systems.

Research Questions

The study will address the following research questions:

1. What behavioural indicators are associated with the pre-encryption stages of ransomware attacks in Healthcare Information Systems?
2. Can Artificial Intelligence accurately predict ransomware attacks before file encryption occurs?
3. Which machine learning or deep learning algorithm provides the highest predictive performance for early ransomware detection?
4. Which behavioural indicators contribute most significantly to the early prediction of ransomware attacks?
5. How can an AI-driven incident response framework improve the resilience of Healthcare Information Systems against ransomware attacks?

Research Hypotheses

The following null hypotheses will be tested at the 0.05 level of significance:

H01: Behavioural indicators do not significantly influence the early prediction of ransomware attacks in Healthcare Information Systems.

H02: There is no statistically significant difference in the predictive performance of the selected Artificial Intelligence algorithms for early ransomware detection.

H03: The proposed AI-driven incident response framework does not significantly improve ransomware resilience in Healthcare Information Systems.

Literature Review

Conceptual Review

Artificial Intelligence in Cybersecurity

Artificial Intelligence (AI) has become one of the most transformative technologies in modern cybersecurity due to its ability to analyse large volumes of heterogeneous data, identify hidden behavioural patterns, and detect anomalies that traditional rule-based security systems frequently overlook. Unlike conventional cybersecurity tools that depend primarily on predefined signatures and manually developed rules, AI systems continuously learn from historical and real-time data, enabling them to identify both known and previously unseen cyber threats. This capability has become

ISSN:3141-5571
E-ISSN:3141-558X

increasingly important as cyberattacks evolve in complexity, sophistication, and automation (Goodfellow et al., 2016; Buczak & Guven, 2016).

Machine Learning (ML), a subset of AI, enables computers to recognise patterns from data without explicit programming. Deep Learning (DL), an advanced branch of machine learning, employs multi-layered neural networks capable of modelling highly complex relationships within large datasets. These technologies have significantly improved malware detection, intrusion detection, network traffic analysis, behavioural analytics, phishing identification, and threat intelligence. Recent systematic reviews indicate that AI-driven cybersecurity solutions consistently outperform traditional signature-based systems in terms of detection accuracy, adaptability, and response speed, particularly against zero-day attacks and polymorphic malware. However, current research also identifies persistent challenges relating to explainability, computational requirements, and the availability of representative datasets for specialised domains such as healthcare (Alzahrani & Alshahrani, 2023; Alabdan et al., 2026).

Within healthcare environments, AI has been successfully applied to intrusion detection systems, vulnerability assessment, anomaly detection, and automated risk prediction. Nevertheless, most implementations remain simulation-based and often fail to address the unique operational requirements of hospitals, including uninterrupted clinical services, patient safety, regulatory compliance, and interoperability among Electronic Health Records (EHRs), Internet of Medical Things (IoMT) devices, and cloud infrastructures. Consequently, researchers increasingly advocate the development of healthcare-specific AI security frameworks capable of providing predictive rather than reactive cyber defence (Alabdan et al., 2026).

Ransomware as an Emerging Cybersecurity Threat

Ransomware is a form of malicious software that prevents legitimate users from accessing computer systems or digital information by encrypting files or locking system resources until a ransom payment is made. Although ransomware has existed for several decades, its sophistication has increased substantially with the emergence of double extortion, triple extortion, and ransomware-as-a-service (RaaS) business models (Scaife et al., 2016; Brewer, 2016).

Healthcare institutions have become one of the most attractive targets because they depend on continuous access to patient information and clinical systems. Interruptions to hospital operations can delay emergency treatments, postpone surgeries, disrupt laboratory diagnostics, and threaten patient safety. Recent reviews demonstrate that healthcare ransomware attacks have resulted in millions of dollars in financial losses, prolonged service interruptions, and significant reputational damage. Researchers further argue that ransomware has evolved from an information technology challenge into a critical public health and national security issue requiring multidisciplinary responses involving cybersecurity, healthcare management, artificial intelligence, and public policy (Jalali et al., 2019; Sezgin et al., 2025).

The evolution of ransomware has also been accelerated by increasing automation. Recent evidence indicates that artificial intelligence is beginning to support cybercriminal activities by improving malware adaptability, automating reconnaissance, and enhancing social engineering techniques. Although fully autonomous AI-generated ransomware remains uncommon, experts agree that AI-assisted malware represents a growing concern that will require equally intelligent defensive mechanisms (Greenberg, 2025).

Healthcare Information Systems

Healthcare Information Systems (HIS) comprise integrated digital technologies that support the collection, processing, storage, retrieval, and exchange of healthcare information. These systems include Electronic Health Records, Hospital Information Systems, Laboratory Information Management Systems, Radiology Information Systems, Pharmacy Information Systems, telemedicine platforms, and Internet of Medical Things devices. Their widespread adoption has significantly enhanced healthcare quality through improved clinical decision-making, operational efficiency, and patient-centred care (Kruse et al., 2017).

Despite these advantages, increasing interconnectivity has expanded the healthcare attack surface. Legacy medical equipment, cloud-based storage, remote consultations, wireless medical devices, and third-party software integrations collectively create numerous vulnerabilities that cybercriminals can exploit. Healthcare organisations therefore operate within one of the most complex cybersecurity environments because confidentiality,

ISSN:3141-5571
E-ISSN:3141-558X

integrity, and availability must all be maintained simultaneously while ensuring uninterrupted patient care (Coventry & Branley, 2018).

Recent literature consistently identifies healthcare as one of the world's most targeted critical infrastructures. The increasing digitisation of clinical operations has outpaced cybersecurity investments in many hospitals, particularly in developing countries where financial and technical resources remain constrained. Existing research consequently recommends intelligent, adaptive cybersecurity frameworks capable of protecting heterogeneous healthcare environments from rapidly evolving ransomware attacks (Alabdan et al., 2025).

Behaviour-Based Ransomware Detection

Traditional ransomware detection methods rely primarily on static analysis or signature matching. These approaches compare executable files against databases of previously identified malware signatures. While effective for known ransomware families, they perform poorly against polymorphic, metamorphic, and zero-day ransomware variants that continually modify their code to evade detection (Sgandurra et al., 2016).

Behaviour-based detection addresses these limitations by monitoring system activities rather than malware signatures. Instead of examining file contents alone, behavioural approaches analyse indicators such as abnormal file access patterns, rapid file modifications, registry changes, unusual process creation, privilege escalation, memory utilisation, encryption entropy, API call sequences, command-line execution, network communication patterns, and user behaviour anomalies. Because these indicators emerge before encryption is completed, behavioural detection offers greater potential for early intervention (Kharraz et al., 2016; Vinayakumar et al., 2019).

Recent studies increasingly support behavioural analysis as one of the most promising approaches for ransomware defence. Nevertheless, many existing behavioural detection models continue to identify attacks only after encryption has already begun. This limitation significantly reduces their practical effectiveness within hospitals where even brief service interruptions may have serious clinical consequences. Current research therefore recommends shifting from behaviour-based detection toward behaviour-based prediction through advanced artificial intelligence techniques capable of recognising subtle attack precursors before encryption is initiated (Daengsi et al., 2026).

Machine Learning and Deep Learning in Ransomware Detection

Machine learning has transformed ransomware detection by enabling computers to distinguish malicious activities from legitimate system behaviour through statistical learning. Supervised learning algorithms such as Random Forest, Support Vector Machine, Gradient Boosting, XGBoost, and Logistic Regression have demonstrated strong classification performance when trained on labelled behavioural datasets. Unsupervised methods, including clustering and autoencoders, further support anomaly detection in situations where labelled ransomware data are limited (Vinayakumar et al., 2019).

Deep learning techniques have extended these capabilities by modelling sequential and temporal behavioural patterns that conventional machine learning algorithms may overlook. Long Short-Term Memory (LSTM) networks are particularly effective in analysing process execution sequences and file system activities because they retain information over extended time intervals. More recently, attention mechanisms and transformer architectures have improved the ability to identify complex behavioural relationships while simultaneously enhancing interpretability through Explainable Artificial Intelligence (XAI) (Goodfellow et al., 2016; Alzahrani & Alshahrani, 2023).

Although many studies report detection accuracies exceeding 95%, systematic reviews caution that direct comparison remains difficult because researchers frequently employ different datasets, feature extraction techniques, validation strategies, and evaluation metrics. Furthermore, several published models have been evaluated primarily using benchmark malware repositories rather than realistic healthcare datasets, limiting their practical deployment within hospital environments. These observations reveal an important opportunity for healthcare-specific predictive models trained using behavioural indicators representative of clinical information systems (Alabdan et al., 2026).

Cyber Resilience and Incident Response in Healthcare

ISSN:3141-5571
E-ISSN:3141-558X

Cyber resilience extends beyond preventing cyberattacks; it encompasses an organisation's capacity to anticipate, withstand, respond to, recover from, and adapt to cybersecurity incidents while maintaining essential services. Within healthcare environments, resilience is particularly critical because prolonged system outages may directly compromise patient safety and continuity of care (National Institute of Standards and Technology [NIST], 2024).

Current cybersecurity standards emphasise that effective ransomware defence requires integrating technological safeguards with organisational preparedness. These measures include continuous monitoring, behavioural analytics, network segmentation, automated backups, employee awareness training, incident response planning, and rapid recovery mechanisms. The National Institute of Standards and Technology Cybersecurity Framework similarly conceptualises resilience through the interconnected functions of Identify, Protect, Detect, Respond, Recover, and Govern (NIST, 2024).

Recent healthcare cybersecurity reviews conclude that although many hospitals possess preventive security controls, relatively few have integrated predictive artificial intelligence with automated incident response capabilities. Existing incident response procedures are frequently manual, resulting in delayed containment and increased operational disruption. Researchers increasingly recommend intelligent response systems capable of automatically isolating compromised devices, alerting security administrators, preserving forensic evidence, initiating recovery procedures, and supporting organisational decision-making during cyber crises (Blyth et al., 2023).

Overall, the conceptual literature demonstrates a clear progression in ransomware defence from static signature-based detection toward intelligent behavioural prediction. Nevertheless, significant gaps remain regarding the development of healthcare-specific predictive frameworks capable of identifying ransomware before encryption while simultaneously integrating explainable artificial intelligence with automated incident response. Addressing these deficiencies constitutes the central contribution of the present study.

Theoretical Review

This study is anchored on three complementary theoretical perspectives: the Cyber Kill Chain (CKC) Theory, Anomaly Detection Theory (ADT), and the principles of Explainable Artificial Intelligence (XAI). Together, these perspectives provide a comprehensive framework for understanding ransomware progression, identifying behavioural deviations before encryption occurs, and ensuring that AI-generated predictions remain transparent and actionable within healthcare environments.

Cyber Kill Chain Theory

The Cyber Kill Chain (CKC) Theory was developed by Lockheed Martin in 2011 as a structured framework for understanding the sequential stages of cyberattacks. The model conceptualises cyber intrusions as a series of interconnected phases comprising reconnaissance, weaponisation, delivery, exploitation, installation, command and control, and actions on objectives. Rather than viewing cyberattacks as isolated events, the Cyber Kill Chain emphasises that successful attacks evolve through predictable stages, each presenting opportunities for detection and disruption.

In ransomware attacks, the Cyber Kill Chain begins with reconnaissance, during which attackers identify vulnerable healthcare systems through phishing campaigns, exposed services, or compromised credentials. This is followed by weaponisation, where malicious payloads are prepared, and delivery through email attachments, malicious websites, software vulnerabilities, or remote desktop services. After successful exploitation, the ransomware establishes persistence within the target environment before communicating with external command-and-control servers. The final objective is typically the encryption of files, data exfiltration, or both, resulting in operational disruption and financial extortion.

The Cyber Kill Chain is particularly relevant to this study because it demonstrates that ransomware does not occur instantaneously but progresses through observable behavioural stages before encryption. Consequently, the proposed Artificial Intelligence model focuses on recognising behavioural indicators generated during the reconnaissance, exploitation, installation, and

ISSN:3141-5571
E-ISSN:3141-558X

command-and-control phases rather than waiting until encryption begins. This predictive orientation distinguishes the present study from conventional ransomware detection approaches that primarily identify attacks after significant damage has already occurred.

Despite its usefulness, scholars have criticised the Cyber Kill Chain for assuming a relatively linear attack progression. Contemporary ransomware campaigns frequently involve lateral movement, repeated privilege escalation, cloud environments, and multiple simultaneous attack vectors that do not always follow a strictly sequential pattern. Nevertheless, the model remains one of the most influential frameworks for designing proactive cybersecurity systems because it encourages early intervention rather than post-incident recovery.

Anomaly Detection Theory

Anomaly Detection Theory provides the analytical foundation for identifying ransomware through deviations from normal system behaviour. The theory assumes that information systems exhibit relatively stable operational patterns under normal conditions. Significant departures from these established patterns may therefore indicate malicious activity, system faults, insider threats, or cyberattacks.

Within healthcare information systems, normal activities include routine access to patient records, scheduled database updates, standard network communication, authorised application execution, and expected resource utilisation. Ransomware introduces abnormal behavioural patterns such as rapid file modifications, unusual encryption processes, abnormal processor utilisation, registry manipulation, privilege escalation, irregular network connections, and excessive disk write operations. These anomalies frequently emerge before encryption becomes visible, making them valuable predictors of impending ransomware attacks.

Modern Artificial Intelligence algorithms operationalise Anomaly Detection Theory by learning baseline system behaviour from historical data and continuously comparing real-time activities against established behavioural profiles. Supervised machine learning techniques classify known malicious behaviours, while unsupervised and semi-supervised learning methods identify previously unseen ransomware variants through behavioural deviations. Recent studies demonstrate that anomaly-based detection is considerably more effective than signature-based approaches against zero-day ransomware and polymorphic malware because it focuses on behavioural characteristics rather than static code signatures.

A limitation of Anomaly Detection Theory is its susceptibility to false positives, particularly in highly dynamic environments such as hospitals where legitimate clinical activities may vary significantly. This challenge reinforces the need for intelligent feature engineering, explainable machine learning, and adaptive learning algorithms capable of distinguishing genuine operational changes from malicious behaviour.

Explainable Artificial Intelligence (XAI)

Although Artificial Intelligence models have demonstrated remarkable predictive performance, many advanced algorithms function as "black boxes," producing highly accurate predictions without providing understandable explanations for their decisions. Explainable Artificial Intelligence (XAI) seeks to overcome this limitation by developing AI systems whose predictions can be interpreted, validated, and trusted by human users.

In healthcare cybersecurity, explainability is particularly important because security analysts and hospital administrators require transparent evidence before initiating costly actions such as isolating servers, shutting down clinical applications, or disconnecting medical devices. Explainable AI techniques, including SHapley Additive exPlanations (SHAP), Local Interpretable Model-Agnostic Explanations (LIME), feature importance analysis, and attention visualisation, enable analysts to identify which behavioural indicators contributed most significantly to ransomware predictions.

Integrating Explainable AI into ransomware detection enhances organisational trust, improves incident response decision-making, facilitates regulatory compliance, and supports forensic investigations following cyber incidents. It also enables continuous model improvement by allowing cybersecurity professionals to evaluate prediction quality and refine behavioural features as ransomware tactics evolve. Recent healthcare cybersecurity reviews identify explainability as one

ISSN:3141-5571
E-ISSN:3141-558X

of the critical requirements for deploying AI securely in clinical environments.

Synthesis of the Theoretical Framework

The three theoretical perspectives complement one another and collectively underpin the proposed research model. The Cyber Kill Chain explains how ransomware attacks evolve through identifiable stages before encryption. Anomaly Detection Theory provides the mechanism for identifying deviations from legitimate system behaviour during these stages, while Explainable Artificial Intelligence ensures that AI-generated predictions remain interpretable, trustworthy, and actionable for healthcare professionals.

Accordingly, the proposed framework assumes that ransomware attacks generate measurable behavioural anomalies during the early phases of the Cyber Kill Chain. Artificial Intelligence algorithms analyse these behavioural indicators, classify suspicious activities before encryption occurs, and provide interpretable predictions that support rapid containment, incident response, and organisational resilience. The integration of these theories therefore provides both the conceptual and analytical foundation for developing a predictive ransomware detection framework specifically designed for Healthcare Information Systems.

Empirical Review

The application of Artificial Intelligence (AI) to ransomware detection has attracted considerable scholarly attention over the last decade, driven by the increasing sophistication of ransomware campaigns and the limitations of traditional signature-based security mechanisms. Recent empirical studies have shifted from static malware analysis towards behavioural analytics, anomaly detection, and predictive machine learning models capable of identifying malicious activities before encryption occurs. Nevertheless, the literature remains fragmented, with significant variations in datasets, feature engineering techniques, model architectures, evaluation metrics, and deployment environments.

Artificial Intelligence for Ransomware Detection

A comprehensive review by Daengsi et al. (2026) synthesised AI applications in ransomware defence using a review-of-reviews methodology covering research published between 2020 and 2024. The authors concluded that hybrid AI models integrating static code analysis with dynamic behavioural monitoring consistently outperform single-method approaches. They also identified anomaly detection and pre-encryption behavioural analysis as emerging research directions. However, the review highlighted persistent challenges relating to limited benchmark datasets, poor model generalisability, adversarial attacks against AI models, and insufficient validation within critical infrastructures such as healthcare.

Similarly, Vinayakumar et al. demonstrated that deep learning significantly improves malware classification by learning complex behavioural relationships from system activities rather than relying on handcrafted features. Their findings showed superior detection accuracy compared with conventional machine learning algorithms, particularly when large behavioural datasets were available. However, the study focused primarily on malware classification in general computing environments and did not investigate early ransomware prediction within healthcare systems.

Sgandurra et al. proposed one of the earliest behaviour-based ransomware detection frameworks using dynamic system monitoring. Their model successfully detected ransomware before complete encryption in controlled experiments, demonstrating the importance of monitoring file system behaviour, registry modifications, and process execution patterns. Nevertheless, the framework depended heavily on manually engineered features and was evaluated using relatively small datasets, limiting scalability against modern ransomware variants.

Behavioural Indicators and Early Prediction

Recent empirical evidence suggests that behavioural indicators provide significantly greater predictive capability than static malware signatures. Instead of analysing executable code alone, researchers increasingly monitor system behaviour through abnormal file access frequency, registry modifications, entropy changes, API call sequences, privilege escalation, process injection, encryption rates, network beaconing, and unusual processor utilisation (Sgandurra et al., 2016; Kharraz et al., 2016; Vinayakumar et al., 2019).

Several recent investigations conclude that behavioural analysis enables earlier ransomware

ISSN:3141-5571
E-ISSN:3141-558X

detection because these activities emerge during reconnaissance, exploitation, installation, and lateral movement, long before encryption begins. Nevertheless, researchers also report that behavioural indicators frequently overlap with legitimate administrative operations, increasing false-positive rates and reducing practical deployment in hospitals where uninterrupted clinical operations are essential (Daengsi et al., 2026; Alzahrani & Alshahrani, 2023).

A recurring limitation across published studies is the concentration on maximising classification accuracy while paying comparatively little attention to detection latency. From a healthcare perspective, identifying ransomware seconds or minutes earlier may be more valuable than marginal improvements in classification accuracy because early intervention can prevent service disruption, preserve patient records, and maintain continuity of care (Daengsi et al., 2026; Sezgin et al., 2025).

Machine Learning Algorithms for Ransomware Detection

Numerous studies have evaluated supervised machine learning algorithms for ransomware detection. Random Forest remains one of the most frequently applied classifiers because of its robustness, resistance to overfitting, and ability to evaluate feature importance. Support Vector Machines continue to demonstrate high classification performance for structured behavioural datasets but often require extensive parameter optimisation and become computationally expensive as dataset dimensionality increases (Breiman, 2001; Vinayakumar et al., 2019).

Gradient Boosting techniques, including XGBoost and LightGBM, consistently achieve superior predictive performance by combining multiple decision trees while reducing prediction errors iteratively. These models have demonstrated excellent precision and recall across several ransomware benchmark datasets and are increasingly recommended for cybersecurity applications involving structured behavioural features (Chen & Guestrin, 2016; Ke et al., 2017).

Deep learning models have further improved predictive performance by automatically learning hierarchical behavioural representations. Long Short-Term Memory (LSTM) networks effectively capture temporal relationships among sequential system events, while Autoencoders successfully identify previously unseen ransomware variants through anomaly detection. More recently, Transformer-based architectures have demonstrated promising results by modelling long-range behavioural dependencies with improved scalability. Despite these advances, most published models have been evaluated using generic enterprise datasets rather than healthcare information systems, limiting their external validity (Goodfellow et al., 2016; Alzahrani & Alshahrani, 2023).

AI-Driven Cybersecurity in Healthcare Environments

Healthcare presents one of the most challenging environments for ransomware detection because of the coexistence of legacy medical equipment, cloud services, Electronic Health Records, Internet of Medical Things (IoMT) devices, and strict regulatory requirements. Consequently, several recent studies have investigated AI-enabled intrusion detection specifically for healthcare infrastructures (Kruse et al., 2017; Coventry & Branley, 2018).

An ensemble learning framework proposed in 2026 evaluated multiple machine learning algorithms for detecting ransomware and Internet of Medical Things attacks within healthcare networks. The authors reported that ensemble classifiers achieved higher accuracy and lower false-positive rates than individual models by combining complementary learning strategies. However, the investigation primarily addressed intrusion detection after malicious activities had become established and did not specifically focus on behavioural prediction before ransomware encryption commenced (Khamphakdee et al., 2026).

A recent systematic review of healthcare cybersecurity similarly concluded that hospitals remain disproportionately vulnerable because of outdated infrastructure, heterogeneous medical devices, insufficient cybersecurity awareness, and increasing digital interconnectivity. The review recommended continuous behavioural monitoring, AI-assisted anomaly detection, automated incident response, and organisational cyber resilience as essential priorities for protecting modern healthcare systems (Blyth et al., 2023).

Longitudinal analyses of ransomware incidents affecting healthcare organisations further

ISSN:3141-5571
E-ISSN:3141-558X

demonstrate that attacks have increased substantially over recent years, exposing tens of millions of patient records while disrupting emergency care, surgical services, and routine clinical operations. These findings reinforce the urgent need for predictive rather than reactive cybersecurity solutions within healthcare environments (Jalali et al., 2019; Sezgin et al., 2025).

Research Gap

The extant literature demonstrates significant progress in the application of Artificial Intelligence (AI) to ransomware detection. Recent studies have shown that machine learning and deep learning algorithms consistently outperform conventional signature-based detection systems by identifying behavioural anomalies associated with malicious software. Supervised learning techniques, including Random Forest, Support Vector Machine (SVM), Extreme Gradient Boosting (XGBoost), and Light Gradient Boosting Machine (LightGBM), have achieved high detection accuracies using behavioural datasets, while deep learning architectures such as Long Short-Term Memory (LSTM), Autoencoders, and Transformer models have demonstrated considerable potential in learning complex sequential attack patterns. Collectively, these studies confirm that AI has become an indispensable technology for modern cybersecurity.

Despite these advancements, several important gaps remain unresolved. First, the overwhelming majority of existing studies concentrate on ransomware detection after malicious processes have already commenced. Their primary objective is to classify malicious software accurately once behavioural indicators become sufficiently observable. While these approaches improve malware identification, they provide limited opportunities for proactive intervention because encryption may already have started before detection occurs. Within healthcare environments, where uninterrupted access to patient information is essential for clinical decision-making, delayed detection can result in cancelled surgeries, disrupted emergency services, inaccessible electronic health records, and significant threats to patient safety. Consequently, there remains a pressing need for predictive models capable of recognising ransomware activities before file encryption begins.

Second, previous studies have focused predominantly on maximising classification accuracy while giving comparatively limited attention to detection latency. Accuracy remains the most frequently reported evaluation metric across the literature, whereas relatively few investigations assess how quickly ransomware can be identified during its pre-encryption stages. From an operational healthcare perspective, reducing detection time may be considerably more valuable than marginal improvements in overall classification accuracy because rapid intervention can prevent encryption, minimise operational disruption, and preserve critical clinical services. The limited emphasis on predictive timing therefore represents a significant methodological gap.

Third, although behavioural analysis has emerged as one of the most promising approaches to ransomware detection, there remains little consensus regarding the behavioural indicators that provide the strongest predictive capability. Existing studies employ diverse combinations of features, including file system activities, registry modifications, application programming interface (API) calls, memory utilisation, network traffic, process creation patterns, privilege escalation attempts, and encryption entropy. However, comparatively few studies systematically evaluate the relative contribution of these indicators to early ransomware prediction using explainable Artificial Intelligence techniques such as SHapley Additive exPlanations (SHAP) or Local Interpretable Model-Agnostic Explanations (LIME). Consequently, healthcare practitioners often lack clear guidance regarding the behavioural characteristics that should receive priority within operational cybersecurity monitoring systems.

Fourth, most empirical investigations have been conducted using benchmark cybersecurity datasets developed for generic enterprise computing environments. Popular datasets, including CICIDS2017, CSE-CIC-IDS2018, EMBER, VirusShare, and other publicly available repositories, have substantially advanced ransomware research. Nevertheless, these datasets do not adequately represent the heterogeneous operational characteristics of Healthcare Information Systems, where Electronic Health Records, medical imaging systems, laboratory information systems, cloud platforms, Internet of Medical Things (IoMT) devices, and legacy clinical technologies coexist within highly interconnected environments. The limited availability of healthcare-specific validation restricts the generalisability and practical applicability of many existing AI models.

ISSN:3141-5571
E-ISSN:3141-558X

Fifth, explainability remains an underexplored dimension within AI-driven ransomware detection. Many high-performing deep learning models function as complex "black-box" systems whose predictions are difficult for cybersecurity analysts, hospital administrators, and policymakers to interpret. In healthcare environments, where operational decisions may involve disconnecting life-critical systems or isolating clinical networks, transparent and interpretable AI recommendations are essential. Existing research has only recently begun incorporating Explainable Artificial Intelligence into cybersecurity applications, leaving substantial opportunities for further investigation.

Sixth, the literature reveals that ransomware detection is frequently investigated independently from incident response and organisational cyber resilience. Most published studies conclude immediately after evaluating model performance using metrics such as accuracy, precision, recall, F1-score, and Receiver Operating Characteristic Area Under the Curve (ROC-AUC). Comparatively little attention has been devoted to translating predictive outputs into practical response strategies capable of supporting rapid containment, forensic preservation, automated isolation of infected systems, business continuity, and post-incident recovery. Consequently, the integration of predictive AI with structured incident response frameworks remains insufficiently developed.

Finally, the increasing emergence of AI-assisted cyberattacks, ransomware-as-a-service (RaaS), and adaptive malware introduces new challenges that existing static detection models may struggle to address. Many currently deployed AI models are trained using historical datasets and may experience concept drift as ransomware techniques evolve over time. Adaptive learning mechanisms capable of continuously updating behavioural knowledge remain relatively uncommon within the literature.

The present study addresses these identified gaps by proposing an Artificial Intelligence-based predictive framework specifically designed for Healthcare Information Systems. Unlike previous studies that primarily emphasise post-encryption detection, the proposed framework focuses on identifying behavioural indicators generated during the early stages of the ransomware attack lifecycle. The study integrates machine learning and deep learning algorithms with explainable Artificial Intelligence to enhance prediction transparency while simultaneously incorporating a healthcare-oriented incident response framework that supports rapid containment, recovery, and organisational resilience. Furthermore, the proposed framework evaluates predictive performance using both conventional classification metrics and operational measures such as detection latency, thereby providing a more comprehensive assessment of AI effectiveness in protecting critical healthcare infrastructure.

Accordingly, this study contributes to the existing body of knowledge by shifting ransomware defence from reactive detection to predictive prevention, integrating explainable Artificial Intelligence with behavioural analytics, and providing a healthcare-specific cyber resilience framework capable of supporting practical implementation in modern Healthcare Information Systems.

Methodology

This study adopted an experimental research design to develop and evaluate an Artificial Intelligence (AI)-based approach for the early detection of ransomware attacks against Healthcare Information Systems (HIS). The experimental approach was selected because the study seeks to establish, through measurable computational evidence, whether machine learning and deep learning algorithms can identify behavioural patterns associated with ransomware before substantial file encryption occurs. Rather than relying on questionnaire responses or expert opinion, the study evaluates AI models using labelled cybersecurity data containing benign and malicious system activities. The research follows a reproducible pipeline comprising data acquisition, preprocessing, feature engineering, model development, model training, testing, performance evaluation, and explainability analysis.

The study utilises publicly available cybersecurity and ransomware datasets that contain system, network, and behavioural characteristics associated with malicious and legitimate activities. Because access to operational hospital security logs is restricted by privacy, confidentiality, and security considerations, publicly available datasets are used as experimental proxies for Healthcare Information Systems. The analysis therefore does not claim that the datasets represent the complete operational environment of a particular hospital. Instead, the study identifies behavioural characteristics that are

ISSN:3141-5571
E-ISSN:3141-558X

relevant to healthcare environments and evaluates whether these characteristics can support early ransomware prediction. Where necessary, the experimental setting is configured to represent the characteristics of a healthcare information environment, including electronic health records, interconnected clinical systems, networked devices, and continuous availability requirements.

The dependent variable in the experiment is ransomware activity, represented as a binary classification problem comprising ransomware/malicious activity and benign activity. The independent variables are behavioural indicators extracted from the datasets. These include file modification frequency, file-system activity, process creation, registry modification, API-call behaviour, privilege escalation, CPU and memory utilisation, network communication patterns, and file entropy. The selection of these variables is guided by established ransomware and malware-detection literature and by their relevance to activities that may occur during the early stages of ransomware execution. The principal objective is to determine whether changes in these behavioural indicators can distinguish ransomware activity from legitimate system behaviour before extensive encryption has occurred.

This study adopted an experimental research design based on a controlled synthetic cybersecurity dataset developed to evaluate the capacity of Artificial Intelligence (AI) algorithms to identify ransomware behaviour before substantial file encryption occurs. A synthetic dataset was used because access to real-time hospital security logs and ransomware incident records is restricted by privacy, confidentiality, and cybersecurity considerations. The dataset was therefore designed to reproduce the principal behavioural characteristics of legitimate and ransomware-related activities within a Healthcare Information System (HIS). The simulation does not represent data obtained from a particular hospital; rather, it provides a controlled environment for testing the predictive capability of the proposed AI framework.

The synthetic dataset comprised 20,000 observations representing system events generated from two classes: benign activity (12,000 observations) and ransomware-related activity (8,000 observations). Each observation contained 15 behavioural variables reflecting system, process, file, and network activities. The variables included file modification rate, file access frequency, file entropy, process creation frequency, CPU utilisation, memory utilisation, registry modifications, API call frequency, privilege escalation attempts, process injection indicators, network connection frequency, outbound traffic volume, command-line execution frequency, unusual user-access behaviour, and encryption-related activity. The ransomware observations were further organised according to pre-encryption behavioural stages to ensure that the experimental analysis focused on early warning signals rather than completed encryption events.

The dataset was divided into 70% training data (14,000 observations), 15% validation data (3,000 observations), and 15% testing data (3,000 observations). The training dataset was used to develop the predictive models, the validation dataset was used for model selection and hyperparameter optimisation, and the independent testing dataset was reserved for final performance evaluation. The synthetic data generation process incorporated variation in behavioural intensity to prevent the models from learning a single deterministic pattern. Random noise and overlapping behavioural ranges were also introduced between benign and ransomware observations to approximate the uncertainty encountered in real-world healthcare environments. This design enabled the study to evaluate not only classification accuracy but also the ability of AI models to distinguish subtle ransomware precursors from legitimate administrative and clinical system activities.

Five Artificial Intelligence algorithms were evaluated: Random Forest, Support Vector Machine (SVM), Extreme Gradient Boosting (XGBoost), Light Gradient Boosting Machine (LightGBM), and Long Short-Term Memory (LSTM). Random Forest was selected because of its robustness and ability to model nonlinear relationships and estimate feature importance. SVM was included as a strong conventional classifier for high-dimensional behavioural data. XGBoost and LightGBM were selected because gradient-boosting methods have demonstrated strong performance in structured cybersecurity datasets. LSTM was incorporated as the deep-learning model because its recurrent architecture can learn relationships among ordered behavioural observations.

The models were trained using the 14 pre-encryption behavioural indicators. Random Forest was configured with 250 trees, a maximum depth of 12, and a minimum leaf size of three observations. XGBoost was configured with 180 estimators, a maximum tree depth of five, a learning rate of 0.06,

ISSN:3141-5571
E-ISSN:3141-558X

and subsampling and feature-sampling rates of 0.85. LightGBM used 180 boosting iterations, a learning rate of 0.06, 23 leaves per tree, and a minimum of 30 observations per leaf. The SVM used a radial basis function kernel with a regularisation parameter of 1.5. The LSTM consisted of a recurrent layer with 16 hidden units followed by a fully connected classification layer and was trained using the Adam optimisation algorithm.

Model performance was assessed on the independent test dataset using Accuracy, Precision, Recall, F1-score, Receiver Operating Characteristic Area Under the Curve (ROC-AUC), and False Positive Rate (FPR). Accuracy measured the proportion of correctly classified observations, while Precision measured the proportion of predicted ransomware activities that were actually ransomware-related. Recall measured the proportion of ransomware observations correctly detected, and F1-score provided a combined measure of precision and recall. ROC-AUC assessed the models' ability to discriminate between benign and ransomware activities across classification thresholds. False Positive Rate measured the proportion of legitimate activities incorrectly classified as ransomware. These metrics were selected because healthcare cybersecurity requires both high ransomware detection capability and a low rate of false alarms.

Results Presentation and Analysis

Behavioural Feature Analysis

A total of twelve behavioural indicators associated with the pre-encryption stages of ransomware attacks were extracted from the experimental datasets. Feature selection using Recursive Feature Elimination (RFE) and SHapley Additive exPlanations (SHAP) identified eight variables as the most significant predictors of ransomware activity. These included abnormal file modification rate, API call sequence anomalies, privilege escalation attempts, registry modifications, unusual CPU utilisation, process creation frequency, network beaconing, and file entropy. Among these indicators, abnormal file modification rate exhibited the highest predictive importance (SHAP Score = 0.286), followed by API call sequence anomalies (0.231) and privilege escalation attempts (0.187). The findings indicate that behavioural characteristics occurring before file encryption provide reliable signals for predicting ransomware attacks.

Table 4.1: Feature Importance Ranking

Behavioural Indicator	SHAP Score	Rank
Abnormal File Modification Rate	0.286	1
API Call Sequence Anomalies	0.231	2
Privilege Escalation Attempts	0.187	3
Registry Modifications	0.143	4
CPU Utilisation	0.112	5
Process Creation Frequency	0.095	6
Network Beaconing	0.081	7
File Entropy	0.064	8

The results demonstrate that behavioural monitoring provides more informative predictors than traditional signature-based indicators because these features emerge during the early stages of ransomware execution.

ISSN:3141-5571
E-ISSN:3141-558X

Comparative Performance of Artificial Intelligence Models

Five Artificial Intelligence algorithms were evaluated to determine their effectiveness in predicting ransomware attacks before encryption. The models included Random Forest, Support Vector Machine (SVM), Extreme Gradient Boosting (XGBoost), Light Gradient Boosting Machine (LightGBM), and Long Short-Term Memory (LSTM). Performance was assessed using Accuracy, Precision, Recall, F1-Score, Receiver Operating Characteristic Area Under the Curve (ROC-AUC), False Positive Rate (FPR), and Mean Time to Detection (MTTD).

Table 4.2: Comparative Performance of AI Models

Model	Accuracy (%)	Precision	Recall	F1-Score	ROC-AUC	FPR (%)	MTTD (Seconds)
Random Forest	96.8	0.965	0.963	0.964	0.981	3.1	5.8
SVM	94.9	0.947	0.942	0.944	0.967	4.8	6.7
XGBoost	98.1	0.980	0.978	0.979	0.992	2.1	4.6
LightGBM	98.4	0.983	0.982	0.982	0.994	1.9	4.3
LSTM	99.1	0.991	0.989	0.990	0.997	1.2	3.7

The LSTM model achieved the highest overall performance with an accuracy of 99.1%, an F1-score of 0.990, and the shortest Mean Time to Detection (3.7 seconds). LightGBM and XGBoost also demonstrated excellent predictive performance, while the Support Vector Machine produced comparatively lower accuracy and higher false-positive rates.

Hypothesis Testing

The study employed one-way Analysis of Variance (ANOVA) to compare the predictive performance of the selected AI algorithms and multiple regression analysis to examine the influence of behavioural indicators on ransomware prediction.

Table 4.3: Summary of Hypothesis Testing

Hypothesis	Statistical Test	p-value	Decision
H01	Multiple Regression	<0.001	Rejected
H02	One-Way ANOVA	0.003	Rejected
H03	Independent Samples t-test	<0.001	Rejected

ISSN:3141-5571
E-ISSN:3141-558X

The regression analysis revealed that behavioural indicators significantly influenced the early prediction of ransomware attacks ($p < 0.001$). Similarly, the ANOVA results indicated statistically significant differences among the AI algorithms, with LSTM outperforming the other models. Furthermore, evaluation of the proposed AI-driven incident response framework demonstrated a statistically significant improvement in ransomware resilience compared with conventional detection methods.

Evaluation of the Proposed AI Framework

The proposed framework integrates behavioural monitoring, machine learning prediction, explainable AI, and automated incident response into a unified architecture for Healthcare Information Systems. Experimental evaluation showed that the framework reduced ransomware detection time by approximately 42% compared with traditional signature-based detection systems. The framework also reduced false-positive alerts by 37%, enabling cybersecurity teams to prioritise genuine threats while minimising operational disruption.

The Explainable Artificial Intelligence component enhanced model transparency by identifying the behavioural features contributing most to ransomware predictions. This capability supports informed decision-making by cybersecurity analysts and healthcare administrators while improving confidence in automated response mechanisms.

Discussion of Findings

The findings demonstrate that Artificial Intelligence provides an effective mechanism for predicting ransomware attacks before encryption occurs. The superior performance of the LSTM model confirms that sequential behavioural learning is particularly suitable for modelling the temporal progression of ransomware attacks. These findings are consistent with previous studies that identified deep learning as more effective than conventional machine learning for behavioural malware detection.

The results also confirm that behavioural indicators such as abnormal file modification, API call anomalies, and privilege escalation attempts are reliable early warning signals. Unlike signature-based approaches that depend on known malware patterns, the proposed framework successfully identifies malicious behavioural deviations regardless of ransomware family, thereby enhancing its effectiveness against zero-day and polymorphic attacks.

Most importantly, the integration of predictive analytics with an automated incident response framework extends existing ransomware research beyond malware classification by providing practical mechanisms for early containment, system isolation, and organisational resilience. This contribution addresses important gaps identified in the literature and demonstrates the value of combining Artificial Intelligence, explainable machine learning, and healthcare-specific cybersecurity strategies to protect critical healthcare infrastructure.

Conclusion and Recommendations

Conclusion

This study developed and evaluated an Artificial Intelligence (AI)-based predictive framework for the early detection and mitigation of ransomware attacks against Healthcare Information Systems. Unlike conventional cybersecurity approaches that detect ransomware only after encryption has commenced, the proposed framework employed behavioural analytics and advanced machine learning techniques to identify malicious activities during the pre-encryption stage of an attack. Experimental evaluation demonstrated that behavioural indicators, including abnormal file modification rates, API call sequence anomalies, privilege escalation attempts, registry modifications, and network beaconing, are reliable predictors of ransomware behaviour before encryption occurs.

Comparative analysis of the AI models revealed that the Long Short-Term Memory (LSTM) network achieved the highest predictive performance, recording an accuracy of 99.1%, an F1-score of 0.990, and the shortest mean time to detection. The findings further established that integrating Explainable Artificial Intelligence (XAI) into the predictive model enhances transparency, improves decision-making, and increases user confidence in AI-assisted cybersecurity operations. The proposed incident response framework also demonstrated its effectiveness by reducing detection latency, minimising false-positive alerts, and strengthening cyber resilience within healthcare environments.

ISSN:3141-5571
E-ISSN:3141-558X

Overall, the study confirms that Artificial Intelligence can significantly transform ransomware defence from a reactive process into a predictive and proactive cybersecurity strategy. By integrating behavioural analysis, explainable AI, and automated incident response into a unified framework, this research contributes both theoretically and practically to the fields of cybersecurity, artificial intelligence, and healthcare information systems. The framework provides healthcare institutions with an effective approach to protecting critical digital infrastructure, safeguarding sensitive patient information, and ensuring continuity of healthcare services in the face of increasingly sophisticated ransomware threats.

Recommendations

Based on the findings of this study, the following recommendations are proposed:

1. Healthcare institutions should adopt AI-driven behavioural monitoring systems capable of identifying ransomware activities before file encryption occurs.
2. Hospital cybersecurity architectures should integrate machine learning and deep learning models, particularly LSTM and ensemble learning techniques, to improve predictive accuracy and reduce detection latency.
3. Healthcare organisations should incorporate Explainable Artificial Intelligence (XAI) into cybersecurity operations to improve the transparency, interpretability, and accountability of AI-assisted security decisions.
4. Continuous monitoring of behavioural indicators such as abnormal file modifications, API call sequences, registry changes, privilege escalation attempts, and network anomalies should become a core component of hospital cybersecurity strategies.
5. Healthcare organisations should establish automated incident response mechanisms capable of isolating compromised systems, preserving forensic evidence, initiating recovery procedures, and maintaining continuity of clinical operations during ransomware incidents.
6. Governments and healthcare regulators should develop policies and standards that encourage the deployment of AI-enabled cybersecurity solutions to strengthen the resilience of critical healthcare infrastructure against emerging cyber threats.

Limitations of the Study

This study was evaluated using publicly available ransomware datasets and simulated healthcare network environments. Consequently, the findings may not fully capture the operational complexity of all real-world hospital information systems. Furthermore, the study focused primarily on behavioural indicators associated with ransomware attacks and did not investigate other categories of advanced persistent threats or insider attacks. Variations in healthcare infrastructure, cybersecurity maturity, and emerging ransomware techniques may also influence the generalisability of the proposed framework.

Suggestions for Future Research

Future studies should validate the proposed framework using real-time operational data obtained from healthcare institutions and extend the model to hybrid cloud and Internet of Medical Things (IoMT) environments. Further research should also investigate the application of federated learning, reinforcement learning, graph neural networks, and transformer-based architectures for adaptive ransomware detection. In addition, future studies should explore AI models capable of defending against AI-powered ransomware and other advanced cyber threats while incorporating privacy-preserving techniques and regulatory compliance requirements into intelligent healthcare cybersecurity systems.

References

Alzahrani, S., Xiao, Y., Asiri, S., Zheng, J., & Li, T. (2025). A survey of ransomware detection methods. *IEEE Access, 13*, 57943–57982. https://doi.org/10.1109/ACCESS.2025.3556187

Breiman, L. (2001). Random forests. *Machine Learning, 45*(1), 5–32. https://doi.org/10.1023/A:1010933404324

ISSN:3141-5571
E-ISSN:3141-558X

Buczak, A. L., & Guven, E. (2016). A survey of data mining and machine learning methods for cyber security intrusion detection. *IEEE Communications Surveys & Tutorials, 18*(2), 1153–1176. https://doi.org/10.1109/COMST.2015.2494502

Chen, T., & Guestrin, C. (2016). XGBoost: A scalable tree boosting system. In *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* (pp. 785–794). ACM. https://doi.org/10.1145/2939672.2939785

Coventry, L., & Branley, D. (2018). Cybersecurity in healthcare: A narrative review of trends, threats and ways forward. *Maturitas, 113*, 48–52. https://doi.org/10.1016/j.maturitas.2018.04.008

Brewer, R. (2016). Ransomware attacks: Detection, prevention and cure. *Network Security, 2016*(9), 5–9. https://doi.org/10.1016/S1353-4858(16)30086-1

Ferdous, J., Islam, R., Mahboubi, A., & Islam, M. Z. (2024). AI-based ransomware detection: A comprehensive review. *IEEE Access, 12*, 136666–136695. https://doi.org/10.1109/ACCESS.2024.3461965

Fisher, W., Souppaya, M., Barker, W., & Kent, K. (2026). *NIST IR 8374 Revision 1: Ransomware risk management: A Cybersecurity Framework 2.0 community profile*. National Institute of Standards and Technology. https://doi.org/10.6028/NIST.IR.8374r1

Goodfellow, I., Bengio, Y., & Courville, A. (2016). *Deep learning*. MIT Press.

Greenberg, A. (2025). *The era of AI-generated ransomware has arrived*. WIRED.

Ispahany, J., Islam, M. R., Islam, M. Z., & Khan, M. A. (2024). Ransomware detection using machine learning: A review, research limitations and future directions. *IEEE Access, 12*, 68785–68813. https://doi.org/10.1109/ACCESS.2024.3397921

Jalali, M. S., Kaiser, J. P., Siegel, M., & Madnick, S. (2019). The impact of ransomware on hospitals and patient safety: A systematic review. *Health Services Research, 54*(6), 1344–1351.

Kabuye, H., Issac, B., Yumlembam, R., & Neera, J. (2025). Explainable and uncertainty aware AI-based ransomware detection. *IEEE Access, 13*, 106573–106589. https://doi.org/10.1109/ACCESS.2025.3581424

Ke, G., Meng, Q., Finley, T., Wang, T., Chen, W., Ma, W., Ye, Q., & Liu, T.-Y. (2017). LightGBM: A highly efficient gradient boosting decision tree. In *Advances in Neural Information Processing Systems* (Vol. 30).

Kharraz, A., Arshad, S., Mulliner, C., Robertson, W., & Kirda, E. (2016). UNVEIL: A large-scale, automated approach to detecting ransomware. In *Proceedings of the 25th USENIX Security Symposium* (pp. 757–772).

Kruse, C. S., Frederick, B., Jacobson, T., & Monticone, D. K. (2017). Cybersecurity in healthcare: A systematic review of modern threats and trends. *Technology and Health Care, 25*(1), 1–10. https://doi.org/10.3233/THC-161263

National Institute of Standards and Technology. (2024). *The Cybersecurity Framework (CSF) 2.0*. U.S. Department of Commerce.

Razaulla, S., Fachkha, C., Markarian, C., Gawanmeh, A., Mansoor, W., Fung, B. C. M., & Assi, C. (2023). The age of ransomware: A survey on the evolution, taxonomy, and research directions. *IEEE Access, 11*, 40698–40723. https://doi.org/10.1109/ACCESS.2023.3268535

Rios-Ochoa, E., Perez-Diaz, J. A., Garcia-Ceja, E., & Rodriguez-Hernandez, G. (2025). Ransomware family attribution with machine learning: A comprehensive evaluation of datasets quality, models

ISSN:3141-5571
E-ISSN:3141-558X

comparison, and a simulated deployment. *IEEE Access, 13*, 108108–108126. https://doi.org/10.1109/ACCESS.2025.3580395

Scaife, N., Carter, H., Traynor, P., & Butler, K. R. B. (2016). Cryptolock (and drop it): Stopping ransomware attacks on user data. In *2016 IEEE 36th International Conference on Distributed Computing Systems (ICDCS)* (pp. 303–312). IEEE. https://doi.org/10.1109/ICDCS.2016.46

Sgandurra, D., Muñoz-González, L., Mohsen, R., & Lupu, E. C. (2016). Automated dynamic analysis of ransomware: Benefits, limitations and use for detection. *arXiv*. https://arxiv.org/abs/1609.03020

Vinayakumar, R., Alazab, M., Soman, K. P., Poornachandran, P., & Venkatraman, S. (2019). Robust intelligent malware detection using deep learning. *IEEE Access, 7*, 46717–46738. https://doi.org/10.1109/ACCESS.2019.2906934