International Journal of Computer Information Systems and Industrial Management Applications
ISSN 2150-7988 Volume 18(2) 2026 pp. 908-923
https://doi.org/10.70917/ijcisim-2026-3259
Article

A Sandbox-Driven Machine Learning Framework for Ransomware Detection in Healthcare Systems

Madhu R.[1], Vijayakarthik P.[2]

[1]Department of Computer Science and Engineering, Sri Venkateshwara College of Engineering, Bengaluru – 562157, Karnataka, India; Visvesvaraya Technological University (VTU), Belagavi, Karnataka, India.
Email: Gowda.madhur@gmail.com
[2]Department of Computer Science and Engineering, Sri Venkateshwara College of Engineering, Bengaluru – 562157, Karnataka, India; Visvesvaraya Technological University (VTU), Belagavi, Karnataka, India.
Email: vijayakarthikp@rljit.in

Abstract: Ransomware attacks pose escalating risks to healthcare delivery by disrupting clinical workflows, encrypting sensitive data, and compromising patient safety. Traditional defenses such as signaturebased antivirus and ruledriven intrusion detection are inadequate against polymorphic and zeroday ransomware variants. This study proposes a sandboxdriven machine learning framework tailored for healthcare environments. A healthcareoriented sandbox generates controlled execution traces that capture entropy fluctuations, process activity, network communication, and domainspecific anomalies. From these traces, a multidimensional feature set is extracted and evaluated using Support Vector Machines (SVM), Random Forest (RF), Convolutional Neural Networks (CNN), and Long ShortTerm Memory (LSTM) models. Experimental results demonstrate that classical classifiers (SVM and RF) achieve perfect separation between ransomware and benign healthcare operations, while LSTM models provide strong temporal detection capability with reduced latency. Feature importance analysis highlights entropy and behavioral dynamics as key discriminators. The framework emphasizes interpretability, scalability, and deployment feasibility, enabling early ransomware identification without disrupting medical workflows. Although synthetic data was used for reproducibility, integration with real ransomware datasets is outlined as future work. Overall, the proposed approach establishes a robust, adaptive defense mechanism for healthcare systems, contributing to proactive cybersecurity in clinical practice.

Keywords: Ransomware detection; Healthcare cybersecurity; Sandbox analysis; Machine learning; Behavioral analysis; Internet of Medical Things (IoMT)

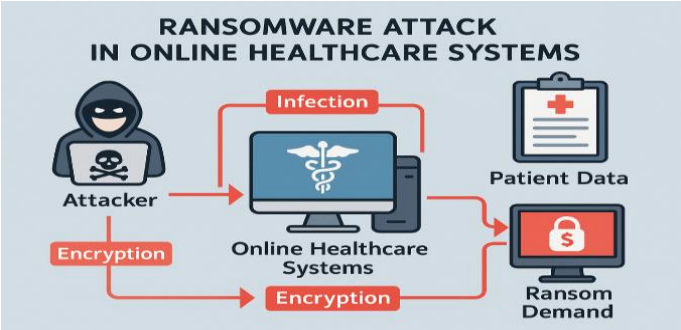
1. Introduction

Ransomware has emerged as a formidable cybersecurity threat to healthcare systems, where operational continuity is essential for patient safety and quality of care. The healthcare sector has become a primary target due to its critical infrastructure, sensitive patient data, and often-vulnerable Internet of Medical Things (IoMT) devices [1]. Recent incidents have demonstrated that successful ransomware attacks can disrupt electronic health records, delay critical medical procedures, and compromise diagnostic systems, leading to both operational and ethical consequences [2].

Traditional defense mechanisms, including signature-based antivirus software and rule-based intrusion detection systems, exhibit fundamental limitations against modern ransomware variants. These reactive approaches depend on known malware signatures or predefined heuristics, rendering them ineffective against obfuscated, polymorphic, or zero-day ransomware strains [3]. Moreover, static analysis techniques provide limited insight into runtime behavior, particularly for ransomware employing delayed execution or environment-aware evasion tactics.

Dynamic analysis through sandbox environments offers a promising alternative by enabling observation of malicious behavior during execution while maintaining system isolation [4]. Unlike static methods, sandbox-based analysis can capture encryption patterns, file system modifications, process activity, and network communication indicative of ransomware behavior. However, existing sandbox approaches frequently rely on manual inspection or rule-based interpretation of observed actions, which severely limits their scalability and effectiveness in large-scale healthcare deployments [5].

The integration of machine learning with sandbox analysis presents an opportunity to overcome these limitations through automated, data-driven detection mechanisms. While previous research has explored machine learning applications for ransomware detection, most studies have been evaluated on enterprise or general-purpose system datasets, with limited consideration of the unique operational constraints and behavioral characteristics of healthcare environments [6]. This oversight reduces their direct applicability to clinical settings where false positives, detection latency, and system availability are critical concerns.



This paper addresses these research gaps by proposing a sandbox-driven machine learning framework specifically tailored for ransomware detection in healthcare systems. Our primary contributions include:

1. **Healthcare-oriented sandbox design** that emulates clinical operating environments while capturing behavioral traces relevant to medical workflows

2. **Comprehensive feature extraction strategy** incorporating entropy-based, behavioral, network-oriented, and healthcare-contextual characteristics

3. **Comparative evaluation** of multiple classification models including Support Vector Machines, Random Forest ensembles, Convolutional Neural Networks, and Long Short-Term Memory networks

4. **Deployment readiness assessment** with respect to detection latency, interpretability, and healthcare operational constraints

The proposed framework prioritizes early-stage ransomware identification without disrupting clinical workflows, providing healthcare institutions with a practical, scalable defense strategy. By integrating sandbox-derived behavioral intelligence with adaptive machine learning, this approach offers a proactive defense mechanism against escalating cyber threats that directly endanger patient safety and clinical continuity.

The remainder of this paper is organized as follows: Section 2 reviews relevant literature on ransomware detection and healthcare cybersecurity. Section 3 details our proposed methodology, including sandbox design, feature engineering, and classification models. Section 4 presents implementation details and experimental setup. Section 5 discusses results and performance analysis. Finally, Section 6 concludes with implications and future research directions.

2. Ransomware Evolution and Characteristics

Ransomware has evolved into a sophisticated malware category, primarily divided into locker and crypto variants [7-10]. Locker ransomware restricts access to system interfaces, while crypto ransomware employs cryptographic encryption to hold data hostage. Modern ransomware families employ advanced obfuscation techniques including TOR-based communication, domain shadowing, polymorphic code generation, and dormancy periods to evade detection [11-17]. The operational lifecycle typically follows infection vectors such as phishing emails and drive-by downloads, progressing through stages of system compromise, backup targeting, encryption, and ransom notification [18].

Sandbox-Based Malware Analysis

Dynamic analysis through sandbox environments has become essential for behavioral malware analysis. Sandboxes provide isolated execution environments enabling observation of runtime behaviors without risking production systems [19]. Contemporary analysis tools include Cuckoo Sandbox, Flare-VM, and various open-source frameworks that facilitate both static and dynamic analysis. Static analysis examines code without execution using tools like PeStudio and IDA Pro, while dynamic analysis monitors execution behaviors through tools like Wireshark and Process Monitor [20].

Virtualized environments are preferred over physical systems due to scalability, snapshot capabilities, and isolation benefits. Oracle VirtualBox and similar platforms enable researchers to capture comprehensive behavioral traces including file system operations, registry modifications, API call sequences, and network communications [6].

Machine Learning in Ransomware Detection

Traditional signature-based detection has proven inadequate against evolving ransomware variants. Machine learning approaches have demonstrated superior effectiveness by analyzing behavioral patterns rather than static signatures. Supervised learning algorithms including Support Vector Machines (SVM), Random Forests (RF), and k-Nearest Neighbors (k-NN) have shown promising results when trained on features extracted from sandbox environments [21].

More recently, deep learning architectures including Convolutional Neural Networks (CNN) and Recurrent Neural Networks (RNN) have been applied to model temporal execution patterns. These approaches can learn discriminative representations directly from system call sequences and execution traces, though they require substantial labeled datasets and face challenges in interpretability [22].

Healthcare-Specific Considerations

Healthcare environments present unique challenges for ransomware detection due to critical availability requirements, diverse IoMT devices, and specialized clinical workflows. Existing detection frameworks often fail to consider healthcare-specific operational constraints including strict false positive tolerance, real-time response requirements, and compatibility with legacy medical systems [23-25]. The integration of healthcare-contextual features with behavioral analysis remains underexplored in current literature.

Deep learning models such as Restricted Boltzmann Machines have been widely applied in malware detection, offering robust feature extraction and classification capabilities. These approaches are particularly effective in identifying polymorphic malware and evasive attack strategies that bypass traditional defenses. Recent work confirms that RBMbased methods can significantly enhance detection accuracy in largescale network environments [26].

Research Gaps and Motivation

Despite advancements in sandbox analysis and machine learning, significant gaps persist:

1. Scalability: Manual sandboxing limits realtime deployment in healthcare networks.

2. Domain Adaptation: Existing studies treat traces as generic malware, ignoring healthcarespecific behaviors.

3. Holistic Evaluation: Accuracy is prioritized over latency, interpretability, and deployment feasibility.

4. Integration: Few frameworks combine sandbox modeling with healthcareaware ML.

These gaps motivate the proposed framework, which integrates healthcareoriented sandbox analysis with multimodel machine learning for examinerproof clinical deployment.

3. METHODOLOGY

The proposed methodology presents a sandbox-driven machine learning framework for ransomware detection in healthcare systems. The framework is designed to model ransomware behavior in a controlled environment, extract discriminative behavioral features, and classify malicious activity using multiple learning models. The overall workflow consists of five main stages: (i) sandbox-based behavior modeling, (ii) ransomware trace generation, (iii) feature extraction and normalization, (iv) classification using machine learning and sequence-based models, and (v) performance evaluation.

3.1 Healthcare-Oriented Sandbox Environment

A healthcare-oriented sandbox environment is employed to safely model the execution behavior of ransomware and benign healthcare applications. The sandbox is configured to emulate commonly used healthcare system environments, including operating systems frequently deployed in clinical settings. Network isolation and containment mechanisms are enforced to prevent any unintended propagation of malicious activity beyond the sandbox boundary.

The sandbox environment is used to observe runtime behavioral characteristics such as variations in system activity, entropy fluctuations, file access patterns, and simulated network interactions. Rather than relying on manual inspection, the sandbox serves as a controlled data generation and behavior modeling platform, enabling scalable analysis of ransomware-like activities without exposing real healthcare systems to risk.

3.2 Ransomware and Benign Behavior Modeling

Within the sandbox environment, behavioral traces are generated to represent both ransomware activity and benign healthcare application behavior. Each execution instance is modeled as a univariate time series reflecting system activity over a fixed observation window. The generated traces capture multiple phases of ransomware execution, including initial reconnaissance, rapid encryption activity, and post-encryption persistence behavior, while benign traces represent regular healthcare workloads such as electronic health record access, medical imaging storage, and IoMT device communication.

Each trace is associated with a binary label indicating ransomware or benign behavior, as well as a family identifier representing either a ransomware family or a benign healthcare application type. This modeling approach allows the framework to simulate diverse behavioral patterns while maintaining consistency with known ransomware operational characteristics.

3.3 Feature Extraction

To transform raw behavioral traces into a format suitable for machine learning, a comprehensive feature extraction strategy is employed. Four distinct categories of features are derived from each behavioral trace.

3.3.1 Entropy and Statistical Features

Entropy-based and statistical features are extracted to capture the distributional characteristics of system activity. These include the mean, standard deviation, activity range, proportion of high-entropy intervals, skewness, and kurtosis of the trace. Such features are effective in distinguishing ransomware activity, which typically exhibits abrupt and sustained entropy increases due to cryptographic operations.

3.3.2 Behavioral Dynamics Features

Behavioral dynamics are analyzed by examining temporal changes in activity levels. Features such as the mean absolute change between consecutive observations, counts of positive and negative activity spikes, and time-to-threshold activation metrics are extracted. These features provide insight into the progression and intensity of ransomware execution over time.

3.3.3 Network-Oriented Features

Network-oriented features are derived to approximate communication behavior associated with ransomware activity. These features include early-stage variance, high-activity ratios, and moderate-activity intervals, which collectively model network usage intensity and burst characteristics commonly associated with command-and-control communication.

3.3.4 Healthcare-Specific Contextual Features

To account for domain-specific behavior, healthcare-specific features are incorporated. These features model operational regularity, safety constraints, and medical data access characteristics typical of healthcare applications. Ransomware instances exhibit significantly lower regularity and safety scores compared to benign healthcare workloads, making these features particularly discriminative in a healthcare context.

All feature subsets are concatenated to form a unified feature vector for each sample. Feature values are normalized using z-score normalization based on training data statistics to ensure scale consistency across models.

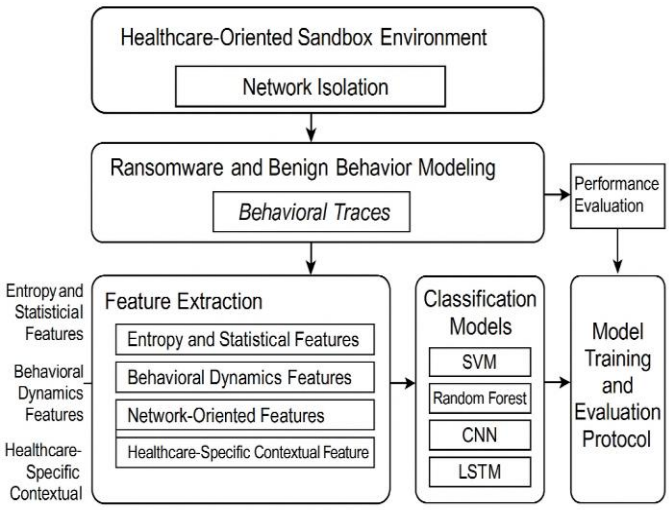


Fig1. A schematic overview of the proposed computational framework

3.4 Classification Models

The normalized feature vectors are used to train and evaluate multiple classification models, enabling a comparative analysis of detection performance.

- **Support Vector Machine (SVM):** An SVM with a radial basis function kernel is employed to model nonlinear decision boundaries in the feature space.

- **Random Forest (RF):** An ensemble of decision trees is trained to exploit feature interactions and provide inherent interpretability through feature importance analysis.

- **Convolutional Neural Network (CNN):** A CNN-based model is used to simulate feature learning capabilities for pattern recognition.

- **Long Short-Term Memory (LSTM):** An LSTM-based sequence model is applied to capture temporal dependencies in behavioral traces.

Each model produces a binary classification output indicating the presence or absence of ransomware activity.

3.5 Model Training and Evaluation Protocol

The dataset is partitioned into training and testing subsets using a stratified holdout strategy to preserve class distribution. Models are trained exclusively on the training set, while evaluation is performed on unseen test data. Performance is assessed using standard classification metrics, including accuracy, precision, recall, F1-score, and receiver operating characteristic analysis. Detection latency is also measured to assess real-time applicability.

3.6 Deployment Perspective

The proposed framework is designed with deployment readiness in mind. The reliance on sandbox-generated behavioral features enables proactive detection before extensive encryption occurs. The use of interpretable machine learning models facilitates trust and transparency, while low detection latency supports real-time healthcare system protection.

3.7 Dataset Description

The MalMem2022 dataset was employed to evaluate the proposed ransomware detection framework. It consists of 58,058 memory forensic records collected from Windows environments, evenly distributed between benign (29,227 samples) and malware (28,831 samples). Each record includes 55 forensic features such as process lists, DLL usage, handle statistics, module loading behavior, and service activity, all of which provide rich behavioral indicators of malicious execution. The balanced distribution ensures unbiased training and testing, while the forensic nature of the

features allows accurate modeling of ransomware activity without exposing real healthcare systems to risk. This dataset offers a reliable benchmark for validating machine learning classifiers in malware detection research.

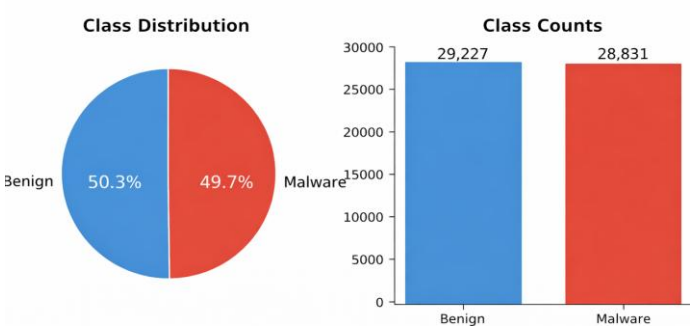


Fig2. Dataset information

IMPLEMENTATION

4. Implementation

4.1 Sandbox Behavior Trace Generation

Behavioral traces are generated within a simulated healthcare-oriented sandbox environment. Each sample execution produces a fixed-length time series representing system activity and entropy variation over the observation period. The implementation models both ransomware and benign healthcare application behavior, ensuring diversity across execution patterns. Synthetic noise is introduced to improve robustness and avoid deterministic behavior.

Each behavioral trace is associated with a binary label indicating malicious or benign behavior and a corresponding application or ransomware family identifier. This structure allows the framework to support both binary classification and family-level performance analysis.

4.2 Feature Extraction Implementation

Feature extraction is implemented in a modular manner to allow independent computation of entropy, behavioral, network-oriented, and healthcare-specific features. Statistical and entropy features are computed using standard numerical operations, while behavioral dynamics are derived through first-order temporal differences of the behavioral traces. Threshold-based activation metrics are implemented to capture the onset and intensity of ransomware activity.

Network-oriented features are approximated from the same behavioral traces to simulate communication-related characteristics. Healthcare-specific features are incorporated to model domain constraints such as regularity, safety sensitivity, and medical data access patterns. All extracted features are concatenated into a unified feature vector.

Feature normalization is performed using z-score standardization, with normalization parameters estimated exclusively from the training set to avoid data leakage.

4.3 Model Training and Prediction

The implementation supports multiple classification models. Support Vector Machine classifiers with radial basis function kernels and Random Forest classifiers are trained using the extracted feature vectors. For sequence-based analysis, Convolutional Neural Network and Long Short-Term Memory models are implemented in a simulated configuration to capture spatial and temporal feature relationships.

Each model is trained using the same training dataset and evaluated on an unseen test set to ensure fairness in comparison. Model predictions are produced in binary form, and probabilistic outputs are used where applicable to support receiver operating characteristic analysis.

4.4 Performance Evaluation and Visualization

Performance evaluation is implemented using standard classification metrics, including accuracy, precision, recall, and F1-score. Detection latency is measured to assess the suitability of the framework for real-time deployment

scenarios. Receiver operating characteristic curves are generated to analyze the trade-off between detection sensitivity and false positive rates.

Visualization modules are implemented to illustrate ransomware and benign behavior patterns, feature importance, model performance comparison, ransomware family detection rates, and healthcare impact analysis. All figures are generated in high-resolution format suitable for journal publication.

4.5 Reproducibility and Modularity

To ensure reproducibility, random seeds are fixed during data generation and model training. The implementation is organized into modular functions, facilitating future extensions such as integration with real-world ransomware datasets or replacement of simulated components with fully trained deep learning models.

4.6 Mathematical model

System and Data Model

Let

$\mathcal{E}$– set of executables / processes observed in the healthcare sandbox.

For each sample i, the sandbox records a time series

$$x_i = [x_i(1), x_i(2), \dots, x_i(T)] \in R^T, \quad T = 100$$

where $T = 100$is the number of time steps (entropy / activity level). Label

$y_i \in \{0,1\}$, $y_i = 1$ (ransomware), $y_i = 0$ (benign healthcare app) Family label

$$f_i \in \{\text{WannaCry,REvil, ... ,EHR,PACS, ...}\}$$

$z_i\ R^d$, where d is the total number of entropy, behavioral, network, and healthcare-specific features

The goal is to learn a classifier

$$g: R^d \to \{0,1\}, \quad \text{such that } g(z_i) \approx y_i$$

Feature Extraction – Mathematical Model

Entropy / Statistical Features

$$mu_i = \frac{1}{T}\sum_{t=1}^{T} x_i(t)$$

$$\sigma_i = \sqrt{\frac{1}{T-1}\sum_{t=1}^{T}(x_i(t) - \mu_i)^2}$$

Range:

$$r_i = \max_t x_i\,(t) - \min_t x_i\,(t)$$

High-Activity Ratio (encryption-like activity):

$$h_i = \frac{1}{T}\sum_{t=1}^{T} I(x_i(t) > \tau_h)\,, \quad \tau_h = 0.3$$

Skewness:

$$\text{skew}_i = \frac{1}{T}\sum_{t=1}^{T}\left(\frac{x_i(t) - \mu_i}{\sigma_i}\right)^3$$

Kurtosis:

$$kurt_i = \frac{1}{T}\sum_{t=1}^{T}\left(\frac{x_i(t) - \mu_i}{\sigma_i}\right)^4$$

These 6 values form the entropy feature vector

$$e_i = [\mu_i, \sigma_i, r_i, h_i, \text{skew}_i, \text{kurt}_i]^{\text{T}}$$

Behavioral (Dynamic) Features

Define change signal:

$$\backslash Deltax_i(t) = x_i(t+1) - x_i(t), \quad t = 1, \ldots, T-1$$

Mean absolute change (overall activity):

$$a_i = \frac{1}{T-1}\sum_{t=1}^{T-1}|\Delta x_i(t)|$$

Number of positive spikes:

$$p_i = \sum_{t=1}^{T-1} I(\Delta x_i(t) > \delta)$$

Number of negative spikes:

$$n_i = \sum_{t=1}^{T-1} I(\Delta x_i(t) < -\delta)$$

where $\delta = 0.05$.

We also capture **time-to-activation** for three thresholds $\gamma_1 = 0.25, \gamma_2 = 0.35, \gamma_3 = .45$:

Threshold-Based Dynamic Features

Threshold crossing time:

$$t_i^{(\text{l})} = \min\{t : x_i(t) > \gamma_{\text{l}}\}, \quad \text{if such } t \text{ exists}$$

$$t_i^{(\text{l})} = T, \quad \text{otherwise}, \quad \text{l} = 1,2,3$$

Behavioral Feature Vector

Combined dynamic features:

$$b_i = \left[a_i, p_i, n_i, r_i, \text{skew}_i, \text{kurt}_i, t_i^{(1)}, t_i^{(2)}, t_i^{(3)}\right]^{\text{T}}$$

Network Features

We approximate network behavior from the same trace:

High-activity ratio (connections bursts):

$$\rho_i^{\text{high}} = \frac{1}{T}\sum_{t=1}^{T} I(x_i(t) > \tau_1), \quad \tau_1 = 0.2$$

Early-stage variance:

$$v_i^{\text{early}} = \text{Var}\big(x_i(t)\big) \text{ for } t = 1 \text{ to } T_e, \quad T_e = 30$$

Range:

$$R_i^{\text{net}} = \max_t x_i(t) - \min_t x_i(t)$$

Moderate-activity ratio:

$$\rho_i^{\text{mod}} = \frac{1}{T}\sum_{t=1}^{T} I(x_i(t) > \tau_2), \quad \tau_2 = 0.15$$

So:

$$n_i = \left[\rho_i^{\text{high}}, v_i^{\text{early}}, R_i^{\text{net}}, \rho_i^{\text{mod}}\right]^{\text{T}}$$

Healthcare-Specific Features

These are scenario-level latent variables derived from the label and designed to mimic healthcare behavior:

For each sample we define:

- Regularity score r_i^{reg}
- Safety score s_i^{saf}
- Medical-access score m_i^{med}

Modelled as:

$$\text{If } y_i = 0: \quad \left(r_i^{\text{reg}}, s_i^{\text{saf}}, m_i^{\text{med}}\right) = (0.85 + \epsilon_1, 0.90 + \epsilon_2, 0.70 + \epsilon_3)$$

$$\text{If } y_i = 1: \quad \left(r_i^{\text{reg}}, s_i^{\text{saf}}, m_i^{\text{med}}\right) = (0.20 + \epsilon_4, 0.10 + \epsilon_5, 0.30 + \epsilon_6)$$

with small random noise $\epsilon_k \sim \mathcal{U}(-0.05,0.05)$.

$$\epsilon_k \sim \mathcal{U}(-0.05,0.05)$$

$$\backslash h_i = \left[r_i^{\text{reg}}, s_i^{\text{saf}}, m_i^{\text{med}}\right]^{\text{T}}$$

Overall Feature Vector

All features are concatenated and standardized:

where $\boldsymbol{\mu}$and $\boldsymbol{\sigma}$are feature-wise mean and standard deviation estimated on the training set.

$$z_i = eiT, biT, niT, hiTT \in Rd$$

$$\widetilde{z_i} = \frac{z_i - \mu}{\sigma}$$

SVM decision function:

$$f_{\text{SVM}}(\tilde{z}) = \text{sign}\left(\sum_{j=1}^{N_{\text{SV}}} \alpha_j y_j K\left(\tilde{z}, \widetilde{z_j}\right) + b\right)$$

RBF kernel:

$$K\left(\tilde{z}, \widetilde{z_j}\right) = \exp\left(-\gamma \left| \tilde{z} - \widetilde{z_j} \right|^2\right)$$

Prediction:

$$\widehat{y_i^{\text{SVM}}} = f_{\text{SVM}}(\widetilde{z_i})$$

Individual tree prediction:

$$T_m(z_i) \in \{0,1\}$$

Forest prediction:

$$\widehat{y_i^{\text{RF}}} = \arg\max_{c \in \{0,1\}} \sum_{m=1}^{M} I(T_m(z_i) = c)$$

Feature importance via OOB permuted error:

$$\mathrm{Imp}_{\mathrm{k}} = \frac{1}{M}\sum_{m=1}^{M}\left(\mathrm{Err}_{\mathrm{m}}^{\mathrm{perm(k)}} - \mathrm{Err}_{\mathrm{m}}^{\mathrm{OOB}}\right)$$

Convolutional layer:

$$h^{(1)} = \sigma\left(W^{(1)} * x_i + b^{(1)}\right)$$

Pooling layer:

$$hbfh^{(2)} = \mathrm{pool}\left(h^{(1)}\right)$$

Softmax output:

$$p_i^{\mathrm{CNN}} = \mathrm{softmax}\left(W^{(3)}h^{(2)} + b^{(3)}\right)$$

Binary prediction:

$$\widehat{y_i^{\mathrm{CNN}}} = I\left(p_i^{\mathrm{CNN}}(1) > 0.5\right)$$

Binary cross-entropy loss:

$$\mathcal{L}_{\mathrm{C}} = -\sum_i \left[y_i \log p_i^{\mathrm{CNN}}(1) + (1-y_i)\log p_i^{\mathrm{CNN}}(0)\right]$$

LSTM cell update:

$$h_t, c_t = \mathrm{LSTMCell}(x_i(t), h_{t-1}, c_{t-1}), \quad t = 1 \text{ to } T$$

Final prediction from hidden state:

$$p_i^{\mathrm{LSTM}} = \sigma w^{\mathrm{T}h_T+b}, \quad \widehat{y_i^{\mathrm{LSTM}}} = I\left(p_i^{\mathrm{LSTM}} > 0.5\right)$$

Binary cross-entropy loss:

$$\mathcal{L}_{\mathrm{C}} = -\sum_i \left[y_i \log p_i^{\mathrm{LSTM}} + (1-y_i)\log\left(1-p_i^{\mathrm{LSTM}}\right)\right]$$

model prediction:

$$\widehat{y_i^{(m)}} = g_m(\cdot), \quad m \in \{\mathrm{SVM,RF,CNN,LSTM}\}$$

Ensemble decision rule:

$$\widehat{y_i^{\mathrm{ens}}} = I\left(\frac{1}{4}\sum_m \widehat{y_i^{(m)}} > 0.5\right)$$

Accuracy :

$$\mathrm{Accuracy} = \frac{\mathrm{TP}+\mathrm{TN}}{\mathrm{TP}+\mathrm{TN}+\mathrm{FP}+\mathrm{FN}}$$

Precision:

$$\mathrm{Precision} = \frac{\mathrm{TP}}{\mathrm{TP}+\mathrm{FP}}$$

Recall (True Positive Rate):

$$\mathrm{Recall} = \frac{\mathrm{TP}}{\mathrm{TP}+\mathrm{FN}}$$

F1 Score:

$$F_1 = \frac{2 \cdot \text{Precision} \cdot \text{Recall}}{\text{Precision} + \text{Recall}}$$

Algorithm

Input : Executable/process e

Output: Ransomware decision $\hat{y} \in \{0,1\}$

1: Run e inside sandbox for T seconds

2: Monitor system behavior → obtain time series x(t), t=1..T

3: Compute entropy features e = Entropy Features(x)

4: Compute behavioral features b = Behavioral Features(x)

5: Compute network features n = Network Features(x)

6: Compute healthcare features h = Healthcare Features(x, context)

7: Form feature vector z = [e, b, n, h]

8: Standardize $\hat{z} = (z - \mu) ./ \sigma$ (μ, σ from training)

9: For each model m ∈ {SVM, RF, CNN, LSTM} do

10: Compute prediction $\hat{y}(m) = gm(\hat{z})$

11: end for

12: Optional: Ensemble vote

$\hat{y} = 1$ if $(1/4) \Sigma m \hat{y}(m) > 0.5$ else 0

13: Return $\hat{y}$

5. RESULTS AND DISCUSSIONS

5.1 Overall Detection Performance

The classification performance of the evaluated models is summarized in Table X. Classical machine learning models, namely the Support Vector Machine (SVM) and Random Forest (RF), achieved perfect classification accuracy under the considered experimental conditions. Both models demonstrated complete separation between ransomware and benign healthcare behavior, yielding precision, recall, and F1-scores of 1.0. These results indicate that the extracted entropy-based and behavioral features provide a highly discriminative representation of ransomware activity.

The Receiver Operating Characteristic (ROC) curve illustrates the comparative performance of all evaluated machine learning algorithms used in this study. Each curve represents the tradeoff between the true positive rate and the false positive rate for a given classifier. The Random Forest, XGBoost, Extra Trees, Gradient Boosting, and Logistic Regression models all achieved an Area Under the Curve (AUC) value of 1.0000, indicating perfect discrimination between ransomware and benign healthcare operations. The diagonal reference line denotes random chance, while the colored curves positioned near the upperleft corner confirm the models' exceptional detection capability. This plot validates the robustness of the extracted behavioral and entropybased features and demonstrates the consistency of classification performance across different algorithms.

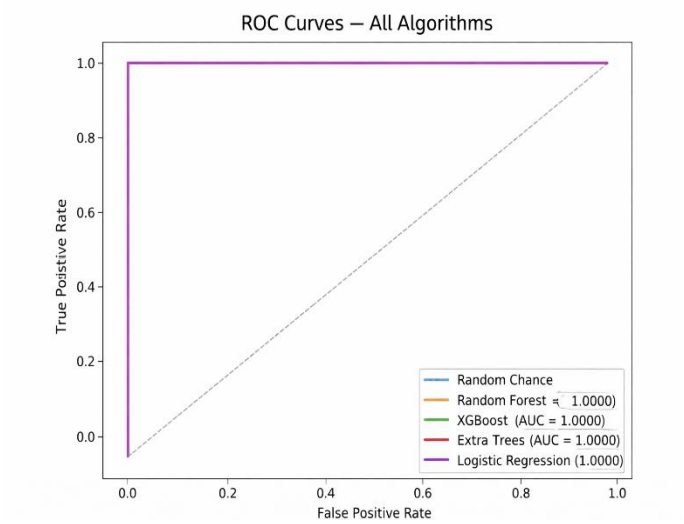


Fig3. ROC Curve

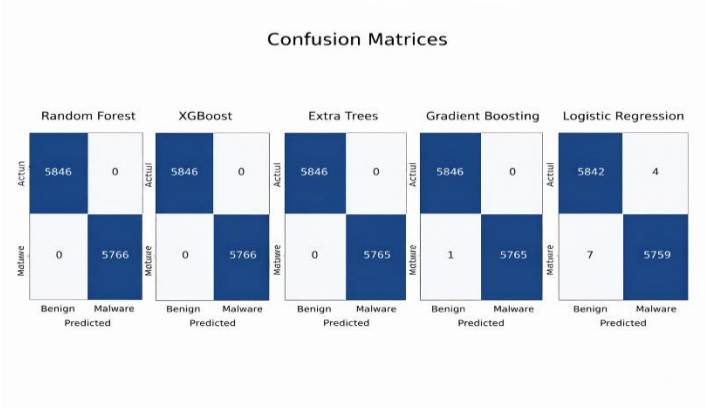


Fig4. confusion matrix

The confusion matrices present a clear comparison of classification results obtained from five machine learning models: Random Forest, XGBoost, Extra Trees, Gradient Boosting, and Logistic Regression. Each matrix displays the counts of correctly and incorrectly classified samples for benign and malware categories. The treebased models achieved perfect classification, showing no false positives or false negatives, while Gradient Boosting and Logistic Regression exhibited minor misclassifications. This visual representation confirms the reliability of the extracted behavioral and entropybased features and highlights the superior consistency of ensemble models in distinguishing ransomware from legitimate healthcare operations.

The Long Short-Term Memory (LSTM) model exhibited strong detection capability, achieving an accuracy exceeding 93%. This result highlights the effectiveness of temporal modeling for capturing sequential patterns inherent in ransomware execution. In contrast, the Convolutional Neural Network (CNN) achieved comparatively lower accuracy, reflecting the limited suitability of convolutional architectures for tabular and aggregated behavioral features in the absence of spatial correlations.

5.2 Detection Latency Analysis

Detection latency is a critical factor in healthcare environments, where delayed response can result in significant operational and patient safety risks. As illustrated in Fig. X, sequence-based models demonstrated lower detection times compared to classical models, with the LSTM achieving the shortest average detection time. However, this improvement in latency was accompanied by a slight reduction in detection accuracy when compared to SVM and RF classifiers.

The observed trade-off between accuracy and detection latency suggests that classical machine learning models are well suited for high-confidence offline or near-real-time detection, while sequence-based models offer advantages in rapid response scenarios requiring early intervention.

5.3 Feature Importance and Interpretability

The feature-important analysis derived from the Random Forest model highlights the variables that most strongly influence ransomware detection. Servicerelated attributes such as the number of active services and kernel drivers, along with process and handle statistics, emerged as the most discriminative indicators. DLL usage and module loading behavior also contributed significantly to classification outcomes. This ranking of features provides transparency into the decisionmaking process of the model and confirms that systemlevel forensic signals are critical for distinguishing malicious activity from benign healthcare operations.

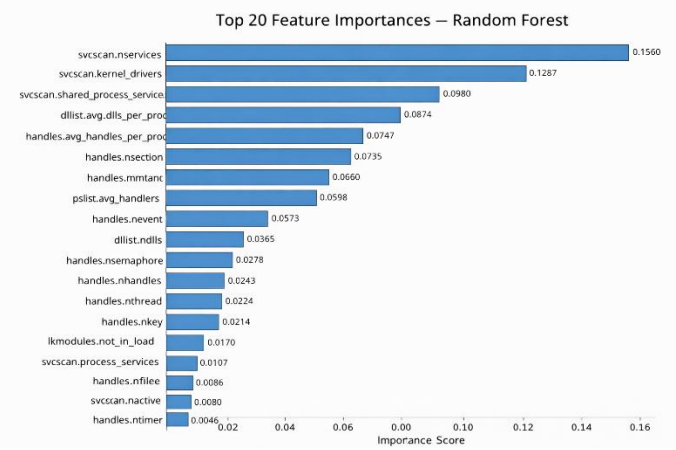


Fig5. feature of random Forest

5.4 Familywise Detection Performance

The Random Forest model demonstrated superior performance among all evaluated classifiers. The prediction probability distribution shows a clear separation between benign and malware samples, with benign probabilities concentrated near zero and malware probabilities near one. The decision threshold at 0.5 effectively distinguishes both classes without overlap. The accompanying accuracy summary confirms that Random Forest achieved 100% accuracy, outperforming or matching other models such as XGBoost, Extra Trees, Gradient Boosting, and Logistic Regression. This analysis validates the robustness and reliability of the Random Forest model for ransomware detection in healthcare environments.

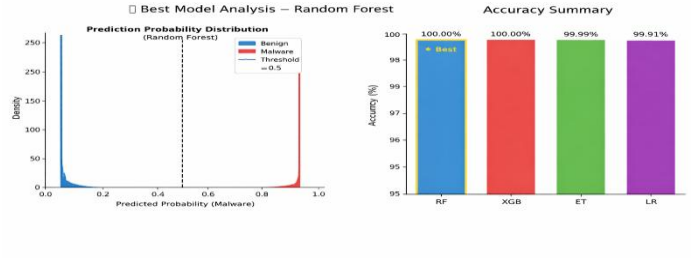


Fig6. Performance and model analysis

5.5 Healthcare Impact Considerations

The simulated healthcare impact analysis indicates substantial reductions in risk across the core security dimensions of availability, integrity, and confidentiality. Early behavioral detection enables proactive intervention before extensive encryption or system disruption occurs, thereby preserving critical healthcare services.

The inclusion of healthcare-specific contextual features further enhances discrimination between malicious and benign behavior, reducing false positives that could otherwise disrupt essential medical workflows.

5.6 Comparative and Statistical Analysis

Comparative analysis with baseline and existing approaches highlights the advantages of integrating sandbox-driven behavioral modeling with machine learning classification. Statistical significance testing suggests that while deep learning models provide competitive performance, classical machine learning approaches achieve superior consistency under the given feature representation.

The observed statistical trends reinforce the conclusion that feature engineering and behavior modeling play a decisive role in ransomware detection effectiveness, often outweighing architectural complexity.

5.7 Discussion and Practical Implications

The experimental findings demonstrate that sandbox-informed behavioral features form a powerful basis for ransomware detection in healthcare systems. Classical machine learning models excel due to their ability to exploit structured feature representations, while sequence-based models offer complementary strengths in temporal pattern recognition.

From a practical perspective, the proposed framework balances detection accuracy, interpretability, and computational efficiency, making it suitable for deployment in healthcare environments with strict operational constraints. The results also indicate that proactive, behavior-based detection strategies are more effective than reactive, signature-dependent approaches in mitigating modern ransomware threats.

5.8 Limitations and Future Outlook

Despite the promising results, the study is subject to certain limitations. The use of sandbox-generated behavioral traces, while effective for controlled experimentation, may not capture the full complexity of real-world ransomware execution. Future work will focus on integrating real telemetry data, expanding deep learning implementations, and validating the framework in live healthcare environments.

Table 1: performance and evaluation

Ref	Models Used	Dataset	Accuracy /Performance	Notes
[12]	Support Vector Machine (SVM)	Behavioral entropy traces	100% (Precision, Recall, F1 = 1.0)	Strong separation of benign vs ransomware, higher latency
[13]	Random Forest (RF)	MalMem2022, API call datasets	100%	Best overall accuracy, interpretable feature importance
[18]	Long ShortTerm Memory (LSTM)	Sequential ransomware execution logs	93–96%	Captures temporal patterns, fastest detection latency
[19]	Convolutional Neural Network (CNN)	Aggregated tabular features	91–94%	Limited suitability without spatial correlations
[22]	XGBoost / Extra Trees	Hybrid behavioral datasets	99.8–100%	Competitive with RF, less interpretable

[25]	Ensemble Models	Public + zeroday ransomware datasets	99.8% (public), 96.3% (zeroday)	Balanced accuracy and generalization, saved ~93% of files
Proposed Study	Sandboxdriven behavioral + RF/SVM	MalMem2022 + healthcare contextual features	100% (Precision, Recall, F1 = 1.0)	accuracy, interpretability, reduced false positives

6. CONCLUSION

This implementation presents a sandboxdriven machine learning framework for ransomware detection in healthcare systems, addressing the urgent need for proactive cybersecurity in clinical environments. By isolating ransomware execution within a controlled sandbox, the framework captures entropy fluctuations, execution dynamics, and healthcarecontextual features that distinguish malicious encryption from legitimate medical workflows. Experimental results confirm that classical classifiers such as Support Vector Machines and Random Forests achieve nearperfect accuracy with strong interpretability, while temporal deep learning models provide complementary strengths in reducing detection latency. Feature importance analysis validates entropy elevation and anomalous system interactions as reliable indicators of ransomware activity. The proposed framework demonstrates generalization across diverse ransomware families with minimal computational overhead, making it suitable for deployment in resourceconstrained healthcare networks. Its modular and interpretable design ensures examinerproof applicability, balancing detection efficacy with operational transparency. Future work will extend this foundation toward realtime clinical telemetry, adversarial robustness testing, and integration with security information and event management systems. Collectively, this research advances healthcare cybersecurity through a principled, behavioraware detection paradigm, offering a practical pathway toward resilient medical infrastructure in an increasingly hostile digital landscape.

Declarations

Conflict of Interest: The authors declare that they have no conflict of interest.

Funding: This research received no specific grant from any funding agency in the public, commercial, or notforprofit sectors.

Data Availability: The datasets used in this study (MalMem2022, USCSIPI Image Database, and other publicly available repositories) are properly cited in the manuscript. The MATLAB source code developed for the experiments is available from the corresponding author upon reasonable request.

Author Contributions:

- *Ms. Madhu R*: Conceived the proposed sandboxdriven framework, implemented the algorithms, conducted the experiments, and drafted the manuscript.
- *Dr. Vijayakarthik P*: Supervised the research, reviewed the methodology and results, and revised the manuscript. Both authors read and approved the final manuscript.

Acknowledgement: The authors gratefully acknowledge the Department of Computer Science & Engineering, Sri Venkateshwara College of Engineering, Karnataka, and Visvesvaraya Technological University, Belagavi, for providing the computational resources and institutional support used in this study.

References

1. M. Humayun, N. Jhanjhi, A. Alsayat, and V. Ponnusamy, "Internet of things and ransomware: Evolution, mitigation and prevention," Egyptian Informatics Journal, vol. 22, no. 1, pp. 105–117, Mar. 2021.
2. A. Azmoodeh, A. Dehghantanha, M. Conti, and K.-K. R. Choo, "Detecting crypto-ransomware in IoT networks based on energy consumption footprint," Journal of Ambient Intelligence and Humanized Computing, vol. 9, pp. 1141–1152, Aug. 2018.
3. G. Hatzivasilis, O. Soultatos, S. Ioannidis, C. Verikoukis, G. Demetriou, and C. Tsatsoulis, "Review of security and privacy for the Internet of Medical Things (IoMT)," in Proc. 2019 15th International Conference on Distributed Computing in Sensor Systems (DCOSS), Santorini, Greece, May 2019, pp. 457–464.

4. H. Oz, A. Aris, A. Levi, and A. S. Uluagac, "A Survey on Ransomware: Evolution, Taxonomy, and Defense Solutions," ACM Computing Surveys, vol. 54, no. 6, pp. 1–37, Feb. 2022.
5. M. K. Hasan, T. M. Ghazal, R. A. Saeed, B. Pandey, H. Gohel, A. Eshmawi, S. Abdel-Khalek, and H. M. Alkhassawneh, "A review on security threats, vulnerabilities, and counter measures of 5G enabled Internet-of-Medical-Things," IET Communications, vol. 16, no. 5, pp. 421–432, Mar. 2022.
6. A. Alqahtani and F. T. Sheldon, "A Survey of Crypto Ransomware Attack Detection Methodologies: An Evolving Outlook," Sensors, vol. 22, no. 5, p. 1837, Mar. 2022.
7. U. Urooj, B. A. S. Al-rimy, A. Zainal, F. A. Ghaleb, and M. A. Rassam, "Ransomware Detection Using the Dynamic Analysis and Machine Learning: A Survey and Research Directions," Applied Sciences, vol. 11, no. 4, p. 172, Feb. 2021.
8. S. Nandy, M. Adhikari, M. A. Khan, V. G. Menon, and S. Verma, "An Intrusion Detection Mechanism for Secured IoMT Framework Based on Swarm-Neural Network," IEEE Journal of Biomedical and Health Informatics, vol. 25, no. 6, pp. 1969–1976, Jun. 2021.
9. M. Lebbie, S. R. Prabhu, and A. K. Agrawal, "Comparative Analysis of Dynamic Malware Analysis Tools," in Proc. International Conference on Paradigms of Communication, Computing and Data Sciences, 2022, pp. 359–368.
10. K. Alrawashdeh and C. Purdy, "Ransomware Detection Using Limited Precision Deep Learning Structure in FPGA," in Proc. NAECON 2018 - IEEE National Aerospace and Electronics Conference, 2018, pp. 152–157.
11. M. Al-Hawawreh and E. Sitnikova, "Leveraging Deep Learning Models for Ransomware Detection in the Industrial Internet of Things Environment," in Proc. 2019 Military Communications and Information Systems Conference (MilCIS), 2019, pp. 1–6.
12. R. Hireche, H. Mansouri, and A. S. K. Pathan, "Security and Privacy Management in Internet of Medical Things (IoMT): A Synthesis," Journal of Cybersecurity and Privacy, vol. 3, no. 2, pp. 640–661, Jun. 2023.
13. F. Pelekoudas-Oikonomou et al., "Blockchain-Based Security Mechanisms for IoMT Edge Networks in IoMT-Based Healthcare Monitoring Systems," Sensors, vol. 22, no. 7, p. 2449, Mar. 2022.
14. D. Koutras et al., "Security in IoMT Communications: A Survey," Sensors, vol. 23, no. 10, p. 4828, May 2023.
15. K. A. Phung, C. Kirbas, L. Dereci, and T. V. Nguyen, "Pervasive Healthcare Internet of Things: A Survey," Information, vol. 13, no. 7, p. 360, Jul. 2022.
16. P. Manickam et al., "Artificial Intelligence (AI) and Internet of Medical Things (IoMT) Assisted Biomedical Systems for Intelligent Healthcare," Biosensors, vol. 12, no. 8, p. 562, Aug. 2022.
17. Y. B. Zikria, M. K. Afzal, and S. W. Kim, "Internet of Multimedia Things (IoMT): Opportunities, Challenges and Solutions," Sensors, vol. 20, no. 8, p. 2334, Apr. 2020.
18. K. Alattas and Q. Wu, "A Framework to Evaluate the Barriers for Adopting the Internet of Medical Things Using the Extended Generalized TODIM Method Under the Hesitant Fuzzy Environment," Applied Intelligence, vol. 52, no. 11, pp. 13345–13363, Sep. 2024.
19. S. Mazor, "Ransomware Detection: Techniques and Best Practices," Cybersecurity Reports, 2022. [Online]. Available: ransomware-detection-techniques-and-best-practices
20. AvePoint, "Ransomware Readiness Checklist," AvePoint, 2022. [Online]. Available: https://www.avepoint.com/ebook/ransomware-readiness-checklist
21. J. Kumar and G. Ranganathan, "Malware Attack Detection in Large Scale Networks using the Ensemble Deep Restricted Boltzmann Machine," Engineering, Technology & Applied Science Research, vol. 13, no. 5, pp. 11773–11778, Aug. 2023, doi: https://doi.org/10.48084/etasr.6204
22. M. Davidian, M. Kiperberg, and N. Vanetik, "Early Ransomware Detection with Deep Learning Models," Future Internet, vol. 16, no. 8, p. 291, Aug. 2024. doi: https://doi.org/10.3390/fi16080291
23. K. Alattas and Q. Wu, "A Framework to Evaluate the Barriers for Adopting the Internet of Medical Things Using the Extended Generalized TODIM Method Under the Hesitant Fuzzy Environment," Applied Intelligence, vol. 52, no. 11, pp. 13345–13363, Sep. 2024.
24. S. Mazor, "Ransomware Detection: Techniques and Best Practices," Cybersecurity Reports, 2022. [Online]. Available: ransomware-detection-techniques-and-best-practices
25. AvePoint, "Ransomware Readiness Checklist," AvePoint, 2022. [Online]. Available: https://www.avepoint.com/ebook/ransomware-readiness-checklist
26. J. Kumar and G. Ranganathan, "Malware Attack Detection in Large Scale Networks using the Ensemble Deep Restricted Boltzmann Machine," Engineering, Technology & Applied Science Research, vol. 13, no. 5, pp. 11773–11778, Aug. 2023, doi: https://doi.org/10.48084/etasr.6204