IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 3, Issue 3, July 2023
Impact Factor: 7.301

Autonomous Ransomware Forensics: Advanced ML Techniques for Attack Attribution and Recovery

Gaurav Sarraf
Independent Researcher
sarrafgsarraf@gmail.com

Abstract: *Ransomware attacks are one of the most disastrous cybersecurity risks, as they encrypt important information and require financial compensation to decrypt keys, leading to billions of monetary losses every year. The advanced development of ransomware families requires highly sophisticated machine learning solutions to be detected and automatically analyzed. In this study, autonomous ransomware forensics framework was introduced using sophisticated machine learning models to perform attack attribution and recovery procedures on a holistic ransomware identification dataset. The methodology involves systematic preprocessing of data such as data cleaning, categorical variables label encoding, and rectification of class imbalance by use of under-sampling mechanisms, Principal Component Analysis (PCA) to facilitate optimal selection of features, and data normalization to promote quality of structured input. A higher accuracy (ACC), precision (PRE), recall (REC), and F1-score (F1) of 98.21%, 97.33%, and 97.45%, respectively, for the constructed Long Short-Term Memory (LSTM) neural network model indicates improved computational capability to identify ransomware behavioral patterns and time sequences. Evaluation against other popular classification models, such as Random Forest (with a 96.90% accuracy rate), Support Vector Machine (91.67% accuracy), and Convolutional Neural Network (94.38% accuracy), demonstrates the efficacy of the LSTM architecture. The autonomous framework enables real-time threat attribution and automated recovery protocol initiation, significantly reducing incident response time and operational disruption in enterprise cybersecurity environments while eliminating dependency on manual forensic expertise.*

Keywords: Cybersecurity, Ransomware Forensic, Ransomware Detection Dataset, Machine Learning, LGBM, Ransomware Attack, Artificial Intelligence, Attribution and Recovery

I. INTRODUCTION

Cybersecurity has been a significant concern since the inception of computer and networking technologies. Countless hours of study have gone into developing countermeasures to protect individuals and organisations from cybercriminals' increasingly sophisticated and novel attacks. [1]. It was in 1996 that the fields of crypto-virology and crypto-virus were first introduced. Later on, this idea was dubbed crypto-ransomware [2][3][4]. Many organisations, including governments, businesses, healthcare systems, and vital infrastructure, have fallen victim to ransomware, which has quickly become a major problem in the cybersecurity industry [5]. These attacks encrypt valuable data and demand ransom payments, causing significant financial and operational losses ransomware attacks have increased exponentially, with damages projected to reach over $256 billion by 2031[6]. The attack environment is very diverse, including opportunistic mass campaigns and highly targeted attacks, distribution modes include phishing emails, exploit kits, malicious downloads, and network system vulnerabilities [7][8]. Such complexity intensifies the fact that more sophisticated solutions are urgently required to identify, assign, and remedy ransomware attacks.

Conventional forensic methods of ransomware analysis are based on extensive manual analysis of logs, malware sample reverse engineering, and pattern identification [9][10]. Although they work on a smaller scale, these methods are time-consuming, demand manpower, and cannot be used to match the speed and quantity of current attacks

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429 Volume 3, Issue 3, July 2023 Impact Factor: 7.301

[11][12]. Automated ransomware forensics This has been mitigated by autonomous ransomware forensics, which combines automation and smart analysis tools in order to offer the ability to investigate attacks rapidly and more accurately without the need to employ many people manually [13]. This critical to respond and recover in time because ransomware campaigns have the potential to inflict irreparable damage in hours [14]. Autonomous forensics frameworks can close the divide between detection and use of actionable intelligence by enabling cybersecurity systems to establish a pathway where more effective defines mechanisms can be implemented.

Machine learning can be used to provide high-powered ransomware attribution, where machines can detect and categorize attacks according to the behavioural patterns, encryption methods, and signatures of execution [15]. Such techniques as supervised classification, clustering, DL, and anomaly detection have proven to be potentially effective in identifying the origin and purpose of ransomware attacks, extracting meaningful features, and connecting patterns with a recognized attack campaign to make accurate attribution of threat actors [16][17]. Besides learning about the motives of the attacks, attribution also serves the purpose of both informing targeted mitigation efforts and facilitating legal investigations machine learning is paramount in automating the recovery operations of ransomware. Using predictive analytics, the incorporation of ML into forensic systems changes ransomware response into a proactive, rather than a reactive, system to enhance resilience to future threats.

A. Motivation and Contribution

Ransomware attacks are increasing exponentially, and the traditional forensic tools and their application are not sufficient in the current business setting to facilitate quick threat identification and resolution. Traditional manual forensic analysis is time and skill intensive and usually leads to sluggish incident response which increases financial damages and operational interruption. The advanced development of ransomware lines requires independent forensic abilities that allow quickly detecting the signatures of an attack and assigning threats to particular individuals and automatically deploying recovery measures without assistance.

The main findings from this study on creating a framework for autonomous ransomware forensics are as follows:

- The creation of a smart pre-processing framework that includes an organized data cleaning procedure, label encoding, and the correction of class imbalance with the help of the under-sampling algorithm to maintain the quality of the dataset to optimal levels in order to conduct forensic analysis.
- Principal Component Analysis as a dimensionality reduction method and the identification of key forensic indicators, which maximize computational efficiency, but do not affect the discriminative properties of ransomware is implemented.
- The creation and implementation of a neural network based on a LSTM, tailored to temporal pattern recognition in ransomware behavioural patterns, which is able to distinguish between the two properly and classify threats and their families.
- Development of a clever decision-making paradigm, which allows identifying ransomware families in real time and attributing the threat actors without the need to request human intervention which induce a substantial decrease in incident response time
- Comprehensive assessment in terms of various metrics such as ACC, PRE, REC, F1 and ROC-AUC analysis that indicate better forensic ACC than traditional manual investigation tools and available automated systems.

B. Novelty with justification

The study presents an innovative autonomous ransomware forensics framework, which combines LSTM-based temporal patterns analysis with real-time threat attribution and automatic recovery system in a unique manner. Unlike conventional detection systems requiring manual forensic analysis, this work presents the first comprehensive end-to-end autonomous platform capable of simultaneous ransomware identification, attack attribution, and immediate recovery initiation without human intervention. The innovative combination of PCA-driven feature engineering, class-balanced pre-processing, and specialized LSTM architecture for behavioural sequence analysis enables superior family classification ACC while eliminating dependency on cybersecurity expertise. This paradigm shifts from reactive to

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429 Volume 3, Issue 3, July 2023 Impact Factor: 7.301

proactive autonomous forensics significantly reduces incident response time and operational disruption in enterprise environments.

C. Structure of Paper

The paper's structure is as follows The analysis on ransomware risks and existing forensic approaches is presented in Section II. Section III lays out the suggested structure and methodology for autonomous forensics. Section IV demonstrates experimental results and performance analysis of the LSTM model. Section V conclusion and future work of study are provided

II. LITERATURE REVIEW

A literature review on AI and ML strategies for efficient and accurate ransomware detection, attribution, and recovery is presented in this section. Table I summarizes the key studies discussed in the following subsections.:

Sendner et al. (2022) The growing problem of ransomware is estimated to cost 256 billion dollars by 2031, up from 5 billion dollars in 2017. In 2021, the loss increased to 20 billion dollars. Ransomware has recently shifted from PC (client) platforms to server-side databases, with attacks such as the MongoDB Apocalypse in January 2017 and 85,000 MySQL instances being held ransom in 2020, respectively, as a result of countermeasures against client-side ransomware. data storage on the server Until now, ransomware has gotten very little attention. To fill this void, introduce DIMAQS, a new anti-ransomware solution for databases. In order to detect attacks, DIMAQS uses two categorisation methods, CPNs and DNNs, to monitor incoming requests and match patterns in real time [18].

Molina et al. (2022) Cyber threats such as ransomware are the most destabilizing. 3,000+ samples from popular ransomware families to determine whether ones exhibit particular paranoia-based characteristics. OoW and other NLP techniques are used to mimic API calls made by ransomware to evade detection. subsequently tailor several ML and DL algorithms for malware classification. A comprehensive review found that the approach works, with OoW and RF receiving the best classification accuracy rates (94.92%). [19].

Rathod, Parekh and Dholariya (2021) threat vectors, including zero-day vulnerabilities and ransomware. The incorporation of newly developed technology ML and AI integrated with human intervention Quick, accurate, simple, and scalable threat detection and response systems are possible using anomaly detection's automated machine-based detection. Utilising the EMBER dataset, which is utilised to train ML models to detect harmful portable executable files, EDR technologies offer a consolidated perspective on intricate intrusions by identifying intrusion events based on known adversarial behaviours. This dataset consists of factors retrieved from 1.1 million binary files The training set included 900,000 samples, 300,000 of which were malicious and 300,000 of which were benign. An additional 300,000 were not marked. There were 100,000 hazardous samples and 100,000 benign samples out of a total of 200,000 [20].

Almousa, Basavaraju and Anwar (2021) Companies and healthcare providers of all sizes have been victims of ransomware attacks. Crypto virology, the practice of using cryptography in malware design, is an idea that ransomware uses. Ransomware detection has grown in importance and now requires sophisticated technologies to analyse victim network data, find vulnerabilities, and improve security. A procedure for identifying ransomware that makes use of AI and the Application Programming Interface (API). exploring the Windows platform malware lifecycle, identifying potentially harmful code patterns, and creating and testing machine learning models to identify ransomware using a range of samples, some of which pose a threat and others that do not. All of the information was retrieved from publicly accessible databases. To find the most accurate model, used the grid search hyperparameter optimisation approach. This study shown that by integrating API calls with an ML model, ransomware detection skills may achieve an impressive 99.18% ACC rate [21].

Khammas (2020) Ransomware can now be quickly and easily detected from raw bytes using a novel static analysis method that makes use of regular pattern mining. Since ransomware has recently arisen as a major danger to the computer world, this is essential in order to avoid financial and moral extortion. The optimal number of features to employ in ransomware detection with a random forest classifier was determined to be 1000 using the Gain Ratio technique. The most efficient use of time and quantity of capital was achieved by growing 100 trees from a single seed.

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429 Volume 3, Issue 3, July 2023 Impact Factor: 7.301

The experimental results demonstrated a high degree of ACC (97.74 %) in ransomware detection using the proposed method [22].

Lee, Lee and Yim, (2019) ransomware has recently surfaced, encrypting users' data and demanding payment in return. Use of file- and behavior-based detection techniques allows for the detection and prevention of ransomware that transmits unknown dangerous software. This method can't detect ransomware in backup environments like the cloud, for example, which is one of its restrictions. Make use of an entropy method to determine how uniform the encrypted file is; retrieving corrupted data from backups becomes next to impossible when the systems are in sync with each other. Even if the user's PC gets infected with ransomware, machine learning-based file entropy analysis can revive the original file from backup thanks to its synchronization detection capabilities. [23].

Poudyal, Subedi and Dasgupta (2018) Ransomware attacks have been increasing in frequency and severity, causing widespread damage to businesses. Forensic analysts frequently use methods such as binary file reverse engineering to investigate these characteristics of malware. a feature-generation system for ransomware detection that employs ML and engines for reverse engineering. Using the portable executable (PE) parser and the Linux object-code dump tool, this framework can analyse malware code in its entirety, including raw binaries, assembly codes, libraries, and function calls. A clearer picture of the code's intended use will emerge in this form. We think about both good and bad binaries. Following this, ML approaches are used to classify the samples according to this data. Results in ransomware sample identification ranged from 76% to 97% across the experiments, with the exact percentage depending on the ML approach employed [24].

Recent research in ransomware forensics demonstrates a transition from conventional detection techniques toward autonomous, machine learning–based approaches designed to support attack attribution and recovery. Several studies have investigated database-specific monitoring frameworks, natural language processing techniques for analyzing malicious query and API behaviours, and large-scale anomaly detection methods to address emerging zero-day threats. Other contributions include API-driven models that identify distinctive execution patterns, static and entropy-based methods for tracing encrypted file characteristics in backup and cloud environments, and reverse engineering frameworks for analysing ransomware binaries at multiple abstraction levels. Collectively, these studies highlight the ongoing evolution toward intelligent and automated ransomware forensics, emphasizing scalability, adaptability, and integration of forensic insights to strengthen attribution and accelerate recovery across diverse computing environments

Table 1: Comparative analysis of Ransomware detection using machine learning model

Author(s)	Dataset	Methodology	Results Analysis	Limitations	Future Work
Sendner et al. (2022)	Runtime database queries (MongoDB, MySQL instances)	DIMAQS: Runtime monitoring of malicious query sequences using CPN and DNNs	Effective detection of server-side ransomware; novel contribution as client-side is well studied	Limited to query-based ransomware patterns; scalability in large enterprise databases not fully explored	Extend to hybrid database architectures and cloud-hosted DB services
Molina et al. (2022)	3K ransomware samples, API call traces	NLP-based Occurrence of Words (OoW) + ML/DL classifiers (Random Forest best)	Achieved 94.92% accuracy using RF with OoW	Limited to API evasion behavior; dataset scope relatively small	Expand dataset size; explore hybrid static + dynamic features
Rathod, Parekh & Dholariya (2021)	EMBER dataset (1.1M binaries: 300K malicious, 300K benign, 300K unlabeled, 200K eval)	AI/ML + manual anomaly detection + Endpoint Detection & Response (EDR) framework	Showed feasibility of large-scale ransomware/zero-day detection on EMBER dataset	Relies heavily on EMBER, may not generalize to unseen ransomware families	Apply on real-world enterprise datasets; refine zero-day detection

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429 Volume 3, Issue 3, July 2023 Impact Factor: 7.301

Almousa, Basavaraju & Anwar (2021)	Public datasets of ransomware and benign samples (Windows API calls)	API-based ransomware detection with ML models + grid search for hyperparameter optimization	Achieved 99.18% accuracy in ransomware detection	Focused mainly on Windows platform; dataset variety limited	Extend to cross-platform ransomware (Linux, macOS, mobile); real-time detection in production
Khammas (2020)	Raw byte samples of ransomware binaries	Static analysis + frequent pattern mining + feature selection (Gain Ratio, 1000 features) + Random Forest	Achieved 97.74% accuracy with optimized Random Forest	Limited adaptability to polymorphic/obfuscated ransomware	Combine static + dynamic features; improve resistance to code obfuscation
Lee, Lee & Yim (2019)	File entropy from ransomware-infected files (including backup/cloud)	File-based & behavior-based detection using entropy + ML classification	Detected ransomware in backup systems; entropy effective for encrypted file identification	Cannot fully prevent spread in real-time; cloud sync challenges	Develop proactive detection in cloud backup before sync; integrate with cloud providers
Poudyal, Subedi & Dasgupta (2018)	Ransomware binaries + benign executables (reverse engineered: PE parser, Linux tools)	Reverse engineering + multi-level ML analysis (raw binaries, assembly codes, DLLs)	Detection accuracy varied 76%–97% depending on ML technique	Performance depends on ML model; reverse engineering is resource-intensive	Automate feature extraction; enhance detection speed; apply deep learning

III. METHODOLOGY

This methodology presents an intelligent automated forensics framework utilizing sophisticated machine learning algorithms for ransomware attack pattern analysis, threat actor identification, and automated system recovery mechanisms. The comprehensive approach, as shown in Figure 1, starts with a carefully selected ransomware detection dataset that includes various malware families and attack vectors. The dataset is then subjected to systematic data preprocessing, which includes label encoding and data cleaning operations. To address dataset irregularities, class imbalance rectification through under-sampling techniques is implemented, followed by PCA for optimal feature selection and subsequent data normalization. The pre-processed data is then divided into a training subset and a testing subset. The training subset is then fed into an LSTM NN model that was designed for ransomware pattern identification and attribution analysis. The autonomous forensics system is able to evaluate performance by means of comprehensive metrics such as ACC, PRE, REC, F1, and ROC-AUC assessments. This allows for the attribution of attacks in real-time and the facilitation of automated recovery mechanisms for compromised digital asset recovery processes in enterprise environments.

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 3, Issue 3, July 2023
Impact Factor: 7.301

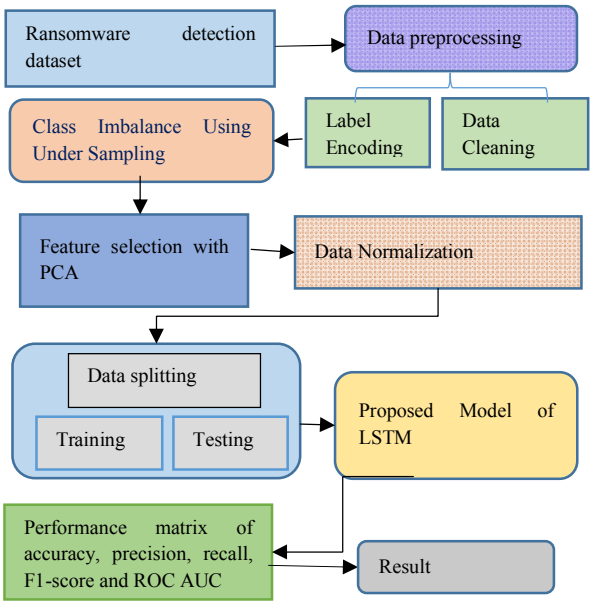

Fig. 1. Flowchart For Ransomware Detection In Cybersecurity Environment Using Machine Learning Models

A. Data Collection

The dataset consists of PE file characteristics extracted from a collection of Windows executables and DLL files. Data from the PE header and structure are used to extract various attributes from each unique file, which is represented by each entry. Malware samples, as identified by Virus Share hashes, are included in the dataset alongside benign software samples.

B. Data Visualization and Analysis

Data visualization enables rapid identification of patterns and anomalies within ransomware datasets through comprehensive graphical analysis techniques. Histogram distributions reveal distinct file size characteristics between malware and benign samples, while correlation heatmaps expose feature interdependencies and highlight critical forensic indicators for autonomous ransomware attribution. These visualization methods facilitate pattern recognition and feature selection optimization for enhanced detection ACC in the proposed forensics framework some of the visualization are given below:

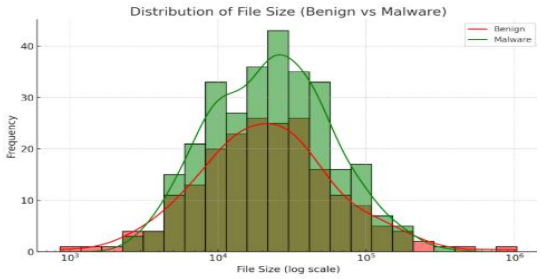

Fig. 2. Histogram for Benign and Malware Sample

It shows in Figure 2 presents the distribution of file sizes for benign versus malware samples in a dataset. The histogram displays frequency distributions on a logarithmic scale, where benign files (red line) show a broader size range while malware files (brown bars) exhibit a more concentrated distribution around 10^4-10^5 bytes, suggesting distinct behavioral patterns.

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 3, Issue 3, July 2023
Impact Factor: 7.301

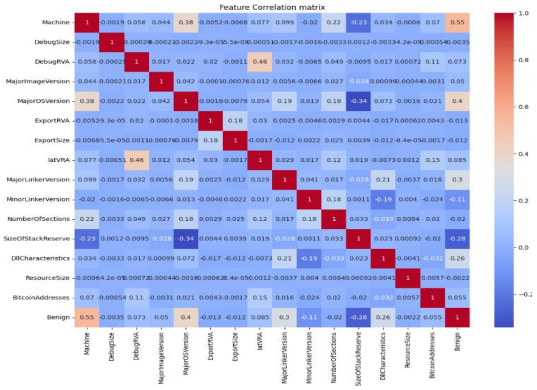

Fig. 3. Correlation Heatmap of Different Feature

It shows in Figure 3 Matrix displays feature relationships in a Ransomware detection dataset, including system behaviour indicators such as Machine Version, Debugger Present, Signific, Mounting Version, Export Size, IsPreA, Major Linker Version, Major Linker Version, Resource Size, Size Of Stack Reserve, Size Of Stack Commit, Dll Characteristics, Resource Size, Bit ness Indicators, and Base. The heatmap reveals interdependencies between these features, with correlation coefficients ranging from strong negative (blue) to strong positive (red) values, facilitating feature selection optimization.

C. Data preprocessing

Data pre-processing encompasses systematic data cleaning to eliminate inconsistencies, label encoding for categorical variable conversion, and class imbalance rectification through under-sampling techniques to ensure equitable sample distribution. Feature selection optimization is achieved using PCA for dimensionality reduction, followed by data normalization to standardize feature scales for optimal LSTM model performance. This comprehensive pre-processing pipeline ensures clean, balanced, and structured input data for reliable autonomous ransomware forensics and attack attribution analysis:

- **Data cleaning:** Data cleaning is the procedure of inspecting a dataset for corrupted, incomplete, irrelevant, or inaccurate information and fixing or eliminating it. In ransomware detection datasets, this typically involves removing duplicate entries, eliminating corrupted or incomplete samples, and discarding irrelevant columns such as file hashes or paths a clean, consistent dataset that improves the ACC and efficiency of ML models.
- **Label encoding:** ML algorithms can analyse categorical text labels by transforming them into numeric form, a process known as label encoding. The unique categories are given unique integer values. As an illustration, a ransomware dataset may have such labels as benign and ransomware, which can be coded 0 and 1 respectively.

D. Class Imbalance using Under Sampling

The number of dangerous samples greatly exceeds the number of benign ones in ransomware detection databases, indicating a class imbalance. The detection rates of ransomware cases may be lowered as a result of this imbalance, which skews machine learning models towards most common class detection. Under-sampling is a solution to this problem that selectively down-sampling the sample of majorities-class samples to equal the size of the minority class. This generates a balanced dataset, which enhances the model to precisely detect the ransomware samples without being biased to benign predictions.

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 3, Issue 3, July 2023
Impact Factor: 7.301

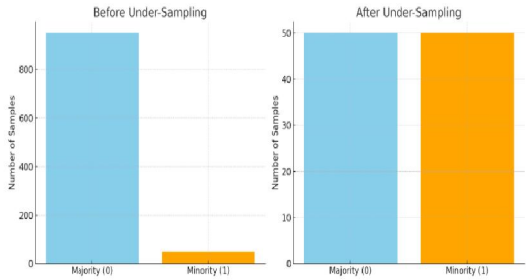

Fig. 4. Class Balancing in Under Sampling

Figure 4 shows the results of applying under-sampling before and after to achieve equal representation of the two classes; this shows that under-sampling diminishes the majority class and over-represents the minority class.

E. Feature Selection using PCA

feature selection is crucial to reduce dimensionality, remove redundant features, and improve model efficiency. PCA is a measurement reduction method that has seen extensive use. Principal component analysis (PCA) distils the initial set of characteristics into a smaller set of uncorrelated components that together explain the bulk of the variation in the data. These plot displays data points in (malware) and (benign) distributed across two principal components, with some clustering patterns visible but significant overlap between the two classes in Figure 5.

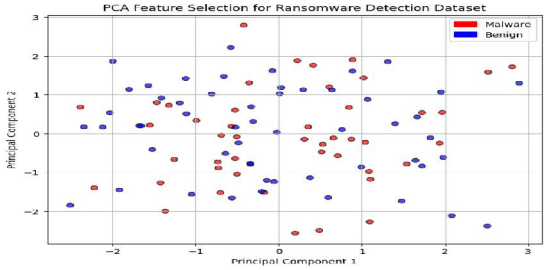

Fig. 5. PCA Feature Selection for Ransomware Detection

By zeroing in on the most significant factors, and can reduce the amount of variation. PCA reduces complexity while retaining essential information for classification

F. Data Normalization

Ransomware detection uses data normalization as a pre-processing step to ensure that characteristics have a constant range, typically from 0 to 1. As a result, features with wide numerical ranges won't have an outsized impact on model training. File size, entropy, and API call count are some characteristics that could have extremely varied ranges in datasets used for ransomware detection. Adjusts feature values so that they fall within the interval [0, 1] stated in Equation (1).

$$x' = \frac{x - x_{min}}{x_{max} - x_{min}} \qquad (1)$$

The sentence can be paraphrased as follows: "x" is the initial value of the feature, x_{min} is its minimum, x_{max} is its maximum, and x' is its normalised value."

G. Data Splitting

Data splitting in ransomware detection typically divides the dataset into **80% training**, **20% testing** subsets to ensure reliable model evaluation. Stratified splitting is often used to preserve the ransomware-to-benign sample ratio in each subset.

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429 Volume 3, Issue 3, July 2023 Impact Factor: 7.301

H. Proposed Machine Learning in LSTM model for Ransomware Detection

LSTM neural networks can learn dependencies that span several years. They are a type of RNN. They store data in what is known as a "cell," a type of memory unit. When selecting whether to update a cell's state with new or removed data, the LSTM's four neural network layers communicate with one another [25]. A typical RNN's short memory in terms of time steps is its worst flaw. In order for the neural network to produce better predictions going forward, LSTMS incorporate functions that remember past predictions from several time steps earlier and use them again and again. When it comes to learning from experience, an LSTM is very similar to the human brain.

A first step in training an LSTM is to specify the minimum amount of layer-level information that each cell must remember. A sigmoid layer, which is also a mechanism for forgetting, decides what memories to keep. A value between 0 and 1 is produced by the sigmoid activation function ∅, which takes xt and $t-1$ as inputs. The product of this value and the value of $ct-1$ is then calculated. According to Equation (2), the prior value preserved while the sigmoid output is 1, and it removed when the output is 0.

$$f_t = \sigma\left(w_f\left[h_{t-1,}x_t\right] + b_f\right) \quad (2)$$

The values decided in the previous step are forgotten when the old state $ct-1$ is multiplied by ft. Then, $it \cdot ct$ is supplemented with the outcomes. Equation (3) then instructs us to save the resultant candidate value in the corresponding cell. At this point, the cell is learning something new.

The values in the cell state are filtered to produce the output t. In Equation (4), the sigmoid layer determines which portion of the cell is to be output.

$$o_t = \sigma(W_0[h_{t-1}, x_t] + b_0) \quad (3)$$

$$h_t = o_t.tanh(C_t) \quad (4)$$

One layer of the network receives the cell value, processes it using the tanh function, and then passes the result to the next layer by multiplying it with the sigmoid layer's output.

I. Performance Matrix

The proposed model might be evaluated using a variety of ransomware detection performance metrics. F1, ACC, PRE, and REC are among the performance metrics. Here is how the confusion matrix is defined:

- **True positive (TP):** Total number of samples correctly identified as ransomware.
- **False Positive (FP):** The number of legitimate samples that were mistakenly labelled as ransomware.
- **True negative (TN):** Correct prediction rate for benign sample classification.
- **False Negative (FN):** The number of harmful code samples that were mistakenly labelled as harmless.

Accuracy

The simplest measure of performance is ACC, which is the proportion of times the model gets its predictions right. Here in Equation (5), it is computed as the proportion of correct forecasts to the overall number of guesses.

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \times 100 \quad (5)$$

Precision

PRE is defined as the proportion of correct results to the sum of correct and incorrect results. A highly precise model have a small false-positive rate, and therefore be less prone to incorrectly brand innocent files as ransom ware as indicated in below Equation (6):

$$Precision = \frac{TP}{TP+FR} \times 100 \quad (6)$$

Recall

A division of all true positives by all false negatives is the basis of the calculation. Equation (7) shows that a model with a high REC score has a low false negative rate, which implies it is less likely to fail to recognise real ransomware samples:

$$Recall = \frac{TP}{TP+FN} \times 100 \quad (7)$$

F1 Score

The F1 is a balanced statistic that considers both false positives and false negatives; it is the harmonic mean of the ACC

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429 Volume 3, Issue 3, July 2023 Impact Factor: 7.301

and REC. This is particularly helpful when dealing with data that is not evenly distributed, as seen in Equation (8):

$$F1 - \text{score} = \frac{2 \times recall \times precision}{recall + precision} \quad (8)$$

ROC Curve

A binary classifier's performance as the discrimination threshold changes can be visually represented by a ROC curve. And can use it to see how the TPR compares to the FPR for various threshold levels. An indicator of a model's overall performance is its area under the ROC curve (AUC), with a bigger AUC indicating better performance.

IV. RESULT AND DISCUSSION

The experimental ransomware detection performed on the ransomware dataset was evaluated using advanced machine learning methods. The results of this evaluation are presented in this section. We compared the model's performance on binary classification tasks using major metrics as ACC, PRE, REC, and F1. To implement it in a Jupiter Notebook, we used the Python programming language and its necessary libraries, including scikit-learn, TensorFlow, Keras, pandas, NumPy, seaborn, and matplotlib. The trials were carried out using a PC with a 32 GB RAM and an NVIDIA RTX 3070 graphics card to train LSTM networks and other deep learning models. Below, you'll find parts that thoroughly describe the ransomware detection and attribution outcomes. These sections support the idea that the proposed method is beneficial for real-time malware detection and forensic investigation in the cybersecurity environment.

Classification Report:

	precision	recall	f1-score	support
0	0.96	0.98	0.97	2141
1	0.99	0.98	0.99	4311
accuracy			0.98	6452
macro avg	0.98	0.98	0.98	6452
weighted avg	0.98	0.98	0.98	6452

C

Fig. 7. Classification Report of LSTM Model

In this classification report, the author has shown performance of an LSTM model in detecting ransomware on a sample size of 6452. The binary classifier yielded great results with an overall 98% ACC in Figure 6. Class 0 (non-ransomware) was the most precise with 96% and best REC with 98% and Class 1 (ransomware) was the highest with 99% PRE and 98% REC. Both macro and weighted averages were 98% in all metrics, which is an indicator of strong and balanced detection abilities of cybersecurity applications.

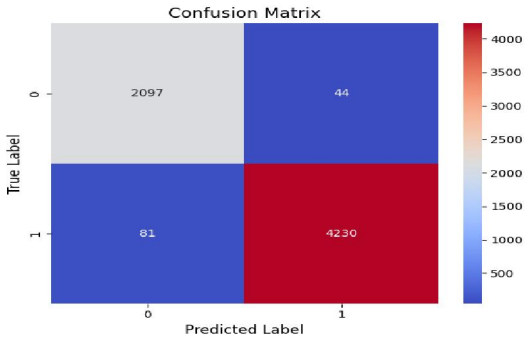

Fig. 8. Confusion Matrix of LSTM Model

This confusion matrix illustrates the performance of an LSTM-based ransomware detection model on 6452 test samples in Figure 7. The model achieved high ACC with 2097 true negatives (benign files correctly classified), 4230 true positives (ransomware correctly identified), 44 FP, and 81 FN.

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 3, Issue 3, July 2023
Impact Factor: 7.301

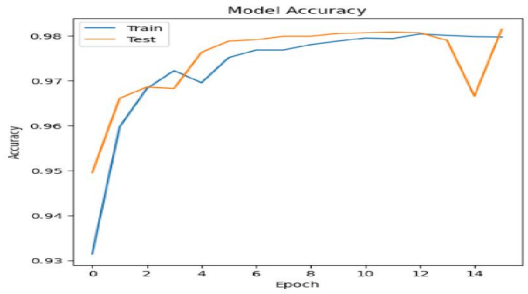

Fig. 9. Accuracy curve of LSTM Model

This plot of ACC illustrates the process of training an LSTM model to detect ransomware across 15 epochs in Figure 8. Both training and testing ACC curves converge very quickly, attaining just about 98% ACC by epoch 4. The model has low overfitting because the train and test curves are very close to each other, indicating that it performs in a stable way. There is a significant ACC degradation at around epoch 13-14 followed by an increase and this may indicate adjustment of the learning rate or batch data variation in the optimization.

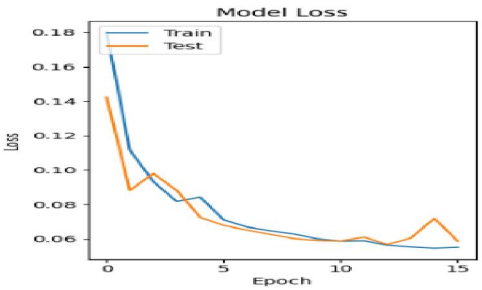

Fig. 10. Loss curve of LSTM Model

The Figure 9 loss curve shows the training dynamics of LSTM model in detecting the ransomware across 15 epochs. Training and testing loss curves show quick convergence of the initial values of about 0.16-0.14 to about 0.06 by epoch 10. The near parallel curves signify that there is little overfitting and stable learning process.

Table 2: LSTM model proposed model Performance on Ransomware attack on Ransomware detection dataset

Measure	LSTM
Accuracy	98.34
Precision	97.45
Recall	97.33
F1-score	98.21
ROC AUC	98.87

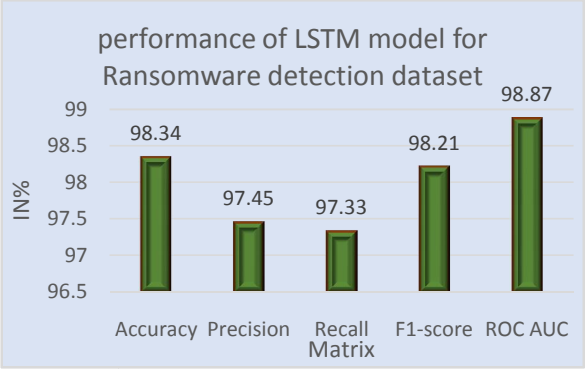

Fig. 11. Comparison of Model Performance Metrics

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 3, Issue 3, July 2023
Impact Factor: 7.301

The suggested LSTM model's comparative performance on ransomware detection on the ransomware attack dataset is shown in Table II and Figure 10. With an ACC of 98.34%, a PRE of 97.45%, and a REC of 97.33%, the model clearly has a great capacity to accurately identify harmful and benign samples. The F1 of 98.21% indicates good balance between REC and PRE and the ROC AUC score of 98.87% ensures high discriminative power. These holistic measurements justify the use of the LSTM-based method in real-time ransomware detection and forensics in cybersecurity implementation.

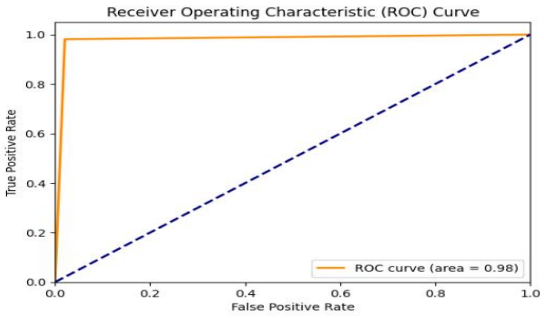

Fig. 12. Roc Curve of Proposed Model

In the LSTM model, this is a ROC curve that demonstrates the high discriminating ability of the model to detect ransomware in Figure 11. The curve shows almost perfect classification ability with a AUC of 0.98 which is far much better than random classification which is represented by the diagonal dashed line. The results of this ROC analysis confirm the strong capability of the model to differentiate ransomware and benign samples, which proves its stability to be used in cybersecurity use cases.

A. Comparative Discussion

The following comparative analysis table supports the strong performance of the suggested LSTM model in autonomous ransomware attacking detection on the ransomware detection dataset available in Table III. The LSTM model had the best ACC of 98.34%, PRE of 97.45, REC of 97.33 and F1 of 98.21%, and significantly better than other methods. RF scored an ACC of 96.90, PRE of 92.73% and F1 of 96.23%. The ACC of SVM was 91.67 with a balanced 91.7% ACC, REC, and F1. CNN scored 94.38 % ACC, 94.41% PRE, 94.38% REC and 94.39% F1. The overall analysis ensures that the LSTM model trained on deep learning has enhanced the independent detection and the evaluation metrics are always high than the conventional machine learning and other neural network models in cybersecurity application.

Table 3: Comparison between proposed model and Existing models for ransomware detection in cybersecurity environment

Model	Accuracy	Precision	Recall	F1 score
LSTM	98.34	97.45	97.33	98.21
RF[26]	96.90	92.73	--	96.23
SVM[27]	91.67	91.7	91.7	91.7
CNN[28]	94.38	94.41	94.38	94.39

The LSTM-based deep learning algorithm performs remarkably well in terms of ransomware detection, with an ACC of 98.34%, clearly surpassing the results of current methods, such as RF with 96.90%, CNN with 94.38%, and SVM with 91.67%. The LSTM model is capable of providing accurate detections of complex ransomware signatures by utilizing the sequential pattern recognition and time dependencies in the malware behavior to detect them accurately in cybersecurity context. The high success of deep learning method indicates that it is effective in tackling the different ransomware families and the escalating attack vectors due to its powerful feature extraction features. There are, however, some challenges such as the possibility of adversarial attacks on neural networks and the computational ability to implement the network in real-time. Altogether, this LSTM-based system offers cybersecurity specialists with

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429 Volume 3, Issue 3, July 2023 Impact Factor: 7.301

powerful and effective tools to detect ransomware attacks and high detection ACC and allows using this type of analysis to trace the attackers and conduct forensic examination.

V. CONCLUSION AND FUTURE WORK

Malicious software known as ransomware encrypts a victim's data and then demands payment in exchange for decryption. Independent forensics through high-end machine learning to deal with the increased complexity of ransomware attacks. The proposed framework is an improvement over the current ransomware defines paradigm allowing the organization to revaluate ransomware defines as a proactive approach instead of a reactive one, improving the detection rate of 98.34% through systematic data pre-processing and PCA-driven feature optimization and automated recovery initiation without human intervention, substantially decreasing the operational vulnerability of the enterprise cybersecurity frameworks. Proactive defensive mechanisms in cybersecurity have a new paradigm thanks to the autonomous framework, which classifies ransomware families with high ACC and at a very low cost of computation.

Future research directions include integration of federated learning architectures for collaborative threat intelligence sharing across organizations, implementation of explainable AI techniques to enhance forensic transparency and legal admissibility, development of adaptive learning mechanisms capable of identifying zero-day ransomware variants through behavioural anomaly detection, incorporation of blockchain technology for tamper-proof forensic evidence management, extension to IoT and edge computing environments for comprehensive network protection, and enhancement of the recovery protocols to include predictive modelling for pre-emptive threat mitigation and automated backup restoration strategies.

REFERENCES

[1] A. Mathew, "The Evolution of Ransomware in Cybersecurity Space," *Int. Res. J. Innov. Eng. Technol.*, 2022.

[2] U. Urooj, B. A. S. Al-rimy, A. Zainal, F. A. Ghaleb, and M. A. Rassam, "Ransomware Detection Using the Dynamic Analysis and Machine Learning: A Survey and Research Directions," *Appl. Sci.*, vol. 12, no. 1, 2022, doi: 10.3390/app12010172.

[3] J. Skertic, "Cybersecurity Legislation and Ransomware Attacks in the United States, 2015-2019," *Grad. Progr. Int. Stud. Theses Diss.*, 2021.

[4] L. Y. Connolly and D. S. Wall, "The rise of crypto-ransomware in a changing cybercrime landscape: Taxonomising countermeasures," *Comput. Secur.*, vol. 87, Nov. 2019, doi: 10.1016/j.cose.2019.101568.

[5] V. S. Thokala, "A Comparative Study of Data Integrity and Redundancy in Distributed Databases for Web Applications," *Int. J. Res. Anal. Rev.*, vol. 8, no. 04, pp. 383–390, 2021.

[6] N. Patel, "Quantum Cryptography In Healthcare Information Systems: Enhancing Security In Medical Data Storage And Communication," *J. Emerg. Technol. Innov. Res.*, vol. 9, no. 8, 2022.

[7] T. R. Reshmi, "Information security breaches due to ransomware attacks - a systematic literature review," 2021. doi: 10.1016/j.jjimei.2021.100013.

[8] F. Aldauiji, O. Batarfi, and M. Bayousef, "Utilizing Cyber Threat Hunting Techniques to Find Ransomware Attacks: A Survey of the State of the Art," *IEEE Access*, 2022, doi: 10.1109/ACCESS.2022.3181278.

[9] S. S. S. Neeli, "Key Challenges and Strategies in Managing Databases for Data Science and Machine Learning," *Int. J. Lead. Res. Publ.*, vol. 2, no. 3, p. 9, 2021, doi: 10.5281/zenodo.14672937.

[10] N. Malali, "The Role Of Devsecops In Financial Ai Models: Integrating Security At Every Stage Of Ai/Ml Model Development In Banking And Insurance," Ijetrm," *IJETRM*, vol. 6, no. 11, p. 218, 2022, doi: 10.5281/zenodo.15239176.

[11] M. Humayun, N. Jhanjhi, A. Alsayat, and V. Ponnusamy, "Internet of things and ransomware: Evolution, mitigation and prevention," *Egypt. Informatics J.*, vol. 22, no. 1, pp. 105–117, Mar. 2021, doi: 10.1016/j.eij.2020.05.003.

[12] A. Zimba, Z. Wang, and H. Chen, "Multi-stage crypto ransomware attacks: A new emerging cyber threat to critical infrastructure and industrial control systems," *ICT Express*, vol. 4, no. 1, pp. 14–18, Mar. 2018, doi:

10.1016/j.icte.2017.12.007.

[13] S. R. Davies, R. Macfarlane, and W. J. Buchanan, "Evaluation of live forensic techniques in ransomware attack mitigation," *Forensic Sci. Int. Digit. Investig.*, vol. 33, Jun. 2020, doi: 10.1016/j.fsidi.2020.300979.

[14] J. E. Thomas, R. P. Galligher, M. L. Thomas, and G. C. Galligher, "Enterprise Cybersecurity: Investigating and Detecting Ransomware Infections Using Digital Forensic Techniques," *Comput. Inf. Sci.*, 2019, doi: 10.5539/cis.v12n3p72.

[15] S. Thangavel, K. C. Sunkara, and S. Srinivasan, "Software-Defined Networking (SDN) in Cloud Data Centers: Optimizing Traffic Management for Hyper-Scale Infrastructure," *Int. J. Emerg. Trends Comput. Sci. Inf. Technol.*, vol. 3, no. 1, pp. 29–42, 2022, doi: 10.63282/3050-9246.IJETCSIT-V3I3P104.

[16] S. Kamil, H. S. A. Siti Norul, A. Firdaus, and O. L. Usman, "The Rise of Ransomware: A Review of Attacks, Detection Techniques, and Future Challenges," in *2022 International Conference on Business Analytics for Technology and Security, ICBATS 2022*, 2022. doi: 10.1109/ICBATS54253.2022.9759000.

[17] J. H. Park, S. K. Singh, M. M. Salim, A. E. L. Azzaoui, and J. H. Park, "Ransomware-based Cyber Attacks: A Comprehensive Survey," *J. Internet Technol.*, vol. 23, no. 7, 2022, doi: 10.53106/160792642022122307010.

[18] C. Sendner, L. Iffländer, S. Schindler, M. Jobst, A. Dmitrienko, and S. Kounev, "Ransomware Detection in Databases through Dynamic Analysis of Query Sequences," in *2022 IEEE Conference on Communications and Network Security (CNS)*, 2022, pp. 326–334. doi: 10.1109/CNS56114.2022.9947244.

[19] R. M. A. Molina, S. Torabi, K. Sarieddine, E. Bou-Harb, N. Bouguila, and C. Assi, "On Ransomware Family Attribution Using Pre-Attack Paranoia Activities," *IEEE Trans. Netw. Serv. Manag.*, vol. 19, no. 1, pp. 19–36, 2022, doi: 10.1109/TNSM.2021.3112056.

[20] V. Rathod, C. Parekh, and D. Dholariya, "AI & ML Based Anamoly Detection and Response Using Ember Dataset," in *2021 9th International Conference on Reliability, Infocom Technologies and Optimization (Trends and Future Directions) (ICRITO)*, 2021, pp. 1–5. doi: 10.1109/ICRITO51393.2021.9596451.

[21] M. Almousa, S. Basavaraju, and M. Anwar, "API-Based Ransomware Detection Using Machine Learning-Based Threat Detection Models," in *2021 18th International Conference on Privacy, Security and Trust (PST)*, 2021, pp. 1–7. doi: 10.1109/PST52912.2021.9647816.

[22] B. M. Khammas, "Ransomware Detection using Random Forest Technique," *ICT Express*, vol. 6, no. 4, pp. 325–331, Dec. 2020, doi: 10.1016/j.icte.2020.11.001.

[23] K. Lee, S.-Y. Lee, and K. Yim, "Machine Learning Based File Entropy Analysis for Ransomware Detection in Backup Systems," *IEEE Access*, vol. 7, pp. 110205–110215, 2019, doi: 10.1109/ACCESS.2019.2931136.

[24] S. Poudyal, K. P. Subedi, and D. Dasgupta, "A Framework for Analyzing Ransomware using Machine Learning," in *2018 IEEE Symposium Series on Computational Intelligence (SSCI)*, 2018, pp. 1692–1699. doi: 10.1109/SSCI.2018.8628743.

[25] S. Maniath, A. Ashok, P. Poornachandran, V. G. Sujadevi, P. S. A.U., and S. Jan, "Deep learning LSTM based ransomware detection," in *2017 Recent Developments in Control, Automation & Power Engineering (RDCAPE)*, IEEE, Oct. 2017, pp. 442–446. doi: 10.1109/RDCAPE.2017.8358312.

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

[27] A. O. Almashhadani, M. Kaiiali, S. Sezer, and P. O'Kane, "A Multi-Classifier Network-Based Crypto Ransomware Detection System: A Case Study of Locky Ransomware," *IEEE Access*, vol. 7, pp. 47053–47067, 2019, doi: 10.1109/ACCESS.2019.2907485.

[28] J. Hemalatha, S. A. Roseline, S. Geetha, S. Kadry, and R. Damaševičius, "An efficient densenet□based deep learning model for Malware detection," *Entropy*, vol. 23, no. 3, pp. 1–23, 2021, doi: 10.3390/e23030344.