SEMARAK ILMU
PUBLISHING
202103268166(003316878-P)

Journal of Advanced Research in Applied Sciences and Engineering Technology

Journal homepage:
https://semarakilmu.com.my/journals/index.php/applied_sciences_eng_tech/index.
ISSN: 2462-1943

Ransomware Classification with Deep Neural Network and Bi-LSTM

Mujeeb ur Rehman Shaikh[1,*], Mohd Fadzil Hassan[2], Rehan Akbar[3], K.S. Savita[4], Rafi Ullah[4], Satria Mandala[5]

[1] Computer and Information Sciences Department, Universiti Teknologi PETRONAS, Seri Iskandar, 32610, Perak, Malaysia
[2] Centre for Research in Data Science (CeRDaS), Universiti Teknologi PETRONAS, 32610 Seri Iskandar, Perak Darul Ridzuan, Malaysia
[3] School of Computing and Information Sciences, Florida International University, Miami, United States of America
[4] Positive Computing Research Centre, Universiti Teknologi PETRONAS, Seri Iskandar, 32610, Malaysia
[5] Human Centric (HUMIC) Engineering & School of Computing Telkom University Bandung, Indonesia

ARTICLE INFO

Article history:
Received 28 October 2023
Received in revised form 3 December 2023
Accepted 4 March 2024
Available online 25 June 2024

Keywords:
Ransomware Classification; Deep Neural Network; Bi-LSTM; Deep Learning; Cybersecurity

ABSTRACT

Malicious attacks, malware, and ransomware families present essential risks to cybersecurity and may result in significant harm to computer systems, data clusters, networks, and mobile apps across a range of industries. Recently, there has been interest in ransomware classification using DNN and Bi-LSTM. DNN, a subset of machine learning techniques, has been found to improve ransomware detection and classification precision and efficacy. Ransomware has been affecting commercial, public, and governmental organizations' networks and computer systems for more than a decade, enabling new dynamic detection techniques to help DNNs detect ransomware. However, deep neural network-based architectures and DL classifiers (such as DNN, and Bi-LSTM classifiers) will be employed to detect ransomware. These networks may learn to correctly identify and categorize new ransomware incidents by integrating various datasets, including known and unknown ransomware samples. The classification of ransomware detection has been thoroughly investigated, and a model incorporating classic DL techniques with DNN and Bi-LSTM-based architecture will be proposed. A model execution experiment will be carried out to facilitate comparative testing of various approaches. This study focuses on the detection and classification of ransomware using DNN and Bi-LSTM. This study provides the groundwork for future investigations into the issues with ransomware detection. To protect against several ransomware attack types, deep neural networks have become an effective tool for ransomware detection. These networks combine machine learning and deep learning techniques.

1. Introduction

Ransomware attacks are the most notable cyber-attacks that have affected organizations around the world in the last five years. Verizon Data Breach Investigation Report (DBIR) 2021 states that 37% of organizations worldwide reported being affected by ransomware also in healthcare industry [1]. The number of ransomware attacks worldwide increasing immensely compared to the previous year

[*] *Corresponding author.*
E-mail address: mujeeb_22007910@utp.edu.my

https://doi.org/10.37934/araset.47.2.266280

by the middle of 2023 [2]. WannaCry's 2017 outbreak brought back the attention to ransomware [3]. The attack highlighted not only the potential dangers of ransomware, but also its cost-effectiveness. The WannaCry assaults were chaos and fear, not monetary gain. Even though the ransomware merely demanded $300, the projected cost of the financial loss was more. Since then, there have been numerous ransomware assaults and variations. The COVID-19 pandemic is also largely to blame for the rise in recent cyberattacks [4,5]. One of the reasons criminal exposures has become more difficult is the use of virtual currencies such as bitcoins in trade, which is almost impossible to track. This model remains valid because attackers are victims of peer pressure and willing to pay any amount to obtain their data. Further, escape technology is rapidly spreading. It is a challenge for antivirus software to adapt to the development of ransomware.

This global economy is beneficial to cybercriminals, as there is a lack of information on spam messages and other mechanisms that allow the spread of extremely high ransomware. In the battle against ransomware, one of the main objectives is to limit file losses if no previous detection has been achieved. The current detection mechanism depends on restricting the number of files lost after encryption by blocking any ransomware-like process (API call, registry key, embedded binary string, etc.). However, there are still risks. It is possible to establish a hypothesis: Before suspicious behavior, there is no alarm. Nevertheless, other measures are mandatory to prevent and limit additional damage and loss of data in systems when warning mechanisms are not detected in the "print" phase, as corporations move to distant work paradigms, workers become more vulnerable to phishing emails and thus introduce security breaches in the organization's defense in contradiction of cyberattacks [6].

The exponential growth in the sophistication and frequency of ransomware attacks necessitates the development of effective pre-encryption detection approaches to identify and mitigate these threats before irreparable damage is done. This research article aims to provide a comprehensive understanding of detection approaches for ransomware detection. It is crucial for identification and mitigation of ransomware attacks, minimizing damage and financial losses. It enables initiative-taking incident responses, reducing downtime and safeguarding sensitive information from encryption and potential data breaches. By strengthening cybersecurity defenses and staying ahead of evolving ransomware threats, organizations can protect their systems, data, and maintain operational continuity focusing on their taxonomy and research directions. By examining existing literature and studies in the field, aim to analyze and identify gaps, assess the current state of knowledge, and suggest future research directions.

This technology evolves, is more concentrated and uses precise noise-free attacks on networks despite changes in technology and some tactics, and cryptographic ransomware has a differentiating feature that distinguishes it from malicious software, the ability and purpose to encrypt victims' data, enabling only malicious actors to decrypt them when payment of ransom [7]. Additionally, the results presented in this research article are based on a comprehensive review of existing literature, including scholarly articles, conference papers, and industry reports. As of our knowledge cutoff in May 2023, considered the most recent advancements in the field. By offering a detailed analysis of pre-encryption detection approaches, their taxonomy, and future research directions, this study article aims to contribute to the development of more effective strategies in the battle against ransomware, Figure 1 illustrates ransomware victimization from 2018 to 2022.

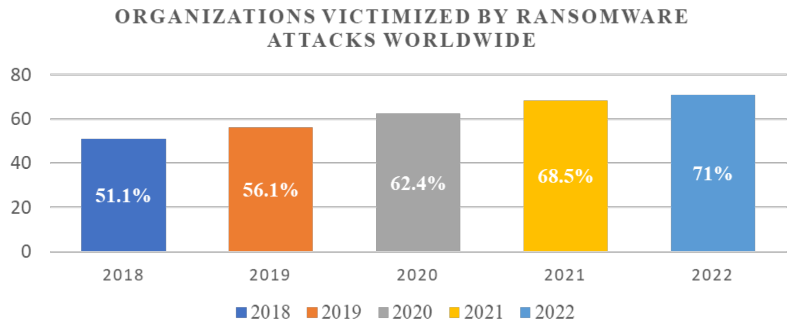


Fig. 1. Ransomware Victimization rate report

Table 1
Traditional vs new generation ransomware

Aspect	Traditional Ransomware	New Generation Ransomware
Encryption Technique	Symmetric encryption algorithms like AES	Asymmetric encryption algorithms like RSA
Infection Vector	Malicious links	Exploits vulnerabilities in software, networks, or devices
Communication	Command-and-control (C&C) servers	Peer-to-peer (P2P) networks
Payment Method	Demands payments in traditional currencies (e.g., Bitcoin)	Demands payments in cryptocurrencies
Targeting	Small to medium-sized businesses (individual users)	Targets larger organizations and critical infrastructure
Evasion Techniques	Bypass signature-based detection	Implements advanced evasion techniques (e.g., polymorphism, encryption) to evade detection
Extortion Tactics	Encrypt files if ransom is not paid	Employ data theft and threat of public exposure
Sophistication Level	Less sophisticated with limited functionalities	More advanced, anti-analysis mechanisms and adaptive behavior

These methods include anomaly detection, behavioral analysis, machine learning, and signature-based detection. With the use of this taxonomy, also intend to establish a systematic framework to compare and evaluate the various methods. Next, will examine each strategy in detail, and their advantages, disadvantages, and performance in various settings. To improve detection accuracy and speed, they include the incorporation of artificial intelligence (AI) techniques like machine learning and deep learning. Findings in this study are based on a thorough analysis of the body of literature, which includes academic articles, conference papers, and business reports [8,9].

The ratio, superiority and costs of malware imposed on the world economy are growing slowly. According to systematic and commercial data, about 1.6 million malware files are created every day, and cybersecurity companies are expected to increase annual costs of global cybercrime by 15% over the next five years, rising from $6 trillion in 2021 to $10.5 trillion by 2025. Table 1 shows the traditional ransomware and new generation ransomware from the various aspects of the ransomware detection.

The impact of cybercrime is not limited to large enterprises, but also to small and medium-sized enterprises, which can suffer significant financial losses. Cybercrime can damage and destroy data, theft of money, productivity, intellectual property theft, and other indirect costs. The growth of cybercrime poses a serious threat to the global economy, underlining the need for effective cybercrime prevention [10]. Despite all the reports and is well cases of ransomware attacks, organizations still survive and keep improving in sophistication and effectiveness. According to 2021 study, 96% of corporations who had previously fallen victim to ransomware groups reported.

As shown in Figure 2, ransomware attacks account for 35%, 33%, and 28% of all cyberattacks in industries such as professional services, government, and health care, respectively, making it the most common attack [11].

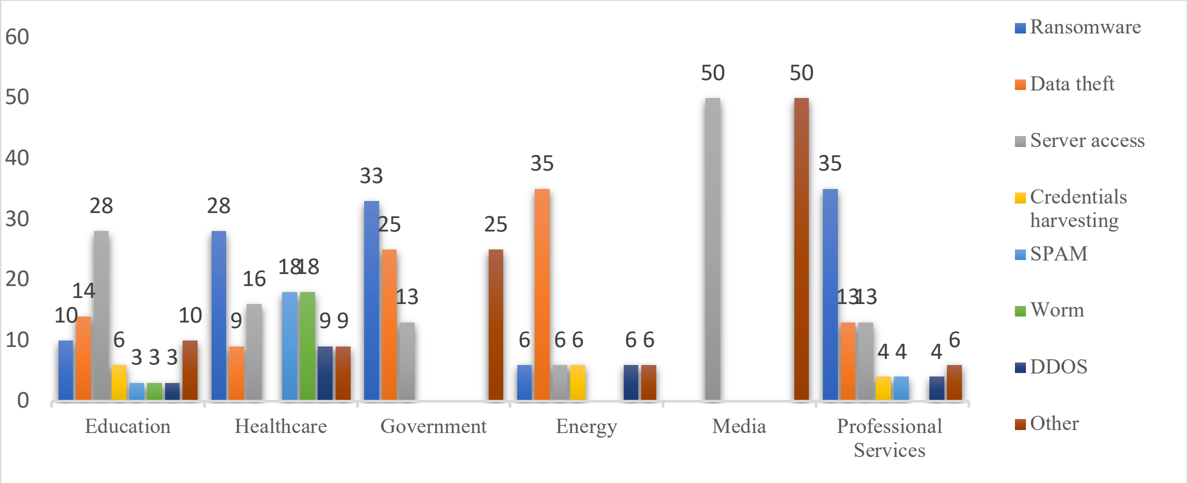


Fig. 2. Types of attacks per industry

However, another trend has been noted in Sophos' research about ransomware. The malicious ransomware actors have shifted from large-scale general and automated attacks to more specific attacks that are carried out with precision and persistence [12]. The analysis of available data on ransomware group operations indicates obvious similarities to the innovative persistent threat mode operations.

1.1 Background of the Research

The risky behavior of crypto-ransomware attacks makes it difficult to handle when scheming a model for detecting kinds of attacks. If the model does not distinguish between benign programs and crypto-ransomware attacks [12,13], there is a high likelihood of false alarms [14]. The behavior of malicious malware and the irreversible nature of the attack made it more difficult to detect [7]. Because of the development of ransomware variants [13], there is a lack of detection solutions to distinguish between legitimate processes and malicious code. Many studies have used a fixed threshold to extract data from crypto-ransomware attacks. The use of cryptographic APIs is also difficult because the API is also used by valid benign programs, leading to high false alarms. The use of cryptographic APIs made detection more difficult [14]. When a design system faces difficulty classifying the encounter process into legitimate, harmless, and malicious programs, the accuracy of the encounter process is low. Models are weak in accuracy when they cannot spot zero-day attacks and cannot deal with the evolving and changing behavior of crypto-ransomware attacks. Crypto-ransomware attacks are irrevocable and are important for the study of cybersecurity [15]. Crypto-ransomware has a long-term effect. Without a decryption key, you cannot deal with a user file attacked by crypto-ransomware [16]. Prior research has attempted to resolve the detection of ransomware attacks in the initial stages before encryption begins. However, these solutions do not solve the dynamic nature of ransomware attacks. The efficiency of zero-day awareness of early detection of Adaptive Crypto-Ransomware is doubled, i.e., the development of Adaptive Technology and accurate detection of attacks of Crypto-Ransomware with the help of Adaptive Online Classifiers [16].

1.2 Related Work

The focus of several research investigations has been how to make Internet-connected devices more secure against virus attacks. Static analysis is a common method for analyzing malware since it identifies malware patterns without running files or looking at the source code [17]. This static analysis method, however, is ineffective in finding malware that is concealed or that displays unusual patterns. On the other hand, dynamic analysis-based malware detection techniques may decipher obscured or unfamiliar malware by tracking system changes and real-time behaviors when files are run [18]. Hybrid analysis, which combines static and dynamic studies, is another efficient method for analyzing malware. This approach, which is frequently used in cybersecurity studies [19] , has the power to accurately detect malware.

1.3 Static Malware Detection (Pre-Execution Analysis)

Static analysis is a method of evaluating software or systems by analyzing their source code or compiled binaries without executing them. In the context of cybersecurity, static analysis is used to examine the code of malware and other malicious code to identify its capabilities and intent, without the need to run the malware. However, because code pathways might not be available during actual execution, static analysis can result in erroneous execution behavior [20].

1.4 Dynamic Analysis (post-execution analysis)

Dynamic analysis is a method of evaluating software or systems by executing them and observing their behavior in the context of cybersecurity, dynamic analysis is used to analyze the behavior of malware and other malicious code by running it in a controlled environment, such as a sandbox. This allows security researchers to observe the malware's behavior, identify its capabilities, and develop countermeasures [21]. Static and dynamic ransomware analysis. Table 2 summarizes the evaluation using the following factors: Speed, safety, ability to analyze obfuscated and polymorphic hardware, level of false positives, and accuracy.

Table 2
Difference between static and dynamic ransomware analysis

Parameters	Degree	Static Analysis	Dynamic Analysis
Speed	High	√	√
	Low		
Safety	Low	√	√
	High		
Obfuscated and polymorphic ransomware	Unable	√	√
	Able		
False positive level	Low High	√	√
Accuracy	Low High	√	√

1.5 Ransomware Research

Locker and crypto ransomware are the two primary categories of ransomware. While leaving the system and files untouched, the Locker ransomware concentrates on changing the user interface. While operating system operations and necessary utilities like input/output tools and desktop apps are disabled by the Locker ransomware, data are left unharmed. On the other side, cryptographic ransomware, sometimes known as crypto ransomware, takes its malicious goal a step further. By encrypting the victims' files and keeping them hostage until a ransom is paid, this kind tries to extort money from its victims. [22]. Figure 3 shows ransomware kill chain cycle.

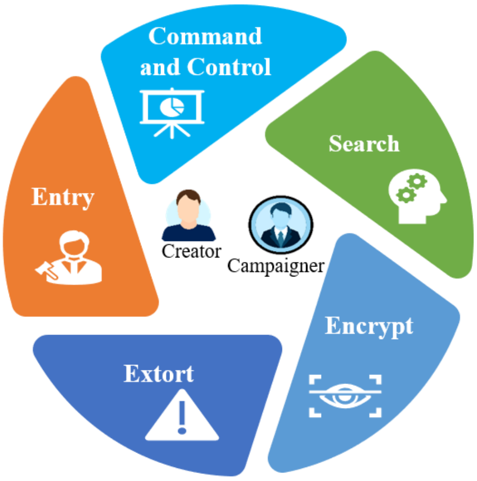


Fig. 3. Ransomware Encryption Steps

- Entry: The malware starts its setup procedures and self-propagation.
- C&C (Command and Control): The malware tries to connect with its command-and-control hub.
- Search: To maximize the possibility that the victim would pay the ransom demands, the ransomware looks for specific files of interest, usually vital data.
- Encrypt: Using encryption keys received from the command-and-control center, the ransomware begins the encryption process on the specified files.
- Extort: The victim is forced to pay a ransom in exchange for the release of the decryption key by the ransomware, which displays an extortion message.

1.6 Research Contribution

The goals of this study have been achieved in significant part. The first significant contribution is identifying critical ransomware attack behaviors through API analysis. The creation of a DNN with BiLSTM architecture to anticipate ransomware before it starts widespread, unauthorized file encryption constitutes the second important contribution. This research focuses on the detection of ransomware assaults, which have permanent effects after encryption and are frequently missed in earlier studies. The research also provides a unique solution by combining the trademark matching strategy with a machine learning technique. The combination of DNN with BiLSTM takes advantage of both deep learning and sequential modeling methods, improving the capacity to identify ransomware. Comparing this method to more conventional ones, it increases detection accuracy and decreases false positives. Additionally, it achieves increased accuracy by expertly combining static and dynamic characteristics, which makes it possible to precisely capture ransomware behaviors. Its versatility is what makes it unique; because of its deep learning capabilities, the model can change along with new ransomware strains, guaranteeing that it always has the latest information. This fusion technique enables real-time detection capabilities, which is feasible. By quickly recognizing and reducing ransomware attacks, this real-time reaction limits the potential damage. Furthermore, it performs well even when dealing with ransomware strains that have never been seen before. It also improves interpretability, which makes it simpler to comprehend detection findings. This function supports forensic investigation and facilitates a better comprehension of the nature of threats discovered. The fusion approach is a flexible and effective tool in the campaign against ransomware because it grows effectively and can handle big datasets and network traffic volumes.

As previously mentioned, most of the prior research has focused on analyzing malware characteristics. Drawing from their analyses, various approaches have been suggested to prevent or identify ransomware. These existing studies according to their primary objectives either preventing ransomware infection or detecting it after it infiltrates the system [23]. A classification diagram depicting the tools employed in the reviewed studies can be observed in Figure 4.

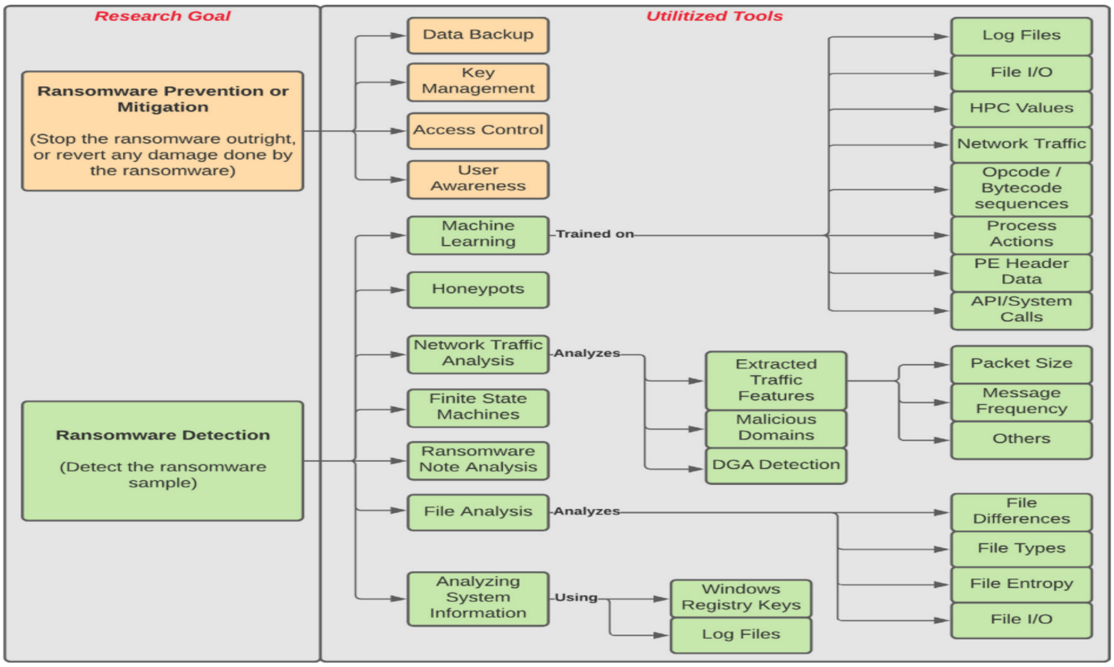


Fig. 4. Ransomware prevention, mitigation, and detection [23]

A fascinating area of computer science, machine learning (ML) has demonstrated its effectiveness in decision-making and image identification [8]. Additionally, deep learning (DL) draws on strong and adaptable models to make it easier to extract crucial insights for challenging jobs. Considering this, DL shows promise in several cybersecurity domains, including malware detection, classification, and analysis; identification and detection of botnets; mitigation of cyberattacks; intrusion detection and prevention; incident response; analysis of network traffic; detection of advanced persistent threats (APTs); identification of cybercriminals; deep packet inspection; and the field of cybersecurity analytics. The possible uses of ML models across several cybersecurity fields are depicted in a taxonomy in Figure 5.

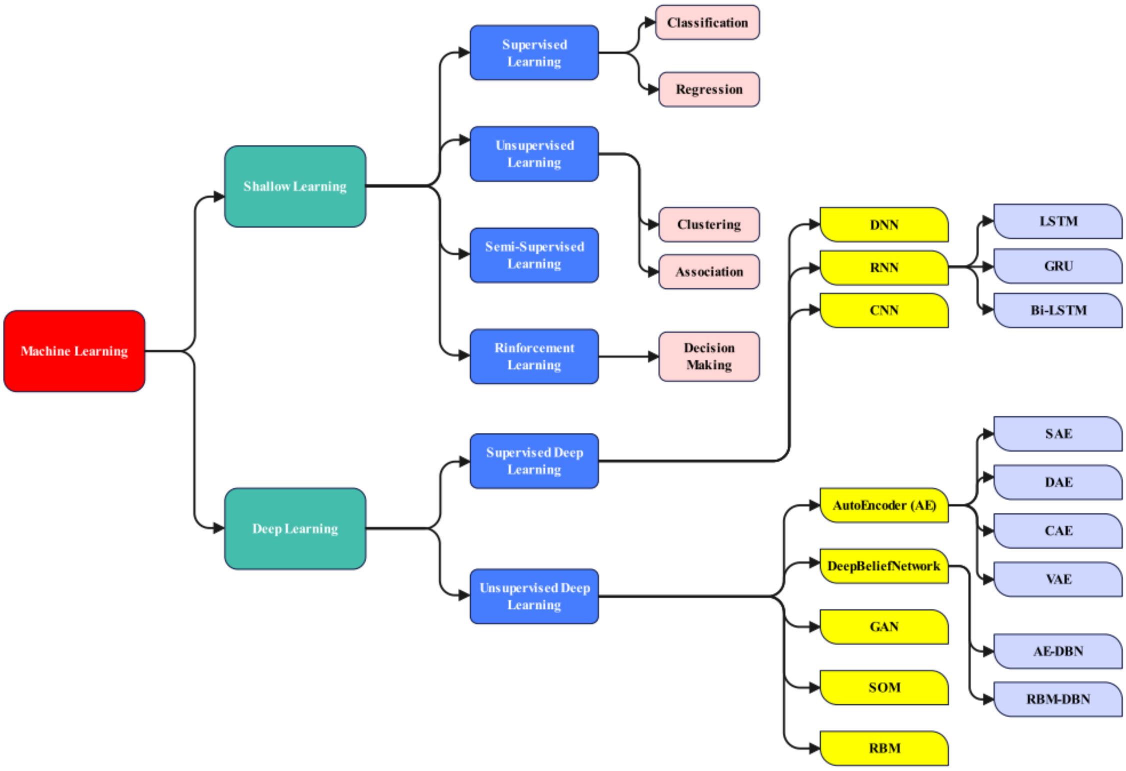


Fig. 5. ML-Based Approaches for Machine learning [8]

1.7 Deep Neural Network

The difficulty of understanding ideas that deep neural networks (DNNs) (as shown in Figure 6) have learnt is discussed in this section. Layers of linked neurons make up DNNs, which use error reverse propagation to learn intricate input-to-output mappings [24]. Understanding abstract ideas represented by top-layer neurons is the aim. The input domain, like as images or text, can be interpreted even if these top-layer neurons are abstract and not easily clear. We will investigate the use of activation maximization to produce interpretable prototypes within the input domain. With DNN it is seen that pattern recognition is highly effective at complex patterns within data, behavior analysis, adaptability, feature extraction and reducing false positives can be trained to distinguish between benign and malicious activity more accurately, reducing the number of false alarms and improving the efficiency of ransomware detection systems. Several limitations including. Limited adaptability, High false alarms with slower response time, and focused on file behavior, network activity and anomaly detection such as unexpected spikes in file encryption or unusual access to sensitive files.

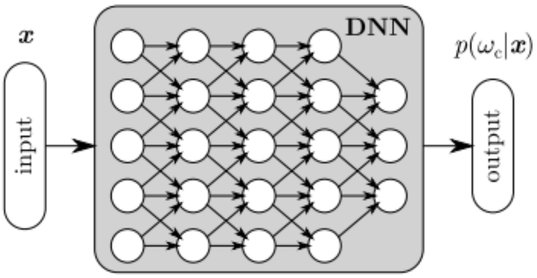


Fig. 6. Deep Neural Networks [24]

1.8Bi-Directional LSTM

Recurrent Neural Networks (RNNs) hold prior outputs for the current step input, in contrast to classic neural networks where inputs and outputs are independent. RNNs use historical context to improve sequence predictions, however they have memory and "vanishing gradient" problems. RNN constraints are solved by Long Short-Term Memory (LSTM), which retains lengthy input sequences and captures connections across dimensions like time [25]. By incorporating information from both forward and backward LSTMs in each step, bi-directional LSTM (Bi-LSTM) enhances LSTM, improving sequence comprehension.

In this study, the Bi-LSTM configuration comprises two components: the forward LSTM and the backward LSTM. Both LSTMs utilize the Sigmoid activation function. The Bi-LSTM model is constructed with the Mean Squared Error serving as the loss function, along with Adam optimization. Additionally, Binary Accuracy is adopted as the metric for evaluation, as highlighted in Figure 7.

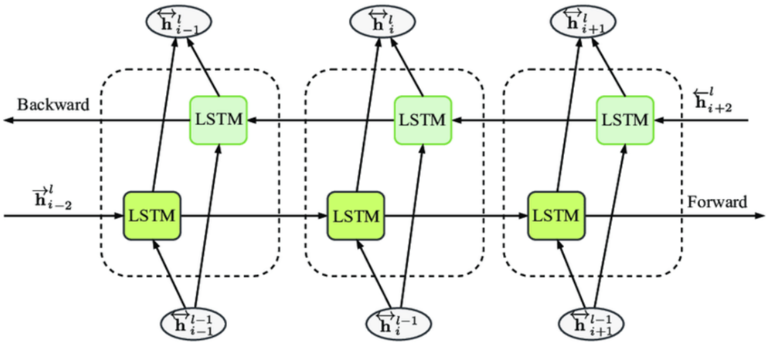

$\overleftrightarrow{h}^l_{i-1}$
$\overleftrightarrow{h}^l_i$
$\overleftrightarrow{h}^l_{i+1}$
Backward
LSTM
LSTM
LSTM
$\overleftarrow{h}^l_{i+2}$
$\overrightarrow{h}^l_{i-2}$
LSTM
LSTM
LSTM
Forward
$\overleftrightarrow{h}^{l-1}_{i-1}$
$\overleftrightarrow{h}^{l-1}_i$
$\overleftrightarrow{h}^{l-1}_{i+1}$

Fig 7. Bi-LSTM Visualization [25]

2. Methodology

Three essential elements make up our thorough categorization system, which is shown in Figure 8.

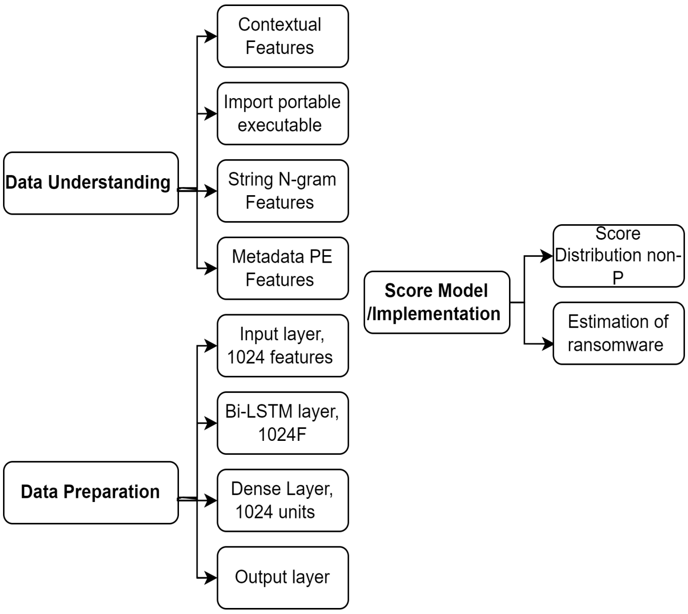


Fig. 8. Ransomware Classification Framework

In the first step, static benign and malicious binaries are mined for four different complementing feature classes. The second essential element is our deep neural network classifier, which consists of an input layer, two Bi-LSTM layers, Dense and an output layer. The third and last element, our score calibrator, is essential in converting neural network outputs into scores that provide a practical approximation of the possibility that the file is indeed ransomware.

We give a detailed analysis of each of these model elements in the sections that follow.
Below illustrates progressively the stages outlined in the journey of this research. Each of the phases has activities to be conducted simultaneously.

2.1 Phase 1: Data Understanding

In the first phase, the task to be performed for the quality of research is to undertake an exhaustive literature search. Therefore, a search will be conducted using six different electronic libraries namely used Google Scholar, IEEE Xplore, Science Direct, ACM, Springer, Web of Science, and other search engines to find pertinent resources. This included book chapters, journal articles, conference papers, e-books, symposiums, and conference proceedings, enabling a thorough search across all published types. Collection of labelled and unlabelled datasets having massive data from the GitHub and Kaggle library after data analysing this data will also be used for technique creation and model development with PE, string N-Gram, metadata PE features.

2.2 Phase 2: Data Preparation

At this phase, the proposed technique for ransomware detection model will be constructed based on the analytical outcome from the Dataset collection. The deep learning, DNN and Bi-LSTM and other methods will be considered for construction. As well as model development phases discussed.

2.3 Phase 3: Model Implementation and Experimentation

After the technique and model construction, the Model will be trained, validated, and evaluated using deep neural network, Bidirectional (DNN, Bi-LSTM), Bi-LSTM is best at handling high-dimensional and large datasets, and it is robust to overfitting and noise. Also, Python programming language will be used for Construction, development, and experimentation. This stage is iterative in nature, it may require revisiting technique construction and model accuracy.

Based on a comprehensive knowledge of ransomware detection as a step followed in a framework, we decided to employ a deep neural network with Bi-LSTM rather than a shallow yet broad neural network. It is recognized that deep architectures can offer improved efficiency in terms of the number of fitting parameters, as illustrated by [26]. This issue is significant in the context of our investigation due to the modest size of our binary sample dataset. When compared to the enormous variety of binaries that may be located inside an organization's network, its size is significantly less. Consequently, we are constrained in how deeply we can investigate the feature space that can be identify through Deep learning classifiers whether it is ransomware or benign if it is benign then inform the user if it found the ransomware quarantine and file will be generated at the end. As illustrated in Figure 9.

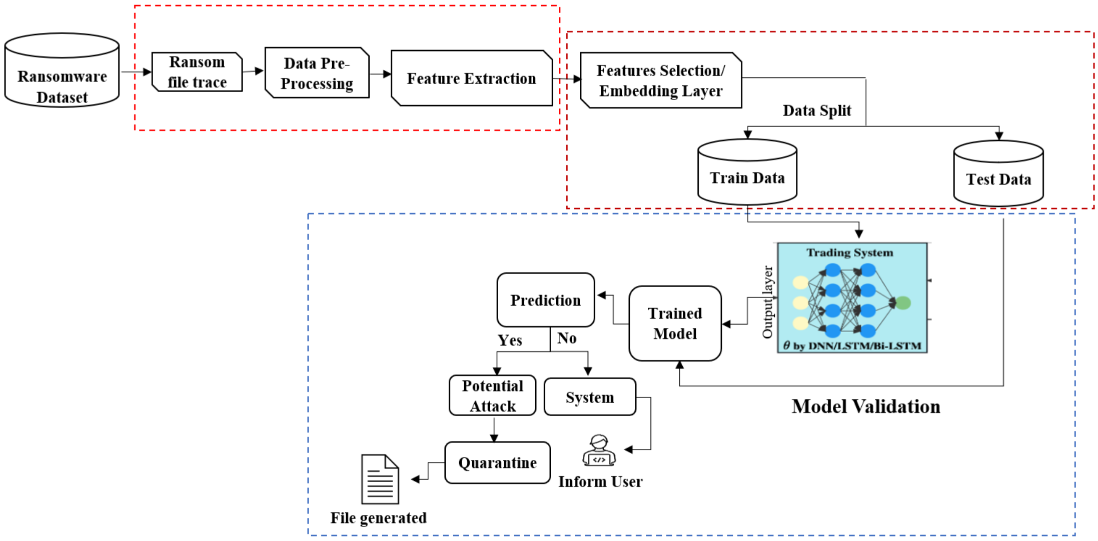


Fig. 9. Proposed Framework

2.4 Motivation of the Research

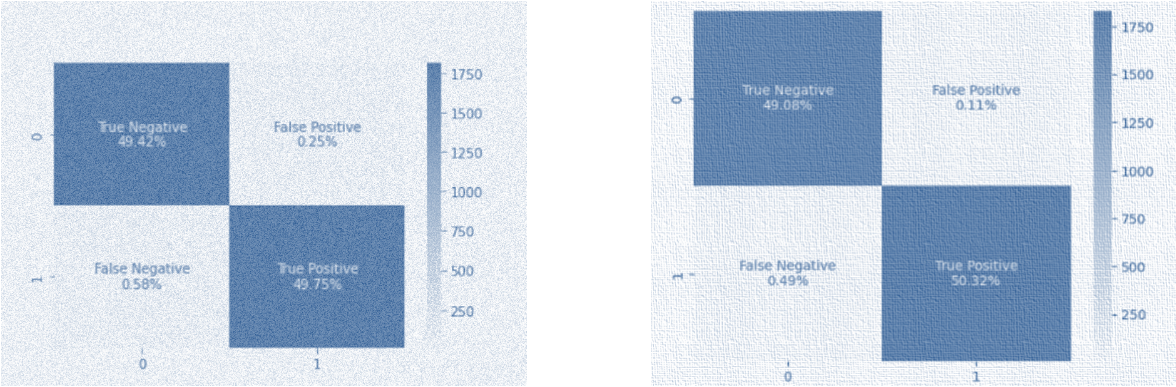
Cyber attackers are constantly changing their strategies in the current digital environment and creating new ways to exploit weaknesses. To prevent ransomware attacks, Deep Neural Networks (DNN) and Bidirectional Long Short-Term Memory (BiLSTM) networks are used as proactive measures to detect and identify previously unknown threats even before they have the chance to carry out encryption. This strategy's main goals are to identify unknown, or zero-day, assaults, increase detection rates, lessen the incidence of false positives and false negatives, and provide quick responses to these unexpected threats. Additionally, this cutting-edge solution provides defense against polymorphic malware, which alters and adapts to take advantage of conventional security mechanisms.

3. Results and Discussion

The modeling techniques, generating test design, building a model, and assessing the model. The process will begin with the generic information extraction from new ransomware and ransomware portable executable. The generic signature information extracted and transformed into a dataset. As a dataset will be split into a 70% training, 30% testing sets according to the standard practice by most researchers [4]. The training set and testing set fed into a supervised machine learning network for supervised training. More data partition is applied for the model training to allow an optimal and accurate classification model to be created [27]. The evaluation of the training based on the proposed evaluation matrices. Figure 10 (a) and (b) shows the flow for the modeling phase and evaluation phase. The model's effectiveness is evaluated using evaluation matrices. Experiments are conducted on 200,000 objects from a malware detection project, using machine learning-based clustering techniques like DNN with BiLSTM. The experiment was conducted on common public datasets: Ransomware pre-encryption detector (PERD) obtained from GitHub [29,30].

For processing the data 80% of the samples are distributed for training the model. Subsequently, both the DNN and Bi-LSTM architectures undergo training for one hundred epochs. Following the completion of training, the models are subjected to testing using the remaining 20% of samples. This test dataset is used for evaluating the efficacy of the deep learning models.

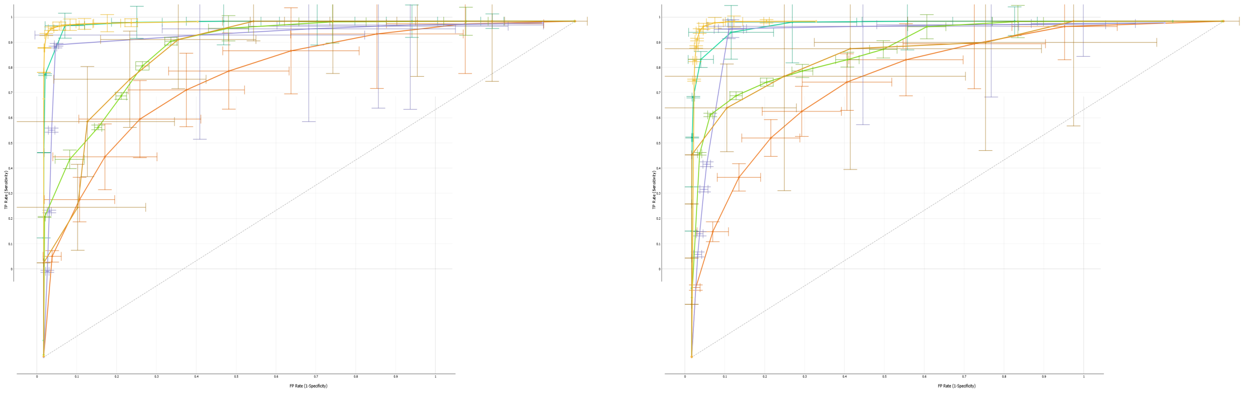
For evaluation purposes, a Confusion Matrix (CM) is employed. This matrix provides a detailed breakdown, expressed as percentages in this study, of True Positives (TP), True Negatives (TN), False Positives (FP), and False Negatives (FN).



(a) (b)

Fig. 10. (a) DNN Confusion Matrix for T=60s (b) Bi-LSTM Confusion Matrix for T=60

Both the Deep Neural Network (DNN) and the Bi-directional LSTM (Bi-LSTM) Confusion Matrix findings are provided. A striking pattern shows that the false negative rates for both DNN and Bi-LSTM drop with increasing time intervals. The Bi-LSTM's achievement of a remarkably low false negative rate, at only 0.58%, is particularly notable. Testing the deep learning models against Zero-Day ransomware variants is part of the assessment process. Through this test, the models' capacity to recognize ransomware variants that were not included in their training data is evaluated. Surprisingly, the performance of both DNN and Bi-LSTM for detection is strong. Bi-LSTM outperformed DNN with an accuracy of 98.9% at T=60s, whereas DNN only managed 99.10% accuracy as consider with [25]. Evaluation on different classifiers can be seen in Table 3. A higher area under the ROC curve generally indicates better model performance. Figure 11 (a) and (b) show the ROC analysis on false positive rates for both signature and benign over specificity. Whereas Figure 12 shows the performance metrics over t=60s, Traditional DNN correlation matrix given in Figure 13.



(a) (b)

Fig. 11. (a) False Positive Rate (Signature) (b) False Positive Rate (Benign)

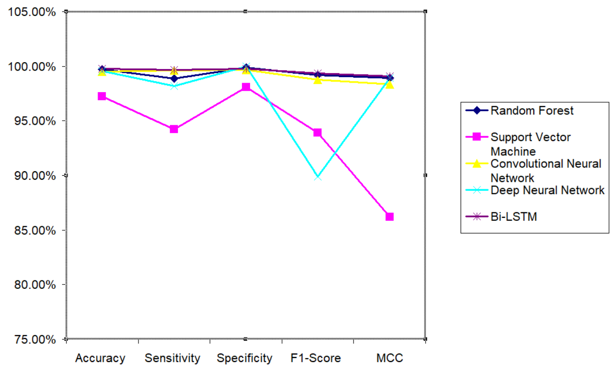


Fig. 12. Performance Metrics over T=60s

Table 3
Evaluation of different classifier

Model	Accuracy	Sensitivity	F1-Score	Specificity	MCC
Random Forest	0.997	0.984	0.984	0.984	0.965
SVM	0.852	0.773	0.776	0.782	0.53
Convolution Neural Network	0.961	0.959	0.958	0.959	0.911
DNN	0.989	0.949	0.95	0.952	0.896
Bi-LSTM	0.998	0.988	0.988	0.988	0.974
AdaBoost	0.997	0.988	0.988	0.988	0.974

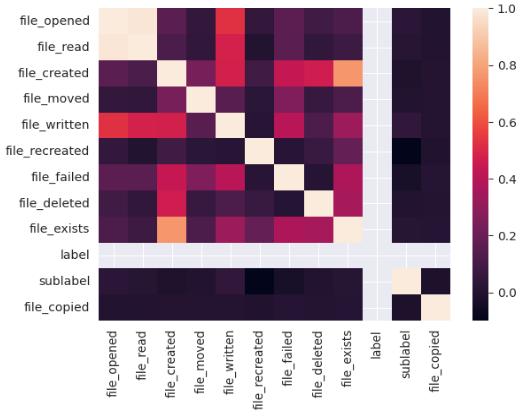


Fig. 13. Traditional DNN correlation matrix

4. Conclusion and Future work

We provide a robust ransomware flow detecting end-to-end trainable DNN-BiLSTM model. how DL techniques relate to malware detection and how the classifier's output may be affected by the dataset that will be selected. Additionally, a variety of datasets will gather, examine, trained on, and verified to better understand how results differ from existing circumstances with just a minor loss of accuracy over time. The model will produce better outcomes. The capacity to extract information about malware classes will be provided using a multi-layer technique, and by enhancing the model with more characteristics, our findings were improved.

Our method utilizes partial flow detection and layer-wise data categorization, boosting real-time capabilities. To increase accuracy and reduce false negatives, new ransomware behavior elements will be added in future updates. The importance of our work lies in early-stage ransomware detection using DNN with Bi-LSTM, with future intentions to expand detection using hybrid models throughout many phases. Our study shows potential for rapid ransomware recognition using DNN and Bi-LSTM models trained over incremental time intervals, given the high stakes for organizations with sensitive data.

Acknowledgements

This research work was supported by the International Collaborative Research Universiti Teknologi PETRONAS Malaysia, under cost centre (015ME0-326).

References

[1] Slayton, Thomas B. "Ransomware: the virus attacking the healthcare industry." *Journal of Legal Medicine* 38, no. 2 (2018): 287-311. https://doi.org/10.1080/01947648.2018.1473186

[2] Kuper, Peter. "The state of security." *IEEE Security & Privacy* 3, no. 5 (2005): 51-53. https://doi.org/10.1109/MSP.2005.134

[3] Akbanov, Maxat, Vassilios G. Vassilakis, and Michael D. Logothetis. "Ransomware detection and mitigation using software-defined networking: The case of WannaCry." *Computers & Electrical Engineering* 76 (2019): 111-121. https://doi.org/10.1016/j.compeleceng.2019.03.012

[4] Razaulla, Salwa, Claude Fachkha, Christine Markarian, Amjad Gawanmeh, Wathiq Mansoor, Benjamin CM Fung, and Chadi Assi. "The age of ransomware: A survey on the evolution, taxonomy, and research directions." *IEEE Access* (2023). https://doi.org/10.1109/ACCESS.2023.3268535

[5] European Union Agency for Cybersecurity, *ENISA threat landscape report 2018 – 15 top cyber-threats and trends*, European Network and Information Security Agency (2019). https://data.europa.eu/doi/10.2824/622757

[6] Dargahi, Tooska, Ali Dehghantanha, Pooneh Nikkhah Bahrami, Mauro Conti, Giuseppe Bianchi, and Loris Benedetto. "A Cyber-Kill-Chain based taxonomy of crypto-ransomware features." *Journal of Computer Virology and Hacking Techniques* 15 (2019): 277-305. https://doi.org/10.1007/s11416-019-00338-7

[7] Scaife, Nolen, Henry Carter, Patrick Traynor, and Kevin RB Butler. "Cryptolock (and drop it): stopping ransomware attacks on user data." In *2016 IEEE 36th international conference on distributed computing systems (ICDCS)*, pp. 303-312. IEEE, 2016. https://doi.org/10.1109/ICDCS.2016.46

[8] Djenna, Amir, Ahmed Bouridane, Saddaf Rubab, and Ibrahim Moussa Marou. "Artificial intelligence-based malware detection, analysis, and mitigation." *Symmetry* 15, no. 3 (2023): 677. https://doi.org/10.3390/sym15030677

[9] Ahsan, Mostofa, Kendall E. Nygard, Rahul Gomes, Md Minhaz Chowdhury, Nafiz Rifat, and Jayden F. Connolly. "Cybersecurity threats and their mitigation approaches using Machine Learning—A Review." *Journal of Cybersecurity and Privacy* 2, no. 3 (2022): 527-555. https://doi.org/10.3390/jcp2030027

[10] M. Hamad and D. Eleyan, "Survey On Ransomware Evolution, Prevention, And Mitigation," *Lume*, vol. 10, no. October, p. 2 (2021) [Online]. Available: http://www.ijstr.org

[11] Begovic, Kenan, Abdulaziz Al-Ali, and Qutaibah Malluhi. "Cryptographic ransomware encryption detection: Survey." *Computers & Security* 132 (2023): 103349. https://doi.org/10.1016/j.cose.2023.103349

[12] Li, Bo, Kevin Roundy, Chris Gates, and Yevgeniy Vorobeychik. "Large-scale identification of malicious singleton files." In *Proceedings of the seventh ACM on conference on data and application security and privacy*, pp. 227-238. 2017. https://doi.org/10.1145/3029806.3029815

[13] Heena, "Advances In Malware Detection- An Overview," (2021) [Online]. Available: http://arxiv.org/abs/2104.01835

[14] Popoola, Segun I., Ujioghosa B. Iyekekpolo, Samuel O. Ojewande, Faith O. Sweetwilliams, Samuel N. John, and Aderemi A. Atayero. "Ransomware: Current trend, challenges, and research directions." In *Proceedings of the World Congress on Engineering and Computer Science*, vol. 1, pp. 169-174. 2017.

[15] Urooj, Umara, Mohd Aizaini Bin Maarof, and Bander Ali Saleh Al-rimy. "A proposed adaptive pre-encryption crypto-ransomware early detection model." In *2021 3rd International Cyber Resilience Conference (CRC)*, pp. 1-6. IEEE, 2021. https://doi.org/10.1109/CRC50527.2021.9392548

[16] Al-Rimy, Bander Ali Saleh, Mohd Aizaini Maarof, and Syed Zainudeen Mohd Shaid. "Ransomware threat success factors, taxonomy, and countermeasures: A survey and research directions." *Computers & Security* 74 (2018): 144-166. https://doi.org/10.1016/j.cose.2018.01.001

[17] Malik, K. A. R. T. I. K., M. A. N. I. S. H. Kumar, M. Sony, R. A. D. H. A. Mukhraiya, P. A. L. A. K. Girdhar, and B. H. A. R. T. I. Sharma. "Static Malware Detection And Analysis Using Machine Learning Methods." *Advances and Applications in Mathematical Sciences* 21, no. 7 (2022): 4183-4196.

[18] Usharani, S., P. Manju Bala, and M. Martina Jose Mary. "Dynamic analysis on crypto-ransomware by using

machine learning: Gandcrab ransomware." In *Journal of Physics: Conference Series*, vol. 1717, no. 1, p. 012024. IOP Publishing, 2021. https://doi.org/10.1088/1742-6596/1717/1/012024

[19] Davies, Simon R., Richard Macfarlane, and William J. Buchanan. "Evaluation of live forensic techniques in ransomware attack mitigation." *Forensic Science International: Digital Investigation* 33 (2020): 300979. https://doi.org/10.1016/j.fsidi.2020.300979

[20] Yamany, Bahaa, Mahmoud Said Elsayed, Anca D. Jurcut, Nashwa Abdelbaki, and Marianne A. Azer. "A new scheme for ransomware classification and clustering using static features." *Electronics* 11, no. 20 (2022): 3307. https://doi.org/10.3390/electronics11203307

[21] Raff, Edward, Richard Zak, Russell Cox, Jared Sylvester, Paul Yacci, Rebecca Ward, Anna Tracy, Mark McLean, and Charles Nicholas. "An investigation of byte n-gram features for malware classification." *Journal of Computer Virology and Hacking Techniques* 14 (2018): 1-20. https://doi.org/10.1007/s11416-016-0283-1

[22] Bhardwaj, Akashdeep. "Ransomware: A rising threat of new age digital extortion." In *Online banking security measures and data protection*, pp. 189-221. IGI Global, 2017. https://doi.org/10.4018/978-1-5225-0864-9.ch012

[23] Beaman, Craig, Ashley Barkworth, Toluwalope David Akande, Saqib Hakak, and Muhammad Khurram Khan. "Ransomware: Recent advances, analysis, challenges and future research directions." *Computers & security* 111 (2021): 102490. https://doi.org/10.1016/j.cose.2021.102490

[24] Montavon, Grégoire, Wojciech Samek, and Klaus-Robert Müller. "Methods for interpreting and understanding deep neural networks." *Digital signal processing* 73 (2018): 1-15. https://doi.org/10.1016/j.dsp.2017.10.011

[25] Jemal, Muna. "Detection of Crypto-Ransomware Attack Using Deep Learning." (2023). https://doi.org/10.1109/DSC61021.2023.10354186

[26] Molina, Ricardo Misael Ayala, Sadegh Torabi, Khaled Sarieddine, Elias Bou-Harb, Nizar Bouguila, and Chadi Assi. "On ransomware family attribution using pre-attack paranoia activities." *IEEE Transactions on Network and Service Management* 19, no. 1 (2021): 19-36. https://doi.org/10.1109/TNSM.2021.3112056

[27] Davies, Simon R., Richard Macfarlane, and William J. Buchanan. "Differential area analysis for ransomware attack detection within mixed file datasets." *Computers & Security* 108 (2021): 102377. https://doi.org/10.1016/j.cose.2021.102377

[28] Panhwar, Ali Orangzeb, Anwar Ali Sathio, Abdullah Lakhan, Muhammad Umer, Rabia Mushtaque Mithiani, and Sanwali Khan. "Plant health detection enabled CNN scheme in IoT network." *International Journal of Computing and Digital Systems* 11, no. 1 (2022): 344-335. https://journals.uob.edu.bh/handle/123456789/4604

[29] Rehman, Mujeeb ur, Rehan Akbar, Mazni Omar, and Abdul Rehman Gilal. "A Systematic Literature Review of Ransomware Detection Methods and Tools for Mitigating Potential Attacks." In *International Conference on Computing and Informatics*, pp. 80-95. Singapore: Springer Nature Singapore, 2023. https://link.springer.com/chapter/10.1007/978-981-99-9589-9_7

[30] M. U. Rehman Shaikh, R. Ullah, R. Akbar, K. S. Savita, and S. Mandala, "Fortifying Against Ransomware: Navigating Cybersecurity Risk Management with a Focus on Ransomware Insurance Strategies," *Int. J. Acad. Res. Bus. Soc. Sci.*, vol. 14, no. 1, pp. 1415–1430 (2024). https://doi.org/10.6007/ijarbss/v14-i1/20566