# الكشف عن برامج الفدية باستخدام التحليل الديناميكي ونماذج MLP وLSTM

# Ransomware detection using Dynamic analysis and LSTM, MLP models

خطة بحث مقدمة لاستكمال متطلبات الحصول على درجة الماجستير في برنامج (تقنية المعلومات/الأمن السيبراني)
في قسم تقنية المعلومات بكلية الحاسب

إعداد

| | |
|---|---|
| إيناس الصحفي | 431214119 |
| مدى الباني | 431214117 |
| اعتدال الرويلي | 431214120 |
| مجد الحريري | 431114118 |
| سليمان الدبيخي | 431114111 |

إشراف

**أسماء الشرجبي**

أستاذ مساعد - قسم تقنية المعلومات

الفصل الدراسي الأول

1444 هـ 2022-2023 م

# Ransomware detection using Dynamic analysis and LSTM, MLP models

## Abstract

The rapid development in the field of information technology, communications, and internet networks has led to the misuse of technology, as systems are designed mostly in a secure manner. On the other hand, attackers design malicious programs to penetrate these systems, and among these programs is what is known as ransomware, which poses a great danger to individuals as well as companies, organizations, etc. The purpose of using ransomware is to obtain money from the victim. Cybersecurity contains a wide range of tools and techniques to detect or deter these attacks. Researchers recommended the use of deep learning techniques due to their high accuracy in predicting malware. The research aims to improve the detection of ransomware by using dynamic analysis and deep learning techniques. To achieve this goal. First, the ransomware files were collected from all datasets used in previous studies which will be arranged, organized, and standardized using Python language.  Then, will Classification of the dataset using dynamic analysis to extract dataset files feature by using Cuckoo Sandbox. Finally, the outputs of dynamic analysis will be training and testing on the Long Short-Term Memory (LSTM) and Multi-Layer Perceptron (MLP) models, to get the highest accuracy in ransomware detection.

**Keywords:** Cybersecurity; Dynamic Analysis; Deep Learning; Malware Detection; Ransomware

## 1. Introduction

With the recent developments that our societies are witnessing, the reliance on computer systems and networks in the current era has become mainly compared to previous years. Systems are mostly designed with a completely secure system from attacks, but despite the comprehensive security of the systems, there are developments in the attack techniques used by the attackers. A cybersecurity contains a wide range of anti-malware tools and techniques to detect ever-increasing malware over a computer system from all attacks and unauthorized access for destruction [1]. Malware contains numerous attacks that are installed to the system with or without the user's knowledge for a specific benefit for the attacker. Some types of malwares include viruses, worms, trojan horses and ransomware [2].

Malware can infect computers and systems in various ways such as tricking the user into opening malicious files or lying to users out of temptation to visit malware sites. Sometimes the malicious program can be downloaded to the CD or the USB, which is inserted into the device, infecting the entire device and may even disable it. One of the most popular and difficult malwares is ransomware, it became ahead of machine learning and deep learning in time, spread, and development, especially with artificial intelligence. Ransomware is a type of malware that can infect devices, spread quickly, and encrypts or disables data.  Its spread and rapid development is a source of great fear in today's internet environment. Studies are increasing on establishing and developing detection techniques. In the related works section, several research works were mentioned that were conducted to detect malware in general on different systems.

## 2. Literature review

### 2.1 Background

### 2.1.1 Malware

In paper [1], cybersecurity is defined as" a set of technologies and processes designed to protect computers, networks, programs, and data from attacks and unauthorized access, alteration, or destruction".  Many cybersecurity scholars believe that malware is the main weapon for carrying out cybersecurity breaches in cyberspace. Malware refers to a large group of attacks that may infect a system by either deceiving users or installing malicious software to the system such as ransomware [2].

### 2.1.1.1 Malware classification

Malware detection and classification fall into a set of lengthy processes through which different techniques are used. Malware is detected with basic steps [3]:

1- Malware analysis using relevant tools.

2- Record the result of the tools and determine the static and dynamic features extracted from them.

3- Train the features identified by machine learning algorithms to categorize and separate harmful and benign programs.

The figure 1 shows the processes for analyzing, detecting, and classifying malware. Malware detection and classification techniques are divided into five groups that include [3]: signature, behavior, guidance, check the form, and deep learning.

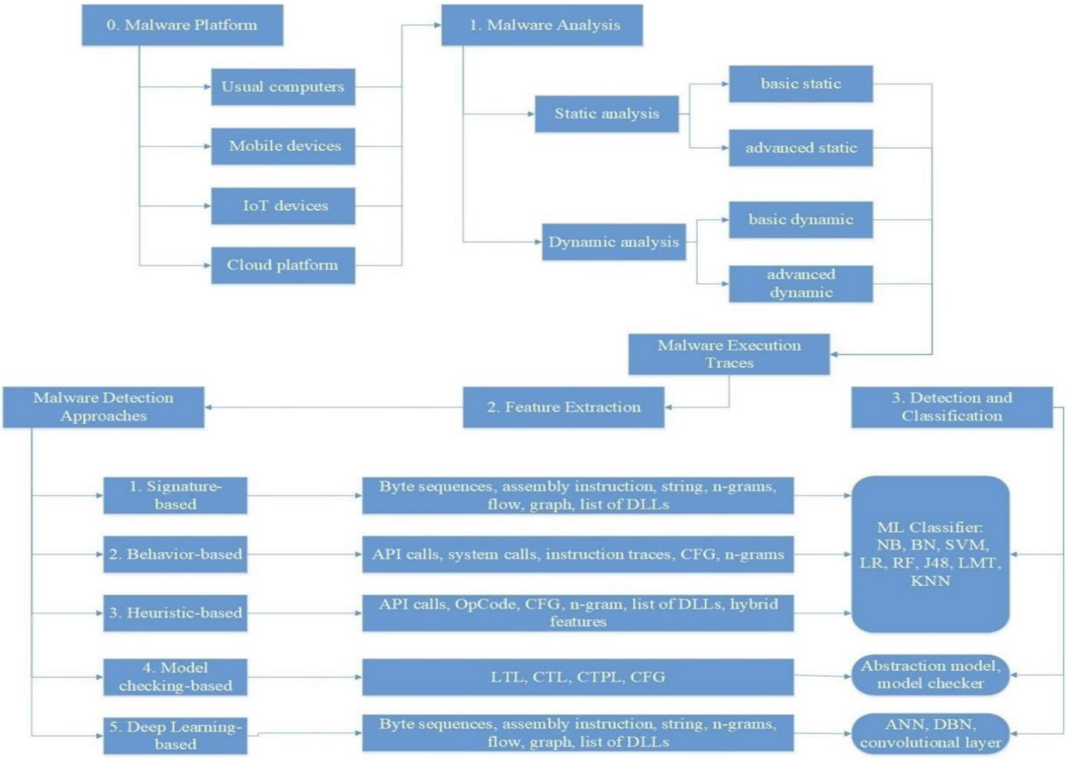

0. Malware Platform
Usual computers
Mobile devices
IoT devices
Cloud platform
1. Malware Analysis
Static analysis
basic static
advanced static
Dynamic analysis
basic dynamic
advanced dynamic
Malware Execution Traces
2. Feature Extraction
Malware Detection Approaches
3. Detection and Classification
1. Signature-based
2. Behavior-based
3. Heuristic-based
4. Model checking-based
5. Deep Learning-based
Byte sequences, assembly instruction, string, n-grams, flow, graph, list of DLLs
API calls, system calls, instruction traces, CFG, n-grams
API calls, OpCode, CFG, n-gram, list of DLLs, hybrid features
LTL, CTL, CTPL, CFG
Byte sequences, assembly instruction, string, n-grams, flow, graph, list of DLLs
ML Classifier: NB, BN, SVM, LR, RF, J48, LMT, KNN
Abstraction model, model checker
ANN, DBN, convolutional layer

*Figure 1 Malware analysis, detection, and classification processes [3].*

### 2.1.2 Ransomware

Ransomware is part of the malware that infects the user's system.  Since 2010 till now, ransomware is malware, and it is very prevalent [4]. It spreads quickly and widely and encrypts user data and makes it inaccessible. Usually, attackers do this for the purpose of extortion to publicly release sensitive information if payment is not made. The attacker demands a ransom payment from the victim to decrypt and access their files. Payment is made using Bitcoin or other untraceable currencies. On the other hand, ransomware locks users' computers using various mechanisms. Getting money is the main purpose of a ransomware attack [5]. There are three ways for ransomware to spread to a device:

- Clicking on an untrusted website.
- Opening pre-injected files.
- Clicking on a malicious link sent to your email.

Ransomware is used to attack a variety of systems, including computers of all kinds, Internet of Things devices, as well as smartphones with various systems and applications. In the past, only specialists and experienced people could use ransomware in attacks, while now those who do not have experience can buy ransomware from the attackers" This practice is known as Ransomware-as-a-Service (RaaS) ", It allows those who are inexperienced to attack using ransomware, The first ransomware attack began in 1989 by the doctor and researcher in AIDS diseases Joseph Pope, the first ransomware program he named AIDS or (PC cyborg), where Joseph gave many floppy disks containing the AIDS program to researchers in the field of his research [6].

In study [7], it was shown that there are four types of ransomwares, which are:

Crypto-malware: this type is one of the most common types due to the fact that it encrypts files and the inability of the user to access them.

Locker: this type completely disables the system, thus preventing the user from accessing software applications or files on the device

Scareware: this type of fraud works as a tool to clean the device from malware, and in fact it is a malicious tool that often locks the device or shows annoying ads on the device screen

Doxware: finally, this type blackmails victim's users into publishing their sensitive information online if they do not pay them.

Ransomware is often analyzed using two techniques: Static analysis is the analysis that is done without the need to run or execute the program. Dynamic analysis technique is done by implementing and understanding the behavior of the program. Dynamic analysis is used to detect malware in many ways, but because there are many paths to execution, not all paths may be covered and may stop working if any of the analysis tools are detected. Some of the important techniques include machine learning, deep learning, data flow analysis and API call binding [8].

#### 2.1.2.1 Ransomware classification

The study [9] was presented competing surveys that include the following:

- A study brief review about ransomware classification, types, enablers, infection vectors, platforms targeted, and life cycle of ransomware attack.
- A study suggested some criteria to avoid ransomware attacks.
- A study that provided a summary of ransomware detection techniques. Various techniques for detecting ransomware and their pros and cons, recent research from 2015 to 2018, and the type of analysis used to detect ransomware are explained. This study presented the approved methodologies for detection, the results generated, and the limitations of each study.
- A study that discussed the approaches and methods of deep learning. Another study discussed the use of deep learning technology in the field of cybersecurity. A brief review on deep learning solutions is described. The questionnaire is illustrated by discussing malware and Android malware detection studies, binary data and traffic analysis, spam and phishing detection, and intrusion detection.

### 2.1.3 Ransomware analysis

Malware analysis in general and specifically the analysis of ransomware is a very important step because it is through it that samples are identified to know the nature of its behavior and the ability to classify it [10].

There are two main types of analysis, static analysis, and dynamic analysis. The analysis is used manually or automatically and often starts with static fundamental analysis and ends with advanced dynamic analysis [11].

#### 2.1.3.1 Static analysis

By static analysis a sample of malware is identified without actually running the code [12]. Static analysis is divided into two main types:

1- Basic static analysis: In this type, malware scans are done in a simple way without looking into the details [10]. Features of this kind are exploited using several tools including PEiD, BinText, MD5deep and PEview [11]. Fundamental static analysis is the first stage in the analysis process.

2- Advanced static analysis: It is required to work on this type to gain more knowledge about malware. In this type of analysis, the code is examined in detail for malware behavior and characteristics using a disassembler to generate assembly codes from machine codes [13], [14].

IDA Pro packet splitter and Hex-Rays translation remover are also widely used for advanced static analysis. Advanced Static Analysis focuses deeply on the characteristics of malware.

#### 2.1.3.2 Dynamic analysis

It is through dynamic analysis that program code is run and executed to identify and evaluate malware behaviors. It is conducted in a closed environment such as using virtual machines. Dynamic analysis is more accurate than static analysis due to the presence of real malware functions. Dynamic analysis is divided into two main parts [3]:

1- Basic Dynamic Analysis: Basic dynamic analysis works to identify malware behaviors using monitoring tools such as Process Monitor, API Monitor, Process Explorer, Regshot, ApateDNS, Wireshark, and Sandboxes [14].

2- Advanced Dynamic Analysis: It uses debugging tools such as OllyDbg and WinDbg to execute each command separately [10].

### 2.1.4 Machine learning

Most cybersecurity companies rely on machine learning to detect malware to find a way because many current cybersecurity applications may not be able to detect new malware activities because they do not recognize this data, so machine learning is a suitable technique. It was found that most recent studies use machine learning and deep learning techniques to analyze and detect malware [15].

In this paper [16], machine learning (ML) is defined as"a branch of AI and is closely related to and often overlaps with computational statistics, which also focuses on prediction-making using computers".

### 2.1.4.1 Deep learning

Deep learning (DL) is a branch of artificial intelligence and is derived from artificial neural networks. Many machine learning techniques have been developed over the past decades by creating algorithms that can learn and improve [7]. DL models " require a large amount of data for each problem domain to construct a data-driven model", DL algorithms "require high computational capabilities to train models with a large amount of data "and One of the advantages of deep learning is that it provides results with high accuracy within a short period [6]. In 2012, researchers showed that deep learning can understand and recognize objects in images just as the human brain can understand [17]. There is a lack of academic work using deep learning to detect and categorize malware so far [3].

## 2.2 Related work

This section presents relevant studies about ransomware attacks against different systems for types of devices with various versions used by people in all fields. Some studies show the experiences conducted by researchers and the accuracy of their results in detecting these attacks and the proposed solutions.

Study researchers in [18] proposed a supervised model for malware detection on Mac OS X using machine learning based on the Radial Base Function (RBF) in the SVM technique. It has a detection accuracy of over 91% and a false alarm rate of 3.9%. On the other hand, the dataset was used based on SMOTE, and it was noted that the detection rate increased by 96.62% and the error rate decreased by less than 4%. They collected 152 datasets from malware that came from [19-20] between Jan 2012 and June 2016.

The researchers in the study [21] proposed the De-LADY deep learning model based on the features of dynamic analysis to detect the malware on android applications. They used 13533 of which 2821 are benign and 10712 are malignant from different sources [22, 23, 24, 25]. They achieved 98.08% and an error rate of 1.92%.

Paper [26] suggested the Mac-A-Mal open-source framework as a kernel extension aiming to monitor malware behavior and attenuate anti-evasion techniques by using static and dynamic analysis. Test framework in three dataset consists of 2,000 macOSmalware collected between January 2017 and July 2017 from VirusTotal Malware Intelligence Services (VTMIS), Macmalware, and Objective-see. The result come up with an analysis of two undetected malware campaigns:

1- OSX/Mughthesec adware campaign: discovered 71 signed archives and 10 Apple legitimate developer certificates.

2- APT32-OceanLotus campaign: catch the second generation of Mac APT32.

paper [27] proposes a framework for detecting and preventing the crypto-ransomware attack in windows 7. That collects user data such as IP addresses, file size, and URLs. Then dynamically analyze current data with previous data by using a deep learning machine. The Dionaea Honeypot is used to get the important data by extracting the network data's properties. Then classification of data into ransomware, benign, and malware. It contains 40,000 ransomware samples and 25,000 samples of benign files. The classification accuracy rate using a machine learning algorithm is 98.45%, with a false-positive rate of 0.01.

In the paper [28], the authors presented empirical research for malware disclosure in Android, where they used a new approach known as DeepAMD to revelation Android malware in both static and dynamic layers by using a deep Artificial Neural Network (ANN), the authors used a dataset from [29, 30] The authors concluded that the accuracy of the technique DeepAMD used is very high, as it achieved results in the fixed layer with a top accuracy of 93.4% for classification of malware and 93.1% accuracy for classification of malware families and its results in the Dynamic layer with a top accuracy of 80.3% for classification of malware class and an accuracy of 59 0% to classify the malware family, In addition, the researchers concluded that DeepAMD is the most effective in detecting malware in Android.

Researchers in the paper [31] proposed a new method to analyze malware dynamically by using hardware events during file execution for the classification model. First, establish the machine learning model by algorithms Convolutional Neural Network (CNN) and Long Short-Term Memory (LSTM). Then, The label of the suspicious sample is then determined by using the voting network between the CNN and LSTM network outputs. The dataset contains 4000 benign and malware files. benign files get from free PortableApps [32] and SnapFiles [33] software sources and malware files taken from the VirusShare [34]. Method accuracy rate 91.63%, recall 91.63%, Precision 91.79% and F-measure 91.61 %.

The paper [35], used a two-stage ransomware detection method, dynamic analysis, and deep learning. First focuses on the Windows API call sequence pattern and builds a Markov model to capture the characteristics of ransomware. This system gives good false positive error (FPR) rate results but poor false negative error (FNR) rates, around 20%. Next, they tried the different machine learning-based classification models to build a Random Forest machine learning model to the remaining data to control both FPR and FNR rates. That gives a good outcome. The datasets used for malware get from and benign data from http://en.softonic.com. Finally, two-stage mixed detection methods can achieve overall accuracy of 97.3% with 4.8% FPR and 1.5% FNR.

The paper [36], the researchers proposed a model based on deep learning to find ransomware targeting smartphones in the Android system. This research measures the performance of the long-term memory (LSTM) algorithm and applies it to a data set of (1509550) from the source [37], Of which the number of good data (1048,514) and malicious data (460976) were tested in a real smartphone .as researchers are the first to apply this algorithm to this Datasets of [37] As a result, the research arrives to classify ransomware With 97.08% accuracy, It turns out that the long-term memory(LSTM) algorithm is very accurate in detecting malicious data in the Android operating system. However, there is a need to apply this algorithm with many deep learning frameworks and compare them to obtain the maximum degree of security.

The paper [38], researchers used a deep learning method to detect and evaluate malware, specifically ransomware, and maintain privacy. The researchers experimented on a large number of ransomware data in a Windows environment, numbering approximately more than 26,300 experimental samples from source [39], which were performed in the Max Pooling (LaMP)model with (LSTM) and (ARL-LSTM) cells in the framework of deep learning, where long-term memory (LSTM )is defined as "A special type of recurrent neural network that can learn long-term dependencies in data because the iteration unit in the model contains a combination of four interaction layers with each other " and Attended Recent Inputs(ARL) has been defined as "providing additional input information in several ways in sequential processing." The researchers Combined (LSTM)cell with the Attended Recent Inputs (ARL) cell to obtain high-accuracy results in ransomware detection. Among the researchers' findings, the (ARL-LSTM) cells are more accurate in detecting ransomware than the (LSTM) cell, as the accuracy rate they reached from (ARL-LSTM) cell is 91% for detecting ransomware and the rate they reached from the (LSTM) cell is 87%. The downside is that the researchers did not mention the type of Windows operating system on which their experiments were conducted.

In this paper [40], the author compared several studies that detect ransomware using machine learning and deep learning. He conducted scientific experiments for these studies trying to discover gaps and challenges and suggested some beneficial solutions, in machine learning author used (Regularised Logistic Regression, Gradient Tree Boosting, Support Vector Machines, J48 Decision Tree, Random Forests, MLP, KNN, LMT, Bayesian networks, AdaBoost, Entropy, Digital DNA Sequencing, GAN, CF-NCF, Markov Chains, HML (Naive Bayes and Decision Trees)) , and in deep learning author used (DNP, LSTM, ANN). The approaches reviewed show high detection rates in the mid to high 90s up to 96.4% as MLP 10 hidden layers showed. Using EldeRan dataset that covers ransomware from 2013 to 2015 and contains 591 ransomware samples and 942 benign samples to training each one of these models, then the author tested it on the new ransomware from 2016 to 2017. The author of this paper used an old dataset of ransomware but using a recent and largest dataset could show better results.

However, this considers a short-term solution for detecting ransomware because of the new ransomware techniques that will develop in the future.

In this study [41], ransomware is detected and classified using Multi-Layer Perceptron (MLP), which is a type of Feed-Forward Networks (FFN). Ransomware is classified by the MLP by the MLP model, which also specifies the family of the ransomware and to where they belong. They used API calls from Cuckoo Sandbox as a dataset. Every experiment is performed with 500 epochs and a learning rate between (0.01 to 0.5). MLP model asserts that it can classify ransomware into its various families with 98% accuracy and can recognize it as benign with 100% accuracy, and Three or more layers will be used in this network model, the first three layers in this model will contain at least one layer as an input layer, one as a hidden layer and one as an output layer, the complexity of the data will determine how many hidden layers it must be. It does not appear that the feature set utilizes any feature reduction approaches to minimize the feature count.

Table 1 summarizes previous studies on malware detection systems, specifically ransomware. The table specifies the model used to detect it, the type of operating system, a short summary, and the limitations of the study.

Table 1: Previous studies summary

| Paper | Year | Model | OS | Dataset | Description | Limitations |
|---|---|---|---|---|---|---|
| [18] | 2021 | De-LADY deep learning | Android | De-LADYs Dataset | - Detection on Android apps.<br>- Detection accuracy is 98.08%. | - The data was not clearly analyzed.<br>- The year in which the dataset was collected was not specified<br>- The false alarm ratewas not specified. |
| [21] | 2021 | Dynamic analysis on crypto ransomware using deep learning | Windows | Dionaea Honeypot | - Classification and prevention ransomware<br>- Detecting accuracy is 98.45% with false-positive rate 0.01 | - Work in windows 7. |
| [26] | 2021 | DeepAMD-using a deep Artificial Neural Network | Android | CIC, Mal2017 and CICnves, Mal2019 | - Revelation Android malware in both static and dynamic layers.<br>- Detection accuracy of over 93.04%. | - There is no way to check the harm of the program before downloading it. |
| [27] | 2021 | Method uses hardware event for dynamic analysis with deep neural networks | Windows | PortableApps, SnapFiles and VirusShare. | - Use CNN and LSTM to create machine learning model<br>- Voting networks get data from CNN and LSTM to determine if it's a malware file.<br>- Method accuracy rate 91.63% | - for windows OS<br>- Low accuracy rate Compared to other solution<br>- Focus only on hardware feature |

| | | | | | | |
|---|---|---|---|---|---|---|
| [28] | 2020 | Two-stage ransomware detection method | Windows | VirusShare and softonic | - Detection method uses dynamic analysis, and deep learning.<br>- Detection accuracy of 97.3% with 4.8% FPR and 1.5% FNR | - normalware can avoid the system |
| [31] | 2020 | Ransomware Detection Using (Logistic Regression-GTB-Random Forests-SVM-J48-Deep Neural Network-MLP10-MLP20-Bayesian Networks) | Windows | EldeRan | - (MLP10) achieved the highest detection rate 96.4%, While (SVM) got the lowest rate 33.1%. | - Unresent dataset. |
| [35] | 2019 | Mac-A-Mal framework as a kernel extension | Mac | Mac-A-Mal | - Analysis malware on macOS.<br>- Detect 71 unknown adware samples. | - Framework detects the malware but not prevention.<br>- unable to include all anti-analysis methods. |
| [36] | 2019 | Ransomware Detection By using long-term memory (LSTM)algorithm | Android | CIC, Mal2017 | - Detection accuracy of over 97.08% | - The researchers did not mention the difficulties they faced. |
| [38] | 2019 | Ransomware Detection By (ARL-LSTM) | Windows | ARI sequence | - (ARL-LSTM) cells are 91% for detecting ransomware While (LSTM)cells are 87%. | - The (LSTM) cells are less accurate than the (ARL-LSTM) cells.<br>- The downside did not mention any type of Windows OS. |
| [40] | 2018 | Radial Base Function (RBF) | Mac | MALWAREHASHES | - Detection on Mac OS X.<br>- Detection accuracy of over 91% False alarm rate of 3.9%. | - The challenges they faced are not mentioned.<br>- The malware classification is complex and unclear. |
| [41] | 2017 | Ransomware Detection and Classification Using (Shallow and deep network) | Windows | Cuckoo Sandbox | - (MLP) 98% accuracy to classify between the ransomware families and classify it as benign with 100% accuracy. | - Did not mentioned how many benign or ransomware samples are used. |

Table 2: Datasets used in previous study

| Dataset | Size | Date | Features | model | Accuracy |
|---|---|---|---|---|---|
| MALWAREHASHES [18] | • 10,712 malicious<br>• 2821 benign | 2021 | contains four categories: banking, gaming, utility, and media player applications | RBF | 98.08% |
| De-LADY [21] | • 40,000 ransomwares<br>• 25,000 benign | 2021 | network features extracted from random forests. features of URL, IP address, File extension, web page. | De-LADY | 98.45% |
| Mac-A-Mal [26] | • 4,354 malwares<br>• 6,500 benign | 2017 | network traffic features (.pcap files). | Cuckoo sandbox | work [17]: In the fixed layer 93.4%, In the Dynamic layer 80.3%<br>work [27]: 97.08% |
| Dionaea Honeypot [27] | • 2000 malware<br>• 2000 benign | 2019 | New and up-to-date dataset. Contains various types of malwares such as Virus, Trojan, Backdoor, Rootkit and Hacking tools. | SVM,RF,GBTA | LSTM 90.60% accuracy. |
| CIC,Mal2017 [29] | • 2,507 ransomwares<br>• 3,886 benign | 2020 | API features & system extension file as dynamic features. | ANN | 97.3% |
| CICnves,Mal2019 [30] | • 582 Ransomware<br>• 942 Benign | 2016 | Static & Dynamic features | CNN and LSTM | 96.4% |
| PortableApps , SnapFiles and VirusShare. [31] | • 2,000 macOS malware<br>• Unknown benign | 2017 | Some of the samples use anti analysis techniques. Many of them execute network operations and write data to the victim machine. | LSTM | 71 signed archives and 10 valid Apple developer certificates were found |
| VirusShare and softonic [35] | • 426 malwares<br>• 5,065 benign | 2019 | Permissions and intents as static features.<br>API calls and all generated log files as dynamic features. | Random Forest | In the fixed layer 93.4%<br>In the Dynamic layer 80.3% |
| ARI sequence [39] | • 26,300 Attended Recent Inputs sequence<br>• | 2015 | API features | ARL-LSTM | 87% |
| EldeRan [40] | • 200,481 malware hashes (MD5)<br>• Unknown benign | 2017 | Not mention | MLP10 | over 91% |
| Cuckoo Sandbox [41] | • 755 Ransomware<br>• 219 Benign | 2017 | API features | MLP | 100% |

## Discussion

- Most of the previous studies focused on the detection of malware in general on the Windows and Android operating systems.
- There are five ransomware detection studies [27], [35],[39], [40], and [41]. These studies did not use sufficient data sets to train and test detection models, as shown in Table 2, which sparked our interest to apply the deep learning algorithms (LSTM) algorithm that used in the study[39]and the (MLP) algorithm that used in studies [40],[41] and its application to a larger data set based on was mentioned in a study [42] the deep learning models require large amounts of data to produce a well-generalized model. While studies' algorithms [27], [35] detect ransomware, while its algorithms are for machine learning.

## 3. Problem definition

It is apparent that the size of the dataset in several previous studies on ransomware detection was insufficient for applying a deep learning detection model, which may have an impact on the efficiency and accuracy of the findings, On the other hand, due to the widespread of ransomware nowadays, the efficiency of machine learning techniques for detecting ransomware has been affected. As a result, deep learning methods with larger data sets are required.

## 4. Aim

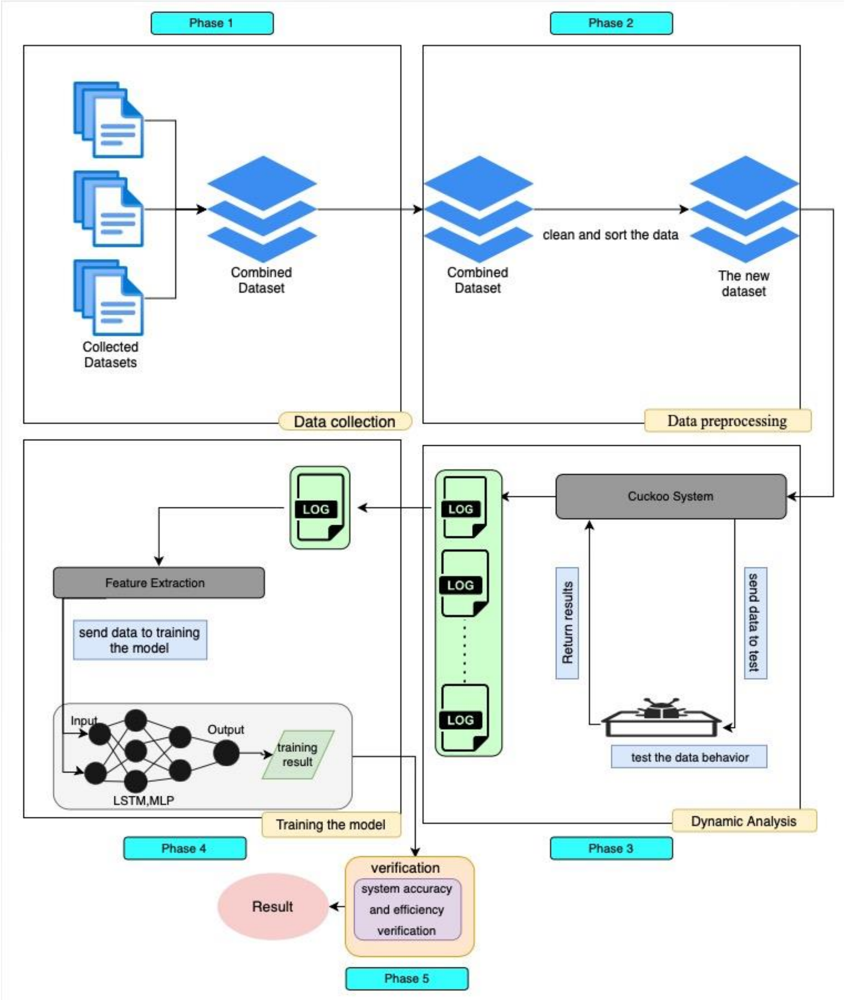
The research aims to improve ransomware detection accuracy and efficiency using dynamic analysis and deep learning models (LSTM, MLP).

## 5. Objectives

1. To build a model to analyze and classify the ransomware datasets using dynamic analysis and deep learning models (LSTM, MLP).
2. To create a large dataset collected from the previous studies datasets.
3. To improve the deep learning model efficiency and accuracy.
4. To evaluate the detection of ransomware.

## 6. Methodology

This research is qualitative study using constructive research methods aiming to improve ransomware detection accuracy and efficiency using dynamic analysis and a deep learning model. It consists of five phases, shown in figure 2.



*Figurer 2 :Methodology*

These phases are explained as follows:

- **Phase one: Data collection**
Collect ransomware files from Dionaea Honeypot (40,000 ransomware files), VirusShare and Softonic (2,507 ransomware files), EldeRan (582 ransomware files), and Cuckoo Sandbox (755 ransomware files) that were used in the previous studies.

- **Phase two: Data preprocessing**
Arrange, organize, and standardize the collected data using Python libraries pandas, matplotlib, numpy, and seaborn.

- **Phase three:  Behavior information collection**
Classification of the dataset using dynamic analysis to extract dataset files feature by using Cuckoo Sandbox.

- **Phase four: Features extraction and training model**

Extract features and apply MLP and LSTM deep learning models to detect ransomware files in the extracted log files from phase three.

- **Phase five: Verification**

Verify accuracy and efficiency for the results of MLP and STML models.

# References

[1] Xin, Y., Kong, L., Liu, Z., Chen, Y., Li, Y., Zhu, H., ... & Wang, C. (2018). Machine learning and deep learning methods for cybersecurity. Ieee access, 6, 35365-35381.
[2] Jang-Jaccard, J., & Nepal, S. (2014). A survey of emerging threats in cybersecurity. Journal of Computer and System Sciences, 80(5), 973-993.
[3] Aslan, Ö., & Yilmaz, A. A. (2021). A new malware classification framework based on deep learning algorithms. Ieee Access, 9, 87936-87951.
[4] D. Palmer. (2017). Ransomware: Security Researchers Spot Emerging New Strain of Malware. Accessed: Mar. 20,2021.[Online].Available:https://www.zdnet.com/article/ransomware-security-researchers-spotemerging-new-strain-of-malware/ .
[5] Zavarsky, P., & Lindskog, D. (2016). Experimental analysis of ransomware on windows and android platforms: Evolution and characterization. Procedia Computer Science, 94, 465-472.
[6] Aldauiji, F., Batarfi, O., & Bayousif, M. (2022). Utilizing Cyber Threat Hunting Techniques to Find Ransomware Attacks: A Survey of the State of the Art. IEEE Access.
[7] Alzahrani, N., & Alghazzawi, D. (2019, November). A review on android ransomware detection using deep learning techniques. In Proceedings of the 11th International Conference on Management of Digital EcoSystems (pp. 330-335).
[8] Choudhary, S. P., & Vidyarthi, M. D. (2015). A simple method for detection of metamorphic malware using dynamic analysis and text mining. Procedia Computer Science, 54, 265-270.
[9] Urooj, U., Al-rimy, B. A. S., Zainal, A., Ghaleb, F. A., & Rassam, M. A. (2021). Ransomware detection using the dynamic analysis and machine learning: A survey and research directions. Applied Sciences, 12(1), 172.
[10] Sikorski, M., & Honig, A. (2012). Practical malware analysis: the hands-on guide to dissecting malicious software. no starch press.
[11] Aslan, Ö. (2017, November). Performance comparison of static malware analysis tools versus antivirus scanners to detect malware. In International Multidisciplinary Studies Congress (IMSC).
[12] Pandey, S. K., & Mehtre, B. M. (2014, May). Performance of malware detection tools: A comparison. In 2014 IEEE International Conference on Advanced Communications, Control and Computing Technologies (pp. 1811-1817). IEEE.
[13] Roseline, S. A., Geetha, S., Kadry, S., & Nam, Y. (2020). Intelligent vision-based malware detection and classification using deep random forest paradigm. IEEE Access, 8, 206303-206324.
[14] Aslan, Ö., & Samet, R. (2017, October). Investigation of possibilities to detect malware using existing tools. In 2017 IEEE/ACS 14th International Conference on Computer Systems and Applications (AICCSA) (pp. 1277-1284). IEEE.
[15] Rathore, H., Agarwal, S., Sahay, S. K., & Sewak, M. (2018, December). Malware detection using machine learning and deep learning. In International Conference on Big Data Analytics (pp. 402-411). Springer, Cham.
[16] Ebert, C., & Louridas, P. (2016). Machine learning. IEEE Software, 33(5), 110-115.
[17] Angelino, E., Yamins, D., & Seltzer, M. (2010, June). StarFlow: A script-centric data analysis environment. In International Provenance and Annotation Workshop (pp. 236-250). Springer, Berlin, Heidelberg.
[18] Pajouh, H. H., Dehghantanha, A., Khayami, R., & Choo, K. K. R. (2018). Intelligent OS X malware threat detection with code inspection. Journal of Computer Virology and Hacking Techniques, 14(3), 213-223.
[19] VirusTotal-Free online virus, malware and URL scanner [Internet]. [cited 2016 Nov 28]. https://www.virustotal.com/
[20] Contagio Malware Dump: Mila. http://contagiodump.blogspot. com/. Accessed 28 Jun 2016
[21] Sihag, V., Vardhan, M., Singh, P., Choudhary, G., & Son, S. (2021). De-LADY: Deep learning based Android malware detection using Dynamic features. J. Internet Serv. Inf. Secur., 11(2), 34-45.
[22] A. F. A. Kadir, N. Stakhanova, and A. A. Ghorbani. An empirical analysis of android banking malware. Protecting mobile networks and devices: challenges and solutions, 209, Nov 2016.
[23] S. Mahdavifar, A. F. A. Kadir, R. Fatemi, D. Alhadidi, and A. A. Ghorbani. Dynamic android malware category classification using semi-supervised deep learning. In Proc. of the 2020 International Conference on Dependable, Autonomic and Secure Computing, International Conference on Pervasive Intelligence and Computing, International Conference on Cloud and Big Data Computing, International Conference on Cyber Science and Technology Congress (DASC/PiCom/CBDCom/CyberSciTech'20), Calgary, Alberta, Canada, pages 515–522. IEEE, August 2020.
[24] C. Mobile. Mobile malware mini dump. Nov 2013.
[25] F. Wei, Y. Li, S. Roy, X. Ou, and W. Zhou. Deep ground truth analysis of current android malware. In Proc. of the 14th International Conference on Detection of Intrusions and Malware, and Vulnerability Assessment(DMVA'17), Bonn, Germany, volume 10327 of Lecture Notes in Computer Science, pages 252–276. Springer, July 2017.
[26] Sahoo, D., & Dhawan, Y. (2022). Evaluation of supervised and unsupervised machine learning classifiers for Mac OS malware detection. In Handbook of Big Data Analytics and Forensics (pp. 159-175). Springer, Cham.
[27] Usharani, S., Bala, P. M., & Mary, M. M. J. (2021). 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.
[28] Imtiaz, S. I., ur Rehman, S., Javed, A. R., Jalil, Z., Liu, X., & Alnumay, W. S. (2021). DeepAMD: Detection and identification of Android malware using high-efficient Deep Artificial Neural Network. Future Generation computer systems, 115, 844-856

[29] Android malware dataset (cicandmal2017 - first part), 2020, https://www. unb.ca/cic/datasets/andmal2017.html (Accessed: 2020-03-12)

[30] Investigation of the android malware (cicinvesandmal2019), 2020, https: //www.unb.ca/cic/datasets/invesandmal2019.html (Accessed: 2020-03-12)

[31] Ghanei, H., Manavi, F., & Hamzeh, A. (2021). A novel method for malware detection based on hardware events using deep neural networks. Journal of Computer Virology and Hacking Techniques, 17(4), 319-331.

[32] PortableApps.com. https:// porta bleap ps. com/ 2019, Accessed 13 Jan 2019

[33] SnapFiles.com. https:// snapf iles. com/ 2019, Accessed 17 Jan 2019

[34] VirusShare.com. https:// virus share. com/ 2019, Accessed 19 Jan 2019

[35] Hwang, J., Kim, J., Lee, S., & Kim, K. (2020). Two-stage ransomware detection using dynamic analysis and machine learning techniques. Wireless Personal Communications, 112(4), 2597-2609.

[36] Bibi, I., Akhunzada, A., Malik, J., Ahmed, G., & Raza, M. (2019, August). An effective Android ransomware detection through multi-factor feature filtration and recurrent neural network. In 2019 UK/China Emerging Technologies (UCET) (pp. 1-4). IEEE .

[37] A. H. a. K. A. F. A. a. T. L. a. G. A. A. Lashkari, "Toward Developing a Systematic Approach to Generate Benchmark Android Malware Datasets and Classification," in 2018 International Carnahan Conference on Security Technology (ICCST), IEEE, 2018, pp. 1--7.

[38] Agrawal, R., Stokes, J. W., Selvaraj, K., & Marinescu, M. (2019, May). Attention in recurrent neural networks for ransomware detection. In ICASSP 2019-2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP) (pp. 3222-3226). IEEE.

[39] Franc¸ois Chollet et al., “Keras,” https://keras. io, 2015.

[40] Fernando, D., Komninos, N., & Chen, T. (2020). A Study on the Evolution of Ransomware Detection Using Machine Learning and Deep Learning Techniques. Iot, 1(2), 551-604. doi: 10.3390/iot1020030.

[41] VinayKumar, R.; Soman, K.P.; Senthil Velan, K.K.; Ganorkan, S. Evaluating Shallow and Deep Networks for Ransomware Detection and Classification. In Proceedings of the 2017 International Conference on Advances in Computing, Communications and Informatics (ICACCI), Udupi, India, 13–16 September 2017.

[42] Marastoni, N., Giacobazzi, R., & Dalla Preda, M. (2021). Data augmentation and transfer learning to classify malware images in a deep learning context. Journal of Computer Virology and Hacking Techniques, 17(4), 279-297.