International Journal of Innovative Research in Engineering
Volume 4, Issue 3(May-June 2023), PP: 31-35.
www.theijire.com ***ISSN No: 2582-8746***

Malware Detection Using Neural Network

Krithika.S[1], Aravind Raj.S[2], Sandeep.S[3], Adhish.M[4], Kaviyarasu.M[5]
[1]Associate professor, Department of Computer Science and Engineering, paavai Engineering College, Namakkal, TN, India.
[2,3,4,5]UG Students, Department of Computer Science and Engineering, paavai Engineering College, Namakkal, TN, India.

How to cite this paper:
Krithika.S[1], Aravind Raj.S[2], Sandeep.S[3], Adhish.M[4], Kaviyarasu.M[5], "Malware Detection Using Neural Network", IJIRE-V4I03-31-35.

https://www.doi.org/10.59256/ijire.2023040251

Abstract: *Malicious assaults, malware, and ransomware families offer serious security challenges for cyber security, and they have the potential to cause catastrophic harm to computer systems, data centres, online, and mobile applications across a wide range of sectors and enterprises. Software (malware) has appeared and is growing in many formats and is becoming increasingly sophisticated. Criminals use them as a tool to infiltrate, steal or falsify information, causing huge damage to individuals, businesses and even threatening national security. It is a complex and varied threat that affects users globally, preventing them from accessing their system or data by locking the system's screen or encrypting and encrypting the users' files unless a ransom is paid. Traditional anti-ransomware technologies are unable to combat newly developed sophisticated assaults. As a result, cutting-edge approaches such as conventional and neural network-based designs can be very beneficial in the creation of unique ransomware solutions. In this project, propose a feature selection-based system for ransomware detection and prevention that uses deep learning methods, including neural network-based designs. We employed Multi-layer Perceptron classifiers on a sample of characteristics to classify malware. Then, to evaluate our proposed technique, we conducted all of the experiments on a single ransomware dataset. In terms of accuracy and precision ratings, the experimental findings show that MLP classifiers outperform other techniques.*

Key Word: *Multi layer perceptron, Deep learning, Pre processing, Prediction.*

I.INTRODUCTION

Malware is software programs designed to harm or perform unwanted actions on a computer system. Malicious software is essentially software like other software on the computer that is used every day and has all the characteristics and properties of normal software, except that it is more malicious. The study listed some common types of malware including Virus, Worm, Trojan Horse, Malicious Mobile Code, Tracking Cookie, Attacker Tool, Phishing, Hoax Virus. According to statistics the situation of malware distribution in 2019 increased by 79% compared to 2018. This is entirely reasonable because hackers used to focus on information systems in the past. This usually chooses to attack the user primarily. Therefore, malware rapidly increases not only in a number of attacks but also its dangerous levels. There are several approaches to detecting malware. The two basic methods used to detect malware are the sign-based detection method and based on behavioural analysis. Methods of detecting malware based on a set of signs have been studied and applied early because of its rapidity and accurate detection capability. Commonly used signs in this method include hash code, IP, Domain or Indicators of compromise. However, the disadvantage of this method isn't able to detect new malware samples that are not in the signature database. In this paper, we propose a method to detect malware based on deep learning techniques. In the paper there are some difficulties in the method of detecting malware based on deep learning. In our study, we propose a malware detection process based on static and dynamic analysis. Finally, to conclude the existence of malware in the system we propose to use deep learning algorithms. In this digital world of Industry 4.0, the rapid advancement of technologies has affected the daily activities in businesses as well as in personal lives. Internet of Things (IoT) and applications have led to the development of the modern concept of the information society. However, security concerns pose a major challenge in realising the benefits of this industrial revolution as cyber criminals attack individual's and networks for stealing confidential data for financialgains and causing denial of service to systems. Such attackers make use of malicious software or malware to cause serious threats and vulnerability ofsystems . A malware is a computer program with the purpose of causing harm to the operating system(OS). Amalware gets different names suchas adware, spyware, virus, worm, trojan, rootkit, backdoor, ransomware and command and control (C&C) bot, basedon its purpose and behaviour. Detection and mitigation of malware is an evolving problem in the cyber security field.Nowadays deep learning has dominated the various computer vision tasks. Not only these deep learning techniques enabled rapid progress in this competition, but even surpassing human performance in many of them. One of these tasks is Image Classification. Unlike more traditional methods of machine learning techniques, deep learning classifiers are trained through feature learning rather than task-specific algorithms. What this means is that the machine will learn patterns in the images that it is presented with rather than requiring the human operator to define the patterns that the machine should look for in the image.

II.EXISTINGSYSTEM

Malicious applications or attacks, malware and ransomware families for instance, consistently endures to pose critical security issues to cyber security and it may cause catastrophic damages to computer systems, data centre, web data, and mobile applications across various industries and businesses. After these attacks, it is incredibly difficult to revert without paying the extortion. Traditional ransomware detection techniques including event-based, statistical-based, and data-centric-based techniques are not adequate to combat. Therefore, implementing the highest level of optimal protection and security by adopting futuristic technology against such advanced malicious attacks should be imperative for the research community. The threats impose by the cyber-attacks due to malicious software (malware) have been increasing drastically with the evolution of information technology. Since people use web applications on a daily basis these malware attacks have become challenging. There have been various attacks affecting confidentiality, integrity and availability of data which has become a major security concern. Though the manual inspection and classification methods seemed to bring up some light to this facet, these methods are no longer considered effective, since they are time consuming and inefficient.

2.1 Disadvantages

- Affecting Confidentiality, integrity has a major
- They are time consuming and inefficient
- Small dataset can generate incorrect predictions

III.PROPOSEDSYSTEM

Ransomware attacks are dangerous and difficult to detect and prevent. Therefore, the task of detecting signs of malware and alerting it for users or the system is very necessary today. One of the most effective malware detection approaches is applying deep learning or deep learning to analyze its behavior. In this project, we will propose a method to use deep learning to detect malicious signs based on their unusual behavior. Accordingly, in our research, we will conduct malicious analysis using static and dynamic analysis methods to detect abnormal behavior and combine them with a Multi-layer perceptron (MLP) to the conclusion on malware behavior. The deep learning models must be trained before it is being tested on a real-life scenario. For the training purpose a recognized data set should be used. However, the problem face by most researchers is that finding out a high-quality dataset. The scarcity of good datasets has set a drawback in this domain. Accuracy can be considered as a proportion of total number of predictions that were correctly identified.

3.1Advantage

- More reliable than the tradition algorithm
- Provide appropriate result with high accuracy
- Capable of solving complex data

IV. FEASIBILITY STUDY

Feasibility study examines the viability or sustainability of an idea, project, or business. The study examines whether there are enough resources to implement it, and the concept has the potential to generate reasonable profits. In addition, it will demonstrate the benefits received in return for taking the risk of investing in the idea. These studies analyze strengths, weaknesses, opportunities, and threats to determine whether the proposals are cost-effective and beneficial to a company's long-term success. Furthermore, investors can benefit from evaluating the problems and solutions listed in the study and determine whether a proposed project is the right choice for a company.

4.1 Technical Feasibility

This assessment is based on an outline design of system requirements, to determine whether the company has the technical expertise to handle completion of the project. When writing a feasibility report, the following should be taken to consideration:

- A brief description of the business to assess more possible factors which could affect the study
- The part of the business being examined
- The human and economic factor
- The possible solutions to the problem

At this level, the concern is whether the proposal is both technically and legally feasible (assuming moderate cost).[citation needed. The technical feasibility assessment is focused on gaining an understanding of the present technical resources of the organization and their applicability to the expected needs of the proposed system. It is an evaluation of the hardware and software and how it meets the need of the proposed system.

V. MODULES DESCRIPTION

5.1 Dataset

The deep learning models must be trained before it is being tested on a real life scenario. For the training purpose a recognized malware data set should be used. However, the problem face by most researchers is that finding out a high quality dataset. The data set used here is taken from the Kaggledataset repository.

5.2 Pre-processing and feature selection

Before training the deep learning models the dataset should be pre-processed and features should be selected. The feature selection was carried out only for the deep learning models and it was done using Classifier. The deep learning models were trained without feature selection. The dataset was imported in the format of data frames, and it was then shuffled. Since the first 500000 data samples were malware and the next 500000 samples were benign, the dataset was shuffled to increase the randomness and prevent the over fitting during training. Then the features and targets were selected from the shuffled dataset.

5.3 Classification

The major efforts in building a robust Multi layer Perceptron are concentrated in the extraction and selection of a set of characterizing and discriminate features. In order to specify and enforce these constraints, we make use of the text classification. From MLP point of view, we approach the task by defining a hierarchical two-level strategy assuming that it is better to identify and eliminate "neutral" sentences, then classify "non-neutral" sentences by the class of interest instead of doing everything in one step. The feature extracted data will be compared to the pre-trained dataset using multi-layer Perceptron, the system classify and predict the malware.

5.4 Malware Prediction

This is the final phase of a deep learning approach. This is where the testing dataset is feed into the model to predict the accuracy of the model which provides an idea of how well the model reacts of unseen data. When feeding the data into the trained model, the testing data should be in the same format of the training dataset. Classifying the input data with the pre-trained dataset and if it is match, the system will predict the malware occurred file to the user model.

VI. SYSTEM DESIGN

6.1 System architecture

A system architecture or systems architecture is the conceptual model that defines the structure, behavior, and more views of a system. An architecture description is a formal description and representation of a system, organized in a way that supports reasoning about the structures and behaviors of the system. System architecture can comprise system components, the externally visible properties of those components, the relationships (e.g. the behavior) between them. It can provide a plan from which products can be procured, and systems developed, that will work together to implement the overall system. There have been efforts to formalize languages to describe system architecture, collectively these are called architecture description languages (ADLs).

6.2 Various organizations define systems architecture in different ways, including:

- An allocated arrangement of physical elements which provides the design solution for a consumer product or life-cycle process intended to satisfy the requirements of the functional architecture and the requirements baseline.
- Architecture comprises the most important, pervasive, top-level, strategic inventions, decisions, and their associated rationales about the overall structure (i.e., essential elements and their relationships) and associated characteristics and behavior.
- If documented, it may include information such as a detailed inventory of current hardware, software and networking capabilities; a description of long-range plans and priorities for future purchases, and a plan for upgrading and/or replacing dated equipment and software
- The composite of the design architectures for products and their life-cycle processes.

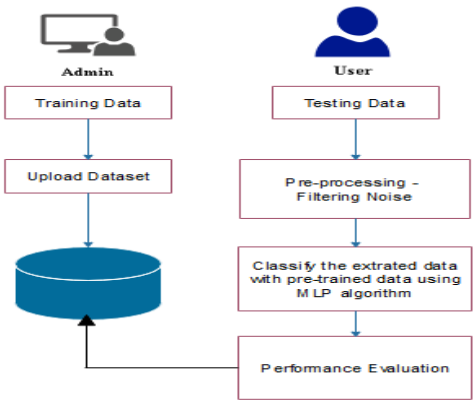

Fig 6.1 System Architecure Diagram

VII. RESULT

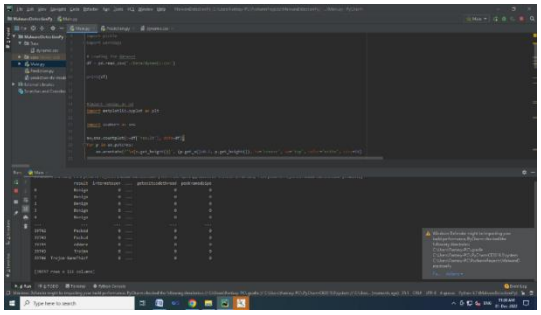

Fig 7.1Main .py file

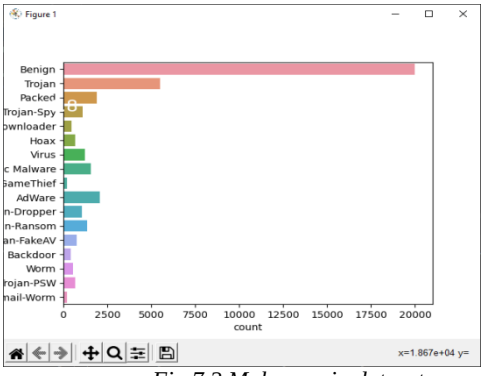

Fig 7.2 Malwares in dataset

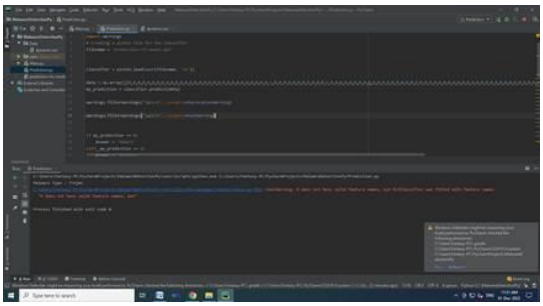

Fig 7.3 Malware Type

VIII.CONCLUSION AND FUTURE ENHANCEMENTS

8.1 Conclusion

Despite the extensive studies and staggering progress that the deep learning approach on malware classification have gained in the recent years; yet it remains a very challenging domain. However, in this domain, it is a necessity to upgrade the existing strategies since the attackers too develop counter measures to overrule the defence strategies. This paper focused on how to overtake such attackers in this race for data and privacy protection with a promising inclusion of deep learning in this domain. The results of this research have also provided an insight to the deep learning approach in malware classification by highlighting the fact that, feature selection can be effectively used for deep learning classifiers to produce outstanding results

8.2 Future Enhancements

In future, prevent the lack of issues and will explore better results with appropriate accuracy. Modifications of the deep learning models is a necessity because over time attackers find mechanisms to overrule the prevailing defence measures. The fine tuning of the deep learning algorithms can produce better results utilizing the features to function at an optimum level. The limitations in datasets create drawbacks in most studies.

References

1. *Shabtai, R. Moskovitch, Y. Elovici, C. Glezer, Detection of malware by applying machine learning classifiers on static features: A stateof- the-art survey, Inf. Secur. Tech. Rep. 14 (1) (2009) 16–29.*
2. *M. Bailey, J. Oberheide, J. Andersen, Z. M. Mao, F. Jahanian, J. Nazario, Automated classification and analysis of internet malware. Recent advances in intrusion detection, Springer, 2007, pp. 178–197.*
3. *U. Bayer, P. M. Comparetti, C. Hlauschek, C. Kruegel, E. Kirda, Scalable, behavior-based malware clustering, in: NDSS, Vol. 9, 2009, pp. 8–11.*
4. *K. Rieck, P. Trinius, C. Willems, T. Holz, Automatic analysis of malware behavior using machine learning,Journal of Computer*

Security 19 (4) (2011) 639–668.https://doi.org/10.3233/JCS-2010-0410

5. *Palahan, D. Babi´c, S. Chaudhuri, D. Kifer, Extraction of statistically significant malware behaviors, in: Computer Security Applications Conference, ACM, 2013, pp. 69–78.*
6. *M. Egele, M. Woo, P. Chapman, D. Brumley, Blanket execution: Dynamic similarity testing for program binaries and components, in: USENIX Security '14, USENIX Association, San Diego, CA, 2014, pp. 303–317.*
7. *M. Lindorfer, C. Kolbitsch, P. M. Comparetti, Detecting environmentsensitive malware, in: Recent Advances in Intrusion Detection, Springer, 2011, pp. 338–357.*
8. *IMPORTANT INFORMATION REGARDING SANDBOXIE VERSIONS. https://www. sandboxie.com/ . [Accessed February 15, 2020].*
9. *Hassan Ramchoun; Mohammed Amine JanatiIdrissi; Mohammed Amine; Youssef Ghanou. Multilayer Perceptron: Architecture Optimization and Training. International Journal of Interactive Multimedia and Artificial Intelligence. 2016, 4, 26-30. https://doi.org/10.9781/ijimai.2016.415*
10. *Do Xuan, Cho, Nguyen, HoaDinh, and Dao, Mai Hoang. APT Attack Detection Based on Flow Network Analysis Techniques Using Deep Learning.Journal of Intelligent & Fuzzy Systems. pp. 1 – 17,. 2020. DOI: 10.3233/JIFS-200694*
11. *Abien Fred Agarap. Deep Learning using Rectified Linear Units (ReLU). arXiv 2018, arXiv:1803.08375.*
12. *KaiboDuan; SathiyaKeerthi, S.; Wei Chu; ShirishKrishnajShevade; AunNeowPoo. Multi-category Classification by Soft-Max Combination of Binary Classifiers.In proceedings ofthe 4th International Workshop, MCS 2003 Guildford, UK, 11–13 June 2003; pp 125–134.https://doi.org/10.1007/3-540-44938-8_13*
13. *HOW TO CREATE A MALWARE DETECTION SYSTEM WITH MACHINE LEARNING. https://www.evilsocket.net/2019/05/22/How-to-create-a-Malware-detection-system-with-Machine-Learning/?fbclid=IwAR1vuaOJA3UryaQATPsqKErktLft2Rt zzAB5kDvgOTo4U3dF4J-Op9te*