International Journal for Modern Trends in Science and Technology, 9(02): 88-91, 2023

ISSN: 2455-3778 online
DOI: https://doi.org/10.46501/IJMTST0902016
Available online at: http://www.ijmtst.com/vol9issue02.html

Robust Intelligent Malware Detection using Deep Learning

Dr.D.Suneetha [1] | K. Anjana Kavya Sri [2] | Md. Rehana Sulthana [2] | M. Venkata Keerthi [2] | K. Sai Charan Gupta [2]

[1]Professor & HOD, Department of CSE, NRI Institute of Technology, India
[2]B.Tech Student, Department of CSE, NRI Institute of Technology, India

To Cite this Article
Dr.D.Suneetha, K. Anjana Kavya Sri, Md Rehana Sulthana, M. Venkata Keerthi, K. Sai Charan Gupta. Research on Robust Intelligent Malware Detection using Deep Learning. International Journal for Modern Trends in Science and Technology 2023, 9(02), pp. 88-91. https://doi.org/10.46501/IJMTST0902016

Article Info
Received: 04 January 2023; Accepted: 01 February 2023; Published: 04 February 2023.

ABSTRACT

As malware assaults develop exponentially, computer users, organizations, and governments face a serious security threat. Static and dynamic analysis of malware signatures and behavior patterns is time-consuming and inadequate at detecting unknown infections. Recent malwares utilize polymorphic, metamorphic, and other evasive strategies to swiftly alter behavior and spawn many malwares. Machine learning algorithms (MLAs) are used to analyze new malware, which are mostly versions of old malware. This needs substantial feature engineering, learning, and representation. Deep learning MLAs eliminate feature engineering. Despite recent findings, training data biases algorithm performance. To improve zero-day malware detection, these technologies must be unbiased and evaluated independently. This study examines traditional MLAs and deep learning architectures for malware detection, classification, and categorization utilizing public and private datasets to address the literature gap. The experimental study uses discontinuous train and test divisions of public and private datasets from distinct timelines. We also present an image processing method with suitable settings for MLAs and deep learning systems. Deep learning architectures surpass standard MLAs in extensive experiments. This research presents real-time malware visual detection utilizing a scalable and hybrid deep learning architecture. Zero-day malware detection is improved by visualization and deep learning architectures for static, dynamic, and image processing hybrid approaches in large data environments.

KEY WORDS: Cyber Security, Cybercrime, Malware detection, Static and Dynamic analysis, Artificial Intelligence, Machine Learning, Deep Learning, Image processing, Scalable and Hybrid framework

1. INTRODUCTION

The fast growth of technology has had an effect on both the day-to-day operations of companies and the day-to-day activities of people living in this digital environment characterized by Industry 4.0. Applications and the Internet of Things (IoT) have been significant contributors to the creation of the contemporary idea of an information society. However, concerns regarding security present a significant barrier to fully realizing the benefits of this industrial revolution. Cybercriminals target individual personal computers as well as networks in order to steal confidential data for the purpose of making financial gains and to cause systems to be denied service. These types of attackers use harmful software, sometimes known as malware, in order to pose major dangers to systems and exploit vulnerabilities [1]. Malware is a term used to refer to a computer software that has been

designed specifically to damage an operating system (OS). According to its function and behavior, a piece of malicious software may be referred to by a variety of names, including adware, spyware, virus, worm, trojan, rootkit, backdoor, ransomware, and command and control (C&C) bot. In the realm of cyber security, the detection and elimination of malicious software is an ongoing challenge. Malware writers are able to increase their ability to avoid being detected as a result of researchers developing new methodologies.

2. LITERATURE SURVEY

As of this day, there has been a significant amount of attention paid to the subject of malware detection in the published literature. However, there are relatively few works that concentrate on the ML methodology that are used, and to the best of our knowledge, none of them give a clear categorization of mobile malware detection systems based on the metrics and ML techniques that are used. This section, which focuses on the years 2017 through 2021 and recognizes such literary contributions in a chronological order, positions these contributions in relation to the work that is now being done. A comprehensive assessment of dynamic mobile malware detection methodologies was provided by Yan et al. [1], which included a variety of criteria and performance evaluation metrics for mobile malware detection. In addition, the authors examined and contrasted the mobile malware detection systems that were available at the time, basing their comparisons on the assessment techniques and outcomes of the study. Finally, the writers discussed several unresolved problems in the industry as well as potential future lines of inquiry. Odusami et al. [2] conducted an assessment of mobile malware detection approaches with the goal of finding gaps and providing information for appropriate steps to be taken against unknown malware. As a result of their research, the authors concluded that strategies that depend on machine learning (ML) to identify malicious applications are more likely to be successful and generate greater detection accuracy than signature-based approaches. Kouliaridis et al. [3] presented a comprehensive analysis of previous research on the subject of mobile malware detection and arranged each of those works according to their own distinct categorization method. To be more specific, the latter organizes the works into categories according to their target platforms, feature selection methods, and detection methodologies, namely signature-based or anomaly-based detection. A detailed study of malware detection methods that make use of ML techniques was published by Liu et al. [4]. The authors conducted research on and provided a summary of a number of important areas, such as sample collecting, data preprocessing, feature selection, machine learning models, algorithm selection, and detection performance. In conclusion, they discussed the shortcomings of the machine learning methodologies and provided their perspectives on some possible possibilities for the future. Gibert et al. [5] evaluated prominent ML approaches for malware detection and in particular, deep learning techniques. The authors examined current trends and advancements in the industry, with a particular emphasis on deep learning schemes, and discussed the research problems and constraints associated with legacy machine learning approaches.

3. PROPOSED SYSTEM

The suggested approach walks one through selecting the most effective machine learning strategies depending on the age of the dataset and the analysis method selected in the first and second steps, respectively. To be more specific, the first two phases include linking the age of the dataset that was used for the assessment to one of the three techniques of analysis. The decisions that were made in the two phases before this one will determine which machine learning classification approaches will be employed in the third step. The last stage is determined by determining whether or not the dataset utilized has an equal number of malicious and benign applications. The answer to this question will indicate whether or not accuracy is a reliable statistic. Even when highly unbalanced datasets are used, the area under the curve (AUC) metric is still the best metric to use since it provides a more definitive and accurate assessment of the models. In general, the area under the receiver operating characteristic curve (AUC) assesses the efficacy of each evaluated method for all conceivable score levels. When a technique is applied to a dataset, the value of AUC is often determined by analyzing the ranking of scores rather than the precise values that are created as a result of applying the method. AUC is not dependent on their being an equal distribution of

positive and negative classes, which is a significant benefit on top of all the other advantages.

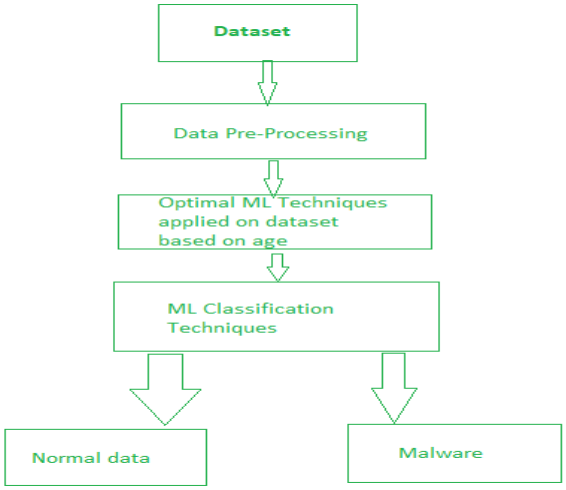

Figure 1: Architecture for Proposed System

4. RESULTS

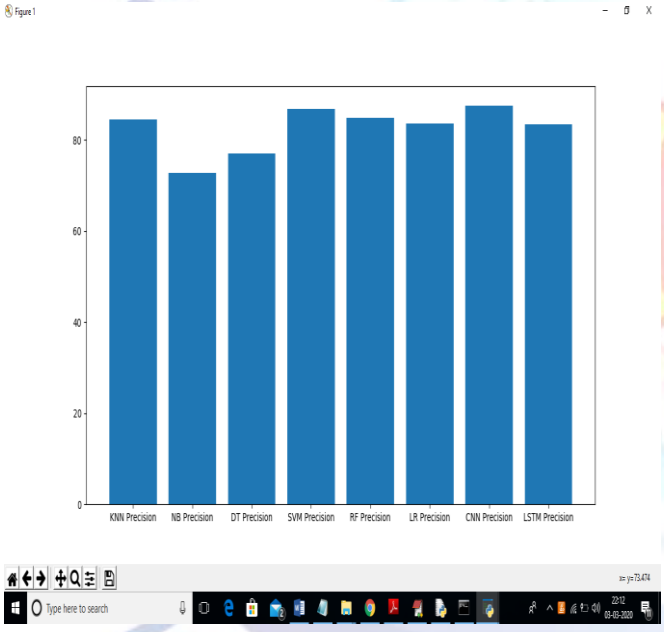

Figure 2: Precision of various algorithms

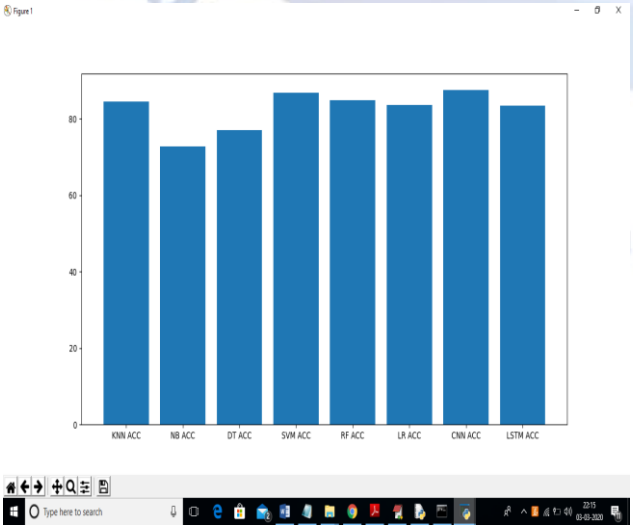

Figure 3: Accuracy of various algorithms

5. CONCLUSION

This article assessed traditional machine learning algorithms (MLAs) and deep learning architectures based on static analysis, dynamic analysis, and image processing for malware detection and built a highly scalable framework called ScaleMalNet to identify, classify, and categorize zeroday malwares. This framework analyzes end-user host malware using deep learning and a two-step procedure. Malware categorization began using a combination of static and dynamic analysis. Image processing categorized malwares in the second step. This research found that deep learning-based techniques outperformed traditional MLAs on benchmark datasets and privately acquired datasets. By adding layers, the framework can evaluate a huge number of malwares in real time. These variances should be examined using additional data characteristics in future study. This work's main discovery, flaw, and future scope are: Two-stage procedure scalable malware detection framework. Deep learning identifies and categorizes malware in the proposed framework. Deep learning architectures beat traditional MLAs in static, dynamic, and image-processing-based malware detection and classification. The dynamic analysis-based malware detection research uses deep learning architectures on domain knowledge derived characteristics. Runtime binary file memory dumps may be mapped into grayscale images to prevent this. In a deep learning-based malware detection investigation, malware pictures were flattened and corrected. The spatial pyramid pooling (SPP) layer may accept photos of any size in future development. This learns features at different sizes and may be placed between the sub sampling layer and the fully linked layer to increase model flexibility. Malimg has very unbalanced malware families. Cost-sensitive methods may address multiclass malware family imbalance. This helps deep learning architectures include cost components during backpropogation learning. The cost component primarily symbolizes classification relevance, giving lower value to classes with more samples and greater value to classes with fewer samples. Deep learning architectures are sensitive to adversaries [50]. The deep learning architectures may be misled by generative adversarial network samples during testing or deployment. The suggested study ignores deep learning architecture resilience. Since malware defection is vital in safety-critical environments, this is a key future path. One misclassification may harm the company.

Conflict of interest statement

Authors declare that they do not have any conflict of interest.

REFERENCES

[1] P. Groves, B. Kayyali, D. Knott, and S. van Kuiken, The`Big Data'Revolution in Healthcare: Accelerating Value and Innovation. USA: Center for US Health System Reform Business Technology Ofce, 2016.

[2] M. Chen, S. Mao, and Y. Liu, ``Big data: A survey,'' Mobile Netw. Appl., vol. 19, no. 2, pp. 171209, Apr. 2014.

[3] P. B. Jensen, L. J. Jensen, and S. Brunak, ``Mining electronic health records: Towards better research applications and clinical care,'' Nature Rev. Genet., vol. 13, no. 6, pp. 395405, 2012.

[4] D. Tian, J. Zhou, Y. Wang, Y. Lu, H. Xia, and Z. Yi, ``A dynamic and self-adaptive network selection method for multimode communications in heterogeneous vehicular telematics,'' IEEE Trans. Intell. Transp. Syst., vol. 16, no. 6, pp. 30333049, Dec. 2015.

[5] M. Chen, Y. Ma, Y. Li, D. Wu, Y. Zhang, and C. Youn, ``Wearable 2.0: Enable human-cloud integration in next generation healthcare system,'' IEEE Commun., vol. 55, no. 1, pp. 5461, Jan. 2017.

[6] M. Chen, Y. Ma, J. Song, C. Lai, and B. Hu, ``Smart clothing: Con- necting human with clouds and big data for sustainable health monitor- ing,'' ACM/Springer Mobile Netw. Appl., vol. 21, no. 5, pp. 825845, 2016.

[7] M. Chen, P. Zhou, and G. Fortino, ``Emotion communi- cation system,'' IEEE Access, vol. 5, pp. 326337, 2017, doi: 10.1109/ACCESS.2016.2641480.

[8] M. Qiu and E. H.-M. Sha, ``Cost minimization while satisfying hard/soft timing constraints for heterogeneous embedded systems,'' ACM Trans. Design Autom. Electron. Syst., vol. 14, no. 2, p. 25, 2009.

[9] J. Wang, M. Qiu, and B. Guo, ``Enabling real-time information service on telehealth system over cloud-based big data platform,'' J. Syst. Archit., vol. 72, pp. 6979, Jan. 2017.

[10] D. W. Bates, S. Saria, L. Ohno-Machado, A. Shah, and G. Escobar, ``Big data in health care: Using analytics to identify and manage high-risk and high-cost patients,'' Health Affairs, vol. 33, no. 7, pp. 11231131, 2014.