Grenze International Journal of Engineering and Technology, June Issue

Securing Healthcare Data from Trojan using Blockchain and Multilayer Perceptron

S Tamil Selvi[1], Dhiviyaabharathi J[2], Rakshith D H[3] and Vaishnavi Devi R[4]
[1-4]Department of Cyber Security, SRM Valliammai Engineering College
Email: tamilselvy@gmail.com, dhiviyaaj@gmail.com, rakshith.durai@gmail.com, vaishnavirkr01@gmail.com

***Abstract*— A computer system or network can be disrupted, damaged, or compromised by malware, a sort of harmful program or software. Malware affects many businesses, including financial institutions and the healthcare industry. The existing model of secure healthcare systems uses blockchain-enabled security frameworks to prevent the system from malware attacks by only providing tamper resistance for healthcare networks. Whereas the proposed method is used to identify malware in intelligent healthcare frameworks, a unique method utilizing the Multilayer Perceptron (MLP) algorithm is presented in the proposed system. The proposed approach uses a benchmark Malimg dataset with the families of Dontovo.A, C2Lop.P, Obfuscator.AD, to train a Multilayer Perceptron model to efficiently detect and mitigate malware threats and the implementation of blockchain technology in the proposed model make the healthcare system as a tamper-proof framework renowned for its heightened security, resilience, and decentralized structure which makes it nearly hard to change once recorded without also changing all blocks that come after. The effectiveness of this method in identifying malware attacks with an emphasis on attachments is proven by thorough testing and analysis. The proposed system is improved to existing techniques in terms of accuracy, as demonstrated by a comparative evaluation.**

***Index Terms*— Healthcare, Malware, Blockchain, Multilayer Perceptron (MLP), Threats.**

I. INTRODUCTION

Digital solutions are being used by the healthcare sector, which is at the forefront of technology development, to improve patient care, expedite administrative procedures, and handle massive amounts of private medical data. This digital transformation has brought numerous benefits, it exposes the sector to a growing array of cybersecurity threats, with malware attacks posing a significant and ever-evolving risk. Malignant programming, at times known as malware, is a general term for various malevolent projects or codes expected to invade, upset, or compromise PC systems and networks. In the healthcare context, these threats can lead to severe consequences, including compromising patient privacy, disrupting critical medical services, and manipulating medical records. The industry must remain vigilant in safeguarding patient data and critical systems from these cybersecurity threats.
Healthcare cybersecurity is of paramount importance as patient data must be kept confidential, accessible only to authorized individuals, and consistently accurate. A malware attack in a healthcare setting can lead to not only financial losses, it endanger patient safety and undermines trust in healthcare institutions. The healthcare industry faces unique challenges, such as strict regulatory requirements (e.g., HIPAA in the United States) and the need for real-time data access in life-critical systems. Safeguarding patient information and ensuring the integrity of

Grenze ID: 01.GIJET.10.2.148

healthcare systems are critical priorities in this environment.
The detection technique comprehensively investigates cutting-edge malware detection techniques tailored specifically for the healthcare sector. By conducting a rigorous review of traditional methods and research endeavors, the study delves into the evolving threat landscape unique to healthcare-related malware and the corresponding defensive measures.
The research categorizes and assesses a wide range of malware detection methods, spanning from conventional signature-based techniques to sophisticated behavior-based, machine learning-based, and blockchain-based strategies. Each approach is evaluated based on its effectiveness, limitations, and relevance in the realm of healthcare cybersecurity. The technique holds significance in its potential to educate healthcare professionals, security experts, and policymakers about the existing landscape of malware threats in the healthcare sector. It aims to provide valuable insights into the available tools and strategies for mitigating these threats. By synthesizing key findings and trends from traditional methods, the research offers crucial insights into state-of-the-art malware detection solutions, vital for safeguarding sensitive information and operations in the healthcare industry. The initial section will provide a concise overview of blockchain technology and elucidate the operational mechanisms of attacks in the healthcare sector. The second segment will conduct an in-depth analysis of prevailing detection and prevention methods within the healthcare industry. This analysis will encompass evaluating the strengths and weaknesses of each technique, identifying prominent trends and challenges in this domain. In the final part, the proposed model will pinpoint the gaps and limitations in existing research, highlighting areas for future exploration. This section will emphasize the need for comprehensive and integrated approaches to detect and counter attacks. It will underscore the importance of developing new techniques that are resistant to evolving attack methods.

II. Literature Survey

Ö. Aslan and A. A. Yilmaz [1] discussed the shift in human activities from physical to virtual realms, accelerated by recent advancements in computer systems and further propelled by the COVID-19 pandemic. Cybercriminals are increasingly targeting virtual vulnerabilities using evolving malware. Researchers used datasets like Malimg, Microsoft BIG 2015, and Malevis to develop a highly accurate malware categorization approach, outperforming many current methods. Their solution achieved 97.78% accuracy on the Malimg dataset.
[2] outlines a novel approach developed by D. Javaheri, P. Lalbakhsh, and M. Hosseinzadeh, introducing an innovative solution to identify elusive and shape-shifting malware programs, addressing the challenge of limited datasets for modelling these types of malwares. Researchers used genetic algorithms and optimization techniques to generate mutated elite malware samples, addressing the lack of training data. This method significantly boosted detection accuracy, especially for uncommon and shape-shifting malware, showcasing effectiveness in tackling elusive threats.
Darem A.A., Abawajy J.H., Al-Rezami A.Y. et al [3] presented a solution addressing the challenges posed by evolving malware variants, which represent significant and ever-changing threats to cybersecurity and have the potential to cause substantial damage to computer systems. Traditional malware detection struggles with the dynamic nature of malware variants, leading to outdated models.
The various capabilities and combination of man-made reasoning (man-made intelligence) applications in the Android working framework, which have created its dominant position in the market, are discussed by Thakur D., Shah B., Abuhmed T, [4]. The feature Fusion-SVM model achieved a high accuracy rate of 93.24% in detecting and classifying complex Android malware using malicious pictures and certificate formats, showcasing its effectiveness against advanced threats.
Cilleruelo C., Enrique-Larriba, and et al[5] address the persistent challenge of Potentially Harmful Apps (PHAs), which pose a threat similar to conventional malware. Researchers propose an innovative machine learning solution to detect Potentially Harmful Apps (PHAs) in the Google Play Store. Their technique accomplished a high 90% exactness rate when tried on a dataset of 91,203 apps. This approach, based on lifespan data, provides a new dimension in the ongoing battle against PHAs in app ecosystems and can complement existing detection models.
A novel approach for automated Android malware detection was put out by brahim, B. Issa, and M. B. Jasser [6]. Researchers conducted experiments focusing on Android malware detection and grouping. Their model outperformed existing methods, achieving a remarkable F1-score of 99.5% for two-class identification and 97% across all five classes. These results highlight the effectiveness of their approach in accurately detecting and classifying Android malware.
A.Al-Hashmi et al.'s [7] novel method for identifying malware programs, many of which are variants of already-existing malware such polymorphic and transformative malware, is introduced. Malware variants often change

behavior to evade security measures, making identification challenging. Researchers incorporated these features into an outrageous angle helping based classifier for independent navigation. Evaluations revealed their model increased detection rates and reduced false negatives, making it reliable for malware variant identification.

A solution for spotting malware behavior is put out by Nguyen H.N. , Abri F., and et al[8]. Static and dynamic exams are the two main approaches used to analyze malicious apps. Static analysis involves parsing a malicious program without executing it, generating artifacts like control-flow graphs. Dynamic analysis, in contrast, requires running the program. The technique introduces MalView, a platform for malware investigation using intelligent visualization. MalView supports behavior analysis and signature-based artifacts, aiding tasks such as characterizing malware, locating payloads, analyzing unknown malware, and examining affected system components.

The work of Almomani I., Alkhayer A., El-Shafai W.[9] on establishing financially savvy profound deep leaning (DL) based calculations for distinguishing Android Malware (AMD) frameworks. Technique introduces a vision-based Android malware detection (AMD) model utilizing sixteen optimized convolutional neural networks (CNNs). Compared to traditional vision-based algorithms, this CNN-based AMD model outperforms previous techniques on benchmark Android datasets.

Pattee J., Lee B.K. [10] propose hardware solutions for detecting malware, a method gaining prominence due to the susceptibility of software-based solutions to intelligent malware compromise. To enhance real-time malware detection accuracy, they propose schemes to improve accuracy, such as introducing additional PMCs, alongside hardware acceleration methods. These combined strategies result in accuracy improvements (5-10%) and faster detection speeds (up to 10%), with a negligible impact on hardware complexity (less than 1% of chip complexity).

The efficacy of several AI (ML) models for malware discovery is explored by H. Manthena, Kimmel J.C. , Abdelsalam M., and Gupta M. [11]. The technique helps to evaluate various ML models, including SVM Direct, SVM-RBF, Irregular Woods, Feed-Forward Brain Organizations, and Convolutional Brain Organizations, for internet spyware identification. To improve interpretability, the Shapley Added explanations (SHAP) method was utilized, utilizing different approaches like KernelSHAP, TreeSHAP, and DeepSHAP. The aim is to make ML models' decisions, particularly in online malware identification.

Jin, Choi J., Hong J.B. , Kim H. [12], emphasises production of malware variations utilizing different evasion techniques to bypass malware detectors. Understanding the specific properties that drive these variations to evade malware detection methods is crucial. The researchers found that utilizing XOR to obfuscate code was the most efficient method. Code obfuscation allows for the creation of malware variants that may avoid detection by an average of 28 distinct malware engines. These discoveries can be useful in creating anti-malware programs that are capable of spotting newly discovered malware variants.

He t., Han C., Isawa R. and et al[13] discuss the enormous social difficulties brought on by the quick rise of recently developed forms of Internet of Things and its variations. Researchers propose an innovative online processing approach for building phylogenetic trees from large malware datasets, addressing time constraints. Using 65,494 IoT malware samples, their method reduced computing time by 97.52% compared to conventional methods. The clustering approach accurately classified family names with a 95.5% accuracy rate and pattern names with a 99.3% accuracy rate. This advancement is crucial for analyzing and classifying massive IoT malware datasets efficiently.

Convolutional Neural Networks (CNNs) are successful in classifying and detecting malware based on perception, according to Belal M.M., Sundaram D.M. [14]. The technique introduces the Butterfly Vision Transformer, a novel structure of perspective-based spyware characterization, identification. By capturing both nearby and worldwide spatial portrayals of malware pictures, B_ViT outperforms other Vision Transformer variants, achieving accelerations of 2.42 times over IEViT and 1.81 times over typical ViT variants. The study highlights B_ViT's effectiveness in detecting surface-based malware and its adaptability to handle polymorphic obfuscation methods, surpassing CNN-based malware classification techniques.

The vital need to safeguard all layers and devices inside a computing environment, especially under edge/haze computing situations, is discussed by Gulatas I., ZaimA.H., et al [15]. IoT devices, vulnerable components in secure data environments, face threats from emerging Edge/Haze computing-based malware. Researchers used analytic methodologies like "Digital Kill Chain" and "Miter ATT&CK for ICS" to characterize these malware families, highlighting their unique characteristics. The study emphasizes the importance of understanding and addressing the challenges posed by Edge/Haze computing malware, particularly in the context of IoT devices.

A. Inference

[1] Leading to an increased focus on cybersecurity as researchers develop highly accurate malware categorization methods. [2] Genetic algorithms are utilized to generate mutated malware samples. [3] A dynamic malware

detection model surpassing traditional methods, new infections with minimal monthly updates. [4] The detecting and classifying complex Android threats, demonstrating its efficacy against advanced malware varieties. [5] Detection of Potentially Harmful Apps (PHAs), offering a lifespan-based approach to enhance the fight against PHAs. [6] an automated Android malware detection method surpassing existing approaches, in accurate malware detection and classification. [7] A novel method incorporating outrageous angle-based classification for identifying malware variants. [8] Facilitating behavior analysis and signature-based artifact examination for malware investigation. [9] Develop a powerful yet efficient model by deep learning for Android malfunctioning software detection. [10] Proposed hardware solutions for malfunctioning software detection, enhancing accuracy and speed through additional PMCs. [11] Evaluate AI models with SHAP to enhance transparency in online malware identification. [12] A framework utilizing XOR obfuscation to create malware variants evading detection for anti-malware development. [13] Proposed an online processing approach for efficiently analyzing large IoT malware datasets. [14] Outperforming CNNs in malware classification by capturing spatial representations and achieving faster detection. [15] Safeguarding IoT devices from Edge/Haze computing-based malware, unique traits and addressing emerging threats.

III. Attacks And Working Mechanism Of Trojan

Here, we discuss various attacks and working mechanisms of trojans and some variants of it.

A. Trojan Attack

Trojan malware, named after the mythological horse, encompasses various types, all of which require user initiation. These malicious programs masquerade as useful utilities or tools but aim to perform nefarious actions such as data deletion, alteration, encryption, or copying, along with activities like file sending and receiving and network slowdowns. Evolving into complex forms like backdoors and downloaders, Trojans include (1) ArcBombs, (2) Backdoors enabling remote control of computers or network spread, (3) Banking Trojans like Emotet and (4) TrickBot target financial data, (5) Clickers for internet site manipulation, (6) DDoS Trojans, causing denial of service attacks. Each type poses significant threats, highlighting the importance of vigilant cybersecurity practices.

B. Selection of Malware and Dataset

The Malimg dataset in the field of malware recognition and arrangement with 25 different families, was used in this work. The malware variations chosen for analysis in the Malimg dataset encompass several prominent families with various variants of malware, notably including C2Lop.P-146 of trojan family, Donvotov.A-162 of trojan downloader family, and Obfuscator.AD-142 of trojan downloader family. The selection of these malware families was motivated by several criteria, including their variety and prevalence in modern cybersecurity environments, as well as their persistence. Ultimately, the requirement for an extensive and representative dataset that reflects actual malware threats faced by intelligent healthcare systems drove the selection of the Malimg dataset and the malware types that it was linked with.

C. Working Mechanism of Trojan

How do Trojans function? - All Trojans comprise of two sections: the server and the client. The client associates with the server with the assistance of the TCP/IP convention. The client might have a UI and a bunch of buttons and info fields for distant organization. The server part is introduced on the casualty's gadget. The server-side cycles (executes) orders from the client and moves different information. When entering the PC, the server side tunes in on a particular port, sitting tight for orders. An assailant pings a port on a contaminated host. Assuming the server part is effectively introduced, it will answer with the PC's IP address and organization name. At the point when the association is laid out, the client begins sending orders to the server.

IV. Methods And Materials

Here, we present the design of a secure healthcare system. The architecture of the system is provided to illustrate the idea of protecting healthcare data transactions and preventing malware attacks.

A. System Design

The architectural framework fig.1 shows a smart healthcare system with three key parts: a strong database server, advanced blockchain tech, and Multilayer perceptron (MLP) algorithm. Together, they protect healthcare data and stop harmful attacks. The system is great at securing normal tasks like sharing data, managing records, and

scheduling appointments. It uses blockchain to make data super secure and transparent. If there's a sneaky attack, the smart deep learning models quickly figure it out and come up with a plan to stop it. This cool approach makes healthcare transactions super safe and helps the whole system handle new online threats. In a world where healthcare data needs strong protection, this setup shows a smart and secure way forward for digital healthcare systems.

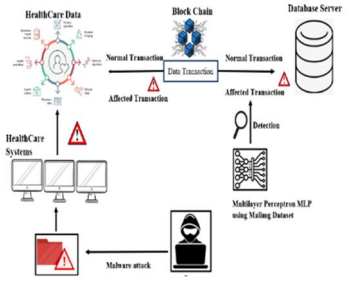


Figure.1 Working model of healthcare system security using blockchain and Multilayer perceptron.

B. Implementation

In the proposed plot, the blockchain empowered conventional malware safeguard framework using Multilayer perceptron deals with the accompanying stages: (i) Healthcare data transaction and storage using blockchain technology, (ii) Classification and behaviour data analysis of malware, (iii) Signature generation of malware, (iv) Malware detection through deep learning, and (v) Mitigation and data recovery.

(i) Healthcare data transaction and storage using blockchain technology

The reinforcements of medical services information of the frameworks are made through blockchain in which information is put away as specific scrambled exchanges. This module explains how blockchain technology can securely store and exchange healthcare data by using decentralized ledgers and encryption methods, ensuring data integrity and privacy while simplifying auditing and enabling interoperability between different systems, although challenges like scalability and regulation still need to be addressed for widespread adoption.

(ii) Classification and behaviour data analysis of malware

In this step, we gather information about different kinds of trojan attacks from various places. This could be from reports, datasets, security databases, or even firsthand experiences. Then, we clean up this information to make sure it's accurate. Next, we create unique signatures for both bad trojan programs and good programs. We also identify special characteristics that can help us detect trojan attacks based on those features. This helps us catch trojans effectively, making our healthcare systems safer from harmful software.

(iii) Signature generation of malware

In this step, we explore a method for generating signatures to identify malware using a machine learning algorithm called Multilayer Perceptron (MLP). Then, we use the MLP algorithm to analyze this data and find patterns that are unique to each type of malware. MLP works by mimicking the way neurons in the human brain function, organizing data into layers of interconnected nodes. These nodes process information and pass it through multiple layers, with each layer extracting increasingly complex features from the data. Through a process called training, the MLP learns to adjust the strength of connections between nodes to minimize errors and improve its ability to classify data accurately. These learned connections and weights form the basis for generating signatures that effectively distinguish between different types of malwares. By using MLP, we can train the malware dataset and create effective signatures that make it easier to detect and protect against harmful software on healthcare networks.

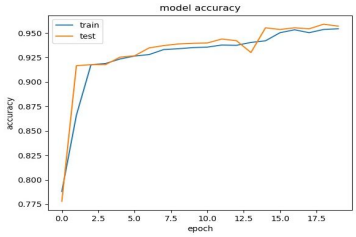


Figure.2 The distinction between testing and training data precision is displayed.

The above fig.2 shows the difference between the accuracy of test and training dataset where the data are trained using MLP. The model is trained till 20 epochs to get better accuracy.

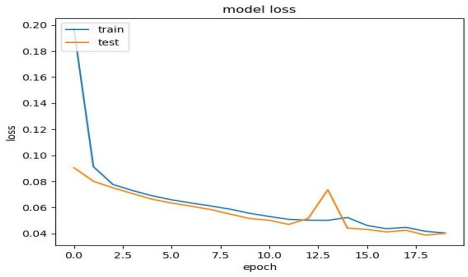


Figure.3 The difference between the test and train accuracy level in loss data graph is displayed.

(iv). Malware detection through MLP

Malware detection through Multilayer Perceptron (MLP) involves training a neural network to recognize patterns indicative of malicious software. The process begins with collecting a diverse dataset containing examples of both benign and malicious files. Features extracted from these files, such as file size, behavior, and code structure, serve as inputs to the MLP model. During training, the MLP learns to distinguish between benign and malicious samples by adjusting the weights and biases of its interconnected layers based on labelled examples. Once trained, the MLP can classify new files as either benign or malicious with high accuracy. This approach offers advantages such as adaptability to evolving malware threats and the ability to detect previously unseen variants.

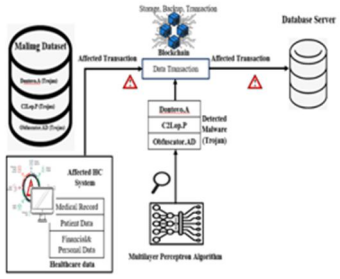


Figure.4 Working model of malware detection system through Multilayer perceptron.

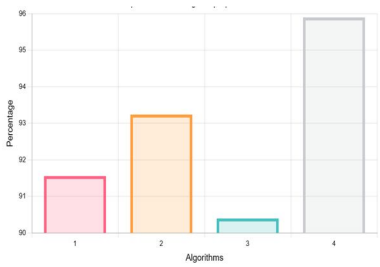
(v) Mitigation and data recovery

Mitigating and recovering from malware in healthcare using Multilayer Perceptron (MLP) involves a few steps. First, we set up systems that use MLP to spot malware by looking for suspicious behaviour in healthcare networks. If malware is detected, we act fast to contain it, isolating infected devices to stop it from spreading. At the same time, we use backup copies to restore any important healthcare data that might have been affected. These MLP-based systems help us quickly find and fix malware issues, so patient care isn't disrupted. We also learn from these incidents to make our systems even better at catching malware in the future, keeping healthcare data safe from evolving threats and concentrating on both detection and prevention frameworks to provide a better security for healthcare system.

V. Comparison Of Existing And Proposed Model

TABLE.I COMPARISON OF EXISTING AND PROPOSED SYSTEM

S.NO.	ALGORITHM	ACCURACY
1.	[1] Recurrent Neural Networks (RNN)	91.56%
2.	[9] Convolutional Neural Networks (CNN)	93.24%
3.	[11] Decision Tree (DT)	90.36%
4.	Proposed Model: Multilayer Perceptron (MLP)	95.69%



Graph.1 - Comparison graph of existing and proposed system

VII. Conclusion

The viability of using blockchain innovation combined with a Multi-Layer Perceptron (MLP) way to deal with the security of medical care information against Trojan assaults. Through exhaustive trial and error and examination, we noticed a huge increase in fruitful Trojan penetration detection with the accuracy of 95.69%, in this manner reinforcing the respectability and privacy of touchy healthcare data. Our outcomes feature the promising capability of coordinating blockchain and MLP techniques as a strong protection component in defending medical services information from developing digital dangers.

VII. Future Enhancement

The major defect of proposed system is considered as the number of epoch while training and testing the dataset using MLP increases, and the time taken for testing and training also increases. To enhance the efficiency of malware detection using an MLP, optimizing the model architecture, employing advanced optimization algorithms, and incorporating techniques like normalization and can reduce training time. We also urge to make our systems even better at malware prevention in the future, keeping healthcare data safe from evolving threats.

References

[1] Ö. Aslan and A. A. Yilmaz, "A New Malware Classification Framework Based on Deep Learning Algorithms," in IEEE Access, vol. 9, pp. 87936-87951, 2021, doi: 10.1109/ACCESS.2021.3089586.

[2] D. Javaheri, P. Lalbakhsh and M. Hosseinzadeh, "A Novel Method for Detecting Future Generations of Targeted and Metamorphic Malware Based on Genetic Algorithm," in IEEE Access, vol. 9, pp. 69951-69970, 2021, doi: 10.1109/ACCESS.2021.3077295.

[3] A. A. Darem, F. A. Ghaleb, A. A. Al-Hashmi, J. H. Abawajy, S. M. Alanazi and A. Y. Al-Rezami, "An Adaptive Behavioral-Based Incremental Batch Learning Malware Variants Detection Model Using Concept Drift Detection and Sequential Deep Learning," in IEEE Access, vol. 9, pp. 97180-97196, 2021, doi: 10.1109/ACCESS.2021.309366.

[4] J. Singh, D. Thakur, T. Gera, B. Shah, T. Abuhmed and F. Ali, "Classification and Analysis of Android Malware Images Using Feature Fusion Technique," in IEEE Access, vol. 9, pp. 90102-90117, 2021, doi: 10.1109/ACCESS.2021.3090998.

[5] C. Cilleruelo, Enrique-Larriba, L. De-Marcos and J. -J. Martinez-Herráiz, "Malware Detection Inside App Stores Based on Lifespan Measurements," in IEEE Access, vol. 9, pp. 119967-119976, 2021, doi: 10.1109/ACCESS.2021.3107903.

[6] M. İbrahim, B. Issa and M. B. Jasser, "A Method for Automatic Android Malware Detection Based on Static Analysis and Deep Learning," in IEEE Access, vol. 10, pp. 117334-117352, 2022, doi: 10.1109/ACCESS.2022.3219047.

[7] A. A. Al-Hashmi et al., "Deep-Ensemble and Multifaceted Behavioral Malware Variant Detection Model," in IEEE Access, vol. 10, pp. 42762-42777, 2022, doi: 10.1109/ACCESS.2022.3168794.

[8] H. N. Nguyen, F. Abri, V. Pham, M. Chatterjee, A. S. Namin and T. Dang, "MalView: Interactive Visual Analytics for Comprehending Malware Behavior," in IEEE Access, vol. 10, pp. 99909-99930, 2022, doi: 10.1109/ACCESS.2022.3207782.

[9] I. Almomani, A. Alkhayer and W. El-Shafai, "An Automated Vision-Based Deep Learning Model for Efficient Detection of Android Malware Attacks," in IEEE Access, vol. 10, pp. 2700-2720, 2022, doi: 10.1109/ACCESS.2022.3140341.

[10] J. Pattee, S. M. Anik and B. K. Lee, "Performance Monitoring Counter Based Intelligent Malware Detection and Design Alternatives," in IEEE Access, vol. 10, pp. 28685-28692, 2022, doi: 10.1109/ACCESS.2022.3157812.

[11] H. Manthena, J. C. Kimmel, M. Abdelsalam and M. Gupta, "Analysing and Explaining Black-Box Models for Online Malware Detection," in IEEE Access, vol. 11, pp. 25237-25252, 2023, doi: 10.1109/ACCESS.2023.3255176.

[12] B. Jin, J. Choi, J. B. Hong and H. Kim, "On the Effectiveness of Perturbations in Generating Evasive Malware Variants," in IEEE Access, vol. 11, pp. 31062-31074, 2023, doi: 10.1109/ACCESS.2023.3262265.

[13] T. He, C. Han, R. Isawa, T. Takahashi, S. Kijima and J. Takeuchi, "Scalable and Fast Algorithm for Constructing Phylogenetic Trees with Application to IoT Malware Clustering," in IEEE Access, vol. 11, pp. 8240-8253, 2023, doi: 10.1109/ACCESS.2023.3238711.

[14] M. M. Belal and D. M. Sundaram, "Global-Local Attention-Based Butterfly Vision Transformer for Visualization-Based Malware Classification," in IEEE Access, vol. 11, pp. 69337-69355, 2023, doi: 10.1109/ACCESS.2023.3293530.

[15] I. Gulatas, H. H. Kilinc, A. H. Zaim and M. A. Aydin, "Malware Threat on Edge/Fog Computing Environments from Internet of Things Devices Perspective," in IEEE Access, vol. 11, pp. 33584-33606, 2023, doi: 10.1109/ACCESS.2023.3262614.