symmetry

MDPI

Article

Malware Analysis and Detection Using Machine Learning Algorithms

Muhammad Shoaib Akhtar and Tao Feng *

School of Computer and Communication, Lanzhou University of Technology, Lanzhou 730050, China
* Correspondence: fengt@lut.edu.cn

Abstract: One of the most significant issues facing internet users nowadays is malware. Polymorphic malware is a new type of malicious software that is more adaptable than previous generations of viruses. Polymorphic malware constantly modifies its signature traits to avoid being identified by traditional signature-based malware detection models. To identify malicious threats or malware, we used a number of machine learning techniques. A high detection ratio indicated that the algorithm with the best accuracy was selected for usage in the system. As an advantage, the confusion matrix measured the number of false positives and false negatives, which provided additional information regarding how well the system worked. In particular, it was demonstrated that detecting harmful traffic on computer systems, and thereby improving the security of computer networks, was possible using the findings of malware analysis and detection with machine learning algorithms to compute the difference in correlation symmetry (Naive Byes, SVM, J48, RF, and with the proposed approach) integrals. The results showed that when compared with other classifiers, DT (99%), CNN (98.76%), and SVM (96.41%) performed well in terms of detection accuracy. DT, CNN, and SVM algorithms' performances detecting malware on a small FPR (DT = 2.01%, CNN = 3.97%, and SVM = 4.63%,) in a given dataset were compared. These results are significant, as malicious software is becoming increasingly common and complex.

Keywords: technological innovation; malicious threats; CNN; SVM; DT; cybersecurity; cyberattack; cyber warfare; cyber threats; suspicious activity

check for updates

Citation: Akhtar, M.S.; Feng, T. Malware Analysis and Detection Using Machine Learning Algorithms. *Symmetry* **2022**, *14*, 2304. https://doi.org/10.3390/sym14112304

Academic Editors: Sergei D. Odintsov and Mihai Postolache

Received: 7 September 2022
Accepted: 26 October 2022
Published: 3 November 2022

1. Introduction

Cyberattacks are currently the most pressing concern in the realm of modern technology. The word implies exploiting a system's flaws for malicious purposes, such as stealing from it, changing it, or destroying it. Malware is an example of a cyberattack. Malware is any program or set of instructions that is designed to harm a computer, user, business, or computer system [1]. The term "malware" encompasses a wide range of threats, including viruses, Trojan horses, ransomware, spyware, adware, rogue software, wipers, scareware, and so on. Malicious software, by definition, is any piece of code that is run without the user's knowledge or consent [2].

In particular, this study demonstrated that detecting harmful traffic on computer systems, and thereby improving the security of computer networks, was possible employing the findings of malware analysis and detection with machine learning algorithms to compute the difference in correlation symmetry (Naive Byes, SVM, J48, RF, and with the proposed approach) integrals.

Malware detection modules are responsible for analysing data they have collected and been trained with to determine whether or not a specific piece of software or network connection constitutes a security concern [3,4]. As an illustration, consider a machine learning system that can explicitly express the principles that underlie the patterns it has observed [5]. Algorithms that have been trained by machine learning systems can improve their ability to predict using feedback regarding how well they performed on previous tasks and using that information to make changes [6].

Worldwide, cybercriminals pose a serious threat to businesses, universities, governments, and individuals through the use of malicious software and the theft of confidential data [7]. Every day, thousands of fraudsters employ harmful software in an attempt to gain access to networks, steal data, or transfer money. As a result, keeping sensitive information safe has become an urgent concern in the scientific world. This study aimed to provide a comprehensive framework for discovering malicious programs and protecting private information from hackers by employing data mining and machine learning classification approaches. In this paper, we analyse signature-based and anomaly-based features to develop a robust and effective approach to malware classification and detection. Experiments have proven that the recommended technique is preferable to alternatives [7].

Modern malware has become increasingly common and complex, posing a major threat to the security of modern websites [8]. Figure 1 depicts types of cyberattacks in the digital world or cyberspace. Malware is software created with the express purpose of causing harm to a computer or network, for example, by monitoring its users or stealing their money. Malware attacks are becoming increasingly common and now even affect IoT devices, medical gear, and environmental and industrial control systems. Modern spyware is notoriously hard to detect, as it constantly updates its code and behaviour. The proliferation of malware has rendered traditional signature-based defenses ineffective. Instead, it is necessary to take a broader range of defensive actions [9].

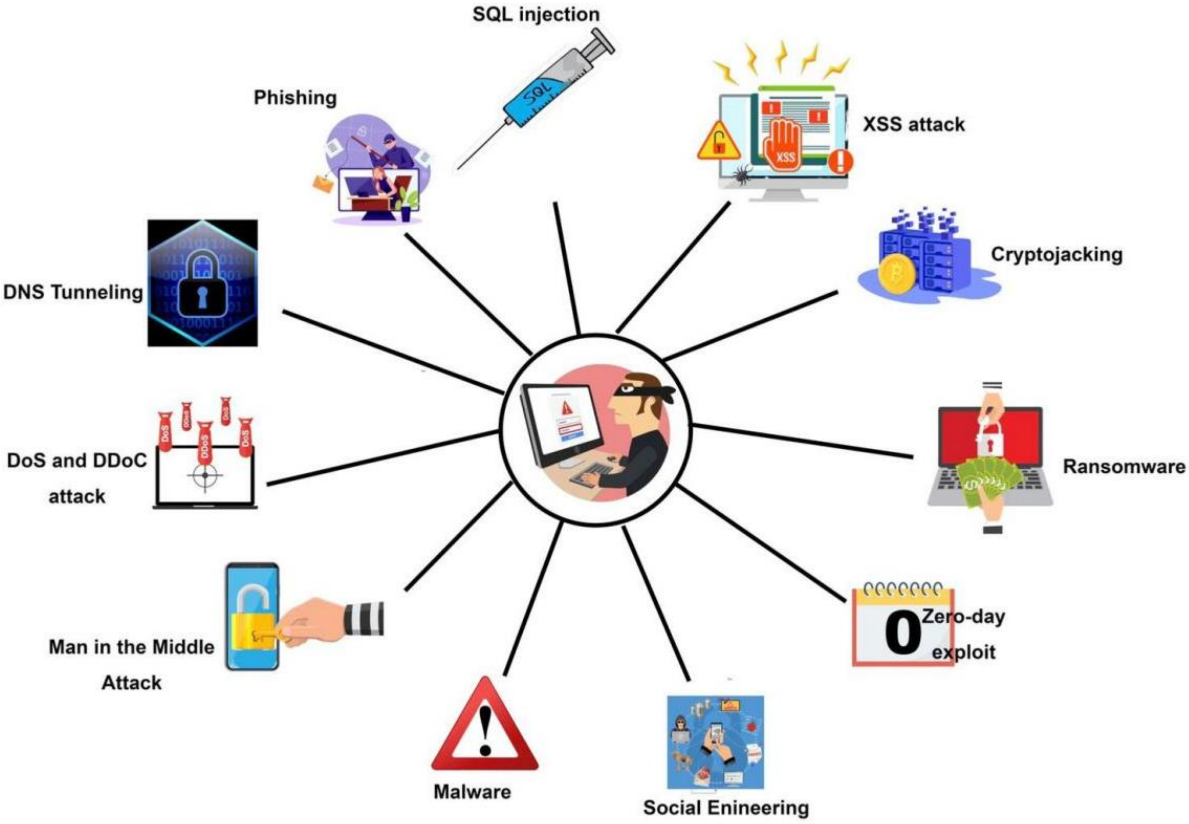

Figure 1. Types of cyberattacks.

Both static and dynamic learning methods may be used to identify behavioural similarities between members of the same family of malware [10]. Unlike static analysis, which examines dangerous files' contents without actually running them, dynamic analysis takes their behaviour into account by tracking data flows, recording function calls, and adding monitoring code to dynamic binaries [11]. Machine learning algorithms may leverage such static and behavioural artefacts to describe the ever-evolving structure of contemporary malware, allowing them to identify increasingly complex malware assaults that could otherwise avoid detection using signature-based techniques. As machine learning-based solutions do not rely on signatures, they are more successful against newly released malware. Deep learning algorithms that can perform feature engineering on their own can be used to obtain and represent features more accurately [12].

Figure 2 illustrates the Martin (2018) Cyber Kill Chain used for cyberattack protection and as for security measure to protect networks. In February of 2020, AWS was the target of a large-scale distributed denial of service attack [13]. The organisation withstood a DDoS attack of 2.3 Tbps, which resulted in a packet forwarding rate of 293.1 Mpps and a request rate of 694,201. Some have claimed it to be the largest known DDoS attack. In July of 2020, three hackers gained access to Twitter and took over a number of prominent users' accounts. President Obama, Amazon's Jeff Bezos, and Tesla's Elon Musk are just a few of the notables whose accounts were hacked. Bitcoin scams uploaded from the stolen accounts generated over $100,000 in profits. Two weeks after these events, the US Justice Department filed charges against three individuals, the youngest of whom was 17 at the time. It was disclosed in 2018 that a cyberattack at Marriott's Starwood Hotels had exposed the personal information of more than 500 million customers [14]. According to data collected by NHS England, the 2017 WannaCry ransomware attack affected more than 300,000 systems in 150 countries and cost billions to fix [15].

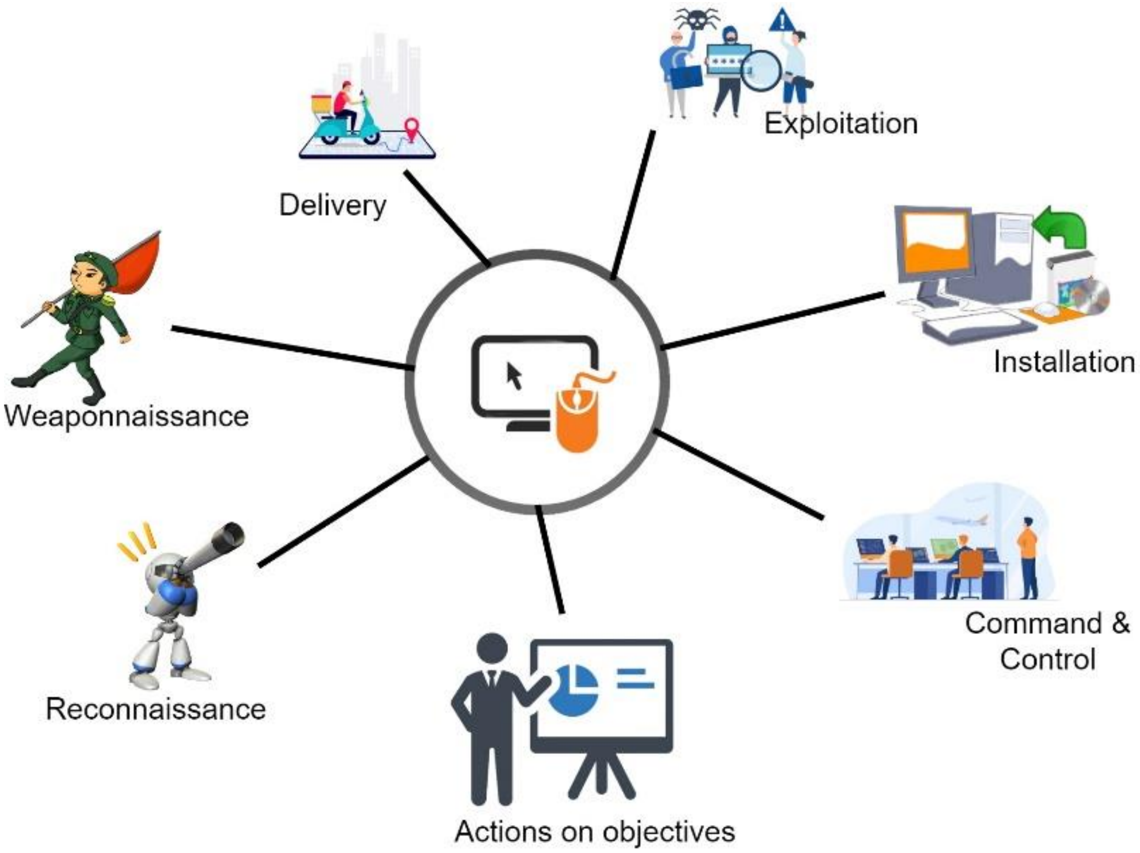

Figure 2. Martin Cyber Kill Chain for prevention of cyber intrusions activity.

As part of its ongoing attempt to destabilize its neighbours, Russia launched a cyberattack on Ukrainian electricity infrastructure in 2017 [16]. This attack showcased Russia's capacity for large-scale cyber warfare for the first time. Despite the fact that it was carried out a full year after Russia's seizure of Crimea, which is widely regarded as the formal beginning of Russia's conflict with the Ukraine, this complex attack was the first successful cyberattack on a power infrastructure [17]. The Russian cyber military unit Sandworm launched an attack on the command centre; the command centre's vulnerability allowed the hackers to seize control of the substation's computer systems, bringing it down. Shortly after, attacks on other substations occurred. It is estimated that between 200,000 and 300,000 people will have ultimately been hurt by the attack [18].

2. Literature Review

The proliferation of computers, smartphones, and other Internet-enabled gadgets leaves the world vulnerable to cyber assaults. A plethora of malware detection methods have arisen in response to the explosion in malware activity. When trying to identify malicious code, researchers use a variety of big data tools and machine learning techniques. Traditional machine learning-based malware detection approaches have a considerable processing time, but may effectively identify newly emerging malware. Feature engineering may become obsolete due to the prevalence of modern machine learning algorithms, such as deep learning. In this study, we examined a variety of malware detection and classification techniques. Researchers have created ways to use machine learning and deep learning to check samples for malicious intent [19].

Armaan (2021) illustrated and tested the accuracy of various models. Without data, no application built for a digital platform can perform its function [20]. There are several cyber risks, so it is essential that precautions be taken to safeguard data. Although feature selection is difficult when developing a model of any sort, machine learning is a cutting-edge approach that paves the way for precise prediction. The approach needs a workaround that is adaptable enough to handle non-standard data. To effectively manage and prevent future assaults, we must analyse malware and create new rules and patterns in the form of creation of malware type as shown in Table 1 [21]. To find patterns, IT security professionals may use malware analysis tools. The availability of technologies that analyse malware samples and determine their level of malignancy significantly benefit the cybersecurity sector. These tools help monitor security alerts and prevent malware attacks. If malware is dangerous, we must eliminate it before it transmits its infection any further. Malware analysis is becoming increasingly popular as it helps businesses lessen the effects of the growing number of malware threats and the increasing complexity of the ways malware can be used to attack [22].

Table 1. Dataset file types.

File Type		No. of Files
Malware	Backdoor	3654
	Rootkit	2834
	Virus	921
	Trojan	2563
	Exploit	652
	Work	921
	Others	3138
Cleanware		2711
Total		17,394

Chowdhury (2018) proposed a viable malware detection approach that uses a machine learning classification technique. We explored whether or not adjusting a few parameters might increase the accuracy with which malware is classified [23]. N-gram and API call capabilities were incorporated into our approach. Experimental evaluation confirmed the efficacy and dependability of our proposed technique. Future work will focus on merging a large number of features to increase detection precision while decreasing false positives. Performance results for competing approaches are shown in Table 2; our Chowdhury [23] approach was clearly superior.

Table 2. Classifiers results comparisons.

Methods	Accuracy (%)	TPR (%)	FPR (%)
KNN	95.02	96.17	3.42
CNN	98.76	99.22	3.97
Naïve Byes	89.71	90	13
Random Forest	92.01	95.9	6.5
SVM	96.41	98	4.63
DT	99	99.07	2.01

At this time, the proliferation of malicious software poses a significant threat to global stability. In the 1990s, as the number of interconnected computers exploded, so did the prevalence of malicious software [23], which eventually led to the widespread distribution of malware. Multiple protective measures have been created in response to this phenomenon. Unfortunately, current safeguards cannot keep up with modern threats that malware authors have created to thwart security programs. In recent years, researchers' focus on malware detection research has shifted toward ML algorithm strategies. In this research paper, we present a protective mechanism that evaluates three ML algorithm approaches to malware detection and chooses the most appropriate one. According to statistics, the decision tree approach has the maximum detection accuracy (99.01%) and the lowest false positive rate (FPR; 0.021%) on a small dataset.

Malware continues to develop and propagate at an alarming rate. Nur (2019) compared three ML classifiers to analyse and quantify the detection accuracy of the ML classifier that used static analysis to extract features based on PE information. As a group, we trained machine learning algorithms to recognise dangerous versus benign information [24]. The DT machine learning method attained 99% accuracy, as illustrated in Table 2 making it the most successful classifier we examined. This experiment demonstrated the potential of static analysis based on PE information and chosen key data features to achieve the highest detection accuracy and the most accurate depiction of malware.

Malicious programs and their threats, or "malware," became increasingly common and sophisticated as the Internet developed. Their rapid dispersion over the Internet has provided malware authors with access to a wide variety of malware generation tools [25]. Every day, the reach and sophistication of malware grows. This study focused on analysing and measuring classifier performance to better understand how machine learning works. Latent analysis extracted features from the recovered PE file and library information; six classifiers based on ML techniques were evaluated. It was recommended that ML systems be trained and tested to determine whether or not a file is harmful. Experimental outcomes verified that the random forest method is preferable for data categorization, with 99.4 percent accuracy. These results showed that the PE library was compatible with static analysis and that focusing on only a few properties could improve malware detection and characterization. The main benefit is that it is less likely that malicious software will be installed by accident, as users can check a file's validity before opening it [26].

3. Research Problem

Malware's potentially harmful components can be detected using either static analysis or dynamic analysis. Static analysis, such as the reverse-engineering method used to disassemble a virus, focuses on parsing malware binaries to discover harmful strings [27]. However, dynamic analysis entails monitoring dangerous software even as it operates in a controlled environment, such as a virtual computer. Both methods have their advantages and disadvantages; however, when analysing malware, it is best to use both [28]. It is possible that reducing the number of dangerous features would improve the accuracy of malware detection. The researcher would then have more time to analyse collected data. We are concerned that a large number of characteristics are being used to detect malware when fewer, more robust characteristics might do the job just as well. The process of choosing which malicious features to implement begins with discovering possible methods or algorithms. We need solutions that can both find malware that has never been seen before and greatly reduce the number of characteristics that are currently needed to do so [29].

H1. *Evaluation of the higher accuracy among three ML methods for malware detection: DT, CNN, and SVM.*

4. Methodology

This research paper introduces the various steps and components of a typical machine learning workflow for malware detection and classification, explores the challenges and limitations of such a workflow, and assesses the most recent innovations and trends in the field, with an emphasis on deep learning techniques. The proposed research methodology of this research study is provided below [30].

To provide a more complete understanding of the proposed machine learning method for malware detection, Figures 3 and 4 illustrate the workflow process from start to finish.

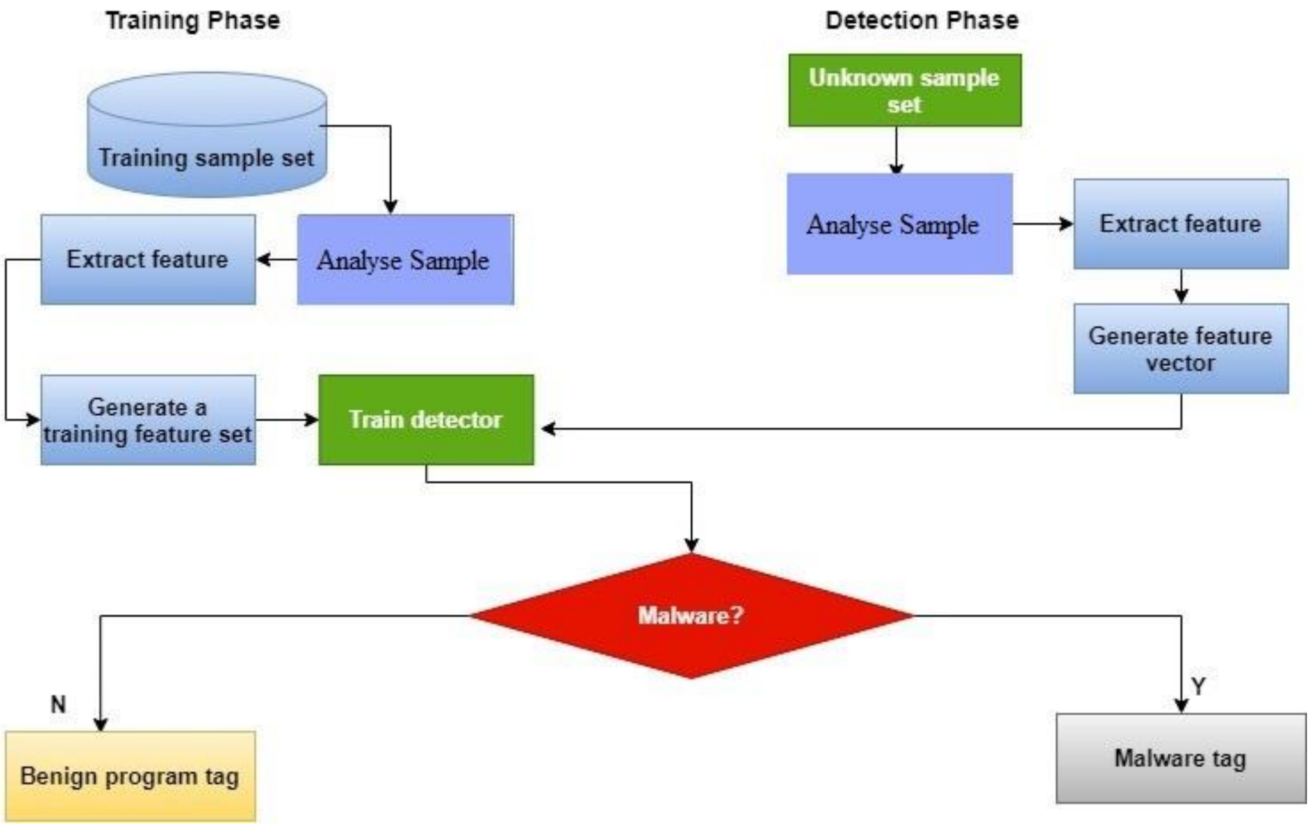

Figure 3. Proposed ML malware detection method.

4.1. Dataset

This study relied entirely on data provided by the Canadian Institute for Cybersecurity. The collection has many data files that include log data for various types of malware [31]. These recovered log features may be used to train a broad variety of models. Approximate 51 distinct malware families were found in the samples. More than 17,394 data points from different locations were included; the dataset had 279 columns and 17,394 rows.

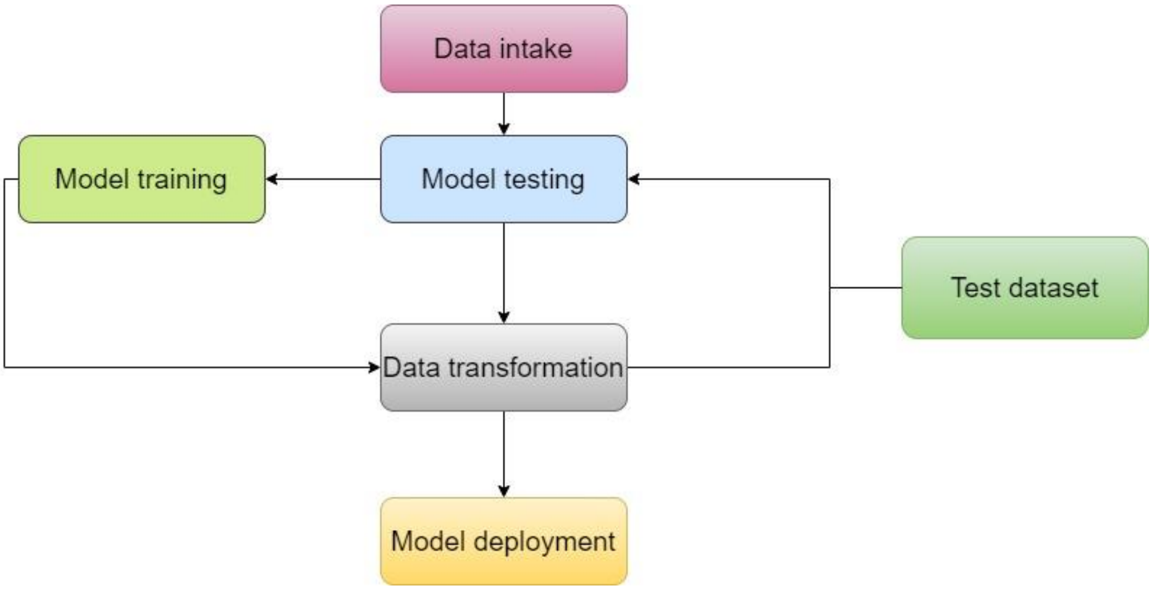

Figure 4. Workflow process illustration.

4.2. Pre-Processing

Data were stored in the file system as binary code, and the files themselves were unprocessed executables. We prepared them in advance of our research. Unpacking the executables required a protected environment, or virtual machine (VM). PEiD software automated unpacking of compressed executables [32].

4.3. Features Extraction

Twentieth-century datasets frequently contain tens of thousands of features. In recent years, as feature counts have grown, it has become clear that the resultant machine learning model has been overfit [33]. To address this problem, we built a smaller set of features from a larger set; this technique is commonly used to maintain the same degree of accuracy while using fewer features. The goal of this study was to refine the existing dataset of dynamic and static features by keeping those that were most helpful and eliminating those that were not valuable for data analysis [34].

4.4. Features Selection

After completing feature extraction, which involved the discovery of more features, feature selection was performed. Feature selection was a crucial process for enhancing accuracy, simplifying the model, and reducing overfitting, as it involved choosing features from a pool of newly recognised qualities. Researchers have used many feature classification strategies in the past in an effort to identify dangerous code in software. As the feature rank technique is very effective at picking the right features for building malware detection models, it was extensively employed in this study [35,36].

5. Results and Discussion

The two main phases of the classification process were training and testing. To train a system, it was sent both harmful and safe files [37]. Automated classifiers were taught using a learning algorithm. Each classifier (KNN, CNN, NB, RF, SVM, or DT) became smarter with each set of data it annotated. In the testing phase, a classifier was sent a collection of new files, some harmful and some not; the classifier determined whether the files were malicious or clean [38].

Logistic Regression

Figure 5 Illustrates that DT had the highest accuracy (99%) and TPR (99.07%), and that FPR had the lowest accuracy (2.01%). It is clear from the confusion matrix that DT had a higher accuracy than all other (KNN, CNN, NB, RF, and SVM) machine learning algorithms or classifiers [39].

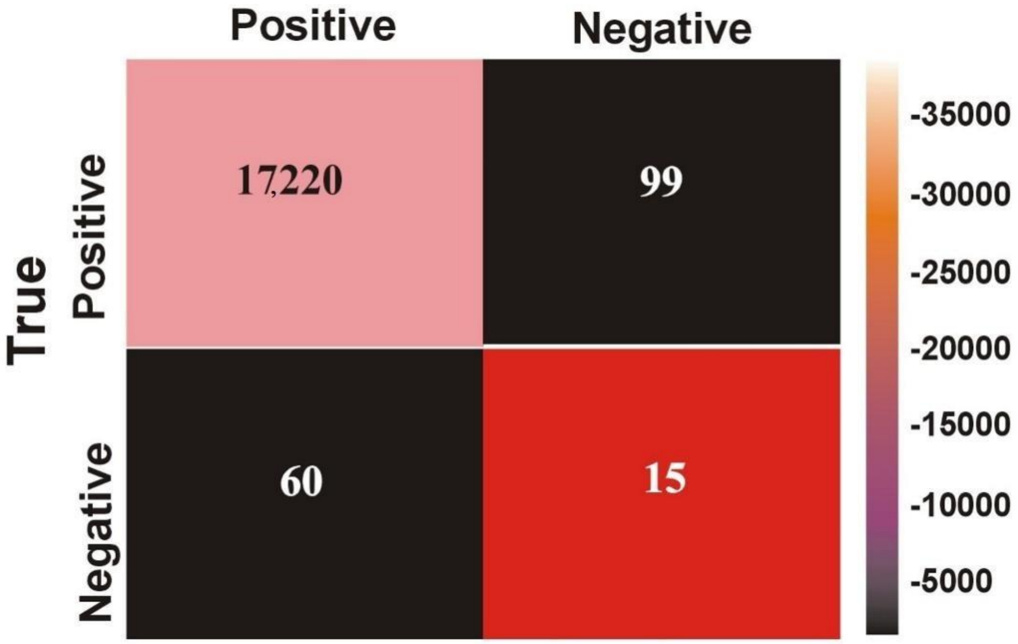

Figure 5. Confusion Matrix.

Our suggested method for malware categorization and detection was experimentally evaluated using the gathered malware and cleanware [40]. We used supervised machine learning algorithms or classifiers (KNN, CNN, NB, RF, SVM, and DT) to examine malware and characterise it.

Through statistical analysis of Table 2's results, we deduced that results of classifiers' accuracy (KNN = 95.02%, CNN = 98.76%, Naïve Byes = 89.71%, Random Forest = 92.01%, SVM = 96.41%, and DT = 99%) showed that DT was the optimal model for the malware detection strategy. Classifiers' TPRs (%) (KNN = 96.17%, CNN = 99.22%, Naïve Byes = 90%, Random Forest = 95.9%, SVM = 98%, and DT = 99.07%) showed that CNN was the second optimal model for the detection and identification of malware, and that SVM was the third optimal model for malware detection. Table 2 shows the classifiers' FPRs (%) (KNN = 3.42%, CNN = 3.97%, Naïve Byes = 13%, Random Forest = 6.5%, SVM = 4.63%, and DT = 2.01%). We presumed that CNN, SVM, DT and KNN classifiers had comparable high accuracy and performance for all intents and purposes. It is clear that using the three most optimal algorithms (DT = 99%, SVM = 96.41%, and CNN = 98.76%), which had a much higher TPR (%) rate and accuracy, to identify malware DT accuracy is highest and DT is better choice for malware detection.

6. Conclusions

This paper demonstrates that academics have recently shown a growing interest in ML algorithm solutions for malware identification. We presented a protective mechanism that evaluated three ML algorithm approaches to malware detection and chose the most appropriate one. The results show that compared with other classifiers, DT (99%), CNN (98.76%), and SVM (96.41%) performed well in terms of detection accuracy. DT, CNN, and SVM algorithms' performances detecting malware on a small FPR (DT = 2.01%, CNN = 3.97%, and SVM = 4.63%,) in a given dataset were compared. In this experiment, we evaluated and quantified the detection accuracy of a machine learning (ML) classifier that used static analysis to extract features based on PE data by comparing it to two other ML classifiers. As a result of our efforts, machine learning algorithms can now identify dangerous versus benign data. The DT machine learning method had the highest accuracy (99%) of any classifier we evaluated. In addition to potentially providing the highest detection accuracy and accurately characterising malware, static analysis based on PE information and carefully selected data showed promise in experimental findings. That we do not have to execute anything to determine if data are malicious is a significant benefit. The three ML models (DT, CNN, and SVM) were trained, tested, and their efficiency compared using the dataset obtained from the Canadian Institute for Cybersecurity.

Author Contributions: M.S.A. and T.F. contributed equally to the study's conception. All authors have read and agreed to the published version of the manuscript.

Funding: National Natural Science Foundation of China (No. 62162039) and (61762060), and the Key Research and Development Program of Gansu Province (No. 20YF3GA016).

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The data used to support the findings of this study are available from the corresponding author upon request.

Conflicts of Interest: The authors declare that they have no conflict of interest.

Abbreviations

CNN	Convolutional Neural Network
FPR	False Positive Rate
RBM	Restricted Boltzmann Machine
DT	Decision Tree
SVM	Support Vector Machine
VM	Virtual Machine

References

1. Nikam, U.V.; Deshmuh, V.M. Performance evaluation of machine learning classifiers in malware detection. In Proceedings of the 2022 IEEE International Conference on Distributed Computing and Electrical Circuits and Electronics (ICDCECE), Ballari, India, 23–24 April 2022; pp. 1–5. [CrossRef]
2. Akhtar, M.S.; Feng, T. IOTA based anomaly detection machine learning in mobile sensing. *EAI Endorsed Trans. Create. Tech.* **2022**, *9*, 172814. [CrossRef]
3. Sethi, K.; Kumar, R.; Sethi, L.; Bera, P.; Patra, P.K. A novel machine learning based malware detection and classification framework. In Proceedings of the 2019 International Conference on Cyber Security and Protection of Digital Services (Cyber Security), Oxford, UK, 3–4 June 2019; pp. 1–13.
4. Abdulbasit, A.; Darem, F.A.G.; Al-Hashmi, A.A.; Abawajy, J.H.; Alanazi, S.M.; Al-Rezami, A.Y. An adaptive behavioral-based increamental batch learning malware variants detection model using concept drift detection and sequential deep learning. *IEEE Access* **2021**, *9*, 97180–97196. [CrossRef]
5. Feng, T.; Akhtar, M.S.; Zhang, J. The future of artificial intelligence in cybersecurity: A comprehensive survey. *EAI Endorsed Trans. Create. Tech.* **2021**, *8*, 170285. [CrossRef]
6. Sharma, S.; Krishna, C.R.; Sahay, S.K. Detection of advanced malware by machine learning techniques. In Proceedings of the SoCTA 2017, Jhansi, India, 22–24 December 2017.
7. Chandrakala, D.; Sait, A.; Kiruthika, J.; Nivetha, R. Detection and classification of malware. In Proceedings of the 2021 International Conference on Advancements in Electrical, Electronics, Communication, Computing and Automation (ICAECA), Coimbatore, India, 8–9 October 2021; pp. 1–3. [CrossRef]
8. Zhao, K.; Zhang, D.; Su, X.; Li, W. Fest: A feature extraction and selection tool for android malware detection. In Proceedings of the 2015 IEEE Symposium on Computers and Communication (ISCC), Larnaca, Cyprus, 6–9 July 2015; pp. 714–720.
9. Akhtar, M.S.; Feng, T. Detection of sleep paralysis by using IoT based device and its relationship between sleep paralysis and sleep quality. *EAI Endorsed Trans. Internet Things* **2022**, *8*, e4. [CrossRef]
10. Gibert, D.; Mateu, C.; Planes, J.; Vicens, R. Using convolutional neural networks for classification of malware represented as images. *J. Comput. Virol. Hacking Tech.* **2019**, *15*, 15–28. [CrossRef]

11. Firdaus, A.; Anuar, N.B.; Karim, A.; Faizal, M.; Razak, A. Discovering optimal features using static analysis and a genetic search based method for Android malware detection. *Front. Inf. Technol. Electron. Eng.* **2018**, *19*, 712–736. [CrossRef]
12. Dahl, G.E.; Stokes, J.W.; Deng, L.; Yu, D.; Research, M. Large-scale Malware Classification Using Random Projections And Neural Networks. In Proceedings of the International Conference on Acoustics, Speech and Signal Processing-1988, Vancouver, BC, Canada, 26–31 May 2013; pp. 3422–3426.
13. Akhtar, M.S.; Feng, T. An overview of the applications of artificial intelligence in cybersecurity. *EAI Endorsed Trans. Create. Tech.* **2021**, *8*, e4. [CrossRef]
14. Akhtar, M.S.; Feng, T. A systemic security and privacy review: Attacks and prevention mechanisms over IOT layers. *EAI Endorsed Trans. Secur. Saf.* **2022**, *8*, e5. [CrossRef]
15. Anderson, B.; Storlie, C.; Lane, T. "Improving Malware Classification: Bridging the Static/Dynamic Gap. In Proceedings of the 5th ACM Workshop on Security and Artificial Intelligence (AISec), Raleigh, NC, USA, 19 October 2012; pp. 3–14.
16. Varma, P.R.K.; Raj, K.P.; Raju, K.V.S. Android mobile security by detecting and classification of malware based on permissions using machine learning algorithms. In Proceedings of the 2017 International Conference on I-SMAC (IoT in Social, Mobile, Analytics and Cloud) (I-SMAC), Palladam, India, 10–11 February 2017; pp. 294–299.
17. Akhtar, M.S.; Feng, T. Comparison of classification model for the detection of cyber-attack using ensemble learning models. *EAI Endorsed Trans. Scalable Inf. Syst.* **2022**, *9*, 17329. [CrossRef]
18. Rosmansyah, W.Y.; Dabarsyah, B. Malware detection on Android smartphones using API class and machine learning. In Proceedings of the 2015 International Conference on Electrical Engineering and Informatics (ICEEI), Denpasar, Indonesia, 10–11 August 2015; pp. 294–297.
19. Tahtaci, B.; Canbay, B. Android Malware Detection Using Machine Learning. In Proceedings of the 2020 Innovations in Intelligent Systems and Applications Conference (ASYU), Istanbul, Turkey, 15–17 October 2020; pp. 1–6.
20. Baset, M. Machine Learning for Malware Detection. Master's Dissertation, Heriot Watt University, Edinburg, Scotland, December 2016. [CrossRef]
21. Akhtar, M.S.; Feng, T. Deep learning-based framework for the detection of cyberattack using feature engineering. *Secur. Commun. Netw.* **2021**, *2021*, 6129210. [CrossRef]
22. Altaher, A. Classification of android malware applications using feature selection and classification algorithms. *VAWKUM Trans. Comput. Sci.* **2016**, *10*, 1. [CrossRef]
23. Chowdhury, M.; Rahman, A.; Islam, R. *Malware Analysis and Detection Using Data Mining and Machine Learning Classification*; AISC: Chicago, IL, USA, 2017; pp. 266–274.
24. Patil, R.; Deng, W. Malware Analysis using Machine Learning and Deep Learning techniques. In Proceedings of the 2020 SoutheastCon, Raleigh, NC, USA, 28–29 March 2020; pp. 1–7.
25. Gavriluţ, D.; Cimpoesu, M.; Anton, D.; Ciortuz, L. Malware detection using machine learning. In Proceedings of the 2009 International Multiconference on Computer Science and Information Technology, Mragowo, Poland, 12–14 October 2009; pp. 735–741.
26. Pavithra, J.; Josephin, F.J.S. Analyzing various machine learning algorithms for the classification of malwares. *IOP Conf. Ser. Mater. Sci. Eng.* **2020**, *993*, 012099. [CrossRef]
27. Vanjire, S.; Lakshmi, M. Behavior-Based Malware Detection System Approach For Mobile Security Using Machine Learning. In Proceedings of the 2021 International Conference on Artificial Intelligence and Machine Vision (AIMV), Gandhinagar, India, 24–26 September 2021; pp. 1–4.
28. Agarkar, S.; Ghosh, S. Malware detection & classification using machine learning. In Proceedings of the 2020 IEEE International Symposium on Sustainable Energy, Signal Processing and Cyber Security (iSSSC), Gunupur Odisha, India, 16–17 December 2020; pp. 1–6.
29. Sethi, K.; Chaudhary, S.K.; Tripathy, B.K.; Bera, P. A novel malware analysis for malware detection and classification using machine learning algorithms. In Proceedings of the 10th International Conference on Security of Information and Networks, Jaipur, India, 13–15 October 2017; pp. 107–113.
30. Ahmadi, M.; Ulyanov, D.; Semenov, S.; Trofimov, M.; Giacinto, G. Novel feature ex-traction, selection and fusion for effective malware family classification. In Proceedings of the sixth ACM conference on data and application security and privacy, New Orleans, LA, USA, 9–11 March 2016; pp. 183–194.
31. Damshenas, M.; Dehghantanha, A.; Mahmoud, R. A survey on malware propagation, analysis and detec-tion. *Int. J. Cyber-Secur. Digit. Forensics* **2013**, *2*, 10–29.
32. Saad, S.; Briguglio, W.; Elmiligi, H. The curious case of machine learning in malware detection. *arXiv* **2019**, arXiv:1905.07573.
33. Selamat, N.; Ali, F. Comparison of malware detection techniques using machine learning algorithm. *Indones. J. Electr. Eng. Comput. Sci.* **2019**, *16*, 435. [CrossRef]
34. Firdausi, I.; Lim, C.; Erwin, A.; Nugroho, A. Analysis of machine learning techniques used in behavior-based malware detection. In Proceedings of the 2010 Second International Conference on Advances in Computing, Control, and Telecommunication Technologies, Jakarta, Indonesia, 2–3 December 2010; pp. 201–203. [CrossRef]
35. Hamid, F. Enhancing malware detection with static analysis using machine learning. *Int. J. Res. Appl. Sci. Eng. Technol.* **2019**, *7*, 38–42. [CrossRef]

36. Prabhat, K.; Gupta, G.P.; Tripathi, R. TP2SF: A trustworthy privacy-preserving secured framework for sustainable smart cities by leveraging blockchain and machine learning. *J. Syst. Archit.* **2021**, *115*, 101954.
37. Kumar, P.; Gupta, G.P.; Tripathi, R. A distributed ensemble design based intrusion detection system using fog computing to protect the internet of things networks. *J. Ambient Intell. Human. Comput.* **2021**, *12*, 9555–9572. [CrossRef]
38. Prabhat, K.; Gupta, G.P.; Tripathi, R. Design of anomaly-based intrusion detection system using fog computing for IoT network. *Aut. Control Comp. Sci.* **2021**, *55*, 137–147. [CrossRef]
39. Prabhat, K.; Tripathi, R.; Gupta, G.P. P2IDF: A Privacy-preserving based intrusion detection framework for software defined Internet of Things-Fog (SDIoT-Fog). In Proceedings of the Adjunct Proceedings of the 2021 International Conference on Distributed Computing and Networking (ICDCN '21), Nara, Japan, 5–8 January 2021; pp. 37–42. [CrossRef]
40. Kumar, P.; Gupta, G.P.; Tripathi, R. PEFL: Deep privacy-encoding-based federated learning framework for smart agriculture. *IEEE Micro* **2022**, *42*, 33–40. [CrossRef]