THE ASIAN BULLETIN OF BIG DATA MANAGMENT
Vol. 4. Issue 3 (2024)
https://doi.org/10.62019/abbdm.v4i3.209

Efficient Malware Investigation and Recognition Using Machine Learning Algorithms

Ali Ahmad Siddiqui*, Israr Ali, Saira Arbab, Shilpa Kumari

Chronicle

Article history

Received: Aug 20, 2024
Received in the revised format: Sept 8, 2024
Accepted: Sept 10, 2024
Available online: Sept 11, 2024

Ali Ahmad Siddiqui, Israr Ali, Saira Arbab, Shilpa Kumari are currently affiliated with Department of Computer Science, Iqra University, Karachi, Pakistan.
Email: alisiddiqui@iqra.edu.pk
Email: Israr.ali@iqra.edu.pk
Email: sairarbab@iqra.edu.pk
Email: shilpa@iqra.edu.pk

Corresponding Author*

Abstract

Malware is unique of the biggest problems that modern internet users have to deal with. Polymorphic malware is a new type of harmful software that is extra pliable than prior peers of bugs. Polymorphic malware continuously alters its signature characteristics in order to evade detection by conventional malware detection techniques. We applied various machine learning algorithms to detect malware or dangerous threats. A high detection ratio meant that the most accurate algorithm had been chosen to be used within the system. One advantage of the confusion matrix is its ability to track false positives and false negatives, providing deeper insights into the system's performance. In particular, it revealed that machine learning algorithms like Naïve Bayes, Support Vector Machine (SVLM), Random Forest (RF), and K-Nearest Neighbor (kNN) can be used to detect harmful traffic on computer systems by calculation changes in correlation patterns. This approach enhances the effectiveness of malware detection and overall security in computer networks. The findings demonstrated that NB (87%), kNN (91.76%), SVM (92.41%), and RF (98.07%) performed well in terms of detection accuracy when compared to other classifiers. These findings are important as malicious software is growing more prevalent and sophisticated.

Keywords: Blockchains, Corporate Governance, Legal Framework, Smart Contract, Legal & Ethical considerations.

INTRODUCTION

Nowadays, the biggest worry in the world of modern technology is cyberattacks. The phrase suggests taking advantage of a system's weaknesses to do evil deeds, such as stealing from, altering, or destroying it. Cyberattacks can take the form of malware. Malware refers to any software or gathering of commands calculated to injury a processor, operator, organization, or processor system (Nikam, U.V et al., 2022). Threats like as bugs, Trojan horses, spyware, rogue software, adware, ransomware, scareware, wipers, and more are all included under the umbrella term "malware." By definition, any cipher that is executed without the user's knowledge or consent is considered malicious software (Akhtar, M.S et al., 2022). This study explicitly demonstrated that it was possible to detect hazardous circulation on processer organizations and so improve the sanctuary of computer nets using machine learning methods to calculate the alteration in correlation symmetry integrals. Malware detection modules are accountable for evaluating gathered and learned data to ascertain the potential security risk associated with a certain software or network connection (Sethi, K et al.,2019), (Abdulbasit, A. et al., 2021). A machine learning scheme that can clearly eloquent the concepts fundamental the designs it has seen (Feng, T.; at al., 2021). Machine learning systems can train algorithms to become more predictive by providing feedback on how well they did on earlier tasks. The algorithms can then use this knowledge to make adjustments (Sharma,

S et al., 2017). Through the deployment of harmful software and the theft of private information, cybercriminals constitute a severe threat to individuals, governments, organizations, and academic institutions worldwide (Chandrakala, D at al., 2021). Each day, thousands of con artists use malicious software to try and access networks, steal information, or send money. Because of this, protecting sensitive data has appeared as a top importance in the technical community. The goal of this research was to contemporary a thorough agenda for using data removal and machine learning organization methods to recognize harmful applications and safeguard confidential data from hackers. In this work, we examine characteristics based on anomalies and signatures to provide a consistent and efficient method for identifying and classifying malware. Studies have demonstrated the superiority of the suggested method over alternatives (Chandrakala, D at al., 2021).

The security of contemporary websites is seriously threatened by the widespread and sophisticated nature of modern malware (Zhao, K et al., 2015). Cyberattack kinds in the digital realm, or cyberspace, are illustrated in Figure 1. Malware is software designed specifically to damage a processor or system for instance, by tracking its users or theft their currency. Malware assaults are getting more frequent and can potentially impact industrial control systems, medical equipment, IoT devices, and environmental settings. Because modern spyware alters its behavior and code frequently, it is infamously difficult to detect. The effectiveness of traditional signature-based security has been undermined by the spread of malware. Rather, a wider variety of protective measures must be taken (Akhtar, M.S et al.,2022). Together stationary and active knowledge techniques can be secondhand to recognize behavioral resemblances among malware belonging to the same family (Gibert, D et al., 2019). Unlike static analysis, which examines the contents of potentially malicious files without executing them, dynamic analysis monitors behavior by recording function calls, tracing data flows, and inserting monitoring code into active binaries. (Firdaus, A et al., 2018).

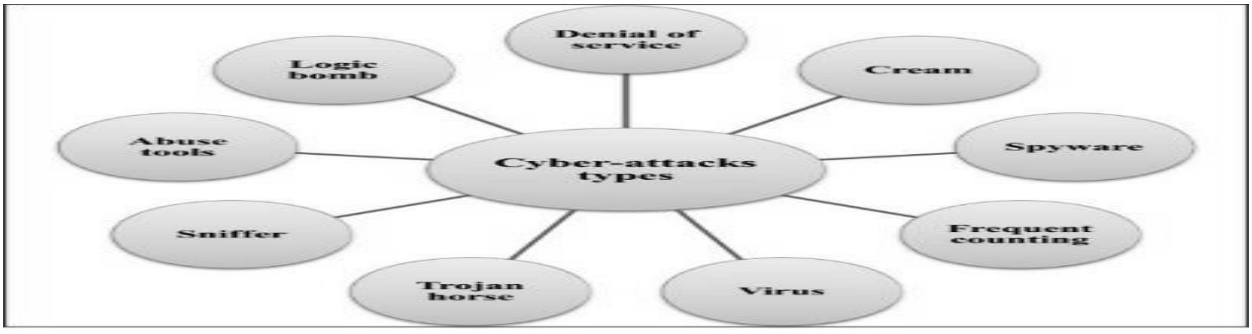

Figure 1.
Digital Realm Kind of Cyberattack.

These static and behavioral artifacts can be leveraged by machine learning procedures to examine the developing construction of modern malware, permitting them to recognize more advanced malware attacks that would else evade discovery by traditional cross practices. Machine learning-based resolutions are more effective against recently released malware since they do not depend on signatures. Accurate feature extraction and representation can be achieved through the use of deep learning algorithms that are proficient of execution piece manufacturing on their own (Dahl, G.E

et al., 2013). The Martin (2018) Cyber Kill Chain, a security mechanism to safeguard networks and prevent cyberattacks, is depicted in Figure 2. A massive distributed denial of service assault targeted AWS in February 2020 (Akhtar, M.S et al., 2021). The organization resisted a 2.3 Tbps DDoS attack, resulting in a 694,201-request rate and a packet promoting rate of 293.1 Mpps. It's been said by some to be the biggest DDoS attack ever recorded. Three hackers obtained access to Twitter in July 2020 and seized control of some well-known users' accounts. Notables whose accounts were compromised included Elon Musk of Tesla, President Obama, and Jeff Bezos of Amazon. Bitcoin scams that were uploaded from the compromised accounts brought in more than $100,000.

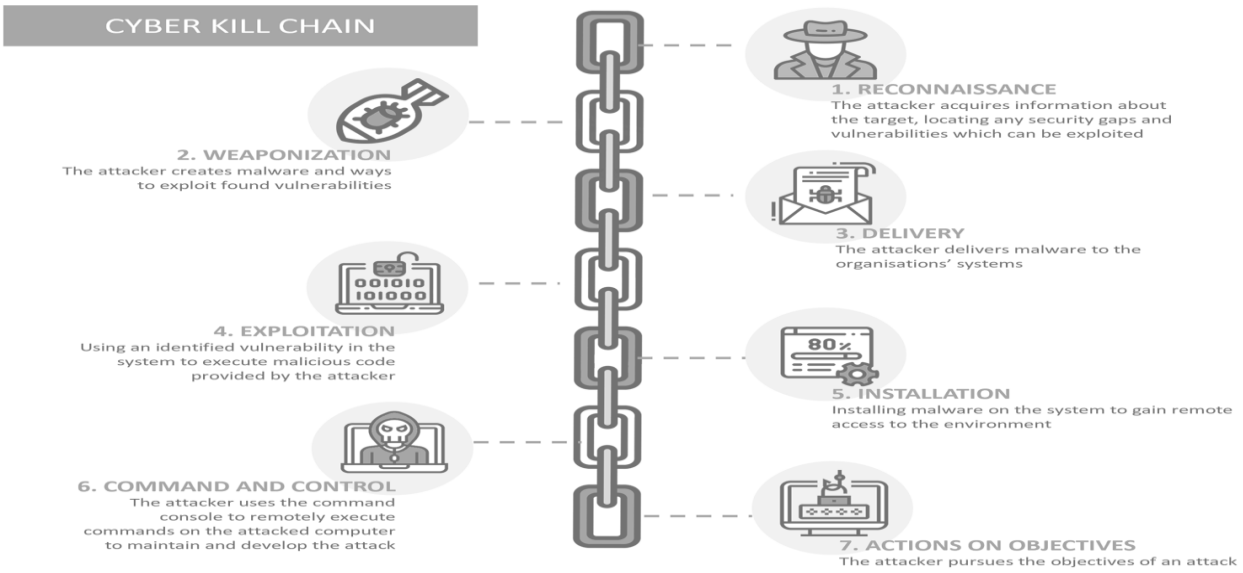

Figure 2.
Cyber-Kill chain.

LITERATURE REVIEW

Due to the general practice of computers, cellphones, and other Internet-enabled devices, cyberattacks are becoming more frequent. The surge in malware activity has led to the emergence of numerous malwares uncovering systems. Researchers' employment a variety of big data knowledge and machine learning tactics toward try then learn hazardous cypher. Though they take an extended time to procedure, outdated machine learning-based malware discovery methods can be valuable in recognizing recently bare malware. Because deep learning and other modern machine learning approaches are so common, feature engineering might eventually become outdated. We looked at an assortment of malware discovery and organization approaches in this research. Researchers have developed methods for detecting malicious intent in samples using deep learning and machine learning (Tahtaci, B et al., 2020). The correctness of several models was assessed and illustrated by Armaan (2021). No request created for a digital phase can purpose deprived of data (Baset, M. et al., 2016). Precautions must be taken to protect data because there are numerous cyber dangers. While creating any kind of model, feature selection is a challenging process,

but machine learning is a pioneering process that opens the door to precise calculation. A flexible solution that can accept non-standard data is required for this method. We must research malware and develop new strategies and frameworks based on malware types (Akhtar, M.S. et al., 2021), to manage and prevent future attacks. IT security experts may employ malware examination tools to look for outlines. The cybersecurity commerce welfares importantly from the arrival of knowledge that examines malware models and quantity their level of distortion. These possessions fund malware spell prevention and security alert monitoring. If malware poses a threat, we have to get rid of it before it spreads its infestation. Malware analysis is flattering progressively prevalent as it assistances organizations reduce the impact of the rising quantity of malware intimidations and the growing complication of the techniques used in malware attacks (Altaher, A. et al., 206). Chowdhury (2018) proposed an effective machine learning-based classification approach for detecting malware detection. In a previous study, (Chowdhury, M et al., 2017) explored whether adjusting certain parameters could enhance the accuracy of malware classification.

The spread of harmful software currently poses a serious threat to international stability. Malicious software became more common as the number of computers connected to the internet increased in the 1990s (Chowdhury, M et al., 2017), which ultimately resulted in the broad dissemination of malware. Numerous safeguards have been developed in reaction to this occurrence. Unfortunately, malware authors have developed new dangers to bypass security programs, and existing safeguards are unable to keep up with them. The focus of academics has changed from malware detection to machine learning algorithm tactics in recent years. In this study, we provide a defense mechanism that selects the best ML algorithm technique for malware detection after evaluating three different approaches.

Malware is still evolving and spreading at a startling rate. To examine and measure the discovery correctness of the ML classifier that extracted features based on PE information using static analysis, Nur (2019) analyzed three ML classifiers. We collectively trained machine learning processes to differentiate between benign and harmful material (Patil, R et al., 2020). This research presented how to get the best discovery correctness and most precise depiction of malware using stationary investigation built on PE material and designated dangerous data rudiments. With the expansion of the Internet, malicious software, commonly referred to as "malware," became more widespread and sophisticated. Its fast-broadcasting crossways the Net providing malware creators through a wide collection of gears for developing such programs (Gavrilu¸t, D. et al., 2009). The complexity and reach of malware continue to grow daily. To improvement a profounder sympathetic of machine learning, the study focused on evaluating and measuring the performance of classifiers. Following the refinement of the PE folder and public library data, the latent analysis identified key features, leading to the evaluation of six machine learning-based classifiers. It was proposed that machine learning (ML) algorithms be qualified and verified to control whether a folder is malicious. Investigational results showed that the random forest algorithm, with an accuracy of 99.4%, is the most effective for data classification. These findings indicated that the PE library could be utilized alongside static analysis, improving malware detection and classification by focusing on a limited set of features. The main benefit is that users can confirm a folder's authenticity beforehand inaugural it, plummeting the danger of

unintentionally installing malicious software (Pavithra, J et al., 2020). Static or dynamic analysis can be secondhand to classify possibly dangerous mechanisms of malware. Parsing malware binaries to find malicious strings is the main goal of stationary examination, which comprises the opposite engineering method used to strip a worm (Vanjire, S. et al., 2021). But dynamic analysis means custody of a judgment on hypothetically harmful software even while its innings in a safe location, like a computer-generated machine. While apiece method has assistances and hitches, it is optional to use both while analyzing malware (Agarkar, S.; et al., 2020). It's feasible that fewer potentially harmful features will increase malware detection precision. Then, the researcher would have more time to examine the information gathered. We are concerned that numerous features are being employed for malware detection when fewer, stronger features might be sufficient. The first phase in determining which hateful features to usage is toward look for potential techniques or algorithms. We require technologies that may meaningfully decrease the number of topographies now required to locate malware and also detect malware that was not ever remained seen beforehand (Sethi, K. et al., 2017).

METHODOLOGY

In this investigation work, the many stages and elements of a standard workflow for machine learning malware discovery and organization are introduced. The limitations and difficulties of this kind of workflow are also examined, and the peak latest progressions and tendencies in the arena with a focus on techniques of deep learning—are evaluated. Below is a description of the research study's suggested methodology (Ahmadi, M. et al., 2016). Figures 3 and 4 illustrate the complete workflow process, providing a clearer sympathetic of the future machine learning-based approach for malware discovery.

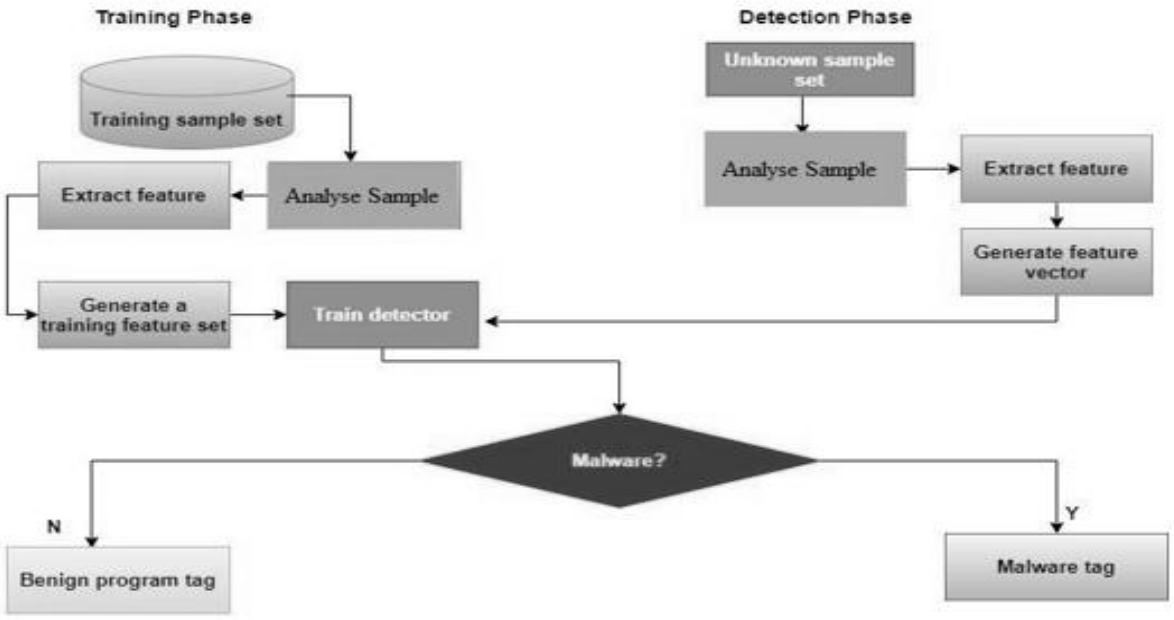

Figure 3.
Entire workflow process.

Data set

This project's primary component is a machine learning model that classifies malware and benign files using a Random Forest classifier tree. 30% of the files in the dataset we are using are innocuous, and 70% are malware. In terms of the splitting process, we separated the information into 70% exercise and 30% tough, after which we used the random forest classifier to identify the key characteristics needed for the classification. Because of this, we chose it for training, and after that, we saved it as a model.pkl and also stored the characteristics that we thought were significant. When extracting the necessary function from any actual file, use pkl to keep track of it.

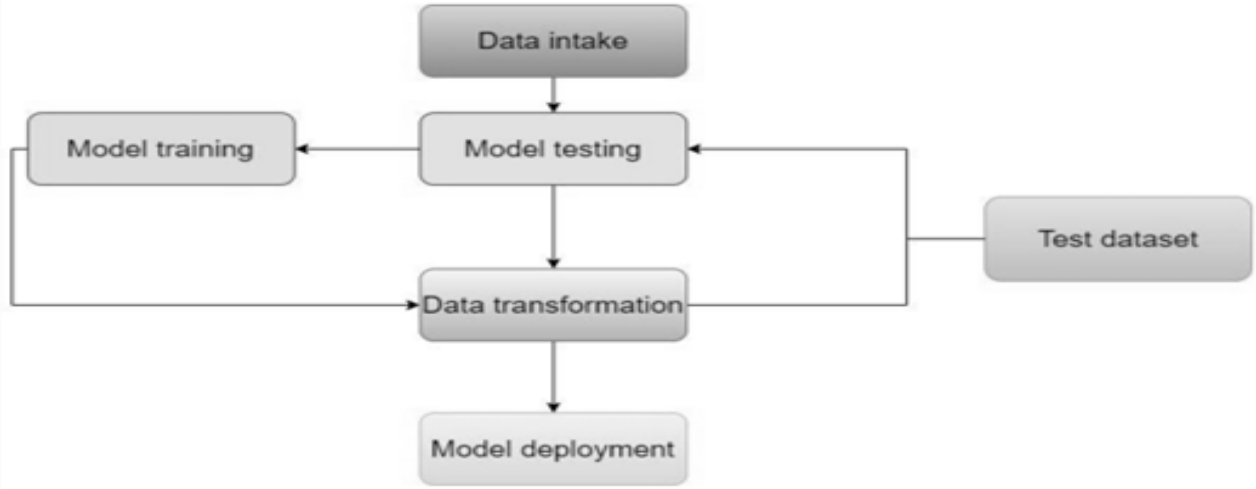

Figure 4.
Entire workflow process after data split.

Pre-processing

The files were unprocessed executables themselves, and the data remained saved in the folder system as a dual cypher. They were ready when we started our examination. A computer-generated machine (VM) or endangered setting was essential to empty the executables. Unloading trodden executables is automated through the PEiD program (Saad, S.; et al., 2019).

Features Extraction

Tens of thousands of features are often included in datasets from the 20th century. It has been evident in recent years that the resulting machine learning classic has been overfitting as feature counts have increased (Selamat, N. et al., 2019). We created a slighter usual of topographies after a bigger set in order to solve this issue; this method is frequently used to preserve a similar level of accurateness with scarcer topographies. This study aimed to improve the current dynamic and static feature dataset by removing features that were not useful for data analysis and retaining the most useful features (Firdausi, I. et al., 2010), (Kumar, P. et al., 2022).

Features Selection

Feature selection was carried out following feature extraction, which required finding supplementary topographies. Selecting topographies after a pond of recently

discovered qualities is known as feature collection, then it is a critical procedure for increasing correctness, shortening the model, and plummeting overfitting. Researchers have used a variety of feature cataloging algorithms in the historical to recognize software cipher that may be hazardous. Because the feature vigorous method is actually active in choosing the right topographies for developing malware detection models, it remained employed lengthily in this work (Hamid, F. et al., 2019), (Prabhat, K. et al., 2021), (Nadeem, G. et al., 2023).

Training and testing set

The training and testing phases are one of the most influential components of research strategy and lay the groundwork for creating and comprehensively evaluating specialized machine learning models meant for precision-driven malware detection. To ensure a solid basis for modeling and evaluation, it has been precisely divided our information set hooked on exercise and testing sets using a 70 to 30 ratio allocation. Notably, this method made it possible to integrate four different data sets, all of which added to a comprehensive perspective on real-world situations.

Machine learning

Due to ML algorithms' significant complex data processing capabilities, they are frequently utilized in malware detection, prediction, and other domains. To find the optimum model for malware detection based on data, four machine learning models were used. The selection of classifier models is limited to those that have been widely recognized and utilized. Naive Byes (NB), Support Vector Machine (SVM), Random Forest (RF), and K-Nearest Neighbor (kNN).

Naïve Bayes

Using probability to forecast an object's probability in a classification job, a Naive Bayes classifier (NB) is a type of probabilistic machine learning model. Equation 2 illustrates the Bayes theorem, establishing the core of the classifier.

$$(y|x)=(x|y)P(y)P(x) \quad (1)$$

In this case, variable x indicates attributes and variable y is the target class of the event. Where P(y) is the prior probability, P(x|y) is the likelihood probability, P(y|x) is the posterior probability, and P(x) is the marginal posterior probability.

Random Forest

Using averaging to increase predictive accuracy and manage over-fitting, the random forest model is a meta-estimator that fits multiple decision tree classifiers on different sub-samples of the dataset. The features are permuted at random after each split. Therefore, even with the same training data, max_features=n_features and bootstrap=False, the best split identified may differ if the improvement of the criterion is the same for various splits listed during the search for the best split. To obtain deterministic behavior during fitting, the random state must be fixed.

K-Nearest Neighbor

A non-parametric, administered knowledge classifier, the k-nearest neighbors (KNN) algorithm uses immediacy to catalog or forecast how a solitary data opinion will be gathered. It is among the most widely used and straightforward regression and classification classifiers in machine learning today. The KNN algorithm, as is widely known, helps find the closest points or groups to a query point. To get the nearest points or groups for a particular query point, we do, however, need a measure.

Support Vector Machine

Robust machine learning algorithms, such as Support Vector Machine (SVM), are applied to regression, linear or nonlinear classification, and outlier identification applications. Among the various uses for SVMs are text classification, image classification, handwriting recognition, spam detection, face detection, gene expression analysis, and anomaly detection. SVMs are adaptable and powerful in a variety of applications because they can handle high-dimensional data and nonlinear relationships.

RESULTS AND COMPARATIVE

The comparative analysis of these four machine learning models reveals distinct strengths and weaknesses in their application to malware detection and showcases the entire accuracy metrics in table 1.

- **Naïve Bayes**

Naïve Bayes are a popular choice for organization tasks due to their interpretability and straightforward nature. In malware detection, NB models build a tree-like structure where each node represents a decision based on feature values, and each branch represents the outcome of that decision. The leaves of the tree correspond to the final classification (malware or benign).

Performance Analysis

Accuracy: The NB model in this study achieved an accuracy of 87%. This figure indicates that the model correctly classified (Kumar, P. et al., 2022), 85% of the samples in the test set. While this is a respectable accuracy rate, it reflects some room for improvement, especially when compared to more complex models like Neural Networks.

Precision: Precision, which measures the proportion of true positive classifications out of all positive classifications made by the model (Kumar, P. et al., 2022), (Nadeem, G. et al., 2023), was recorded at 82%. This implies that 82% of the samples classified as malware by the NB model were indeed malicious. A precision of 82% suggests that the model performs well in minimizing false positives, which is crucial for reducing the number of benign files incorrectly flagged as malware.

Recall: Recall, the metric representing the proportion of actual malware samples correctly identified by the model, was 85%. This value indicates that the NB model was able to detect 78% of the actual malware instances. Although this is a decent recall rate, it reveals that the model missed a proportion of malware samples, which could be problematic in scenarios where detecting every instance of malware is critical.

- **Support Vector Machines (SVM)**

Support Vector Machines (SVM) are a powerful class of supervised learning algorithms that work well for both linear and non-linear classification problems. SVM aims to find the optimal hyperplane that separates different classes with the maximum margin. This approach is particularly useful in high-dimensional spaces.

Performance Analysis

Accuracy: The SVM model achieved an accuracy of 92.41%, surpassing the Decision Tree in performance. This higher accuracy reflects the model's ability to effectively separate malware from benign files with greater precision.

Precision: The precision of the SVM model was 91%. This means that 85% of the samples classified as malware were correctly identified, demonstrating a strong ability to reduce false positives compared to the Decision Tree model.

Recall: The recall for the SVM model was 89%, which is slightly better than the NB model's recall but on par with its precision. This designates that while the SVM model achieves fine in classifying actual malware occurrences, it still misses a portion of malware samples.

- **K-Nearest Neighbor**

KNN is a statistical method used for binary classification that models the probability of a class label based on input features. Despite its name, logistic regression is used for classification rather than regression.

Performance Analysis

Accuracy: The kNN and SVM models. This suggests that kNN achieved the accuracy level of 91.76% in distinguishing between malware and benign software in this study.

Precision: The precision of the kNN model was 89%. This indicates that 89% of the malware predictions were accurate, reflecting a moderate ability to minimize false positives.

Recall: The recall for the kNN model was 91%, the lowest among the models evaluated. This means that the kNN model detected only 91% of the actual malware samples, which could lead to a higher rate of undetected malware instances.

- **Random Forest (RF)**

Rando Forest (RF) are a lesson of machine learning mockups enthused by the construction and purpose of the decision tree. They involve of manifold sheets of consistent trees that process input data through complex transformations to produce predictions.

Performance Analysis

Accuracy: The Rando Forest model achieved an impressive accuracy of 98%, the highest among the models evaluated. This indicates that the RF model effectively differentiates between malware and benign software, capturing complex patterns and relationships in the data.

Precision: The precision of the RF model was 95%, reflecting a high rate of accurate malware predictions and a low rate of false positives. This high precision underscores the model's ability to correctly identify malware samples with minimal errors.

Recall: The recall for the RF model was 99%, also the highest among the models assessed. This suggests that the RF model was able to identify 99% of the actual malware samples, making it highly effective at detecting malware instances.

Table 1.
Accuracy metric of entire ML models.

Model	Accuracy	Precision	Recall	F1-Score
SVM	92	91	89	90
kNN	91	89	91	90
NB	87	82	85	80
RF	98	95	99	100

CONCLUSION

This research highlights the increasing interest in machine learning (ML) algorithm solutions for malware identification among academics in recent times. We devised a defense mechanism that selected the best of three machine-learning algorithm approaches for malware detection. In a specific dataset, the malware detection performances of the NB, RF, and KNN algorithms on a modest were examined. In this experiment, we compared a ML classifier with two other ML classifiers to assess and enumerate the discovery accurateness of the ML classifier that extracted features based on PE data using static analysis. Our work has enabled machine learning algorithms to distinguish between benign and harmful data. Of all the classifiers we tested, the RF machine learning approach had the best accuracy (99%). Static examination found happening PE information and properly chosen data presented potential in trial results, not just offering the highest detection accuracy and precisely characterizing malware. One important advantage is that we can detect whether or not the data is malicious without having to run any code. Via the dataset acquired from the Canadian Institute for Cybersecurity, the four machine learning models (NB, kNN, RF, and SVM) were trained, and evaluated, and their efficacy was compared.

DECLARATIONS

Acknowledgement: We appreciate the generous support from all the contributor of research and their different affiliations.
Funding: No funding body in the public, private, or nonprofit sectors provided a particular grant for this research.
Availability of data and material: In the approach, the data sources for the variables are stated.
Authors' contributions: Each author participated equally to the creation of this work.
Conflicts of Interests: The authors declare no conflict of interest.
Consent to Participate: Yes
Consent for publication and Ethical approval: Because this study does not include human or animal data, ethical approval is not required for publication. All authors have given their consent.

REFERENCES

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]

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.

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.

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]

Akhtar, M.S.; Feng, T. An overview of the applications of artificial intelligence in cybersecurity. *EAI Endorsed Trans. Create. Tech.* **2021**, *8*, e4. [CrossRef]

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]

Akhtar, M.S.; Feng, T. Deep learning-based framework for the detection of cyberattack using feature engineering. *Secur. Commun. Netw.* **2021**, *2021*, 6129210. [CrossRef]

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]

Akhtar, M.S.; Feng, T. IOTA based anomaly detection machine learning in mobile sensing. *EAI Endorsed Trans. Create. Tech.* **2022**, 9, 172814. [CrossRef]

Altaher, A. Classification of android malware applications using feature selection and classification algorithms. *VAWKUM Trans. Comput. Sci.* **2016**, *10*, 1. [CrossRef]

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.

Baset, M. Machine Learning for Malware Detection. Master's Dissertation, Heriot Watt University, Edinburg, Scotland, December 2016. [CrossRef]

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]

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.

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.

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.

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]

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]

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]

Gavrilu¸t, 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.

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]

Hamid, F. Enhancing malware detection with static analysis using machine learning. *Int. J. Res. Appl. Sci. Eng. Technol.* **2019**, *7*,

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]

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]

Nadeem, G., & Anis, M. I. (2024). Investigation of bovine disease and events through machine learning models. Pakistan Journal of Agricultural Research, 37(2), 102-114.

Nadeem, G., Rehman, Y., Khaliq, A., Khalid, H., & Anis, M. I. (2023, March). Artificial Intelligence based prediction system for General Medicine. In 2023 4th International Conference on Computing, Mathematics and Engineering Technologies (iCoMET) (pp. 1-6). IEEE.

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]

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.

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]

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]

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.

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]

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.

Saad, S.; Briguglio, W.; Elmiligi, H. The curious case of machine learning in malware detection. *arXiv* **2019**, arXiv:1905.07573.

Selamat, N.; Ali, F. Comparison of malware detection techniques using machine learning algorithm. *Indones. J. Electr. Eng. Comput. Sci.* **2019**, *16*, 435. [CrossRef]

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.

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.

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.

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.

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.

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.

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.