International Journal For Recent Developments in Science & Technology
2581-4575
Crossref
A Peer Reviewed Research Journal

AUTOMATED ANDROID MALWARE DETECTION USING OPTIMAL ENSEMBLE LEARNING APPROACH FOR CYBER SECURITY

[1]SIRIGA PRASAD,[2]GUDUGUNTLA PAVAN KUMAR,[3]SURYA ALEKYA,[4]DEVIREDDY VEEKSHITHA REDDY,[5]MS B MADHURI

[1,2,3,4]Students, Department of computer Science And Engineering, Malla Reddy Engineering College (Autonomous),Hyderabad Telangana, India 500100
[5]Assistant Professor, Department of computer Science And Engineering, Malla Reddy Engineering College (Autonomous),Hyderabad Telangana, India 500100

ABSTRACT

The paper titled "Automated Android Malware Detection Using Optimal Ensemble Learning Approach for Cyber Security" addresses the growing challenges in detecting malware due to the evolving nature of malicious software. With advancements in technology, malware variants have become more sophisticated, utilizing advanced packing and obfuscation techniques that complicate their classification and detection. Traditional machine learning (ML) methods are often inadequate for identifying new and complex malware, making it necessary to explore alternative solutions. The paper proposes a novel technique called Automated Android Malware Detection using Optimal Ensemble Learning Approach for Cybersecurity (AAMD-OELAC), which focuses on the automated classification and identification of Android malware. This technique utilizes an ensemble learning process involving three machine learning models: Least Square Support Vector Machine (LS-SVM), kernel extreme learning machine (KELM), and Regularized Random Vector Functional Link Neural Network (RRVFLN). To enhance the detection performance, the method incorporates a Hunter-Prey Optimization (HPO) approach for optimal parameter tuning of the models. The experimental results demonstrate the effectiveness of the AAMD-OELAC method, showing superior performance compared to existing approaches in malware detection.

Keywords: Android malware, ensemble learning, cybersecurity, LS-SVM, KELM, RRVFLN, Hunter-Prey Optimization, malware detection, machine learning, optimization.

I.INTRODUCTION

Cybersecurity has become a critical focus for network engineers and computer scientists due to the increasing threats posed by malware. As technology advances and integrates into various aspects of life, the prevalence of malware, particularly Android malware, has surged. Android, being one of the most widely used operating systems, has become a major target for cyberattacks. This has led to a growing concern about the identification and management of malware applications that often have numerous parameters, making detection increasingly difficult. To address this challenge, researchers have focused on developing more effective techniques for detecting Android malware, which has become a significant area of study in cybersecurity Malware authors employ various methods to evade detection, such as code obfuscation, which complicates traditional

detection techniques. This has led to the need for new approaches to efficiently identify, deactivate, or remove Android malware. Several detection mechanisms have been proposed, including static, dynamic, and hybrid analysis methods. Static analysis helps identify harmful behaviors of apps by examining their features without needing to deploy the actual application. However, this method is often undermined by obfuscation techniques. Dynamic analysis, on the other hand, monitors applications at runtime to detect malicious activities. While static analysis can find malware through source code examination, dynamic analysis helps detect it during execution. With Android developers and users increasingly vulnerable to malware, the development of more effective detection systems is essential. This study focuses on leveraging machine learning (ML) and deep learning (DL) methods to recognize malicious Android Application Packages (APKs) and identify vulnerable code segments in software. The paper introduces the Automated Android Malware Detection using Optimal Ensemble Learning Approach for Cybersecurity(AAMD-OELAC) technique, which aims to enhance malware detection. The AAMD-OELAC approach begins with data preprocessing and employs an ensemble learning process that combines three machine learning models: Least Square Support Vector Machine (LS-SVM), kernel extreme learning machine (KELM), and Regularized Random Vector Functional Link Neural Network (RRVFLN). To further optimize the performance of these models, the Hunter-Prey Optimization (HPO) algorithm is used for parameter tuning, leading to improved malware detection results. The method is thoroughly tested through experimental analysis to demonstrate its superiority over existing approaches. The main contributions of the paper are as follows: the introduction of an intelligent AAMD-OELAC technique, which integrates data preprocessing, ensemble learning, and HPO-based hyperparameter tuning for enhanced Android malware detection. The technique incorporates a classification process using the LS-SVM, KELM, and RRVFLN models, and the combination of ensemble learning with the HPO algorithm improves the accuracy of malware detection. By utilizing multiple classifiers and optimization strategies, the proposed model can effectively identify malicious behaviors and patterns in Android applications, setting it apart from existing methods in the literature.

II. LITERATURE SURVEY

Rathore et al. (2023): This study explores the challenges in Android malware detection, particularly focusing on adversarial attacks in reinforcement learning-based evasion tactics. It discusses the need for effective defenses against evolving malware, highlighting the superiority of adversarial approaches in detecting and defending against Android malware. The paper emphasizes the role of reinforcement learning in both evasion strategies and defense mechanisms, offering valuable insights into the dynamic nature of Android malware detection and the continuous need for innovation in detection methods to keep up with evolving threats.

Wang et al. (2022): This paper proposes a hybrid approach for Android malware detection, utilizing both static and dynamic tactics based on app permissions. The

authors argue that the permissions requested by an app can provide critical insights into its malicious behavior, enhancing detection accuracy. By combining different detection strategies, this approach leverages the strengths of both static analysis and runtime monitoring, addressing the limitations of individual methods and improving the overall effectiveness of malware detection in Android environments.

Albakri et al. (2023): This research presents a metaheuristic-based deep learning model for Android malware detection and classification. The authors combine optimization techniques, such as genetic algorithms, with deep learning models to enhance malware detection performance. The paper focuses on improving classification accuracy and detection efficiency by incorporating metaheuristics, which allow the model to adapt to complex and dynamic malware patterns, demonstrating the potential for hybrid solutions in addressing the challenges of Android cybersecurity.

Ibrahim et al. (2022): This study focuses on an automatic Android malware detection method that combines static analysis and deep learning techniques. The authors propose a deep learning model that processes static features of Android applications to identify malicious behavior. By leveraging the power of deep learning in feature extraction and classification, this method enhances the accuracy of malware detection while reducing the computational burden associated with dynamic analysis. The paper highlights the advantages of static analysis in detecting known malware signatures and the role of deep learning in identifying sophisticated threats.

Hammood et al. (2023): This paper introduces a machine learning-based adaptive genetic algorithm (AGA) for Android malware detection, particularly in the context of auto-driving vehicles. The AGA model adapts and evolves to detect malware specific to the automotive sector, which poses unique cybersecurity challenges. By integrating genetic algorithms with machine learning, the approach dynamically selects features and optimizes detection accuracy, ensuring that the system can handle the complexities of malware targeting autonomous vehicle systems.

Bhat and Dutta (2022): This study presents a multi-tiered feature selection model for Android malware detection, utilizing feature discrimination and information gain. The authors propose a model that focuses on selecting the most informative features from large datasets to enhance the accuracy and efficiency of malware detection. The model's ability to prioritize features based on their discriminative power allows for improved detection performance, making it suitable for large-scale malware classification tasks where feature selection plays a critical role in reducing computational costs and improving accuracy.

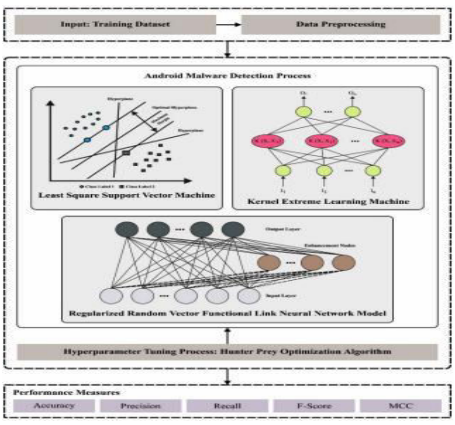
Wang et al. (2023): This paper surveys various deep learning-based techniques for Android malware detection. It provides a comprehensive overview of the application of deep learning models in the detection of Android malware, analyzing different architectures, training methods, and feature extraction strategies. The survey emphasizes the growing importance of deep learning in automating malware detection, offering a detailed comparison of different models and approaches, and

International Journal For Recent Developments in Science & Technology
2581-4575
Crossref
A Peer Reviewed Research Journal

highlighting the challenges and future directions for enhancing the effectiveness of deep learning in Android malware detection.

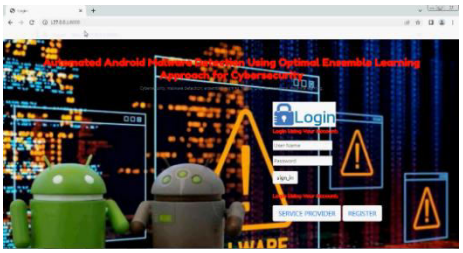
III.PROPOSED METHODOLOGY

The proposed methodology in this paper introduces the Automated Android Malware Detection using Optimal Ensemble Learning Approach for Cybersecurity (AAMD-OELAC) technique. The process begins with data preprocessing, preparing the dataset for malware detection. The core of the detection process utilizes ensemble learning, combining three machine learning models: Least Square Support Vector Machine (LS-SVM), Kernel Extreme Learning Machine (KELM), and Regularized Random Vector Functional Link Neural Network (RRVFLN). These models work together to improve the accuracy of malware detection by leveraging their individual strengths. Additionally, the Hunter-Prey Optimization (HPO) algorithm is applied for optimal parameter tuning of the three models, further enhancing the detection performance. To validate the effectiveness of AAMD-OELAC, a comprehensive experimental analysis is conducted, demonstrating its superiority over other existing methods.

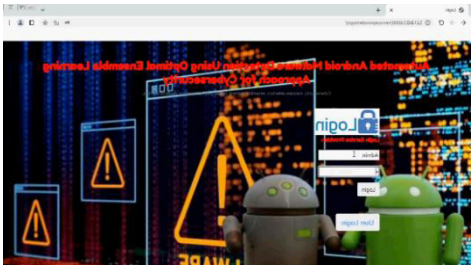


IV.WORKING METHODOLOGY

The methodology for the Automated Android Malware Detection system follows a robust process, integrating machine learning and deep learning models to detect Android malware based on URLs. It is divided into two main parts: the service provider side (SProvider) and the remote user side (RUser), both interacting with a Django backend.

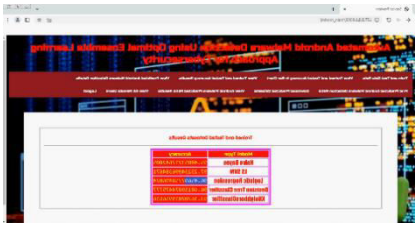


Service Provider (SProvider): The service provider manages the system, providing functionalities to train and evaluate malware detection models, track prediction ratios, and manage user data. The process starts with login functionality for the service provider.



Once logged in, the system allows viewing of remote users, trends, and charts detailing detection ratios and accuracy. The core of the service provider's tasks involves training the malware detection models using data from URLs. A variety of machine learning models are employed for this task, including Naive Bayes, SVM, Logistic Regression, Decision Trees, and K-Nearest Neighbors. These models are trained on a dataset (URLs classified as malware or normal), and their accuracy is

calculated and stored in the detection_accuracy model.



A Voting Classifier is used to combine predictions from these models for better overall detection performance. Additionally, malware detection results are stored and can be exported as Excel files for further analysis.

Remote User (RUser): Remote users register and log in to the system, and they can input URLs to check for malware. After entering data, the system preprocesses it and applies the trained models to predict whether the URL is normal or malware. The prediction is returned to the user along with the associated details (IP address, URL, etc.). The system uses the same machine learning models trained by the service provider, including Naive Bayes, SVM, Logistic Regression, and Decision Trees, and evaluates them using metrics like accuracy, confusion matrix, and classification report. A final prediction is made using a VotingClassifier, and the result (Normal or Malware) is stored in the malware_detection model.

V.CONCLUSION

The conclusion of this study highlights the successful development of the AAMD-OELAC (Automated Android Malware Detection using Optimal Ensemble Learning Approach with Hyperparameter Optimization) technique, which significantly enhances the accuracy and efficiency of Android malware detection. The approach integrates data preprocessing, ensemble classification, and hyperparameter optimization (HPO) for parameter tuning, combining three powerful machine learning models: LS-SVM, KELM, and RRVFLN. By applying ensemble learning and HPO, the technique achieves superior detection accuracy compared to traditional methods, confirming its effectiveness in identifying and classifying Android malware. The simulation results substantiate the advantages of the AAMD-OELAC method, showcasing its supremacy over existing malware detection techniques. The study's experimental analysis demonstrates that the ensemble learning framework, combined with optimized model parameters, leads to improved detection performance. This method is expected to play a crucial role in automated malware detection systems, making it more reliable and robust.

Looking ahead, the study suggests areas for further research. Future work could focus on refining detection capabilities to identify more subtle and sophisticated malware behaviors, which could involve the development of more advanced techniques. Additionally, exploring privacy-preserving methods such as secure multi-party computation or federated learning could help create collaborative malware detection systems that protect user privacy while maintaining high detection accuracy. These future directions would further strengthen the

potential of the AAMD-OELAC approach in real-world cybersecurity applications.

VI.REFERENCES

[1] H. Rathore, A. Nandanwar, S. K. Sahay, and M. Sewak, ‘‘Adversarial superiority in Android malware detection: Lessons from reinforcement learning based evasion attacks and defenses,’’ Forensic Sci. Int., Digit. Invest., vol. 44, Mar. 2023, Art. no. 301511.

[2] H. Wang, W. Zhang, and H. He, ‘‘You are what the permissions told me! Android malware detection based on hybrid tactics,’’ J. Inf. Secur. Appl., vol. 66, May 2022, Art. no. 103159.

[3] A. Albakri, F. Alhayan, N. Alturki, S. Ahamed, and S. Shamsudheen, ‘‘Metaheuristics with deep learning model for cybersecurity and Android malware detection and classification,’’ Appl. Sci., vol. 13, no. 4, p. 2172, Feb. 2023.

[4] M. Ibrahim, B. Issa, and M. B. Jasser, ‘‘A method for automatic Android malware detection based on static analysis and deep learning,’’ IEEE Access, vol. 10, pp. 117334–117352, 2022.

[5] L. Hammood, İ. A. Doğru, and K. Kılıç, ‘‘Machine learning-based adaptive genetic algorithm for Android malware detection in auto-driving vehicles,’’ Appl. Sci., vol. 13, no. 9, p. 5403, Apr. 2023.

[6] P. Bhat and K. Dutta, ‘‘A multi-tiered feature selection model for Android malware detection based on feature discrimination and information gain,’’ J. King Saud Univ.-Comput. Inf. Sci., vol. 34, no. 10, pp. 9464–9477, Nov. 2022.

[7] D.Wang, T. Chen, Z. Zhang, and N. Zhang, ‘‘A survey of Android malware detection based on deep learning,’’ in Proc. Int. Conf. Mach. Learn. Cyber Secur. Cham, Switzerland: Springer, 2023, pp. 228–242.

[8] Y. Zhao, L. Li, H. Wang, H. Cai, T. F. Bissyandé, J. Klein, and J. Grundy, ‘‘On the impact of sample duplication in machine-learning-based Android malware detection,’’ ACM Trans. Softw. Eng. Methodol., vol. 30, no. 3, pp. 1–38, Jul. 2021.

[9] E. C. Bayazit, O. K. Sahingoz, and B. Dogan, ‘‘Deep learning based malware detection for Android systems: A comparative analysis,’’ Tehnički vjesnik, vol. 30, no. 3, pp. 787–796, 2023.

[10] H.-J. Zhu, W. Gu, L.-M. Wang, Z.-C. Xu, and V. S. Sheng, ‘‘Android malware detection based on multi-head squeeze-and-excitation residual network,’’ Expert Syst. Appl., vol. 212, Feb. 2023, Art. no. 118705.

[11] K. Shaukat, S. Luo, and V. Varadharajan, ‘‘A novel deep learning-based approach for malware detection,’’ Eng. Appl. Artif. Intell., vol. 122, Jun. 2023, Art. no. 106030.

[12] J. Geremias, E. K. Viegas, A. O. Santin, A. Britto, and P. Horchulhack, ‘‘Towards multi-view Android malware detection through image-based deep learning,’’ in Proc. Int.Wireless Commun. Mobile Comput. (IWCMC), May 2022, pp. 572–577.

72516 VOLUME 11, 2023

IEEE Transaction on Machine Learning,Volume:11,Issue Date:11.July.2023

[13] J. Kim, Y. Ban, E. Ko, H. Cho, and J. H. Yi, ‘‘MAPAS: A practical deep learning-based Android malware detection system,’’ Int. J. Inf. Secur., vol. 21, no. 4, pp. 725–738, Aug. 2022.

[14] S. Fallah and A. J. Bidgoly, ‘‘Android malware detection using network traffic based on sequential deep learning models,’’ Softw., Pract. Exper., vol. 52, no. 9, pp. 1987–2004, Sep. 2022.

[15] V. Sihag, M. Vardhan, P. Singh, G. Choudhary, and S. Son, ‘‘De-LADY: Deep learning-based Android malware detection using dynamic features,’’ J. Internet Serv. Inf. Secur., vol. 11, no. 2, p. 34, 2021.