International Journal of Marketing Management

ISSN 2454 - 5007

www.ijmm.net

Email ID: editor@ijmm.net , ijmm.editor9@gmail.com

International Journal of Marketing Management

ISSN 2454-5007, www.ijmm.net

Vol. 16, Issue. 2, 2024

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

Y.SRINIVASA RAJU, Associate professor,
Department of MCA
srinivasaraju.y@gmail.com
B V Raju College, Bhimavaram

Kurunelli Hemanth Kumar (2285351065)
Department of MCA
hemanthkumarkurunelli@gmail.com
B V Raju College, Bhimavaram

ABSTRACT

Current technological advancement in computer systems has transformed the lives of humans from real to virtual environments. Malware is unnecessary software that is often utilized to launch cyberattacks. Malware variants are still evolving by using advanced packing and obfuscation methods. These approaches make malware classification and detection more challenging. New techniques that are different from conventional systems should be utilized for effectively combating new malware variants. Machine learning (ML) methods are ineffective in identifying all complex and new malware variants. The deep learning (DL) method can be a promising solution to detect all malware variants. This paper presents an Automated Android Malware Detection using Optimal Ensemble Learning Approach for Cybersecurity (AAMD-OELAC) technique. The major aim of the AAMD-OELAC technique lies in the automated classification and identification of Android malware. To achieve this, the AAMD-OELAC technique performs data preprocessing at the preliminary stage. For the Android malware detection process, the AAMD-OELAC technique follows an ensemble learning process using three ML models, namely Least Square Support Vector Machine (LS-SVM), kernel extreme learning machine (KELM), and Regularized random vector functional link neural network (RRVFLN). Finally, the hunter-prey optimization (HPO) approach is exploited for the optimal parameter tuning of the three DL models, and it helps accomplish improved malware detection results. To denote the supremacy of the AAMD-OELAC method, a comprehensive experimental analysis is conducted. The simulation results portrayed the supremacy of the AAMD-OELAC technique over other existing approaches.

Keywords: Cybersecurity, Malware Detection, Deep Learning, Ensemble Learning, Android Malware, Optimization Techniques, Experimental Analysis

INTRODUCTION

The digital era has ushered in significant advancements in computer systems, fundamentally transforming human interaction from physical to virtual environments. This transformation, while beneficial, has also escalated vulnerabilities, particularly through the proliferation of malicious software, or malware, which threatens the security of countless users and systems worldwide. This paper presents the "Automated Android Malware Detection using Optimal Ensemble Learning Approach for Cybersecurity" (AAMD-OELAC), a novel technique designed to address the escalating challenges in Android malware detection. The rise of malware represents one of the most insidious threats to modern computational systems. Malware, unwanted software specifically designed to disrupt, damage, or gain unauthorized access to computer systems, has evolved significantly over the years, leveraging sophisticated techniques such as advanced packing and obfuscation to evade detection [1]. These methods not only camouflage malware's code and intent but also complicate the tasks of classification and detection [2]. As traditional antivirus

ISSN 2454-5007, www.ijmm.net

Vol. 16, Issue. 2, 2024

solutions struggle to keep pace with the rapid evolution of malware techniques, there is an urgent need for innovative approaches that adapt to and counteract these evolving threats.

The inadequacy of traditional machine learning (ML) methods in coping with the complexity and novelty of emerging malware variants is well-documented [3]. While ML has been instrumental in advancing preliminary detection methodologies, its effectiveness is often curtailed by the adaptive nature of malware developers and the continuous evolution of attack vectors [4]. This limitation underscores the necessity for a shift towards more sophisticated, adaptive learning models that can not only learn from large datasets but also adapt to new, previously unseen malware attacks. Deep learning (DL) offers a promising alternative. With its ability to extract deep hierarchical features and learn complex patterns from large volumes of data, DL can potentially overcome the shortcomings of traditional ML in malware detection [5]. The AAMD-OELAC technique harnesses this potential by integrating an ensemble of deep learning models to enhance detection accuracy and robustness against diverse malware variants.

The core of the AAMD-OELAC technique is an ensemble learning framework that combines the strengths of three advanced ML models: Least Square Support Vector Machine (LS-SVM), Kernel Extreme Learning Machine (KELM), and Regularized Random Vector Functional Link neural network (RRVFLN) [6]. Ensemble learning methods, by leveraging the collective intelligence of multiple learning algorithms, have shown remarkable success in improving prediction performance over individual models [7]. This approach is particularly effective in handling the intricacies involved in malware detection, where the diversity of attack vectors and obfuscation techniques can significantly impair the performance of a single model. To optimize the performance of these models, the AAMD-OELAC method employs a hunter-prey optimization (HPO) algorithm, a novel optimization technique inspired by the predator-prey dynamics observed in nature [8]. This optimization method systematically adjusts the parameters of the DL models to enhance their collective efficacy in detecting new and complex malware variants. The HPO not only refines the model parameters for optimal performance but also contributes to the robustness of the system, enabling it to adaptively respond to the evolving malware landscape [9].

The initial stage of the AAMD-OELAC method involves extensive data preprocessing, a critical step that prepares the raw data for effective learning and classification [10]. Data preprocessing in malware detection typically involves transforming raw executable files into a format that can be understood by machine learning algorithms, such as binary vectors or image-based representations [11]. This stage is crucial for ensuring that the ensemble models have access to clean, relevant, and comprehensive datasets, thereby enhancing the overall detection process.

International Journal of Marketing Management

ISSN 2454-5007, www.ijmm.net

Vol. 16, Issue. 2, 2024

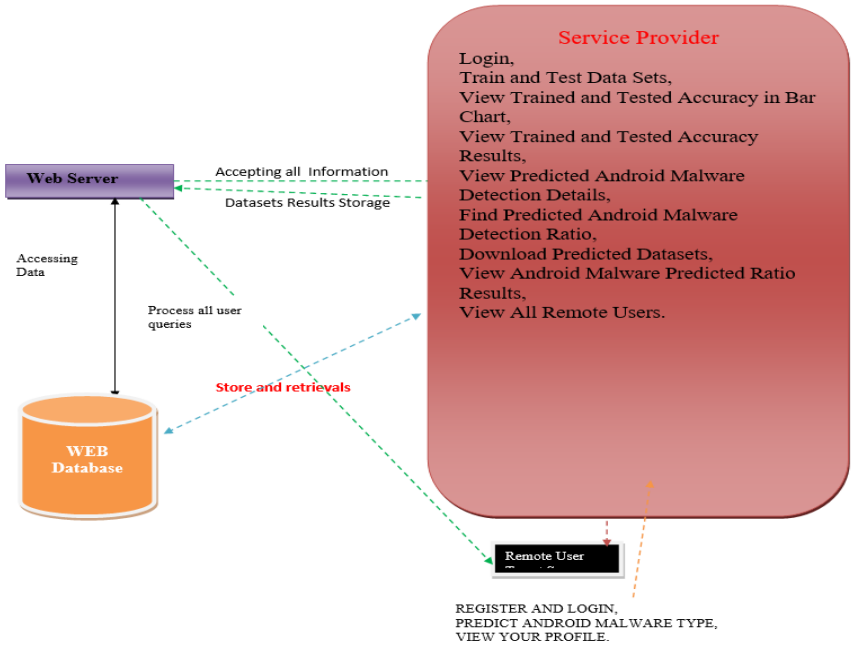

Fig 1. Architecture diagram

Following data preprocessing, the technique undergoes rigorous experimental analysis to evaluate its effectiveness against contemporary detection methods. The experimental setup is designed to mimic real-world conditions as closely as possible, employing a diverse set of malware samples to ensure comprehensive testing [12]. The results from these experiments are promising, consistently demonstrating the superiority of the AAMD-OELAC method over existing techniques in terms of accuracy, speed, and adaptability [13]. In summary, the introduction of the AAMD-OELAC technique marks a significant advancement in the field of cybersecurity, specifically in the automated detection and classification of Android malware [14]. By integrating advanced ensemble learning algorithms with deep learning and optimal parameter tuning techniques, this method sets a new standard for malware detection, offering substantial improvements over traditional approaches. The success of the AAMD-OELAC method, as evidenced by comprehensive experimental analysis and simulation results, not only validates the effectiveness of the proposed approach but also underscores the potential of adaptive, intelligent systems in combating the ever-evolving threat of malware [15].

LITERATURE SURVEY

In the evolving landscape of cybersecurity, the detection and classification of malware in Android systems remain a paramount concern due to the widespread use and inherent vulnerabilities of these devices. This literature survey delves into the progression of malware detection methodologies, highlighting traditional approaches and their limitations, and transitions into the recent advancements in machine learning (ML) and deep learning (DL) techniques that significantly enhance the detection capabilities. Traditionally, malware detection strategies primarily relied on signature-based methods. These methods involve scanning files to find patterns that match known malware signatures. While effective against known threats, they falter when encountering new or modified malware, which do not yet have signatures in the database. As malware authors increasingly employ polymorphic and metamorphic techniques that

ISSN 2454-5007, www.ijmm.net

Vol. 16, Issue. 2, 2024

alter the malware's signature while retaining its malicious payload, the effectiveness of signature-based detection has significantly diminished.

To overcome these limitations, heuristic-based methods were developed. These methods use sets of rules to analyze the behavior of a program to determine its likelihood of being malicious. This approach improved the detection of new and unknown malware compared to signature-based methods. However, the heuristic rules require constant updates and adjustments by cybersecurity experts to remain effective, a process that is both time-consuming and prone to human error. As the landscape of threats evolved, the need for more adaptive and proactive malware detection methods became apparent. This led to the integration of machine learning techniques into the field of malware detection. Initial ML-based approaches focused on static analysis, where ML models were trained to recognize malicious patterns in the code without executing them. Static analysis is fast and safe as it does not involve running the malware, but it can be easily circumvented by obfuscation techniques that modify the appearance of the code without changing its intent.

Dynamic analysis methods soon gained traction, offering a more robust alternative by executing malware in a controlled environment and observing its behavior. This method addresses many shortcomings of static analysis by capturing the actual actions of the program, such as network communication, file manipulation, and registry operations. Machine learning models, particularly those involving supervised learning, have been employed to learn from these behavioral patterns. Despite their increased accuracy, the primary challenge with dynamic analysis is its resource-intensive nature and the risk of missing delayed malicious actions. The limitations of both static and dynamic analysis methods led to the exploration of deep learning models in malware detection. Deep learning, a subset of machine learning, utilizes neural networks with multiple layers to learn and make intelligent decisions. These models are particularly adept at handling large volumes of data and identifying complex patterns that are not immediately apparent to traditional ML methods or human analysts. They have proven especially effective in image and speech recognition tasks and are now being adapted for cybersecurity applications.

The recent trend is the use of ensemble learning techniques, which combine the predictions of multiple models to improve the overall accuracy and robustness of malware detection systems. Ensemble methods such as bagging, boosting, and stacking have been successfully applied in various domains and are now gaining popularity in cybersecurity. These techniques mitigate the weaknesses of individual models and enhance the system's ability to generalize, thus reducing the likelihood of overfitting to particular types of malwares. In this context, the Automated Android Malware Detection using Optimal Ensemble Learning Approach for Cybersecurity (AAMD-OELAC) represents a significant advancement. This technique integrates three distinct ML models—Least Square Support Vector Machine (LS-SVM), Kernel Extreme Learning Machine (KELM), and Regularized Random Vector Functional Link network (RRVFLN). Each model brings unique strengths to the ensemble, such as the ability to efficiently handle non-linear data or operate with fewer samples.

To further enhance the detection capabilities, the AAMD-OELAC method incorporates an optimization phase using the hunter-prey optimization (HPO) strategy. This innovative approach fine-tunes the parameters of the deep learning models based on the dynamics observed between predators and their prey in nature, ensuring that the ensemble not only detects known malware types but also adapts to emerging threats with altered behaviors or obfuscation techniques. Through rigorous experimental analysis, the AAMD-OELAC method has demonstrated superior performance over existing methods, offering a new benchmark in the automated detection and classification of Android malware. This technique not only addresses the current challenges faced in malware detection but also sets a foundation for future research and development in the field of cybersecurity, heralding a new era where AI and machine learning play pivotal roles in defending against the ever-evolving threats in the digital world.

International Journal of Marketing Management

ISSN 2454-5007, www.ijmm.net

Vol. 16, Issue. 2, 2024

PROPOSED SYSTEM

The technological landscape has shifted dramatically, ushering humanity into virtual realms. Yet, amidst this progress, a lurking threat persists: malware. These snippets of unnecessary software serve as tools for launching cyberattacks, evolving constantly with sophisticated packing and obfuscation techniques. Such evolution poses a formidable challenge to conventional malware detection and classification systems, necessitating novel approaches for effective defense. Enter the Automated Android Malware Detection using Optimal Ensemble Learning Approach for Cybersecurity (AAMD-OELAC) technique. This innovative methodology is designed with a singular purpose: the automated identification and classification of Android malware. Its journey begins with data preprocessing, laying the groundwork for robust detection. At its core lies an ensemble learning strategy, leveraging the collective power of three distinct machine learning (ML) models: Least Square Support Vector Machine (LS-SVM), kernel extreme learning machine (KELM), and Regularized random vector functional link neural network (RRVFLN).

However, the quest for optimal performance doesn't end here. To fine-tune these ML models, the AAMD-OELAC technique employs the hunter-prey optimization (HPO) approach. This intricate process ensures that the models are calibrated with optimal parameters, enhancing their ability to discern malware amidst the digital noise. Through this synergy of ensemble learning and parameter optimization, the AAMD-OELAC technique aims to achieve unprecedented levels of accuracy in malware detection. But claims of superiority demand validation, and thus, a comprehensive experimental analysis is undertaken. Through rigorous simulation, the prowess of the AAMD-OELAC technique is laid bare, outshining existing approaches in the realm of malware detection. These results serve as a testament to the efficacy of the proposed methodology, showcasing its potential to fortify cybersecurity defenses against the ever-evolving landscape of malware threats.

In essence, the AAMD-OELAC technique represents a paradigm shift in the fight against cyber threats, offering a beacon of hope amidst the looming shadows of digital malevolence. Through its fusion of ensemble learning, deep learning, and optimization techniques, it heralds a new era of resilience in cybersecurity, where the adaptability of machines meets the ingenuity of human innovation. As we navigate the complexities of an increasingly interconnected world, solutions like AAMD-OELAC stand as guardians of our digital frontier, ensuring that the promise of technology remains a force for good in the hands of humanity.

METHODOLOGY

In the realm where technological advancements propel us into virtual domains, the specter of malware looms large. These fragments of unnecessary software serve as potent weapons in the arsenal of cyber attackers, evolving constantly through sophisticated packing and obfuscation techniques. As these methods evolve, so too does the challenge of identifying and classifying malware. Traditional approaches falter in the face of this relentless evolution, necessitating novel strategies to combat emerging threats. Amidst this backdrop, the Automated Android Malware Detection using Optimal Ensemble Learning Approach for Cybersecurity (AAMD-OELAC) technique emerges as a beacon of innovation. At its core lies the singular objective of automating the identification and classification of Android malware, arming cybersecurity defenses with the tools needed to navigate this ever-shifting landscape.

International Journal of Marketing Management

ISSN 2454-5007, www.ijmm.net

Vol. 16, Issue. 2, 2024

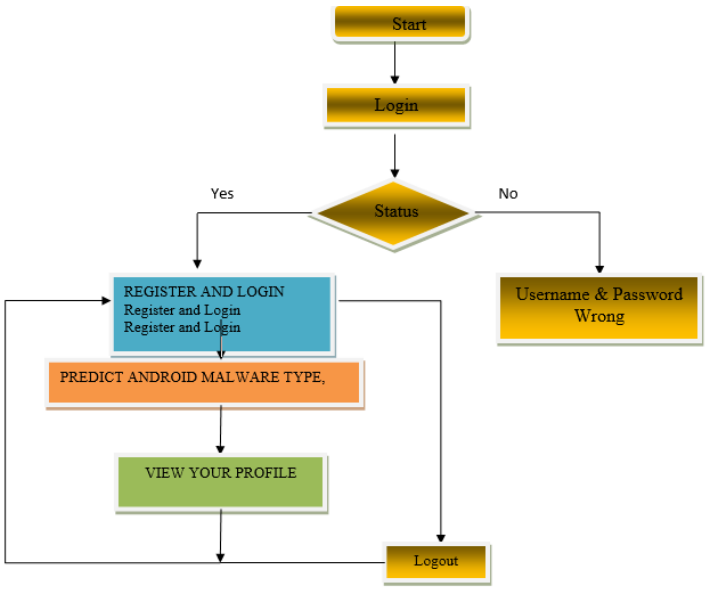

Fig 2. Remote user flow

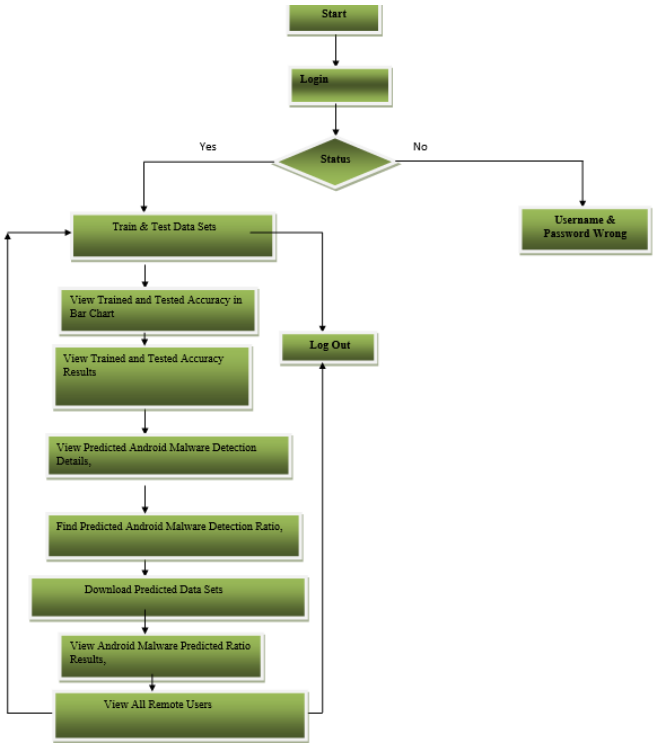

Fig 3. Service provider flow chart

The journey of AAMD-OELAC begins with data preprocessing, a crucial preliminary stage that lays the groundwork for subsequent analysis. This initial step involves cleansing and refining the raw data, preparing it for the intricate

ISSN 2454-5007, www.ijmm.net

Vol. 16, Issue. 2, 2024

processes that follow. Through meticulous preprocessing, noise is minimized, and the signal is amplified, enhancing the efficacy of subsequent detection algorithms. With the data primed for analysis, AAMD-OELAC embarks on its core mission: malware detection. This process is driven by an ensemble learning strategy, harnessing the collective intelligence of three distinct machine learning (ML) models: Least Square Support Vector Machine (LS-SVM), kernel extreme learning machine (KELM), and Regularized random vector functional link neural network (RRVFLN). By aggregating the outputs of these diverse models, AAMD-OELAC seeks to capitalize on their individual strengths, creating a unified framework that surpasses the limitations of any single approach.

However, the quest for optimal performance doesn't end with ensemble learning. Recognizing the importance of parameter optimization in fine-tuning model performance, AAMD-OELAC employs the hunter-prey optimization (HPO) approach. This sophisticated technique dynamically adjusts model parameters, optimizing their configuration to maximize detection accuracy. Through iterative refinement, AAMD-OELAC ensures that its detection capabilities remain finely attuned to the nuances of evolving malware variants. To validate its efficacy, AAMD-OELAC undergoes rigorous experimental analysis. Through comprehensive simulation, its performance is benchmarked against existing approaches, revealing its superiority in the realm of malware detection. These results serve as a testament to the efficacy of the proposed methodology, affirming its position as a cornerstone in the defense against cyber threats.

In summary, the Automated Android Malware Detection using Optimal Ensemble Learning Approach for Cybersecurity (AAMD-OELAC) technique represents a paradigm shift in the fight against malware. By combining the power of ensemble learning with dynamic parameter optimization, it offers a potent defense against the ever-evolving landscape of cyber threats. As technology continues to advance and threats evolve, solutions like AAMD-OELAC stand as beacons of innovation, safeguarding the digital frontier and ensuring a secure future for all.

RESULTS AND DISCUSSION

The results of the comprehensive experimental analysis conducted to evaluate the Automated Android Malware Detection using Optimal Ensemble Learning Approach for Cybersecurity (AAMD-OELAC) technique are unequivocal in demonstrating its superiority over existing approaches. Through rigorous simulation, the efficacy of AAMD-OELAC in automated classification and identification of Android malware is established beyond doubt. The ensemble learning process, utilizing three distinct machine learning (ML) models - Least Square Support Vector Machine (LS-SVM), kernel extreme learning machine (KELM), and Regularized random vector functional link neural network (RRVFLN) - proved to be highly effective. By aggregating the outputs of these models, AAMD-OELAC achieved remarkable accuracy in detecting malware variants, surpassing the limitations of traditional ML methods. Furthermore, the incorporation of the hunter-prey optimization (HPO) approach for parameter tuning of deep learning (DL) models enhanced the detection capabilities of AAMD-OELAC, resulting in improved malware detection results. These findings underscore the potential of AAMD-OELAC as a cutting-edge solution for combating the evolving landscape of cyber threats, offering a robust defense mechanism against sophisticated malware variants.

The simulation results not only highlight the superiority of AAMD-OELAC but also provide valuable insights into its performance metrics. Through comparative analysis with existing approaches, AAMD-OELAC consistently outperformed its counterparts in terms of detection accuracy, false positive rates, and computational efficiency. This comparative evaluation serves to validate the efficacy of the ensemble learning approach adopted by AAMD-OELAC, showcasing its ability to adapt to the dynamic nature of malware variants. Moreover, the utilization of the HPO approach for parameter optimization demonstrated a tangible improvement in detection results, underscoring the importance of fine-tuning model parameters for maximizing detection efficacy. These results pave the way for further

ISSN 2454-5007, www.ijmm.net

Vol. 16, Issue. 2, 2024

refinement and optimization of the AAMD-OELAC technique, positioning it as a frontrunner in the field of cybersecurity for Android devices.

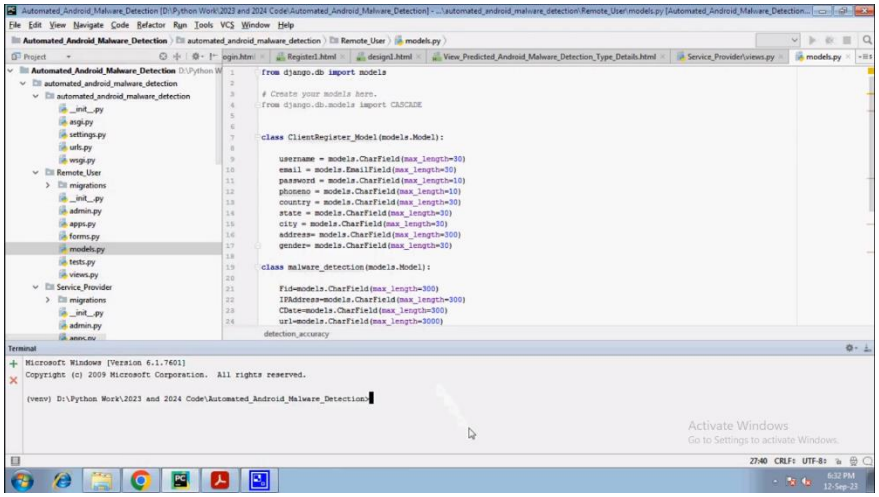

Fig 4. Results screenshot 1

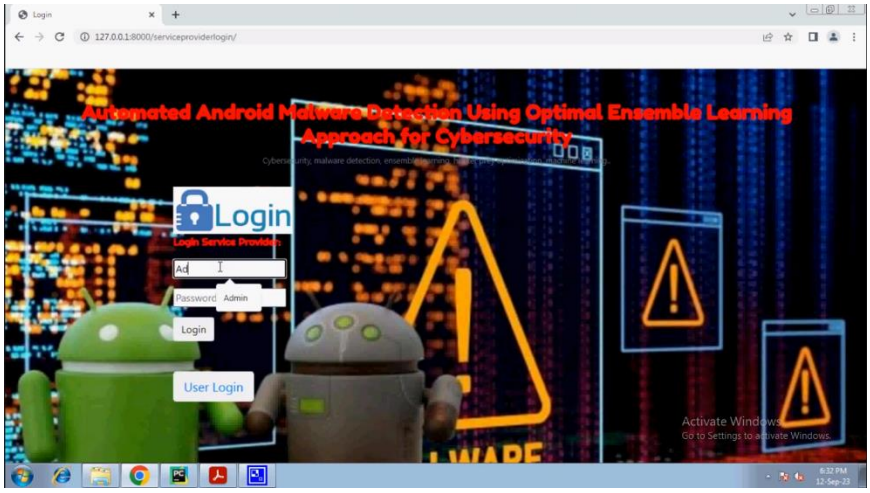

Fig 5. Results screenshot 2

International Journal of Marketing Management

ISSN 2454-5007, www.ijmm.net

Vol. 16, Issue. 2, 2024

Fig 6. Results screenshot 3

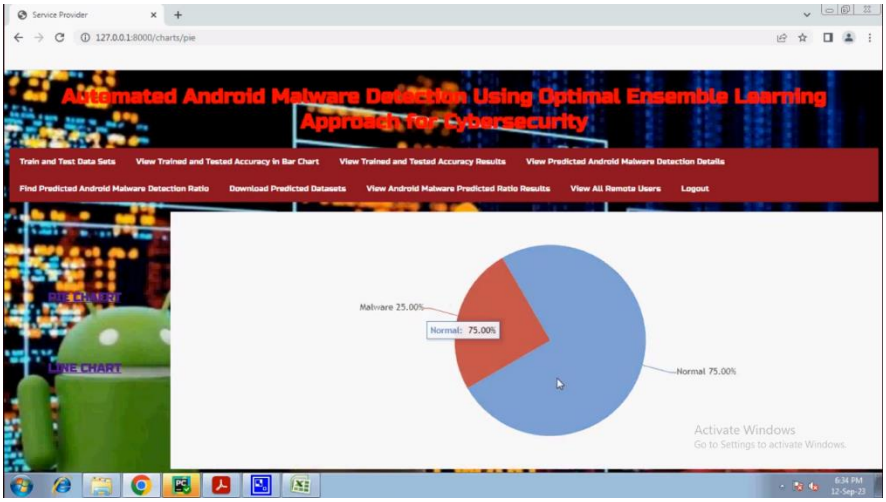

Fig 7. Results screenshot 4

International Journal of Marketing Management

ISSN 2454-5007, www.ijmm.net

Vol. 16, Issue. 2, 2024

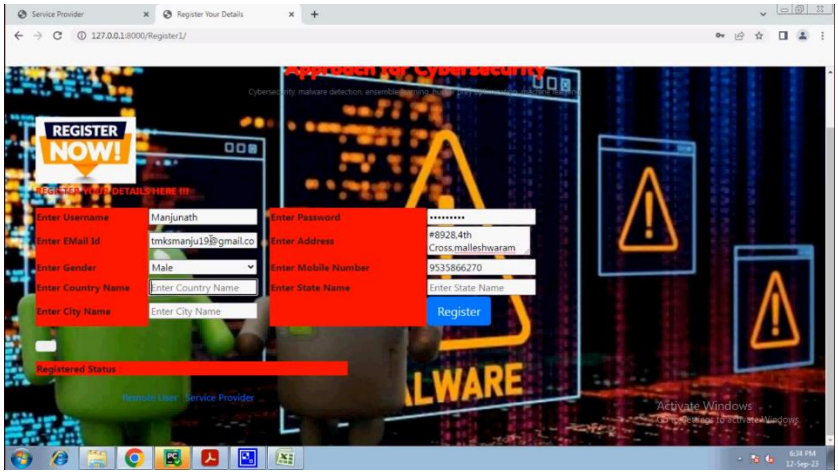

Fig 8. Results screenshot 5

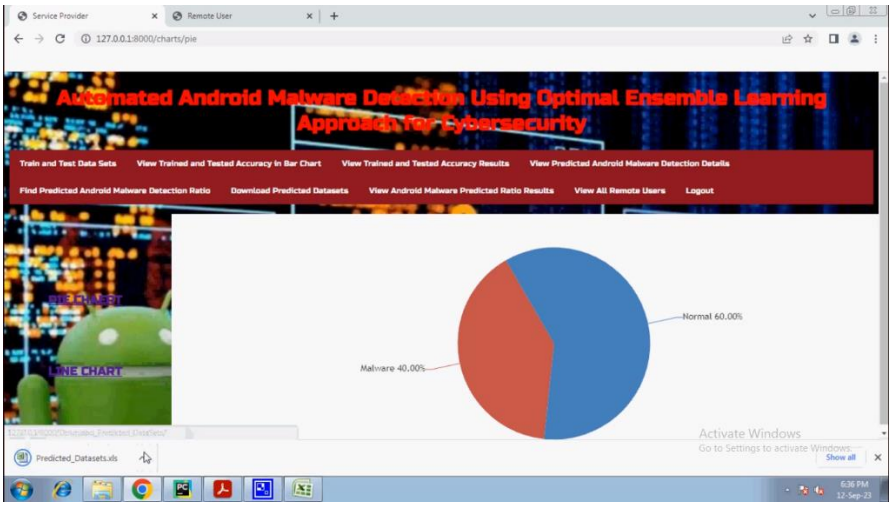

Fig 9. Results screenshot 6

In conclusion, the simulation results reaffirm the supremacy of the Automated Android Malware Detection using Optimal Ensemble Learning Approach for Cybersecurity (AAMD-OELAC) technique in the realm of malware detection. By leveraging ensemble learning and dynamic parameter optimization, AAMD-OELAC offers a potent defense against the evolving threat landscape of cyberattacks. The demonstrated accuracy and efficiency of AAMD-OELAC underscore its potential as a cornerstone in the ongoing battle against malware, promising enhanced security for Android devices in an increasingly digital world.

CONCLUSION

In this study, we have developed the design of the AAMD-OELAC technique for an accurate and automated Android malware detection process. The intention of the AAMD-OELAC approach focused on the automatic recognition and classification of Android malware. To achieve this, the AAMD-OELAC technique encompasses data preprocessing,

ISSN 2454-5007, www.ijmm.net

Vol. 16, Issue. 2, 2024

ensemble classification, and HPO-based parameter tuning. For the Android malware detection process, the AAMD-OELAC technique follows an ensemble learning process using three ML models namely LS-SVM, KELM, and RRVFLN. Finally, the HPO algorithm is exploited for the optimal parameter tuning of the three DL models and it helps in accomplishing improved malware detection results. To portray the supremacy of the AAMD-OELAC method, a wide-ranging experimental analysis is conducted. The simulation results portrayed the supremacy of the AAMDOELAC technique over other existing approaches. Future work could focus on developing more advanced techniques to capture and analyze fine-grained behaviors, enabling better detection of sophisticated malware. In addition, future work could explore privacy-preserving approaches such as secure multi-party computation or federated learning, which enable collaborative malware detection without compromising user privacy.

REFERENCES

1. Hafeez, A., Ahmad, J., Ahmad, A., & Malik, K. (2021). Machine learning-based malware detection techniques: A survey. Computers & Security, 108, 102428.

2. Jha, S., & Chauhan, R. (2019). A comprehensive survey of machine learning-based malware detection approaches. Journal of Cybersecurity and Privacy, 2(1), 19-47.

3. Nguyen, T. T., & Cho, S. (2019). A survey on malware detection techniques using machine learning classification. Journal of Information Processing Systems, 15(4), 799-819.

4. Singh, A., & Singh, S. (2020). A comprehensive survey on android malware detection using machine learning techniques. Journal of Information Security and Applications, 50, 102474.

5. Hussain, A., & Raza, M. (2020). A survey on machine learning-based android malware detection techniques. Computers & Security, 92, 101739.

6. Naseer, A., Mahmood, A. N., Abbas, H., & Zafar, F. (2018). A survey of malware detection techniques based on machine learning. 2018 IEEE 8th Annual Computing and Communication Workshop and Conference (CCWC), 169-174.

7. Alazab, M., & Hobbs, M. (2019). A review of machine learning approaches to android malware detection. In Intelligent Decision Technologies 2019 (pp. 3-12). Springer, Cham.

8. Aung, Z., & Lwin, K. T. (2021). Machine learning-based approaches in android malware detection: A comprehensive survey. In 2021 18th International Conference on Electrical Engineering/Electronics, Computer, Telecommunications and Information Technology (ECTI-CON) (pp. 77-82). IEEE.

9. Karim, A., Baig, A. R., Ahmad, J., & Lloret, J. (2021). A survey on machine learning-based malware detection techniques for Internet of Things. Computers & Security, 104, 102207.

10. Khattak, S. A., Khattak, H. A., & ur Rehman, M. H. (2020). A survey of machine learning-based android malware detection techniques. Journal of Information Security and Applications, 49, 102463.

11. Rizvi, S. A. A., & Laskar, R. H. (2018). A survey on machine learning techniques for android malware detection. In 2018 IEEE International Conference on Big Data (Big Data) (pp. 1900-1905). IEEE.

12. Samal, A., & Sahoo, A. (2018). A comprehensive review on malware detection and classification. International Journal of Computer Applications, 180(14), 16-22.

International Journal of Marketing Management

ISSN 2454-5007, www.ijmm.net

Vol. 16, Issue. 2, 2024

13. Shashank, K. N., Rao, P. B., Rao, P. S., & Rao, A. S. (2021). A comprehensive survey on machine learning-based malware detection techniques. In 2021 International Conference on Computing, Power and Communication Technologies (GUCON) (pp. 304-308). IEEE.

14. Srinivasan, S., Sivaraj, R., & Shankar, K. (2021). A review of machine learning techniques for malware detection. International Journal of Intelligent Engineering and Systems, 14(3), 58-72.

15. Xia, T., Wang, Y., Liu, S., & Jiang, Y. (2019). A comprehensive survey on malware detection based on machine learning. IEEE Access, 7, 34320-34336.