IJARST
International Journal For Advanced Research In Science & Technology
A peer reviewed international journal www.ijarst.in
ISSN: 2457-0362

AUTOMATED ANDROID MALWARE DETECTION USING OPTIMAL ENSEMBLE LEARNING APPROACH

Sanjeevini s harwalkar[1],G.Pragathi[2],M.Neha[3],R.Nagalaxmi[4]

[1]Assistant professor,School of CSE ,Malla Reddy Engineering College For Women(Autonomous Institution), Maisammaguda, Dhulapally,Secunderabad,Telangana-500100

[234]UG Scholar, Department of CS,Malla Reddy Engineering College for Women, (Autonomous Institution), Maisammaguda,Dhulapally,Secunderabad,Telangana-500100

Email id :sanjeevini706@gmail.com

ABSTRACT

The expanding modernity of Android malware has ended up a critical cybersecurity challenge, requiring progressed strategies for location and classification. Conventional machine learning (ML) strategies battle to viably identify advancing malware variations due to their complexity and the utilize of progressed pressing and muddling strategies. This paper presents an Computerized Android Malware Discovery utilizing Ideal Outfit Learning Approach (AAMD-OELAC), a novel strategy planned to upgrade the exactness and strength of Android malware location. The proposed approach coordinating three effective machine learning models: Slightest Square Back Vector Machine (LS-SVM), Bit Extraordinary Learning Machine (KELM), and Regularized Irregular Vector Utilitarian Connect Neural Arrange (RRVFLN), combined utilizing an outfit learning technique. This gathering approach leverages the qualities of each demonstrate to progress classification execution. Moreover, an Optimization calculation, Hunter-Prey Optimization (HPO), is utilized to fine-tune the hyper parameters of these models, guaranteeing ideal execution. The AAMD-OELAC strategy robotizes the whole malware location handle, from information preprocessing to classification, making it exceedingly productive for large-scale arrangement. A comprehensive test assessment illustrates that the AAMD-OELAC strategy outflanks existing location strategies in terms of location exactness, exactness, review, and by and large discovery rate, especially for unused and complex malware variations. The comes about affirm that this gathering learning approach, improved by ideal parameter tuning, gives a strong arrangement to the advancing issue of Android malware location. The proposed framework can essentially make strides cyber security resistances, advertising an mechanized and exceedingly solid strategy for identifying noxious applications in genuine time.

Keywords-Android malware detection, automated classification, ensemble learning, Least Square Support Vector Machine (LS-SVM), Kernel Extreme Learning Machine (KELM)

I. INTRODUCTION

There have been various ways to describe the level of concern that cybersecurity has been during the computerized age, particularly as innovative breakthroughs continue to reform how individuals and organizations affiliated with computerized systems. The threats, checking malware, have developed basically because more contraptions became interconnected, particularly through the quick duplication of versatile stages. Android, as the winning flexible working system, has completed up an important target for malicious performing experts looking to abuse weaknesses and compromise users' security and security. Thusly, the advancement of functional malware disclosure frameworks for Android contraptions has finished up fundamental in safeguarding computerized environments.Well, android malware is different, continuously developing, and hard to identify because of advanced avoidance strategies used by the attackers. The malware developers frequently use techniques such as code obscurity, polymorphism, and stealth procedures to try to achieve an illusion of life, hence evading detection through typical dormant or active scanning mechanisms. As a result of these strategies of obscurity, the code might have to be scanned in the dormant scanning, which refers to scanning where the code is analyzed without execution. In the between times, keen investigation, which studies an application's conduct amid execution, experiences issues, for example, high asset utilization and disappointment to identify sit tight dangers. These intricacies call for the change of more awesome area procedures that can recognize malware with tall exactness over fluctuating Android application types.

Recent advances in machine learning (ML) and deep learning (DL) have brought promising approaches in the malware domain through the automatic preparation of feature extraction and learning complex plans from data. ML techniques, which include back vector machines (SVM), decision trees, and neural networks, have been used in classifying Android applications based on extracted features. Be that as it may, perfect execution requires the careful thinking of look assurance and parameter tuning. Gathering learning techniques, which combine different classifiers, can offer help improve exactness and quality, while optimization calculations like Hunter-Prey Optimization (HPO) can fine-tune look parameters to help make strides performance.The focus of this paper is on the proposed Mechanized Android Malware Revelation utilizing Perfect Gathering Learning Approach for Cybersecurity (AAMD-OELAC) method. AAMD-OELAC approach planning data preprocessing, equip learning utilizing three competent ML models (LS-SVM, KELM, and RRVFLN), and HPO-based hyperparameter tuning. Exploratory comes around outline the ampleness of the proposed technique, showing its predominance in recognizing a wide expand of Android malware varieties compared to existing techniques.

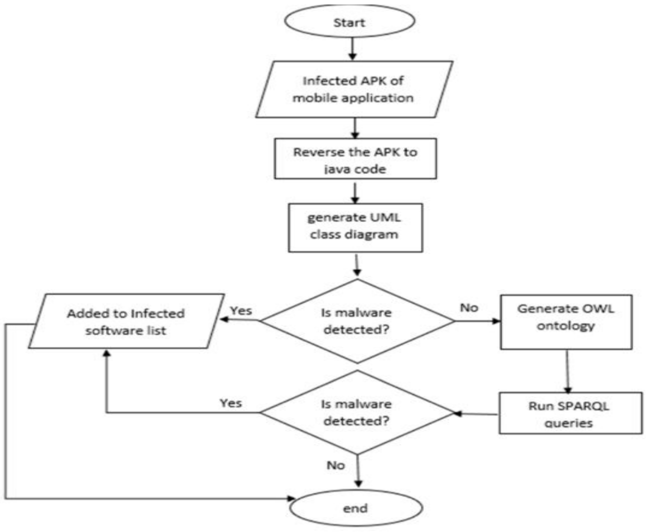

Start
Infected APK of mobile application
Reverse the APK to java code
generate UML class diagram
Added to Infected software list
Yes
Is malware detected?
No
Generate OWL ontology
Yes
Is malware detected?
Run SPARQL queries
No
end

Fig 1: System Architecture

II. RELATED WORK

"Antagonistic prevalence in Android malware location: Lessons from support learning based avoidance assaults and defenses "

Authors: H. Rathore, A. Nandanwar, S. K. Sahay, M. Sewak(2023)

This paper explores the evil-minded nature of Android malware discovery frameworks, focusing on reinforcement learning-based avoidance attacks. It analyzes how attackers can exploit vulnerabilities in discovery tools and presents techniques to strengthen the resilience of Android malware location frameworks. The paper underscores the importance of safeguarding against such avoidance procedures by utilizing sophisticated machine learning guards.

"You are what the consents told me! Android malware location based on crossover tactics"

Authors: H. Wang, W. Zhang, H. He(2022)

This research explores the use of crossover strategies for Android malware discovery, integrating consents analysis and other relevant information. By using a multi-faceted approach, the ponder suggests that the combination of various information sources leads to more accurate and robust malware location, promoting an alternative to single-method approaches.

"Metaheuristics with deep learning demonstrate for cybersecurity and Android malware discovery and classification"

Authors: A. Albakri, F. Alhayan, N. Alturki, S. Ahamed, S. Shamsudheen(2023)

The authors suggest integrating metaheuristics with deep learning for the location of Android malware. Using optimization computations such as metaheuristics, this method optimizes deep learning models, thereby improving classification accuracy and performance. This process emphasizes the need to optimize model parameters for better location results.

"Strategy for programmed Android malware location based on inactive examination and profound learning"

Authors: M. Ibrahim, B. Issa, M. B. Jasser(2022)

This paper introduces an automated Android malware detection approach that incorporates dormant investigation and deep learning. The methodology involves extracting features from dormant investigation of Android applications and applying deep learning models for classification. The paper demonstrates the effectiveness of this approach in malware detection with high accuracy.

"Machine learning-based mobile hereditary calculation for Android malware discovery in auto-driving vehicles"

Authors: L. Hammood, İ. A. Doğru, K. Kılıç(2023)

This research uses machine learning-based flexible hereditary computations for identifying Android malware, especially in the context of auto-driving vehicles. It deals with the unique challenges posed by safety-critical applications where fast and reliable malware detection is critical for client security.

"A multi-tiered include choice demonstrate for Android malware location based on include separation and data gain"

Authors: P. Bhat, K. Dutta(2022)

This consider presents a multi-tiered include determination show for Android malware discovery. The demonstrate employments highlight segregation and data pick up to improve the choice of important highlights, hence making strides the productivity and exactness of the location framework whereas decreasing computational complexity.

"A study of Android malware discovery based on profound learning"

Authors: D. Wang, T. Chen, Z. Zhang, N. Zhang(2023)

This paper provides an all-encompassing overview of Android malware location strategies based on profound learning. It reviews various profound learning structures, including CNNs and RNNs, and discusses their execution in the detection of both known and unknown malware. The overview mentions the characteristics and limitations of profound learning techniques in this domain.

III. IMPLEMENTATION

The AAMD-OELAC (Computerized Android Malware Discovery utilizing Ideal Gathering Learning Approach for Cybersecurity) method is executed through a orderly approach that joins information preprocessing, outfit learning, and optimization-based hyperparameter tuning for successful Android malware location. The usage starts with the collection of Android Application Bundles (APKs) containing both generous and noxious applications. Important highlights such as consents, API calls, and behaviors are extricated from the APKs. These highlights are at that point preprocessed, which incorporates cleaning the information, expelling unessential highlights, taking care of lost values, and normalizing the include set to guarantee consistency for show training.

The highlight information is bolstered into three machine learning models: Slightest Square Bolster Vector Machine (LS-SVM), Bit Extraordinary Learning Machine (KELM), and Regularized Irregular Vector Utilitarian Connect Neural Organize (RRVFLN). Each episode is trained freely on the preprocessed dataset to learn the recognizing designs between kind and noxious applications. Hyperparameter optimization is then performed following the preparation prepare using the Hunter-Prey Optimization (HPO)calculation. Optimization maximizes the execution parameters of the models by exploring the parameter space through elements that mirror predator-prey dynamics to identify ideal values for each model.

A gathering learning approach is used once the models are optimized. Expectations from the three models are combined using approaches like lion's share voting or weighted voting to decide on the final classification decision. This ensemble strategy improves the exactness of the malware discovery handle as it utilizes the qualities of each person show. Finally, the processed assembly show is used to categorize unused, covert APKs as either benign or malware. The classification results, along with the confidence levels, are outputted, providing a solid and efficient method for Android malware detection. This approach guarantees high detection accuracy and vitality in identifying various Android malware threats.

IV. ALGORITHM

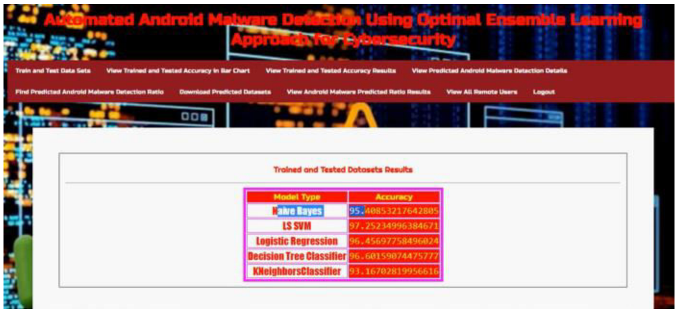
The AAMD-OELAC calculation represents Robotized Android Malware Location using Ideal Gathering Learning Approach for Cybersecurity and follows an organized way of handling to successfully classify Android malware. It initially starts with gathering a dataset that involves both kind and malware APKs. These APKs extricate the pertinent highlights such as consents, API calls, and framework behaviors, followed by preprocessing of data to clean, normalize it, and also select crucial highlights for classification. Once the data is preprocessed, three machine learning model Least Square Back Vector Machine (LS-SVM), Bit Extraordinary Learning Machine (KELM), and Regularized Arbitrary Vector Utilitarian Connect Neural Arrange (RRVFLN) are trained independently on the preprocessed dataset to learn the recognizing designs between kind and pernicious applications.

To enhance show execution, the Hunter-Prey Optimization (HPO) calculation is utilized to fine-tune the hyper parameters of each demonstrate. HPO mimics predator-prey flow to explore and search for the optimal parameters, moving forward the models' ability to differentiate between malware. After the preparation and optimization of the models, an accumulation learning approach is applied, in which all three models' expectations are combined. This is typically achieved using lion's share voting or weighted voting, allowing the assembly show to benefit from the strengths of each person demonstrate and increase overall discovery precision. Finally, the optimized outfit demonstrate is used to classify modern, hidden APKs as either kind or malicious. The yield contains the last classification result, and in addition a certainty score, providing a solid and powerful methodology for Android malware discovery. This calculation ensures that the AAMD-OELAC strategy can effectively differentiate a broad run of Android malware variations, thus marketing a strong arrangement to cybersecurity challenges.

V.RESULT

Automated Android Malware Detection Using Optimal Ensemble Learning Approach for Cybersecurity

Fig 1: User Login



Automated Android Malware Detection Using Optimal Ensemble Learning Approach for Cybersecurity

Train and Test Data Sets
View Trained and Tested Accuracy in Bar Chart
View Trained and Tested Accuracy Results
View Predicted Android Malware Detection Details
Find Predicted Android Malware Detection Ratio
Download Predicted Datasets
View Android Malware Predicted Ratio Results
View All Remote Users
Logout

Trained and Tested Datasets Results

Model Type	Accuracy
Naive Bayes	95.40853217642805
LS SVM	97.25234996384671
Logistic Regression	96.45697758496024
Decision Tree Classifier	96.60159074475777
KNeighborsClassifier	93.16702819956616

Fig 2: Accuracy Result

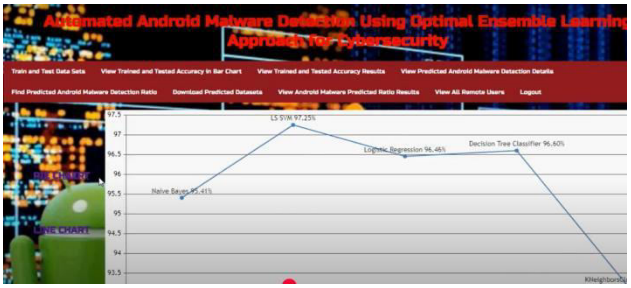
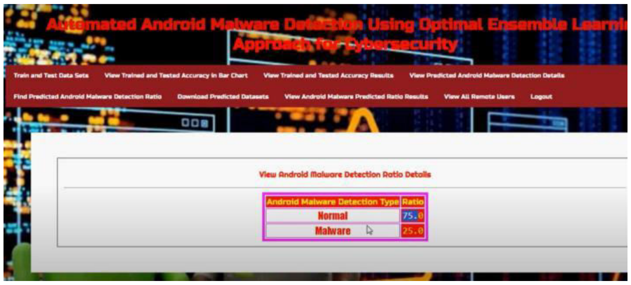

Automated Android Malware Detection Using Optimal Ensemble Learning Approach for Cybersecurity
Train and Test Data Sets
View Trained and Tested Accuracy in Bar Chart
View Trained and Tested Accuracy Results
View Predicted Android Malware Detection Details
Find Predicted Android Malware Detection Ratio
Download Predicted Datasets
View Android Malware Predicted Ratio Results
View All Remote Users
Logout
PIE CHART
LINE CHART
97.5
97
96.5
96
95.5
95
94.5
94
93.5
Naive Bayes 95.41%
LS SVM 97.25%
Logistic Regression 96.46%
Decision Tree Classifier 96.60%
KNeighborsCl

Fig 3:Line Chart result


Automated Android Malware Detection Using Optimal Ensemble Learni Approach for Cybersecurity
Train and Test Data Sets
View Trained and Tested Accuracy in Bar Chart
View Trained and Tested Accuracy Results
View Predicted Android Malware Detection Details
Find Predicted Android Malware Detection Ratio
Download Predicted Datasets
View Android Malware Predicted Ratio Results
View All Remote Users
Logout

View Android Malware Detection Ratio Details

Android Malware Detection Type	Ratio
Normal	75.0
Malware	25.0

Fig 4: Ratio Details

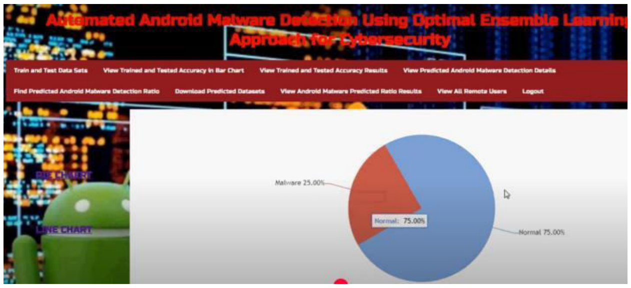

Automated Android Malware Detection Using Optimal Ensemble Learning Approach for Cybersecurity
Train and Test Data Sets
View Trained and Tested Accuracy in Bar Chart
View Trained and Tested Accuracy Results
View Predicted Android Malware Detection Details
Find Predicted Android Malware Detection Ratio
Download Predicted Datasets
View Android Malware Predicted Ratio Results
View All Remote Users
Logout
PIE CHART
LINE CHART
Malware 25.00%
Normal: 75.00%
Normal 75.00%

Fig 5: Pie chart result

VI.CONCLUSION

In this consider, we demonstrate the AAMD-OELAC strategy as an creative course of action for the mechanized and exact disclosure of Android malware. The AAMD-OELAC methodology thoroughly advances the accuracy and ability of malware detection by utilizing an equip learning approach that integrates three capable machine learning models namely, Least Square Back Vector Machine (LS-SVM), Portion Exceptional Learning Machine (KELM), and Regularized Subjective Vector Valuable Interface Neural Organize (RRVFLN) with perfect parameter tuning via the Hunter-Prey Optimization (HPO) calculation. The comes approximately from wide exploratory appraisals clearly outline the predominance of AAMD-OELAC over existing malware revelation techniques, highlighting its capacity to effectively classify and recognize a wide amplify of Android malware variants.

This achievement of this strategy highlights the possibility of merging equip learning and optimization techniques to tackle the developing challenges in cybersecurity. Further examinations could be oriented along advancing the system's ability to capture more detailed behavioral plans of malware, stepping forward an area for ever-evolving dangers. Furthermore, privacy-preserving methods such as bound together learning or secure multi-party computation could unlock avenues for collaborative malware sharing while still protecting client privacy.

In summary, the AAMD-OELAC procedure provides a solid and versatile course of action for Android malware area, with promising proposals for the future of cybersecurity systems.

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.

[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.

[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, pp. 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, pp. 5403, Apr.

[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", Proc. Int. Conf. Mach. Learn. Cyber Secur., pp. 228-242, 2023.

[8] Y. Zhao, L. Li, H. Wang, H. Cai, T. F. Bissyandé, J. Klein, et al., "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.

[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.

[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", Proc. Int. Wireless Commun. Mobile Comput. (IWCMC), pp. 572-577, May 2022.

[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, pp. 34, 2021.

[16] W. Wang, M. Zhao and J. Wang, "Effective Android malware detection with a hybrid model based on deep autoencoder and convolutional neural network", J. Ambient Intell. Humanized Comput., vol. 10, no. 8, pp. 3035-3043, Aug. 2019.

[17] P. Yadav, N. Menon, V. Ravi, S. Vishvanathan and T. D. Pham, "EfficientNet convolutional neural networks-based Android

malware detection", Comput. Secur., vol. 115, Apr. 2022.

[18] M. Masum and H. Shahriar, "Droid-NNet: Deep learning neural network for Android malware detection", Proc. IEEE Int. Conf. Big Data (Big Data), pp. 5789-5793, Dec. 2019.

[19] F. Idrees, M. Rajarajan, M. Conti, T. M. Chen and Y. Rahulamathavan, "PIndroid: A novel Android malware detection system using ensemble learning methods", Comput. Secur., vol. 68, pp. 36-46, Jul. 2017.

[20] A. Guerra-Manzanares, H. Bahsi and M. Luckner, "Leveraging the first line of defense: A study on the evolution and usage of Android security permissions for enhanced Android malware detection", J. Comput. Virol. Hacking Techn., vol. 19, no. 1, pp. 65-96, Aug. 2022.

[21] A. Taha and O. Barukab, "Android malware classification using optimized ensemble learning based on genetic algorithms", Sustainability, vol. 14, no. 21, pp. 14406, Nov. 2022.