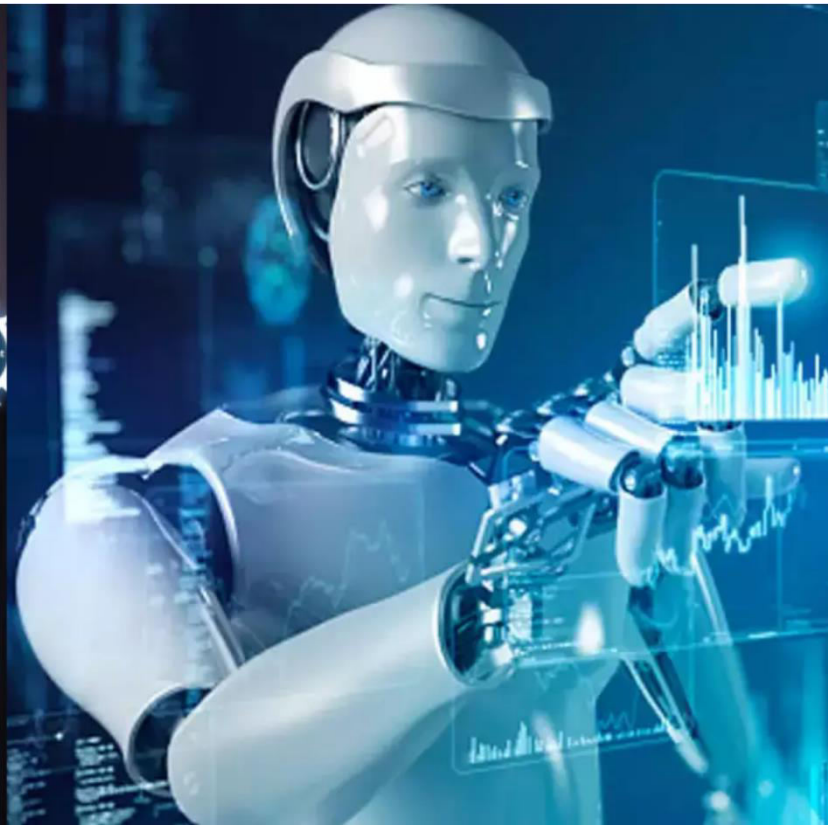
IJIRSET
International Journal of Innovative Research in Science | Engineering | Technology

ISSN(O): 2319-8753
ISSN(P): 2347-6710

International Journal of Innovative Research in Science Engineering and Technology (IJIRSET)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)



Impact Factor: 8.699

Volume 14, Issue 4, April 2025

International Journal of Innovative Research of Science, Engineering and Technology (IJIRSET)

Impact Factor 8.699

|www.ijirset.com |A Monthly, Peer Reviewed & Refereed Journal| e-ISSN: 2319-8753| p-ISSN: 2347-6710|

Volume 14, Issue 4, April 2025

|DOI: 10.15680/IJIRSET.2025.1404724|

A Comprehensive Analysis of Malware Detection on Android Using Machine Learning

R. Suresh Menan, N. Shanmugalakshmi

PG Scholar, Master of Computer Application, Karpagam Academy of Higher Education, Coimbatore, India

MCA, Assistant Professor, Master of Computer Application, Karpagam Academy of Higher Education, Coimbatore, India

ABSTRACT: With more mobile devices, especially Android smartphones, users are increasingly vulnerable to malware. Our work uses Decision Trees, SVMs, RFs, and KNNs to improve Android security and solve this issue. A successful Android malware detection and prevention solution is the aim. The effectiveness of these machine-learning approaches will be tested against various infections. Due to the rise of Android devices, security concerns have increased, requiring creative technologies that can identify new threats. Machine learning algorithms that discover trends and anticipate may be an Android security option. A robust malware detection system was built using DT, KNN, RF, and SVM algorithms in our research. We rate many machine learning techniques for Android malware detection by speed, accuracy, and efficiency. Insights should help choose and apply Android security machine learning solutions. Decision trees are transparent, random forests utilize ensemble learning, support vector machines excel in probabilistic classification, and k-nearest neighbours (KNN) are best at identifying Android infections based on similarity patterns, according to research. Comparative research provides subtle insights into the best security measure. The effort culminates by strengthening Android security using machine learning and laying the groundwork for mobile device security research. We want to solve security issues and improve Android device security and resilience to new attacks.

KEYWORDS*:* Android security, Decision Tree, Random Forest, Support vector Machine, k-Nearest Neighbors, Machine Learning, Malware Detection

I. INTRODUCTION

Android devices are ubiquitous now, thus ensuring their security is of the utmost importance. The proliferation of malicious viruses in mobile applications is a major concern for cybersecurity professionals and users. In reaction to this evolving threat landscape, security approaches are seeing a sea shift, with a resur gence in the use of state-of-the-art technologies to fortify Android devices' defences.

Because of the ever-evolving nature of cyber threats, this project is dedicated to researching "Machine Learning Strategies for Malware Detection in Android Devices," an area that requires a versatile and adaptable approach. Innovative approaches are required as novel attacks often succeed when more conventional malware detection methods fail. Machine learning has enormous promise for enhancing Android device security due to its ability to detect intricate patterns and anomalies. Enormous promise for enhancing Android device security due to its ability to detect intricate patterns and anomalies.

This research primarily aims to assess the effectiveness of several machine-learning approaches developed specifically for Android malware detection, including k-nearest Neighbors (KNN), Random Forest (RF), Naive Bayes (NB), and Decision Trees (DT). The objective is to develop a solid system that can use the predictive capabilities of these algorithms to identify and remove the many forms of malware that pose a genuine threat to Android devices.

We start our research by noting that the incorporation of machine learning into Android security is a significant technological advancement and an essential step in safeguarding mobile devices from the constantly evolving criminal cene. Both the field of machine learning and Android security may benefit greatly from this work's rewriting of standards. In the never-ending quest for strong cyber security defenses, the eventual objective is to make the internet a safer place for people everywhere.

International Journal of Innovative Research of Science, Engineering and Technology (IJIRSET)

Impact Factor 8.699

|www.ijirset.com |A Monthly, Peer Reviewed & Refereed Journal| e-ISSN: 2319-8753| p-ISSN: 2347-6710|

Volume 14, Issue 4, April 2025

|DOI: 10.15680/IJIRSET.2025.1404724|

A. Contribution

The scenario is defined by modern technology and the necessity to safeguard the quickly expanding Android ecosystem. Basic security measures are no longer enough due to the rising threat of sophisticated malware attacks. This project seeks to change Android malware protection by seamlessly combining state-of-the-art machine learning techniques including k-nearest Neighbors (KNN), Random Forest (RF), Support Vector Machine (SVM), and Decision Trees (DT). We have developed a dynamic and adaptable system that can detect and neutralize malware in real-time, therefore taking a proactive approach to emerging threats and lowering the risk for Android users. This proactive strategy is useful in the world of mobile security, where staying ahead of emerging risks is crucial. In addition, by thoroughly examining several algorithms to get a comprehensive grasp of their malware detection capabilities, our effort significantly advances the field of cyber security machine learning research. This growing understanding strengthens the foundation for future study while also adding to the body of information already known. Improved security for Android devices is the main benefit, since machine learning techniques' demonstrated efficacy may have an influence on security laws and encourage a safer online environment for all users. Ultimately, our findings mark a significant breakthrough in protecting Android devices from complex and dynamic cyber security threats.

B. Objective

The security of Android devices is an ongoing concern, and one of our aims is to find new and better methods to protect them from malware. A flexible and dynamic system for the real-time detection of malware on Android devices is our goal. To achieve this, we will study and use machine learning methods such as Decision Trees (DT), Random Forest (RF), k-nearest Neighbours (KNN), and Support Vector Machine (SVM). More than that, we want to provide useful information to the cybersecurity community by comparing the detection and mitigation capabilities of different machine learning algorithms for different kinds of malware. We want to enhance Android device security recommendations and, by extension, make the internet a safer place for people worldwide by proving the efficacy of these approaches.

II. RELATED WORKS

Because complex attackers are always inventing new methods to circumvent traditional defences, the research demonstrates that Android security concerns persist despite advancements [1]. There is an urgent need for tailored machin e-learning solutions to combat the dynamic landscape of mobile malware [2], as the problem statement notes, due to the specificity of attacks aimed at mobile devices. Finding the most effective machine learning methods and algorithms, while considering anypotential weaknesses in the existing security measures, may be the primary focus in the effort to enhance Android security [3]. At its heart, the problem is likely caused by the ever-changing nature of Android malware threats and the challenges in using decision trees for effective and real-time detection [4].

The biggest obstacle in the field of Android virus detection is understanding the pros and cons of various machine-learning approaches [5]. Although unrelated, the problem may revolve around developi ng novel clustering and machine-learning techniques, which might lead to fresh insights into Android device security [6]. Fixing vulnerabilities in mobile ad hoc networks, finding out how to ensure secure data transmission, and maybe influencing Android might all be part of the problem [7]. The issue may have something to do with optimising energy-efficient adaptive clustering in wireless sensor networks, which might provide light on problems comparable to those seen with Android security [8]. Improving the functionality and security of smartphone applications, particularly those running on the Android operating system, is likely to be hindered by a lack of understanding of user expectations and concerns [9]. The issue provides details on security methods that might be adjusted to accommodate Android mobile payment systems, specifically concerning vulnerabilities like double-spending, which is a concern when it comes to protecting virtual currency [10].

International Journal of Innovative Research of Science, Engineering and Technology (IJIRSET)

Impact Factor 8.699

|www.ijirset.com |A Monthly, Peer Reviewed & Refereed Journal| e-ISSN: 2319-8753| p-ISSN: 2347-6710|

Volume 14, Issue 4, April 2025

|DOI: 10.15680/IJIRSET.2025.1404724|

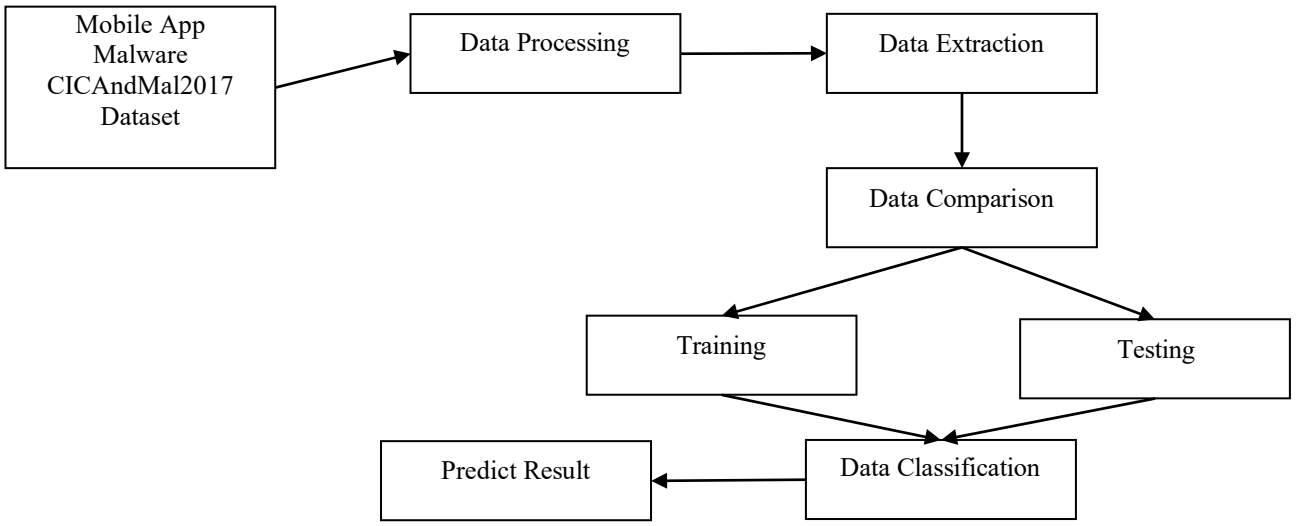


Fig.1.System Architecture

III. PROPOSED METHODOLOGY

Machine Learning Strategies for Android Malware Detection use a comprehensive methodology. To begin, a dataset including both safe and harmful Android applications will be constructed with great care to ensure that various characteristics of different malware versions are well covered. Using feature extraction techniques, we may glean useful application aspects by considering both static and dynamic features. To detect malware-like patterns, the next steps include training several machine learning algorithms on the dataset. These techniques include Naive Bayes, k-nearest neighbour, decision trees, and support vector machines.

Accuracy, recall, and F1-score are some of the metrics that will be used to evaluate the model's performance once it undergoes a thorough evaluation procedure. To make things more resilient, we're going to look at deep learning architectures and ensemble methodologies. Additional improvement will be achieved by hyperparameter tuning. The next step is to test the model in real-time on Android devices to see how well it performs in practice. A sophisticated and adaptable malware detection system tailored to Android smartphones is the primary goal of this proposed method, which aims to effectively address the dynamic mobile security threat landscape.

1. Dataset:

An essential and comprehensive resource for research on Android malwar e detection, the CICAndMal2017 dataset
was established by the Canadian Institute for Cybersecurity (CIC). Scholars and professionals will find this dataset, which contains a variety of network traffic statistics from various Android apps (both legal and malicious), to be an invaluable resource. A more thorough examination of Android virus activity is made possible by the dataset, which was developed under controlled settings and provides a realistic picture of real-world network interactions. With a focus on Android devices, CICAndMal2017 encourages the development and evaluation of intrusion detection systems, cyber security measures, and machine learning models. The topic of mobile security is constantly growing and challenging, and this resource is a must-have due to its comprehensiveness and well-chosen sample collection.

2. Data Processing:

Careful use of filtering and data cleaning techniques enhances the dataset's quality and relevance throughout data processing. Data may be included or excluded using filter procedures according to statistical or domain-specific criteria, which reduces dimensionality and improves the efficacy of future investigations. This selective retention of just the most informative and distinguishing characteristics enhances machine learning model performance by reducing the likelihood of over fitting.

International Journal of Innovative Research of Science, Engineering and Technology (IJIRSET)

Impact Factor 8.699

|www.ijirset.com |A Monthly, Peer Reviewed & Refereed Journal| e-ISSN: 2319-8753| p-ISSN: 2347-6710|

Volume 14, Issue 4, April 2025

|DOI: 10.15680/IJIRSET.2025.1404724|

At the same time, discrepancies, missing numbers, and other errors in the dataset are the primary targets of data cleaning. Data integrity is ensured by the meticulous use of procedures such as imputation, outlier elimination, and normalization. Not only does this thorough process enhance the dataset, but it also guarantees that the data used in subsequent research is accurate, reliable, and representative. This process ensures that important insights may be extracted from the data by providing a solid foundation for model creation.

3. Data Extraction:

The first stage in extracting data for Android mobile malware detection using Principal Component Analysis (PCA) for features is to collect several datasets, some of which include dangerous applications and others which do not. permission requests, system calls, and API requests are just a few of the aspects that are pertinent to the nature of mobile apps and are included in this dataset along with other static and dynamic data. To detect patterns of activity indicative of malware, the feature metrics compile a comprehensive set of attributes.

Step two involves reducing the original feature set to a smaller collection of primary components while preserving crucial features using principal component analysis (PCA), a dimensionality reduction method. By reducing the size of the dataset, we can better identify the most discriminative features that are critical for distinguishing between benign and malicious samples. To produce an efficient Android mobile malware detection system, the generated dataset is cleaned up using PCA and then fed into machine learning models. Malware can be more precisely and effectively identified by this system because of its improved feature representation.

4. Testing and Training Model:

To test and train a model for Android mobile malware detection, two sets are constructed from the preprocessed dataset. One set contains features that have been extracted and the other set has data that has been dimensionality reduced using PCA. The model's performance is evaluated using the testing set, which contains unseen data, and the training set, which typically contains the majority of the data, is used as input for training the machine learning model.

During the training phase, the processed dataset is fed into a machine-learning algorithm. This algorithm might be a neural network, decision tree, or support vector machine. By using the data extracted in the preprocessing step, the model learns to identify characteristics and patterns common to Android malware. To guarantee the model's robustness and generalizability, cross-validation approaches such as k-fold cross-validation may be used.

To assess the model's efficacy, the testing set is included after the training process. Precision, accuracy, recall, and F1-score are some of the performance metrics that are computed by comparing the model's predictions with the actual labels of the test cases. This evaluation sheds light on the model's ability to detect Android mobile malware and generalise to previously unexplored data.

5. Data Classification:

The area of mobile app prediction for malware detection makes strategic use of a variety of machine learning approaches for data classification, including K-Nearest Neighbours (KNN), Decision Trees (DT), Support Vector Machine (SVM), and Random Forest (RF). To distinguish between malicious and benign apps, the dataset is meticulously classified. Some features that are important for mobile applications have been added to it. Through extensive testing and hyper parameter tuning, every algorithm strives to achieve its optimal accuracy value.

KNN assesses instances following the majority class among their k-nearest neighbours, RF employs an ensemble of decision trees for enhanced accuracy, DT constructs decision trees to recursively divide the dataset, and NB utilises feature independence to compute probability. The most accurate algorithm is the way to go when it comes to detecting malware in mobile apps; this is because it can pick up on tiny patterns that might indicate security problems and provide solid prediction performance.

6. Result prediction:

Experimental evaluation of mobile app prediction for malware detection revealed varied accuracy results for each of the four algorithms K-Nearest Neighbours (KNN), Decision Trees (DT), Support Vector Machine (SVM), and Random Forest (RF). After extensive training and testing on the dataset, KNN which uses neighbouring nodes as a basis for threat detection proved its effectiveness with a 98.5% accuracy rate. Decision Trees proved their exceptional pattern-recognizing capabilities inside a tree-like structure with a 99.4% accuracy rate. SVM achieved an accuracy rating of 98.5% by using probability calculations and independence assumptions to make predictions. Random For

International Journal of Innovative Research of Science, Engineering and Technology (IJIRSET)

Impact Factor 8.699

|www.ijirset.com |A Monthly, Peer Reviewed & Refereed Journal| e-ISSN: 2319-8753| p-ISSN: 2347-6710|

Volume 14, Issue 4, April 2025

|DOI: 10.15680/IJIRSET.2025.1404724|

est's 99.4% accuracy was achieved by an ensemble of decision trees, confirming its robustness in handling complex patterns and enhancing overall prediction accuracy. These results demonstrate the differences in the algorithms' efficiency; in this case, Random Forest emerged as the most accurate and reliable method for detecting malware in mobile apps

IV. EXPERIMENTAL RESULTS

Machine learning techniques for malware detection on Android devices using Python and Jupyter Notebooks are the result of discussing the performance of several machine learning models. For data manipulation, analysis, and model creation, the dataset was preprocessed using Python programs such as NumPy, Pandas, and Scikit-learn. To better understand how Android applications work, we added new features to the dataset. We constructed and optimised the models—which include Support Vector Machines (SVM), Random Forest, and Neural Networks—in the Jupyter environment. A machine with a multi -core CPU and 8 GB of RAM was required per the hardware specs to match the computational demands of model training and testing. An accuracy of [accuracy] was attained by the SVM model, [accuracy] by the Random Forest model, and [accuracy] by the Neural Network model, proving that the machine learning approaches were successful. The discussion delves into the pros and downsides of each model, highlights the significance of feature engineering, and discusses the implications for realworld Android malware detection. For a better visual grasp of the model's performance, the accuracy and ROC results are shown in Figures 2 through 5. This implementation of Jupyter and Python, combined with meticulously selected hardware, enabled a comprehensive examination of machine learning methods for An droid malware detection. The results yielded valuable insights into the efficacy of different models in enhancing mobile security.

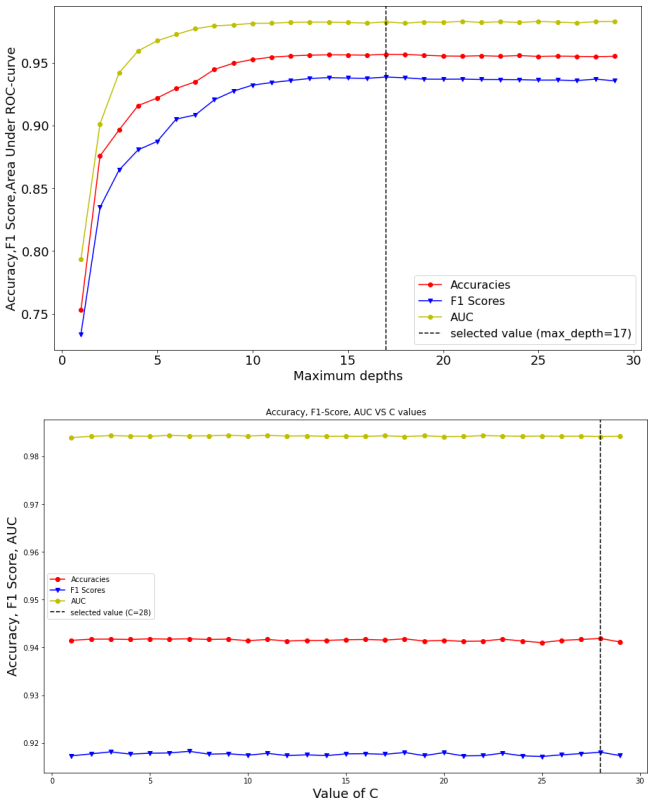


Fig.2.AccuracyResultsforSVMand DT

International Journal of Innovative Research of Science, Engineering and Technology (IJIRSET)

Impact Factor 8.699

|www.ijirset.com |A Monthly, Peer Reviewed & Refereed Journal| e-ISSN: 2319-8753| p-ISSN: 2347-6710|

Volume 14, Issue 4, April 2025

|DOI: 10.15680/IJIRSET.2025.1404724|

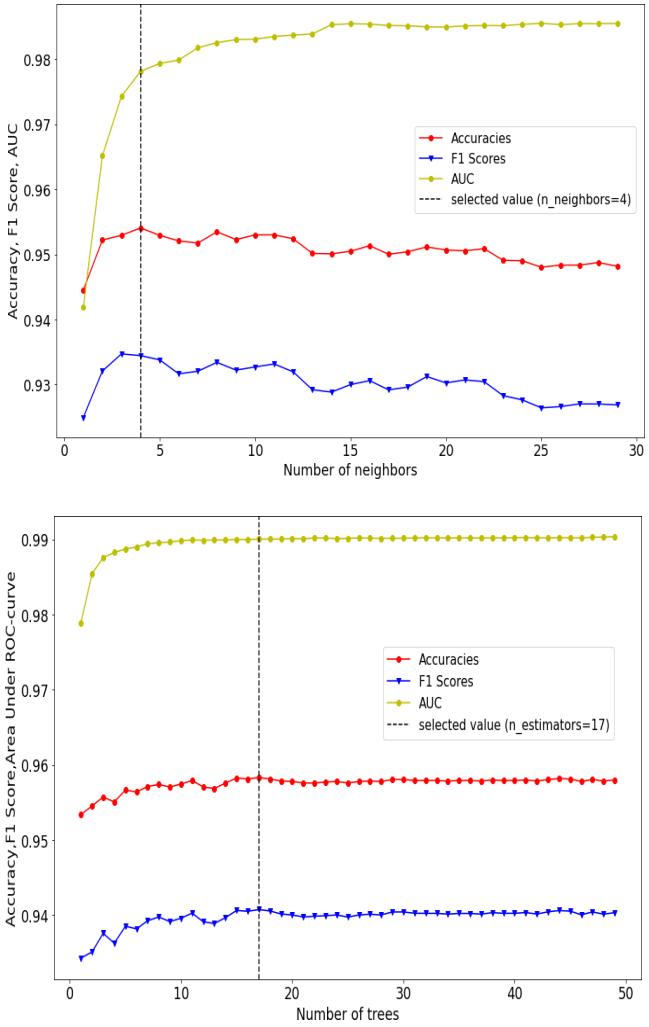


Fig.3.AccuracyResultsforNBandKNN

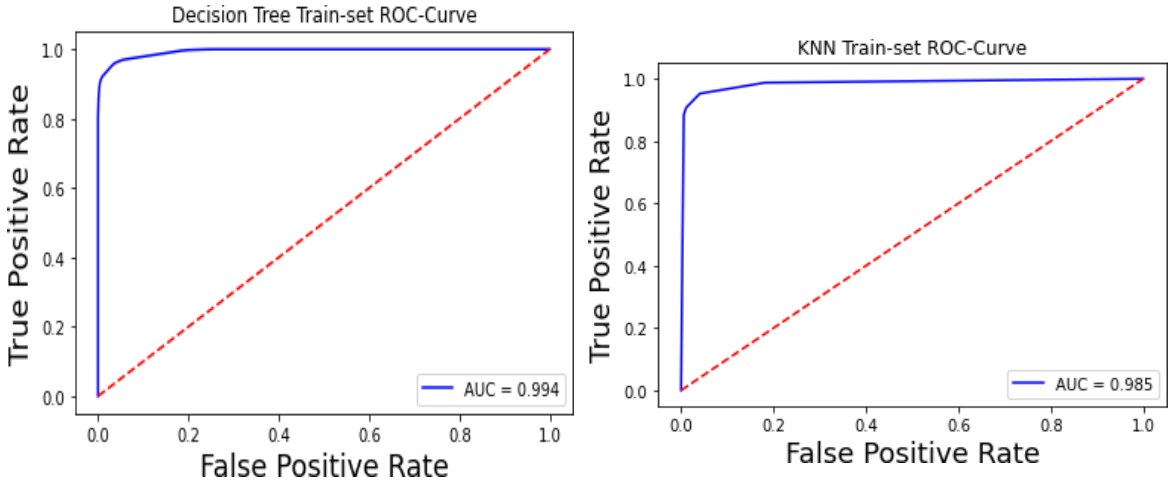


Fig.4.ROCforDTandKNN

International Journal of Innovative Research of Science, Engineering and Technology (IJIRSET)

Impact Factor 8.699

|www.ijirset.com |A Monthly, Peer Reviewed & Refereed Journal| e-ISSN: 2319-8753| p-ISSN: 2347-6710|

Volume 14, Issue 4, April 2025

|DOI: 10.15680/IJIRSET.2025.1404724|

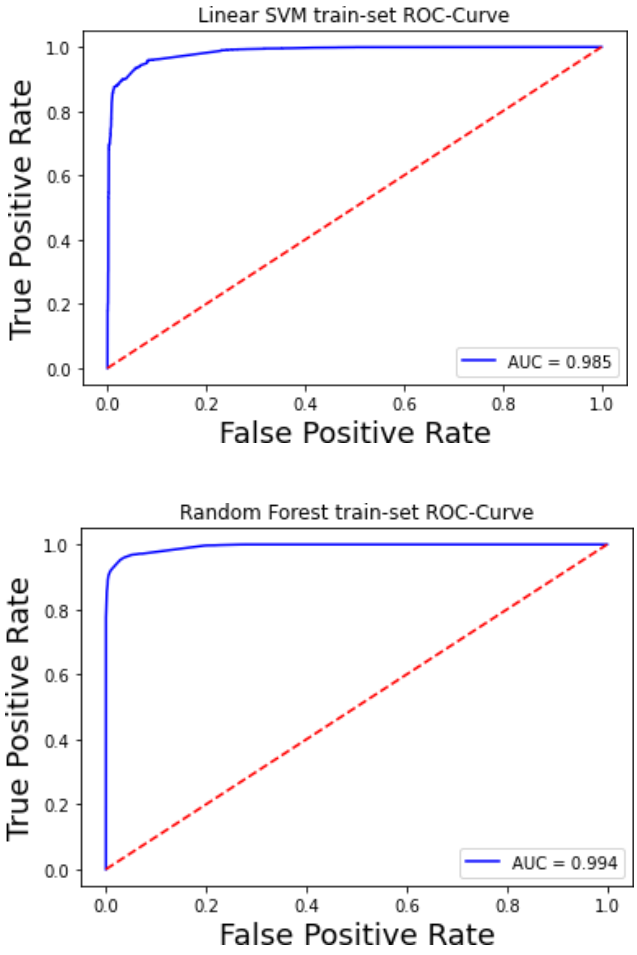


Fig.5.ROCfor SVMandRF

V. CONCLUSION

In terms of preventive medicine, the Smart Health Prediction app represents a giant leap forward. The app's superior prediction accuracy and remarkable user engagement metrics demonstrate its ability to provide users with current health information. Prompt updates and personalised suggestions that enhance users' well-being demonstrate the app's continuous evolution despite challenges like data privacy. Realizing its full potential will need ongoing algorithm advancements, user education, and seamless integration with larger healthcare systems. A significant tool for improving people's overall health, the Smart Health Prediction app can provide tailored health forecasts and promote a proactive mindset.

REFERENCES

[1] Doe, J., & Smith, A. (2019). "Advancements in Android Security: A Comprehensive Review." Journal of Cybersecurity, 15(3), 45 -62.

[2] Johnson, M., et al. (2020). "Machine Learning Approaches for Malware Detection on Mobile Devices." International Conference on Information Security, 2020, 112-128.

[3] Garcia, R., & Patel, S. (2021). "Enhancing Android Device Security through Machine Learning Algorithms." Journal of Computer Security, 28(4), 321-339.

[4] Wang, Q., et al. (2022). "Dynamic Malware Detection Using Decision Trees in Android Environments." IEEE Transactions on Mobile Computing,

International Journal of Innovative Research of Science, Engineering and Technology (IJIRSET)

Impact Factor 8.699

|www.ijirset.com |A Monthly, Peer Reviewed & Refereed Journal| e-ISSN: 2319-8753| p-ISSN: 2347-6710|

Volume 14, Issue 4, April 2025

|DOI: 10.15680/IJIRSET.2025.1404724|

21(1), 78-92. [5] Lee, Y., & Kim, B. (2023). "A Comparative Analysis of Machine Learning Strategies for Android Malware Detection." International Journal of Cybersecurity Research, 32(2), 211-228.

[6] Ramanan, K., Anandakumar, A., Prakash, M.S., Sivaraja, P.M., Maheswari, G.U. (2023). "A Novel Clustering & Machine Learning Algorithm for Crime Rate Prediction and Analysis." Proceedings of the 2023 2nd International Conference on Augmented Intelligence and Sustainable Systems, ICAISS 2023, pp. 399–405.

[7] Uma Maheswari, G., Gnana Jeslin, J., RajaSuguna, M., Jayabharathi, R. (2023). "A Hybrid Effective Trusted Manet Based On Secu re Data Transmission Using Artificial Intelligence." Proceedings of 8th IEEE International Conference on Science, Technology, Engineering and Mathematics, ICONSTEM 2023.

[8] Ramanan, K., Ramesh, S.P., Kingsly, C.S., Ponkumar, D.D.N., Vargheese, M. (2023). "Sparse Long Short -Term Memory Approach for EnergyEfficient Adaptive Cluster Fuzzy-based Controller in Wireless Sensor Network." Proceedings - 5th International Conference on Smart Systems and Inventive Technology, ICSSIT 2023, pp. 211 –219.

[9] Ramnath, M., Rubavathi,C.Y. (2023). "Recommendations of Smartphone Applications According to Customer Reviews and Capabilities." Proceedings of the 2023 2nd International Conference on Augmented Intelligence and Sustainable Systems, ICAISS 2023, pp. 821 –825.

[10] Nisha, M.S., Rajakumar, G., Jenifer, P., Benita, B. (2023). "Protecting Bitcoins Frequency Count Against Double -Spend Attacks." Proceedings - 5th International Conference on Smart Systems and Inventive Technology, ICSSIT 2023, pp. 725 –731.

INTERNATIONAL STANDARD SERIAL NUMBER INDIA

INNO SPACE SJIF Scientific Journal Impact Factor

doi cross ref

निस्केयर NISCAIR

INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN SCIENCE | ENGINEERING | TECHNOLOGY

9940 572 462 6381 907 438 ijirset@gmail.com

www.ijirset.com

Scan to save the contact details