International Journal of Smart Sensor and Adhoc Network

Volume 3
Issue 4 *Role of Emerging & Intelligent Technologies for the Society.*

Article 9

January 2023

Effective Secure Data Agreement Approach-based cloud storage for a healthcare organization

Ramya Thatikonda
DEPARTMENT OF INFORMATION TECHNOLOGY, UNIVERSITY OF THE CUMBERLANDS, ramya.t0211@gmail.com

Adithya Padthe
Department of Information Technology, University of the Cumberlasnds, adithya.padthe@gmail.com

Srinivas Aditya Vaddadi
Department of Information Technology, University of the Cumberlands, vsad93@gmail.com

Pandu Ranga Rao Arnepalli
Department of Information Technology, University of the Cumberlands, parnepalli92@gmail.com

Recommended Citation

Thatikonda, Ramya; Padthe, Adithya; Vaddadi, Srinivas Aditya; and Arnepalli, Pandu Ranga Rao (2023) "Effective Secure Data Agreement Approach-based cloud storage for a healthcare organization," *International Journal of Smart Sensor and Adhoc Network*: Vol. 3: Iss. 4, Article 9.
DOI: 10.47893/IJSSAN.2023.1232
Available at: https://www.interscience.in/ijssan/vol3/iss4/9

Effective Secure Data Agreement Approach-based cloud storage for a healthcare organization

[1] Ramya Thatikonda, [2] Adithya Padthe, [3]Srinivas Aditya Vaddadi, [4] Pandu Ranga Rao Arnepalli

School of Computer and Information Sciences

University of the Cumberlands, USA

[1]ramya.t0211@gmail.com, [2]Adithya.padthe@gmail.com, [3]Vsad93@gmail.com , [4]parnepalli92@gmail.com

Abstract:

In recent days, there has been a significant development in the field of computers as they need to handle the vast resource using cloud computing and performing various cloud services. The cloud helps to manage the resource dynamically based on the user demand and is transmitted to multiple users in healthcare organizations. Mainly the cloud helps to reduce the performance cost and enhance data scalability & flexibility. The main challenges faced by the existing technologies integrated with the cloud need to be solved in managing the data and the problem of data heterogeneity. As the above challenges, mitigation makes the services more data stable should the healthcare organization identify the malware. Developed countries are utilizing the services through the cloud as it needs more security. In this work, a secure data agreement approach is proposed as it is associated with feature extraction with cloud computing for healthcare to examine and enhance the user parties to make effective decisions. The proposed method classifies into two components. The first component deals with the modified data formulation algorithm, used to identify the relationship among variables, i.e., data correlation, and validate the data using trained data. It helps to achieve data reduction and data scale development. In the second component, Feature selection is used to validate the model using subset selection to determine the model fitness based on the data. It is necessary to have more samples of different Android applications to examine the framework using factors like data correctness and the F-measure. As feature selection is a concern, this study focuses on Chi-square, gain ratio, information gain, logistic regression analysis, OneR, and PCA.

Keywords: Cloud Computing, Cybersecurity, Decision making, Data correlation, Factor analysis, and Resource management.

1. INTRODUCTION

Due to the rise in smart phone usage, the wireless network has recently expanded and integrated into daily life. Because of people's physical and mental reliance on mobile devices, multiple programs are necessary to perform various functions such as banking, chatting, social networking, etc. To use the specified mobile applications, a strong internet connection and a lot of mobile data storage—where the cloud plays an important role—are required. Malware detection serves a crucial role and continuously gathers crucial data because there is a reason for downloading different programs [1]. According to statistical analysis, a small number of new threats will be introduced by mobile applications. The mobile user is thus put through a great deal of hardship, and new intruder risks are a major worry.

Threat detection and identification are critical for mobile apps and may be addressed through feature selection. Before selecting features, the datasets can be trained using machine learning techniques. Because there are two primary classes in feature selection, the feature and subset selection approaches are prioritized. When it comes to feature selection, below properties play critical role in defining the model outputs and performances.

- **Gain Ratio:** It may be decided based on the important facts of the decision learning tree. It reduces the tendency based on multiple-valued characteristics and is based on the selection of the attributes utilizing branch range and counts [2].
- **Chi-Squared Test:** Categorical data are considered as the input variables when handling the classification problem. The statistical test process is then used to construct the output variables. The exam has a conditional and non-conditional strategy depending on the input quality.
- **Information Ratio:** Assessing financial data in light of the adjustment risk connected to certain limits is critical. To calculate the surplus data that will be acquired and to offer data identification in relation to the issue, establish data consistency [3].
- **One R:**OneR is a data categorization approach that employs the one rule technique to predict data. One rule performs prediction based on rule formation when calculating the overall error rate. Depending on the data aim, the table frequency is considered for projection development [4].
- **Principal Component Analysis (PCA):** A particular number of variable correlations are transformed into variable in correlations using a simple statistical approach known as the transform method. The main purposes of it are to investigate different prediction models and data identification.
- **Logistic Regression Analysis:** It refers to logical regression, with the inputs including the data weights and their value coefficients. The aforementioned input values can forecast the data to generate the prediction value output, which can be made using binary values of 0s and 1s [5]. For the data subset feature,
- **Feature selection correlation:** The statistical test must be assessed using different machine learning algorithms for this feature selection technique to be faster to execute [6]. According to the machine learning algorithm, this technique helps to improve data performance by removing unwanted and unnecessary data.
- **Rough Set Analysis:** Rough datasets are employed in this study based on data uncertainty. The research contributes to the calculation of object attributes and the building of upper and lower data sets. Because of the magnitude and complexity of real-world applications, data fluctuates. Even yet, finishing the data analysis and managing the implementation [7] can keep the data consistent and reduce its size.
- **Consistency subset evaluation approach:** It includes feature selection, allowing a more efficient technique with reduced dimensionality. Data categorization assists in selecting characteristics with the highest degree of accuracy and the least quantity of data [8]. The two strategies employed in this feature selection are candidate selection subset and feature space search.
- **Filtered subset evaluation:** This filtering approach is a statistical tool for investigating the relationship between the input and output data generated at the destination [9]. The score produced from the input data will be used to pick the filter during the filtering procedure.

Using the traditional machine learning approaches, malware may be identified and analyzed. This study proposes a modified cloud-based malware identification technique to locate and assess dangerous software based on the cloud concept. This can function based on the application process interface. The proposed method would typically execute the machine learning algorithm through an iOS or Android-compatible app. In this methodology, machine learning will support supervised, unsupervised, and hybrid methodologies [21]. The recommended approach will achieve high accuracy in performance analysis while identifying malware based on the vast datasets obtained and data from sources [25].

The suggested cloud-based malware identification method can accurately locate and examine the program while finding the malware based on sizable datasets. The suggested work, in this case, involves three steps described below.

Since all malware-based software is hosted there, input is gathered from a variety of sources. The malware scanner is used to find mobile application files and identify them. It is necessary to do the feature extraction related to API requests and application authorization. This extract feature, or feature selection based on machine learning, will assist in carrying out any test described in Section 1 by helping to conduct the test. The machine learning technique will be used to find the relevant characteristic with the aid of the malware detection model. Applications based on reality are scrutinized to confirm the virus Real-based apps are examined based on several characteristics, such as accuracy and F-measure based on the P value and t value, to authenticate the virus. The datasets are the foundation for the deployment, detection rate, and data accessibility.

The thorough literature review and proposed framework from Section 2 are presented next. The section illustrates the issue statement, and Section 4 illustrates the research technique [27]. Section 5 presents the performance analysis, and Section 6 presents the study conclusion and upcoming work.

2. LITERATURE REVIEW

Here, many research breaches are featured together with a discussion of the current malware-based detection approach. The suggested work then resolves the noted violations.

Using packet networking, the malware detection model is connected to cloud computing [11]. Data mining is used to identify packets, which is the input, and minimize packet knowledge, which aids in determining if malware has been found or not [10]. The learning algorithm will learn the input dataset while data mining will evaluate the data extraction. Therefore, the SMMDS-based malware detection model will use machine learning. The program will be detected using a detection method based on malware during installation. This strategy will contain a set of guidelines that must be followed in mobile apps.

It is proposed that the framework extract the feature and run it during the installation of mobile apps. Next, components are taught using machine learning algorithms to categorize data, assisting in malware detection. This strategy limits the number of system resources and data loading time. When an irregularity in a mobile application is found, the detection-based malware detection approach is applied [12]. Certain machine learning algorithms, such as naive Bayes and logistics, compute the data accuracy rate to operate feature information. Specific limits, such as CPU utilization, memory, battery, and training data, are also neglected in this manner.

The malware detection model, which is based on a Gaussian mixture, employs a machine learning-based feature selection concept [13]. It is possible to extract feature sets based on CPU utilization, battery life, data memory, and other factors. The disadvantage of the proposed paradigm is that it does not permit the cloud server, which is a remote server. The mobile phone danger has yet to be represented in mobile behavior when infected with malware. the mechanism governs how long a mobile device's battery lasts under malware assault [14]. Given the constraints, more accurate models necessitate an accurate model to detect the virus [24].

Malware detection and power saving will be more effective and efficient with a cloud-based detection technique [15]. Based on power-saving parameters, the machine learning detection model performs better in this situation than the cloud-based detection model, as shown in Table 1. The details about the existing works on healthcare are outlined below.

Table. 1. Existing Works based on healthcare Organization

S.NO.	TITLE OF THE PAPER	TECHNIQUES USED	MERITS	GAP IDENTIFIED
1.	On lightweight mobile phone application certification [11]	In-depth Android security analysis - a collection of rules to match malware attributes	To mitigate malware based on surety rules. Users will feel more at ease downloading software with less virus targeting.	Keep up the security requirements engineering process to find more guidelines to stop malware.
2.	Evaluation of machine learning classifiers for mobile malware detection [12]	anomaly-based approach with machine learning classifiers enables secure data-sensitive intelligent malware detection	Good selection of suitable network features for inspection of malware detection. To find the best classifier based on true-positive rate (TPR) data.	Utilizing cloud-based machine learning classifiers, real-time mobile malware detection is created.
3.	Android Malware Detection via a Latent Network Behavior Analysis [13]	Using a system for automatically detecting malware, the network spatial properties of Android apps that have been extracted, and independent component analysis (ICA),	Accepting polymorphism to determine the spatial characteristics' domain name resolution behavior. Automatic identification of Android apps that are harmful.	It is for public Android Malware app datasets. Popular apps collected from the Android Open Market. Effectiveness and Identification of Android malware.
4.	A Gaussian Mixture Model for Dynamic Detection	A combination of probabilistic models, such as the Gaussian mixture model, is	The effectiveness of model-based clustering in identifying apps with unusual behavior.	Decentralized data management System

	of Abnormal Behavior in Smartphone Applications [14]	used in smartphone applications to identify anomalous behavior dynamically.	a Gaussian mixture model for estimating behavior model applications	
5.	Power-based malicious code detection techniques for smartphones [15]	Detecting bad code behavior using two smartphone-specific techniques based on unique power consumption patterns depending on time and location.	Recognizing the existence of rootkits and malware. It is reducing battery usage dramatically. Increase scanning speed noticeably and security measures.	Accuracy of power consumption-based detection techniques Concerning the power signature.
6.	Power-aware anomaly detection in smartphones: An analysis of on-platform versus externalized operation [16]	On-platform tactics outperform machine learning-based detection methods in terms of power usage.	Trade-offs for power usage when using anomaly-detecting components. scenarios where one device and the cloud are involved.	Enable distributing computational tasks to improve the complete battery life of devices.

Its limitation is that actual time applications must be considered based on power-saving requirements.

When information is transferred across mobile devices, information security is considered with deep learning utilizing an ML approach based on Knowledge Decision Database (KDD) [16]. Security knowledge must be fostered to mitigate the risks associated with information transmission [17, 27].

An ICFS approach to detect the malware associated with feature selection and classifier based on machine learning [18].The naive Bayes approach was proposed based on machine learning and constructed the data pattern to identify the malware application and unlabelled data [19, 20]. The mechanism for performing data breaching is user privacy to obtain data privacy [21, 22, 26]. It can achieve resource protection and data related to mobile applications.

Deploy the malware detection-based model by considering the vast datasets based on a machine learning approach. F-measure and data accuracy based on various machine learning classifications are the parameters for model analysis.

3. PROBLEM DEFINITION

The IoT environment contributes to a significant upgrade in wireless signals, which advances mobile technology. As the usage of mobile devices increases, this is important. Threat developers are also actively propagating malware daily to damage user privacy and trust, which is eventually necessary for mobile users. While operating the mobile device, there are certain considerations to keep in mind, such as CPU deployment, phone battery, data memory, data correctness based on F-measure, and other machine learning-related tests.

4. RESEARCH METHODOLOGY

The recommended cloud-based malware classification system can detect and scrutinize the application to achieve high precision while distinguishing the malware created on vast datasets. Due to the fact that all malware-based apps are kept there, the reaction is gathered from a variety of sources. The malware electronic scanner finds mobile application papers and makes them visible. API requests will be phased along with feature isolation and application agreement. According to Figures 1 and 2, this extract feature, or feature selection developed using machine learning, will get better at passing any test.

The systematic approach to solving an issue through data collection, analysis using multiple methodologies, and generating conclusions from the data acquired is known as the methodology in research. Data are quantified in the quantitative investigation, and the conclusions from the intended population are generalized. This study employs statistical analysis and objective analysis.

A thorough researcher description and in-depth observation are prerequisites for qualitative research. The context and interpretation of the data obtained using this approach are provided by qualitative research.To establish trust, accountability, reciprocal trust, and fairness in the study, researchers must conform to standard norms and circumstances. Any study project must consider ethical factors, and researchers must be cautious of data privacy, security, and integrity. The approaches used in this study take ethical concerns and select the best research methodology into account.

Research for any form of scientific inquiry should not just be "methodologically guided," but the chosen technique should also be relevant to the researcher's philosophical viewpoint and the social science phenomena being studied. This study methodology is guided by all the above principles and rules for better results and performance under ideal conditions.

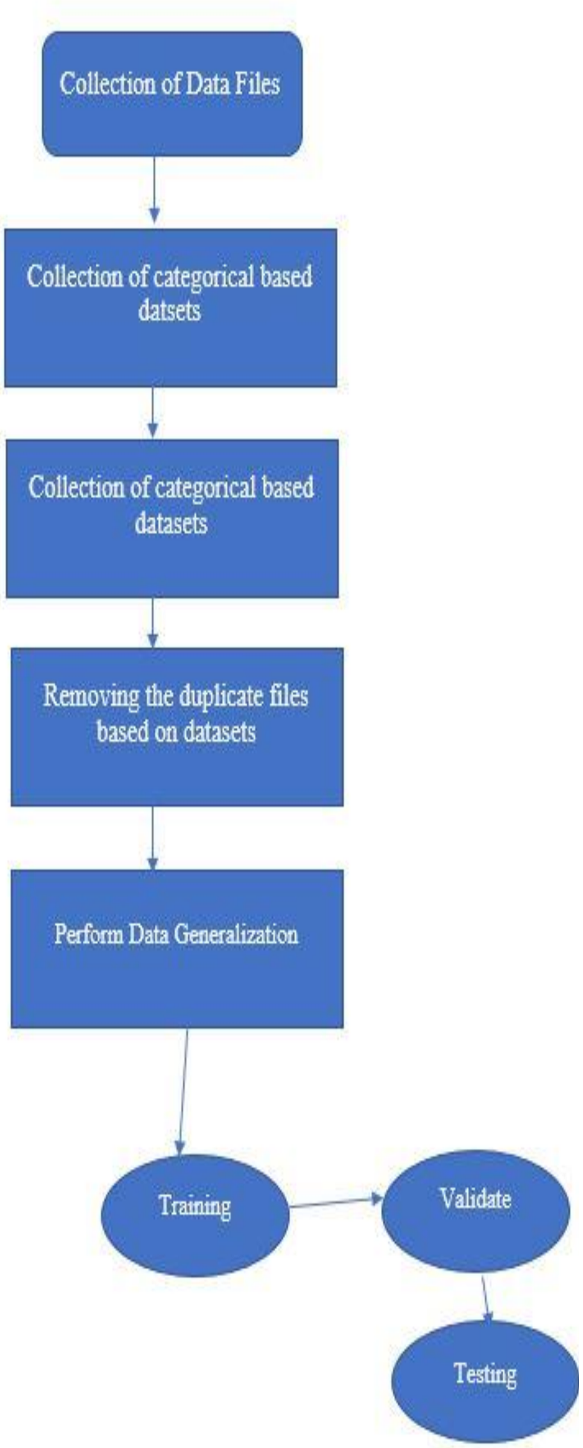
Collection of Data Files
Collection of categorical based datsets
Collection of categorical based datasets
Removing the duplicate files based on datasets
Perform Data Generalization
Training
Validate
Testing

Figure1. Feature Extraction Process

Algorithm 1: Dataset formulation based on machine Learning

```
// Collection of Data Files
Input: Input files
Output: Extracted feature selection
      of categorical-based datasets (.apk Files)
      Deselecting duplicate files or datasets
      based on inputs.
      Consider all the files with ordinary, trojan,
      backdoors,worms, botnet, and spyware
      classes.
      Perform Data Specifications using feature
      selection.
          Training datasets.
          Verify the dataset
          Testing the data source
          Final Outcome and reporting

// Extracting the Feature from the dataset
     o   Gathering the distinctive data samples
     o   Extract the API Calls and file
         permissions.
     o   to revoke the resource's authorization
         to use.
     o   Execute the data you've gathered, then
         extract the file.
    If (API Calls)
    {     1
          Else
    0     }
    Extracted the feature data.
```

The malware recognition simulation facilitates distinguishing the notable feature, which will be presented based on a machine learning algorithm [23]. To validate the malware, real-time applications are examined based on numerous factors such as precision and F-measure created on the P value and t value. The implementation, recognition rate, and data accessibility are centered on the datasets.

Algorithm 2: Feature Selection

Input: Feature Extraction
Output: Malware Discovery Final Prototype

```
    Tolocate appropriate datasets using feature extraction.
    Categorical datasets are clarified as classified and unclassified errors.
    // Feature Ranking
    Training Datasets
    Used to apply filtering technique to perform ranking:
        • Mutual Information
        • Relief -F
    The combination of two filtering methods and the ranking features needs to be used as below.

// Feature Subset Selection
    Boost the Feature selection process
    Based on the training of the dataset, run the below if-else algorithms.
    If (Identifying the feature indices)
    {
        Select the feature set based on the training set
    Else
        Classify the data based on a machine-learning algorithm
    }
    To anticipate and confirm whether malware has been detected or not.
```

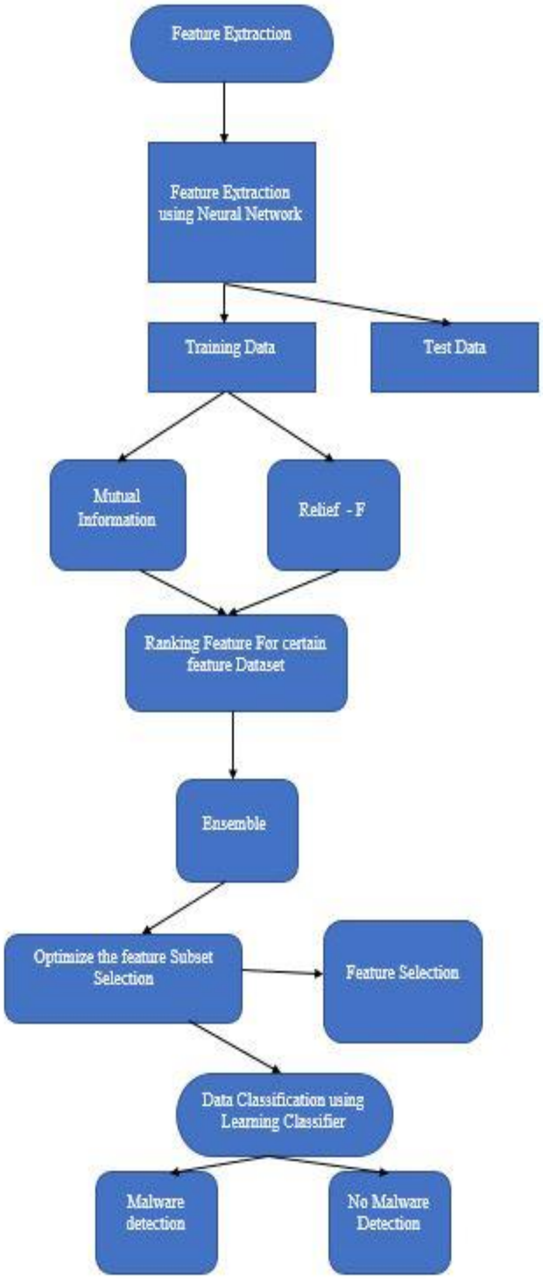
Feature Extraction
Feature Extraction using Neural Network
Training Data
Test Data
Mutual Information
Relief - F
Ranking Feature For certain feature Dataset
Ensemble
Optimize the feature Subset Selection
Feature Selection
Data Classification using Learning Classifier
Malware detection
No Malware Detection

Figure 2. Feature Ranking and Feature Subset Selection

5. PERFORMANCE ANALYSIS

Researchers mainly use a performance analysis approach as a process to evaluate the performance of a system or application. These often offer a place to start before giving advice on the underlying problem or effects. Academicians may need to attempt several different techniques before they succeed since they are each best suited for dealing with specific groups of problems. Tables 1 and 2 depict the data precision and overall implementation time centred on the disparity number of epochs. This study significantly highlights the method's performance for better results and data quality standards. The tables below highlight the outcomes in detail.

Table 1. Data Accuracy based on the number of Epochs

S.No.	Proposed Secure DA	RF	SVM
10	0.985	0.968	0.945
20	0.965	0.952	0.915
30	0.954	0.947	0.895
40	0.925	0.898	0.865
50	0.918	0.875	0.845

Table 2. Total execution time based on the number of Epochs

S.No.	Proposed Secure DA	RF	SVM
10	0.0045	0.0024	0.0018
20	0.0041	0.0023	0.0015
30	0.0038	0.0019	0.0013
40	0.0036	0.0018	0.0011
50	0.0031	0.0014	0.0009

6. CONCLUSION

The results of this study suggest that future research should concentrate on improving the efficiency of search approaches and approximation methods for difficult optimization problems in large-scale building, as well as reducing the time and effort needed for such jobs. Assessing the robustness of optimum techniques for enhancing building performance stability also requires more effort.Thesuggested cloud-based malware detection approach can identify and examine the application to reach high levels of accuracy while identifying the malware constructed on enormous datasets. The highlighted datasets to differentiate the malware program are shown, and the chosen datasets propose the form of a malware recognition approach. The feature selection process makes it easier to use fewer included datasets while still performing better. The virus may then be precisely located using data categorization errors by the featured set. Use a machine learning method after that to demand better exposure and cost calculation. Using cutting-edge blockchain technology and AI can improve cloud storage and privacy.

References

[1]. Singh KU, Gupta PK, Ghrera SP (2015) Performance evaluation of AOMDV routing algorithm with local repair for wireless mesh networks. CSI Trans ICT 2(4):253–260.

[2]. Novakovic J (2010) The impact of feature selection on the accuracy of Naı¨ve Bayes classifier. In: 18th Telecommunicationsforum TELFOR, vol 2, pp 1113–1116.

[3]. Plackett RL (1983) Karl Pearson and the chi-squared test. Int StatRev/Revue Int Stat 51(1):59–72

[4]. Wang W, Wang X, Feng D, Liu J, Han Z, Zhang X (2014) Exploring permission-induced risk in android applications for malicious application detection. IEEE Trans Inf ForensSecur9(11):1869–1882

[5]. Cruz C, Erika A, Ochimizu K (2009) Towards logistic regression models for predicting fault-prone code across software projects. In: Proceedings of the 2009 3rd international symposium on empirical software engineering and measurement. IEEE ComputerSociety, pp 460–463

[6]. Hall MA (1999) Correlation-based feature selection for machine learning (Doctoral dissertation, The University of Waikato, Dept. of Computer Science)13. Pawlak Z (1982) Rough sets. Int J Comput Inf Sci 11(5):341–356.

[7]. Dash M, Liu H (2003) Consistency-based search in featureselection. ArtifIntell 151(1–2):155–176

[8]. Kohavi R, John GH (1997) Wrappers for feature subset selection.ArtifIntell 97(1–2):273–324

[9]. Arp D, Michael S, Malte H, Hugo G, Konrad R, Siemens CERT (2014) Drebin: effective and explainable detection of androidmalware in your pocket. In: NDSS, vol 14, pp 23–26

[10]. Cui B, Jin H, Carullo G, Liu Z (2015) Service-oriented mobile malware detection system based on mining strategies. PervasiveMob Comput 24:101–116

[11]. Enck W, Ongtang M, McDaniel P (2009) On lightweight mobile phone application certification. In: Proceedings of the 16th ACM conference on Computer and communications security. ACM,pp 235–245

[12]. Narudin FA, Ali F, Nor BA, Abdullah G (2016) Evaluation of machine learning classifiers for mobile malware detection. SoftComput 20(1):343–357

[13]. Wei T-E, Mao C-H, Jeng AB, Lee H-M, Wang H-T, Wu D-J (2012) Android malware detection via a latent network behaviour analysis. In: 2012 IEEE 11th international conference on trust, security and privacy in computing and communications. IEEE,pp 1251–1258

[14]. El Attar A, Khatoun R, Lemercier M (2014) A Gaussian mixture model for dynamic detection of abnormal behaviour in smartphone applications. In: 2014 global information infrastructure and networkingsymposium (GIIS). IEEE, pp 1–6

[15]. Dixon B, Mishra S (2013) Power-based malicious code detection techniques for smartphones. In: 2013 12th IEEE international conference on trust, security and privacy in computing andcommunications. IEEE, pp 142–149

[16]. Suarez-Tangil G, Tapiador JE, Peris-Lopez P, Pastrana S (2015) Power-aware anomaly detection in smartphones: an analysis of on-platform versus externalized operation. Pervasive Mob Comput18:137–151

[17]. Dash, B., Ansari, M. F., Sharma, P., & Ali, A. (2022). Threats and Opportunities with AI-based Cyber Security Intrusion Detection: A Review. International Journal of Software Engineering & Applications, 13(5), 13–21. https://doi.org/10.5121/ijsea.2022.13502

[18]. Chen PS, Lin S-C, Sun C-H (2015) Simple and effective method for detecting abnormal internet behaviours of mobile devices. InfSci 321:193–204

[19]. Quan D, Zhai L, Yang F, Wang P (2014) Detection of malicious android apps based on the sensitive behaviours. In: 2014 IEEE 13th international conference on trust, security and privacy incomputing and communications. IEEE, pp 877–883

[20]. Ng DV, Hwang J-IG (2014) Hwang. Android malware detection using the dendritic cell algorithm. In: 2014 International conference on machine learning and cybernetics, vol 1. IEEE,pp 257–262

[21]. Sharma, P., Dash, B., & Ansari, M. F. (2022). Anti-phishing techniques – a review of Cyber Defense Mechanisms. IJARCCE, 11(7). https://doi.org/10.17148/ijarcce.2022.11728

[22]. Tong F, Yan Z (2017) A hybrid approach of mobile malwaredetection in Android. J Parallel DistribComput 103:22–31

[23]. Jakka, G., Yathiraju, N., & Ansari, M. F. (2022). Artificial Intelligence in Terms of Spotting Malware and Delivering Cyber Risk Management. *Journal of*

Positive School Psychology, *6*(3), 6156-6165.

[24]. Ansari, M., Dash, B., Sharma, P., &Yathiraju, N. (2022). The Impact and Limitations of Artificial Intelligence in Cybersecurity: A Literature Review. *International Journal of Advanced Research in Computer and Communication Engineering*, *11*(9), 81–90. https://doi.org/10.17148/IJARCCE.2022.11912

[25]. Md Haris Uddin Sharif, & Mehmood Ali Mohammed. (2022). A literature review of financial losses statistics for Cyber Security and future trend. *World Journal of Advanced Research and Reviews*, *15*(1), 138–156. https://doi.org/10.30574/wjarr.2022.15.1.0573

[26]. Dash, B., Ansari, M. F., Sharma, P., & Swayamsiddha, S. (2022). FUTURE READY BANKING WITH SMART CONTRACTS - CBDC AND IMPACT ON THE INDIAN ECONOMY. International Journal of Network Security and Its Applications, 14(5). https://doi.org/10.5121/ijnsa.2022.14504

[27]. Ansari, M. F. (2021). The relationship between Employee's Risk Scores and the Effectiveness of the AI-Based Security Awareness Training Program. Retrieved February 4, 2022