IJCRT.ORG ISSN : 2320-2882

INTERNATIONAL JOURNAL OF CREATIVE RESEARCH THOUGHTS (IJCRT)

An International Open Access, Peer-reviewed, Refereed Journal

ENHANCED CLOUD SECURITY SOLUTIONS: INTEGRATING HYBRID FEATURES SELECTION AND MACHINE LEARNING CLASSIFICATION FOR ADVANCED INTRUSION DETECTION SYSTEMS

G Gayathri[1], Pavithra A [2]

[1]M.Sc., Department of Computer Science and Engineering, Dr. M.G.R Educational and Research Institute, Chennai, India

[2]Faculty, Centre of Excellence in Digital Forensics, Dr. M.G.R Educational and Research Institute, Chennai, India

ABSTRACT

Recent IT infrastructure now includes cloud computing heavily since it offers scalability and flexibility. However, the increasing reliance on cloud services also attracts malicious activities and cyber threats. To protect cloud environments in this situation, an efficient Intrusion Detection System (IDS) is important. So, In this proposed System a better Cloud Intrusion Detection System design that makes use of a machine learning classifier and a hybrid feature selection method. The suggested system improves both the accuracy and effectiveness of intrusion detection by utilizing label encoding, correlation analysis, and the Extra Tree algorithm. The suggested system has been verified by the UNSW-NB15 dataset Gather a diverse and representative dataset of cloud security incidents and non-incidents. Understand the characteristics of the dataset, including feature types (categorical, numerical) and potential challenges Exploratory data analysis (EDA) is used by data scientists to analyze and investigate data sets and list their primary attributes, frequently using data visualization. The UNSW-NB15 datasets have validated the proposed method, yielding accuracies of more than 98% and 99% in the multi-class classification scenario, respectively. It was found that an intrusion detection system would function better if it had less informative features.

Keywords

Intrusion Detection System, Machine Learning, Extra Tree Algorithm, UNSW-NB15 Dataset, Exploratory Data Analysis (EDA)

I. INTRODUCTION

Nowadays, the progress in digital technologies has led to an explosive growth of Cloud computing (CC) applications in different fields due to its services (SaaS, PaaS, and IaaS) and advantages such as expandability, availability, low cost, and so on. On the other hand, this has increased dangers and generated a huge market for cybersecurity. According to this research, companies and organizations faced 50 million cyber assaults in 2010, and by 2019, that figure had increased to 900 million, and the figure is still continuously rising.

Both individuals and enterprises have suffered serious damage and big financial losses as a result of these cyberattacks. Based on recent Juniper research, the expense of security breaches is forecast to increase from USD three trillion annually to over USD five trillion in 2024[1]. These immense economic losses made users apprehensive about storing their data in the cloud; thus, the primary goal of the cloud service provider (CSP) is to allay their fears by investing in cybersecurity solutions, you can maintain their data and offer the highest level of protection. In 2022, the worldwide cybersecurity industry was estimated to be worth USD 202.72 billion. From 2023 to 2030, it is anticipated to increase at a CAGR of 12.3%.

The cloud is composed of three primary network types: virtual, internal, and external. Communication among virtual machines (VMs) running on the same physical server/infrastructure is allowed through the virtual network [2]. Various cloud components, such as network servers, management systems, and storage systems, can connect over the internal network. The external network serves as the main communication channel between the cloud user (front end) and the CSP (back end). When taken as a whole, these networks make it possible for clients to successfully get cloud services.

Hence it is necessary to protect the network against any potential threats. To solve various security challenges, the cloud uses a range of cybersecurity techniques, including firewalls, intrusion prevention systems (IPS), intrusion detection systems (IDS), etc. Recently, network threats have worsened due to a lack of adequate counter-security actions. Consequently, IDS is implemented in the cloud model to combat security concerns. among the biggest obstacles to the achievement of cloud computing is guaranteeing data safety and confidentiality, as these offerings are transmitted through the internet network.

The primary obstacle to cloud security is identifying and stopping network intrusions Given that the network serves as the cloud's backbone, any network vulnerabilities have an immediate impact on the overall security of the cloud. Conventional solutions such as firewalls and even traditional signature-based intrusion detection techniques are no longer effective in confronting intruders, as their non-deterministic nature makes them unsuitable for cloud environments. Therefore, it is crucial to develop anomaly-based IDSs using ML models with high accuracy, an elevated DR, and a minimal FAR before implementing this IDS on each server in cloud computing to monitor network traffic for detecting attacks.

II. REVIEW OF LITERATURE

Halit Bakir and Ozlem Ceviz 12 April [3] Empirical Enhancement of Intrusion Detection System: The hybrid characteristic choice method with an evolutionary algorithm-based hyperparameter tuning mechanism is presented in this paper as a comprehensive approach to enhancing intrusion detection systems (IDSs). An intrusion detection system (IDS) that can comprehend attacks in the Internet of Things environments is proposed by the study. It is based on device learning The hybrid characteristic choice approach is centered on pinpointing the essential elements of the task and applying genetic algorithms to optimize the hyperparameter. As the study is empirical, it can be subjected to rigorous testing and quantitative evaluation of the techniques' efficacy. The approach's viability is confirmed by the results, which show a gradual improvement in IDS performance, particularly in detection latency.

D.jayalatchumy, Rajakumar Ramalingam., et al., 01 March 2024[4] Improved crow search-based Feature Selection And Ensemble Learning for IoT Intrusion Detection. In this paper, they present that Network intrusion detection within the Internet of Things (IoT) framework has posed enormous demanding situations in current decades. A huge sort of machine-getting-to-know tactics are brought in community intrusion detection. The current methodologies generally lack consistency in reaching the finest overall performance throughout diverse elegance categorization tasks. The gift examination elucidates imposing a unique intrusion gadget with the number one goal of enriching the efficacy of community intrusion detection. In the preliminary phase, it's far vital to appoint information-denoising methodologies to efficiently address the difficulty of information imbalance. In the following step, the improved Crow seek set of rules is used to decide the maximum good-sized capabilities that are useful resources in classifying intrusion attacks. In the very last phase, the ensemble classifier takes the chosen capabilities as enter to categorize the same old and invader labels. The gift paintings introduce an ensemble mechanism that incorporates 4 wonderful classifiers. The evaluation of the proposed technique is demonstrated on denoised datasets, NSL-KDD and UNSW-NB15 in particular. According to the experimental results, the developed method achieves remarkable accuracy for the NSL-KDD and UNSW-NB15 datasets, respectively, of 99.4% and 99.2%.

Mubasher Malik, Hamid Ghous ., et al., 01 March [5] Intelligent Intrusion Detection System For Internet Of Things Using Machine Learning Techniques. The Internet of Effects (IoT) and the way it can automate numerous kinds of tasks without any need for mortal involvement are bandied in the paper. It discusses IoT integration of artificial intelligence and intrusion discovery systems. The study evaluates several types of classifiers, including the LGBM, and chooses the way they are doing to achieve high delicacy rates.

Mohamad Bakro, Rakesh Ranjan Kumar., et al., 26 June [6] An Improved Design For A Cloud Intrusion Detection System Using Hybrid Features Selection Approach With ML Classifier. In this paper, they present an improved cloud IDS designed by incorporating the synthetic minority over-sampling technique (SMOTE) to address the imbalanced data issue. For feature selection, they proposed to use a hybrid approach that includes three techniques: information gain (IG), chi-square (CS), and particle swarm optimization (PSO). Finally, the (random forest) RF model is utilized for detecting and classifying various types of attacks. The suggested system has been verified by the UNSW-NB15 and Kyoto datasets, achieving accuracies of over 98% and 99% in the multi-class classification scenario, respectively. It was noticed that an intrusion detection system with fewer informative features would operate more effectively.

Noah Oghenefego Ogwara and Krassie Petrova and Mee Loong Yang 27 Feb[7] Towards The Development Of A Cloud Computing Intrusion Detection Framework Using An Ensemble Hybrid Feature Selection Approach. The study implies a framework for enhancing hybrid ensemble feature selection-based intrusion detection systems in cloud computing environments, resulting in high classification accuracy in binary and multiclassification detection engines.

Achmad Akbar Megantara and Tohari Ahmad 02 November [8] A hybrid machine learning method for increasing the performance of network intrusion detection system. In this paper, they proposed that the internet has reconstituted many different industries, including education, healthcare, and some financial technology. However, this also poses a risk of cyberattack. This study proposes a hybrid machine learning method by combining feature selection method, supervised learning representation, and data reduction such as unsupervised learning to build a suitable model. It works by selecting relevant and significant features using a feature importance decision tree-based method with recursive feature elimination and detecting anomaly/outlier data using the local outlier factor (LOF) method. The experimental results show that the proposed method achieves the highest accuracy in detecting r2l (i.e., 99.89%) and keeps higher for other attack types than most other research in the NSL-KDD dataset. Therefore, it has a more stable performance than the others. More challenges are experienced in the unsw-nb15 dataset with binary classes.

Ammar Aldallai and Faisal Alisa 3 December [9] Effective intrusion detection system to secure data in the cloud using machine learning. In this article, a hybrid intrusion detection system is proposed. It combines support vector machine (SVM) and genetic algorithm (GA) methods with an improved adaptive function developed to evaluate the accuracy of the system. Both algorithms, GA and SVM, were executed in parallel to achieve two optimal objectives simultaneously: obtaining the best subset of features with maximum accuracy. In this scenario, an SVM was employed using different values of hyperparameters of the kernel function, gamma, and degree. The results were compared with KDD CUP 99 and NSL-KDD. The results show that the proposed model significantly outperforms these benchmarks by 5.74%. This system will be effective in cloud computing as it is expected to provide a high level of symmetry between information security and detection of malicious

III. RESEARCH METHODOLOGY

Before deploying an IDS (Intrusion Detection System) in the cloud, process some of the datasets used to train it. So some of the datasets are quite large and encompass a wide range of attacks, along with a significant amount of unrelated information. Therefore, the most relevant features will later be used to train the classifier [10]. This classifier will then differentiate between benign packets and diverse types of attacks. So, this section provides brief theoretical details of the vision. With the help of the suggested system has been verified by the UNSW-NB15 dataset Gather a diverse and representative dataset of cloud security incidents and non-incidents. Understand the characteristics of the dataset, including feature types (categorical, numerical) and potential challenges Exploratory data analysis (EDA) is used by data scientists to analyze and investigate data sets and list their primary attributes, frequently using data visualization. Then proceed to the algorithm which is used with the help of the Extra Tree Algorithm for the Best Accuracy [11] At last, the prediction stage was conducted and several standard performance metrics such as precision, accuracy, and error in classification were considered for the computation of performance efficacy of this model. And the pre-processed data are trained and input that was given by the user goes to the trained dataset.

High-dimensional data and the inherent drawbacks of each standalone feature selection method are the driving forces behind the creation of a hybrid feature selection strategy [12]. Complex patterns in a high-dimensional dataset are frequently present and may not be sufficiently picked up by a single feature selection technique. There are specific benefits and drawbacks to each approach. Enhanced this project's accuracy by adding a few new techniques with the growing popularity of cloud computing, it is imperative to guarantee that security measures are strong. The purpose of this project is to create a cloud security framework that is optimized for effective intrusion detection by fusing machine learning classification and hybrid feature selection techniques [13]. The framework increases threat detection and mitigation by leveraging extra tree algorithms and cloud-based data analytics, thereby improving the overall resilience of cloud infrastructures. As shown in fig1

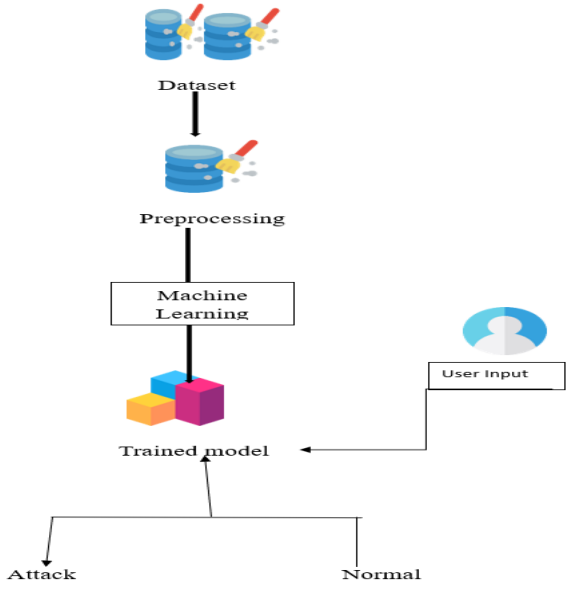

Fig 1 Process of building an ML model.

3.1 Data Collection and Understanding

The UNSW-NB15 dataset has confirmed the recommended system. Compile a varied and inclusive dataset of events and non-events related to cloud security. Recognize the features of the dataset, such as its feature types (numerical, category), and any potential issues. Data scientists often use data visualization together with exploratory data analysis (EDA) to identify and catalog the key features of data sets.

3.2Algorithm Implementation

The Classification Algorithms produce the most optimal outcomes and forecast the Cloud Intrusion Detection System using the Extra Tree Algorithm. The best-performing models are indicated in the above results due to their low error rate.

3.3Prediction

Several standard performance metrics such as accuracy, precision, and error in classification have been considered for the computation of performance efficacy of this model. Pre-processed data are trained and input given by the user goes to the trained dataset.

3.4 Extra Tree Algorithm:

Similar to the random forest algorithm, the extra trees algorithm generates a large number of decision trees, but it does so randomly and without replacement for each tree. This generates a dataset with distinct samples for every tree. For every tree, a certain number of features are also chosen at random from the entire set of features. The random selection of a splitting value for a

feature is the most significant and distinctive feature of extra trees. The algorithm chooses a split value at random rather than employing Gini or entropy to split the data and calculate a locally optimal value. The trees become diverse and uncorrelated as a result. Thus, the datasets are built in Python and using Anaconda. Python distribution Anaconda offers a command-line interface called Command Prompt. This is a dedicated terminal that has the Anaconda environment, which includes Python and several data science libraries, pre-configured. On to the testing process, using the anaconda command prompt to open the created three modules, and the command as shown in Fig 1

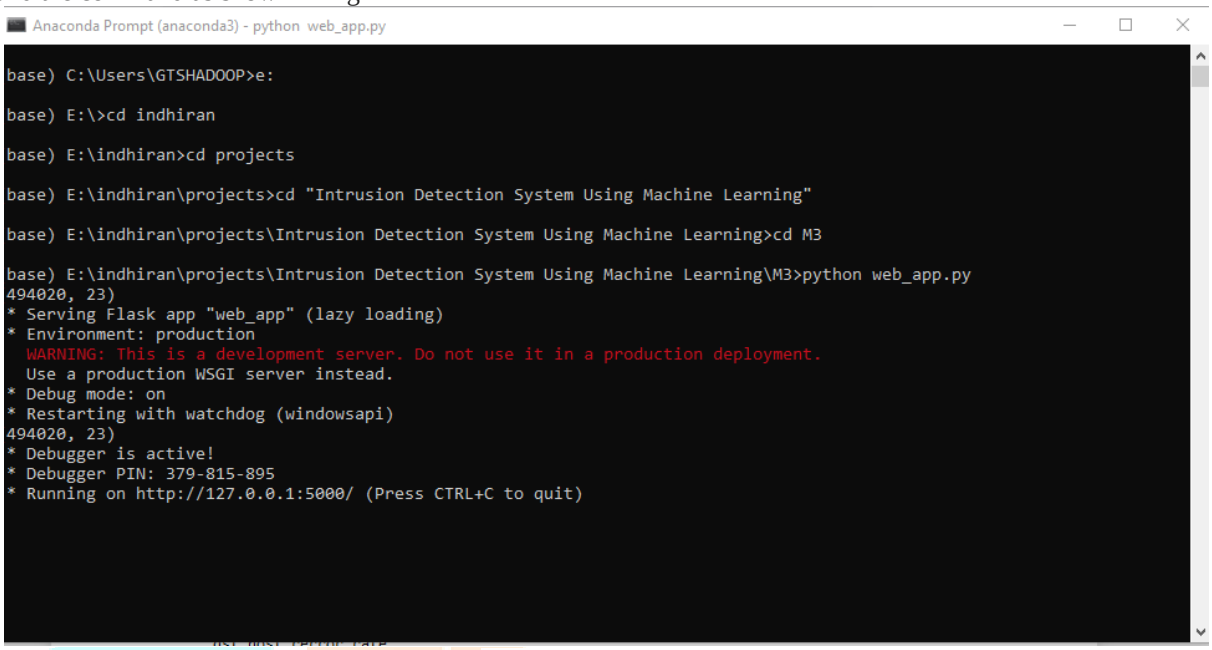

Fig 2 Anaconda prompt

In the below figure shows the row number that needs to be scanned the detection those data are collected from UNSW-NB15 and Kyoto and those data are already trained by the classifications [14]. Before opening this page, the Jupiter notebook should run a code and, in that book, it will generate an IP address. Then copy and paste the link in Chrome it will redirect it to the webpage which is shown below. As shown in the below figure, the details can be manually entered on that web page. This is the home page of the Intrusion Detection System.

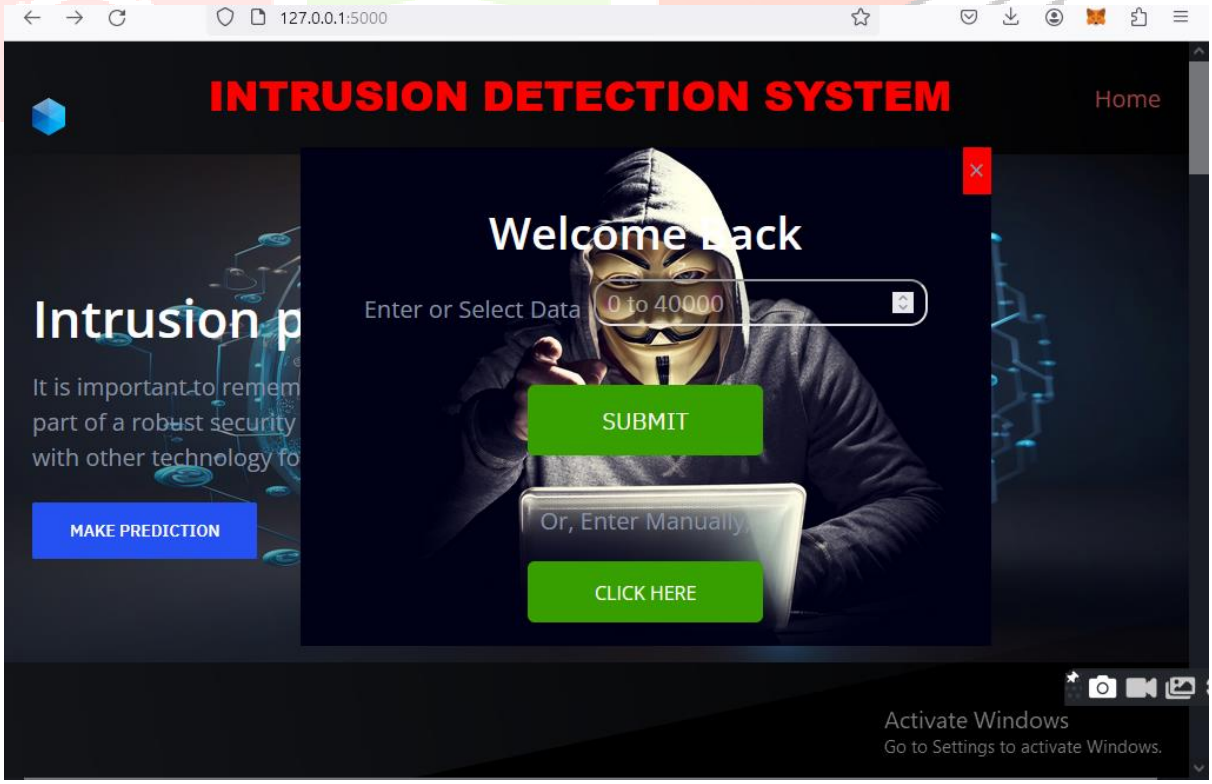

Fig 3 Displaying the home page of the Intrusion Detection system

The figure shows the data of the row that has been given in the above screenshot data it can be given manually or else if the given row number, all the details will come automatically because the detection system is trained like that with the help of the methodology. After giving the details shown in the below figure click the Predict tab and it will scan the data.

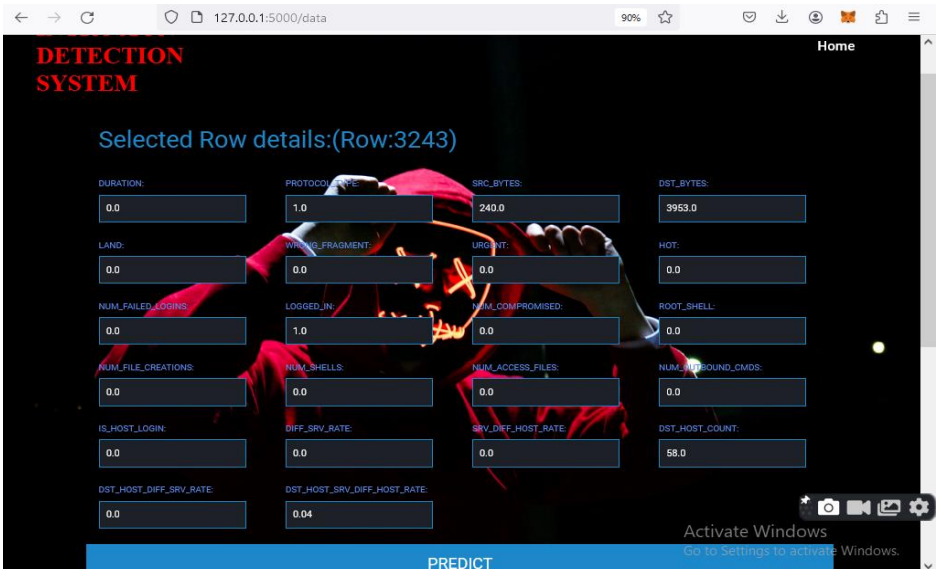

Fig 4 Representation of entering the dataset value

The figure shows the result of the data that have been entered either manually or automatically. In this figure, it is showing that there is no virus in the cloud data. If the data is safe and if there is no one is attacking the data then the result will be shown like this.

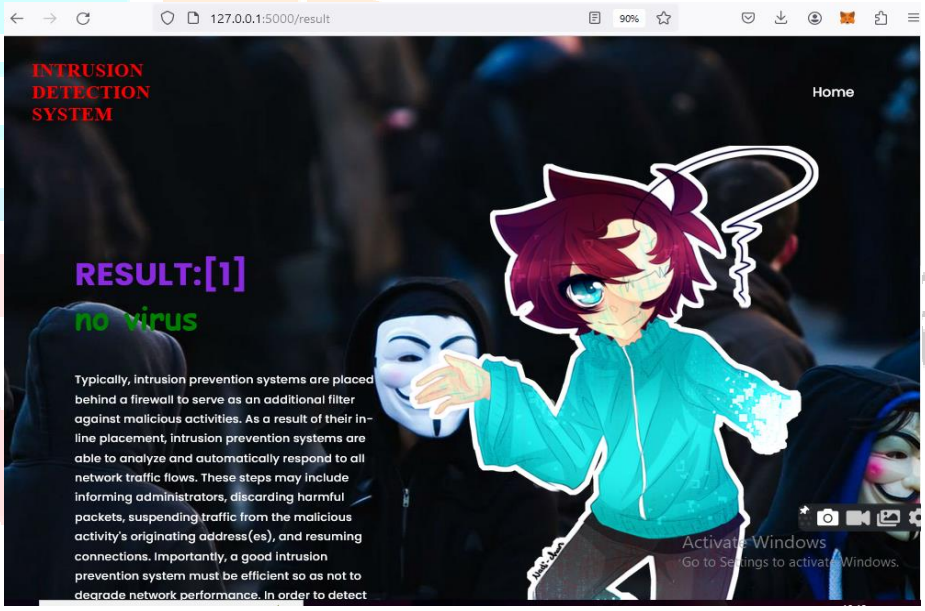

Fig 5 Displaying the result of Intrusion Detection as no virus

The below figure shows that if the data is affected by any attacks like DDOS attack, phishing attack, smishing, DOS attack, session hijacking, etc. if the virus attacks the user data it will give the alert message to the user shown in the below figure and also it will give an explanation for example if session hijacking attacks the computer and it will exfoliate the web session control mechanism, which is normally managed for a session token.

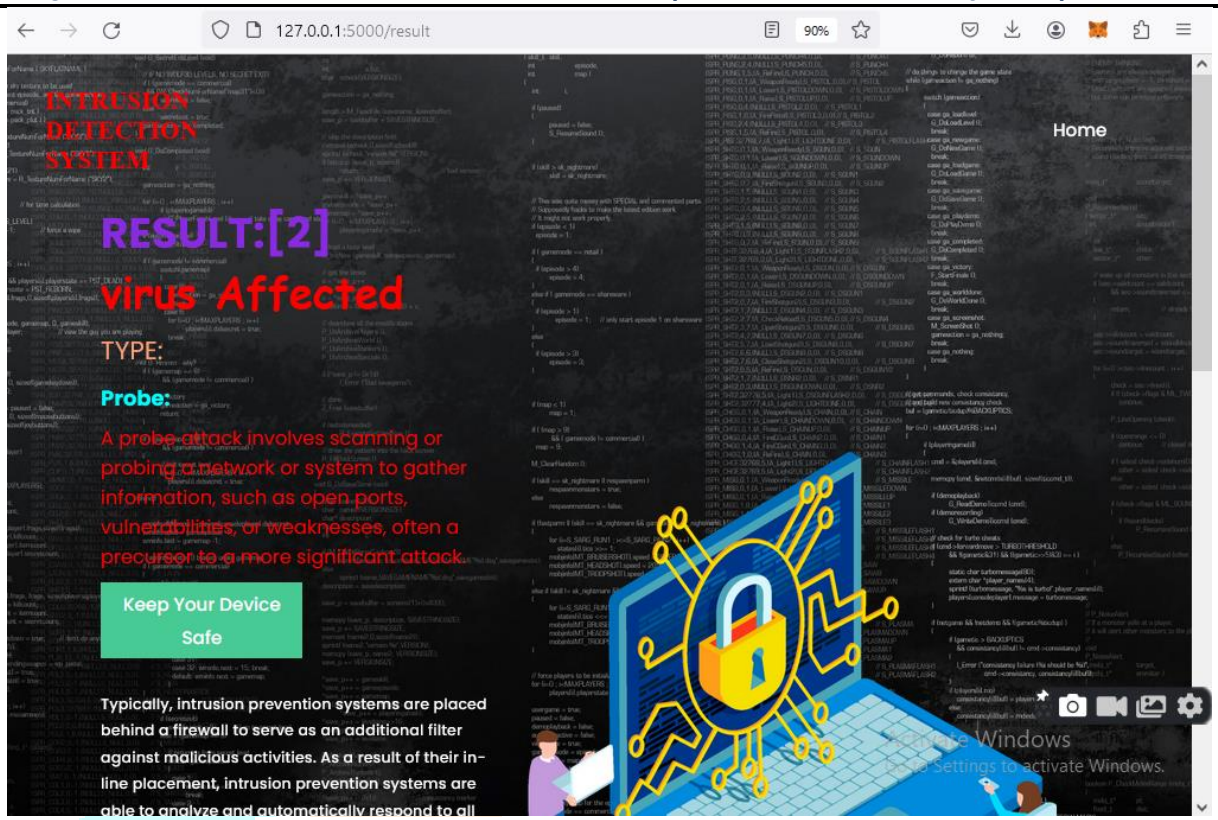

Fig 6 Displaying the result of Intrusion Detection as virus-affected

CONCLUSION

The proposed cloud-based intrusion detection system (IDS) emerges as a robust and effective solution amidst the escalating cyber threats in cloud computing environments. By addressing the critical issues of data confidentiality and integrity, the system employs a well-crafted ensemble model integrating machine learning models, such as Data preprocessing, feature selection, and classification. The utilization of the extra tree algorithm for optimal weight generation enhances the model's predictive accuracy.

The most significant achievement was the substantial reduction in overhead time, a critical metric for IDS systems. The proposed hybrid feature selection method, combined with extra tree algorithm-based hyperparameter tuning, resulted in over 40% and 98% reduction in training time for XGBoost and RF-based IDS, respectively, in both binary and multi-class detection processes. To validate the efficiency of this approach across diverse datasets and test it on the dataset, achieving a 100% F1 score in detecting attacks. These findings have crucial implications for the development of effective IDS systems, enabling the identification of optimal hyperparameters and a reduction in feature dimensions for enhanced model efficiency and performance. Looking ahead, future research could explore alternative hyperparameter optimization techniques and feature-select

REFERENCES

[1]. Megantara, A.A., Ahmad, T. A hybrid machine learning method for increasing the performance of network intrusion detection systems. *J Big Data* **8**, 142 (2021)

[2]. Aldallal, A.; Alisa, F. Effective Intrusion Detection System to Secure Data in Cloud Using Machine Learning. Symmetry **2021**, 13, 2306

[3]. Noah Oghenefego Ogwara Krassie Petrova, Mee Loong Yang, "Towards the Development of a Cloud Computing Intrusion Detection Framework Using and Ensemble Hybrid Feature Selection Approach", Journal of Computer Network and Communications, vol. 2022, Article ID 5988567, 16 pages, 2022.

[4]. M. Bakro et al., "An Improved Design for a Cloud Intrusion Detection System Using Hybrid Features Selection Approach With ML Classifier," in IEEE Access, vol. 11, pp. 64228-64247, 2023,

[5]. Bakır, H., Ceviz, Ö. Empirical Enhancement of Intrusion Detection Systems: A Comprehensive Approach with Genetic Algorithm-based Hyperparameter Tuning and Hybrid Feature Selection. *Arab J Sci Eng* (2024)

[6]. D. Jayalatchumy, R. Ramalingam, A. Balakrishnan, M. Safran and S. Alfarhood, "Improved Crow Search-Based Feature Selection and Ensemble Learning for IoT Intrusion Detection," in IEEE Access, vol. 12, pp. 33218-33235, 2024,

[7]. Megantara, A.A., Ahmad, T. A hybrid machine learning method for increasing the performance of network intrusion detection systems. *J Big Data* **8**, 142 (2021).

[8]. Kumar, Rakesh Ranjan, et al. "OPTCLOUD: An Optimal Cloud Service Selection Framework Using QoS Correlation Lens."Computational Intelligence and Neuroscience 2022 (2022): n. pag.

[9]. M. Baker, S. K. Bisoy, A. K. Patel, and M. A. Naal, ''Performance analysis of cloud computing encryption algorithms,'' in Advances in Intelligent Computing and Communication, in Lecture Notes in Networks and Systems, vol. 202. Singapore: Springer, 2021, pp. 357–367, doi: 10.1007/978-981-16-0695-3_35.

[10]. Akbar, Muhammad Azeem, et al. "Prioritization-based Taxonomy of Cloud-based Outsource Software Development Challenges: Fuzzy AHP analysis." *Appl. Soft Comput.* 95 (2020): 106557.

[11]. Benmessahel, I., Xie, K. & Chellal, M. A new evolutionary neural network based on intrusion detection systems using multiverse optimization.Appl Intell 48, 2315–2327 (2018)

[12]. Yang, Y.; Zheng, K.; Wu, C.; Yang, Y. Improving the Classification Effectiveness of Intrusion Detection by Using Improved Conditional Variational AutoEncoder and Deep Neural Network.*Sensors*2019,

[13]. Ahmad, Zeeshan, et al. "Network intrusion detection system: A systematic study of machine learning and deep learning approaches."Transactions on Emerging Telecommunications Technologies32 (2020): n. pag.

[14]. Rashid, Md. Mamunur et al. "A tree-based stacking ensemble technique with feature selection for network intrusion detection." *Applied Intelligence* 52 (2022)