Industrial Engineering Journal
ISSN: 0970-2555
Volume : 53, Issue 12, No.2, December : 2024

ENHANCED INTRUSION DETECTION IN CLOUD COMPUTING BASED ON MACHINE LEARNING

Devanshi Raghuwanshi M.Tech. Scholar Department of CSC Technocrats Institute of Technology and Science Bhopal, India Email: devanshi13399@gmail.com
Dr Vikas Gupta Prof. and Director Technocrats Institute of Technology and Science Bhopal, India
Email: vikasgupta.bhopal@gmail.com

Abstract-
Cloud Intrusion detection is one of the important security problems in today's cyber world. A significant number of techniques have been developed which are based on machine learning approaches. So for identifying the intrusion we have designed the machine learning algorithms. By using the algorithm we find out intrusion and we can identify the attacker's details also. IDS are mainly two types: Host based and Network based. A Host based Intrusion Detection System (HIDS) monitors individual host or device and sends alerts to the user if suspicious activities such as modifying or deleting a system file, unwanted sequence of system calls, unwanted configuration changes are detected. A Network based Intrusion Detection System (NIDS) is usually placed at network points such as a gateway and routers to check for intrusions in the network traffic. In this paper, KDD cup IDS dataset was taken from dataset repository. Then, we have to implement the pre-processing techniques. Then, we have to implement the different machine learning algorithms such as support vector machine and decision tree. The experimental results shows that the accuracy for above mentioned machine learning algorithms.

Keywords-
Cloud Intrusion Detection, Decision Tree, Support Vector Machine (SVM), Machine Learning

I INTRODUCTION

The customer can benefit from a multitude of services offered by cloud computing, including infrastructure, applications, and storage capacity. A cloud user mostly uses the internet to access and modify hardware and software to suit their needs. Although there are many advantages to using cloud computing, there are also drawbacks and difficulties. Cloud computing presents a number of issues, including load balancing, privacy, security, and performance management. The most significant issue among these is security because user data and apps are located on cloud infrastructure. Policies and procedures for safeguarding cloud-based data, applications, and infrastructure from intrusions and assaults are included in cloud computing security. Additionally, according to [1] and [2] it guards against SQL injection, cross-site scripting, data manipulation, leakage, and flooding attacks. Furthermore, cloud providers and users frequently report security problems brought on by different types of assaults. For instance, VUPEN Security found a virtual machine (VM) escape exploit in [3] Similar to this, Dropbox was the target of a distributed denial of service (DDoS) attack in 2013, which was publicized by ENISA [4]The attack caused a complete disruption of service for all subscribers for a period of fifteen days. Furthermore, according to Symantec (2015), around 450 vulnerabilities, including zero-day vulnerabilities, were revealed in January 2015. Cloud users experienced more than 650 million cyberattacks in 2018. In addition, 2019 saw a rise in advanced phishing efforts, IoT assaults, distributed denial of service (DDoS) campaigns, targeted ransomware, and attacks against cloud services and containers.

System for detecting intrusions (IDS)

Moreover, there are three primary categories under which cloud services fall: Software-as-a-Service (SaaS), Platform-as-a-Service (PaaS), and Infrastructure-as-a-Service (IaaS). To ensure user security, each of these services and methods has unique vulnerabilities and problems that need to be resolved [5].For instance, according to publications [6].PaaS is susceptible to phishing attacks, Man-in-the-

Middle attacks, and port scanning attacks; SaaS is susceptible to DoS/DDoS attacks, authentication attacks, and SQL injection attacks; and IaaS is vulnerable to attacks on virtual machine images, virtual network assaults, hypervisor attacks, domain name system (DNS) poisoning, ARP or IP spoofing, and cross-site scripting attacks, and data attacks [7]Thus, a system that can offer security against intrusions and malevolent actions is required. Intrusion detection systems were created in order to offer cloud security because of this. Host-based and network-based IDS are the two primary categories of IDS. An operating system is monitored by a HIDS, while suspicious activity in network traffic is observed by an NIDS. In order to identify internal changes made by insiders, a HIDS monitors system calls, critical files, and apps on each machine. HIDS is frequently used to alert the network manager to unusual activity. Conversely, an NIDS monitors both internal and external cloud networks by operating at key network nodes. It keeps an eye on every device linked to the network in order to see any suspicious activity or malicious activity [8]. Each method has benefits and drawbacks of its own. Therefore, total cloud security can be achieved by combining these strategies. While NIDS keeps an eye out for external attacks on network-connected devices including firewalls, routers, switches, and print servers, HIDS protects the equipment against insider threats.

II RELATED STUDY

Several sectors have embraced Cloud Computing (CC) due to its inherent characteristics, such as scalability and flexibility. However, despite these advantages, security concerns remain a significant challenge for cloud providers. CC introduces new vulnerabilities, including unauthorized access, data breaches, and insider threats. The shared infrastructure of cloud systems makes them attractive targets for attackers. The integration of robust security mechanisms becomes crucial to address these security challenges. One such mechanism is an Intrusion Detection System (IDS), which is fundamental in safeguarding networks and cloud environments. An IDS monitors network traffic and system activities. [9-10]

In recent years, researchers have explored the use of Machine Learning (ML) and Deep Learning (DL) approaches to enhance the performance of IDS. ML and DL algorithms have demonstrated their ability to analyze large volumes of data and make accurate predictions. By leveraging these techniques, IDSs can adapt to evolving threats, detect previous attacks, and reduce false positives. This article proposes a novel IDS model based on DL algorithms like the Radial Basis Function Neural Network (RBFNN) and Random Forest (RF). The RF classifier is used for feature selection, and the RBFNN algorithm is used to detect intrusion in CC environments. Moreover, the datasets Bot-IoT and NSL-KDD have been utilized to validate our suggested approach. To evaluate the impact of our approach on an imbalanced dataset, we relied on Matthew's Correlation Coefficient (MCC) as a normalized measure. Our method achieves accuracy (ACC) higher than 92% using the minimum features, and we managed to increase the MCC from 28% to 93%. The contributions of this study are twofold. Firstly, it presents a novel IDS model that leverages DL algorithms, demonstrating an improved ACC higher than 92% using minimal features and a substantial increase in MCC from 28% to 93%. Secondly, it addresses the security challenges specific to CC environments, offering a promising solution to enhance security in cloud systems. By integrating the proposed IDS model into cloud environments, cloud providers can benefit from enhanced security measures, effectively mitigating unauthorized access and potential data breaches. The utilization of DL algorithms, RBFNN, and RF has shown remarkable potential in detecting intrusions and strengthening the overall security posture of CC.[11-15]

Rajeev Kumar Tiwari et.al. (2023) the introduction of cloud computing (CC) has brought about a revolutionary change in the way information technology (IT) services are managed and supplied. These services are now able to be managed and delivered with flexibility, and they offer exceptional scalability and cost efficiency. In order to protect sensitive data and services, organizations are going to need to implement powerful intrusion detection systems (IDS) as they gradually become more dependent on cloud-based structures. This is because the attack surface is going to increase. Cloud-based intrusion detection systems (IDS) are able to safeguard against unauthorized access and other

harmful acts by proactively recognizing and responding to security threats and data breaches. This is accomplished through the utilization of developed machine learning (ML) methods and anomaly detection systems. In light of this, the purpose of this research is to introduce a unique Harris Hawks Optimizer with Deep Learning aided Intrusion Detection Approach (HHODL-IDA) for the Common Core environment. The automatic intrusion detection procedure is utilized by the HHODL-IDA technique, which results in an increase in the level of security within the CC environment. One of the primary applications of the HHODL-IDA technique is the application of min-max normalization, which guarantees consistent data pretreatment. In addition, the HHODL-IDA method makes use of stacked auto encoders (SAE), which are capable of deriving complicated features and enabling accurate detection of anomalies. In the long run, the HHO method will be able to be utilized for the purpose of optimizing the hyper parameters of the SAE method. It is possible to study the experimental evaluation of the HHODL-IDA technique using benchmark IDS datasets. On the basis of the extensive findings, it can be deduced that the HHODL-IDA technique achieves superior performance in terms of intrusion detection in comparison to other contemporary models in the CC environment.[16]

Govinda Rajulu. G et.al. (2022) among the most difficult components of cloud security, the investigation of intrusion detection systems is widely recognized as being among the most hard. Denial of service attacks, scanning, injection of malicious code, viruses, worms, and the breaking of passwords are examples of security breaches that are becoming increasingly widespread in the context of cloud computing. In the case that these attacks are not stopped in a timely manner, the company's reputation may be put in peril, and it may end up suffering financial damage as a result. Taking this into consideration, it is of the utmost importance to protect the cloud from threats of this nature. Within the scope of this investigation, we offer solutions to protect cloud-based systems by incorporating a number of the most effective ways of intrusion detection. When it comes to intrusion detection systems (IDS), we have been concentrating in primarily on two most important aspects: the efficiency of the detection methods and the rate of detection. The utilization of machine learning in conjunction with parallelization is something that everybody recommends doing in order to overcome these challenges. Consequently, the value of the human expert is diminished as a result of the fact that approaches that are based on machine learning get information from the data itself. It is possible to recognize intrusions through the utilization of a wide range of techniques, including statistical models, safe system approaches, neural networks, and a variety of other ways. An approach to intrusion detection that is based on existing intrusion detection systems and the various architectures that those systems employ is going to be shown in this study with the intention of enhancing the accuracy of intrusion detection in cloud computing. This will be accomplished by presenting a novel approach to intrusion detection. [17]

III PROPOSED WORK

In this system, the KDD cup IDS dataset was taken as input. The input data was taken from the dataset repository. Then, we have to implement the data pre-processing step. In this step, we have to handle the missing values for avoid wrong prediction, and to encode the label for input data. Then, we have to split the dataset into test and train. The data is splitting is based on ratio. In train, most of the data's will be there. In test, smaller portion of the data's will be there. Training portion is used to evaluate the model and testing portion is used to predicting the model. Then, we have to implement the classification algorithm (i.e.) machine learning. The machine learning algorithms such as SVM and DT. Then, the system can store the encrypted data in cloud for security purpose. Finally, the experimental results shows that the performance metrics such as accuracy, precision and recall.

Industrial Engineering Journal
ISSN: 0970-2555
Volume : 53, Issue 12, No.2, December : 2024

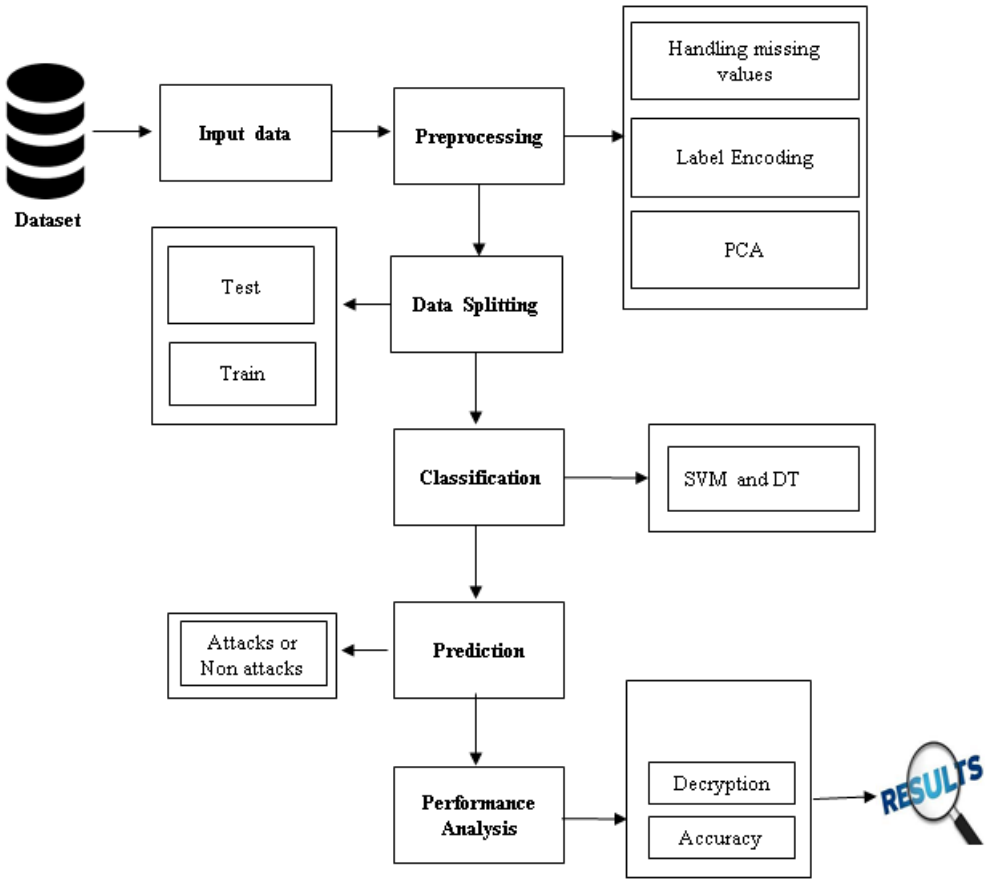

Figure 1 : system architecture

Data selection

- The input data was collected from dataset repository.
- In our process, the **KDD cup IDS** dataset is used.
- The data selection is the process of detecting the attacks.
- The input dataset was taken from dataset repository such as UCI repository.
- The dataset contains the information such protocol, duration, host error rate, label and so on.

Data Preprocessing

- Data pre-processing is the process of removing the unwanted data from the dataset.
- Pre-processing data transformation operations are used to transform the dataset into a structure suitable for machine learning.
- This step also includes cleaning the dataset by removing irrelevant or corrupted data that can affect the accuracy of the dataset, which makes it more efficient.
- Missing data removal
- Encoding Categorical data
- Missing data removal: In this process, the null values such as missing values and Nan values are replaced by 0.
- Missing and duplicate values were removed and data was cleaned of any abnormalities.
- Encoding Categorical data: That categorical data is defined as variables with a finite set of label values.
- That most machine learning algorithms require numerical input and output variables.

Feature Extraction

- PCA is used to reduce the dimensionality of a dataset while retaining as much variance (information) as possible.

- It transforms the original variables into a new set of orthogonal (uncorrelated) variables called principal components (PCs). These components are ordered by the amount of variance they explain in the data.
- **Dimensionality Reduction**: Reduces the number of variables (features) in a dataset while preserving as much information as possible.
- **Noise Reduction**: Filters out noise by focusing on the principal components with the most variance.
- **Visualization**: Visualizes high-dimensional data in a lower-dimensional space (e.g., 2D or 3D) for easier interpretation and exploration.
- **Feature Extraction:** Identifies patterns and relationships among variables, making it easier to interpret and analyze complex data.

Data Splitting

- During the machine learning process, data are needed so that learning can take place.
- In addition to the data required for training, test data are needed to evaluate the performance of the algorithm in order to see how well it works.
- In our process, we considered 70% of the input dataset to be the training data and the remaining 30% to be the testing data.
- Data splitting is the act of partitioning available data into two portions, usually for cross-validator purposes.
- One Portion of the data is used to develop a predictive model and the other to evaluate the model's performance.
- Separating data into training and testing sets is an important part of evaluating data mining models.
- Typically, when you separate a data set into a training set and testing set, most of the data is used for training, and a smaller portion of the data is used for testing.

Classification

This study focuses on enhancing cloud computing security by comparing two prominent machine learning techniques—Decision Tree and Support Vector Machine (SVM)—for intrusion detection. Decision Tree is known for its simplicity and interpretability, making it effective in identifying and classifying different types of intrusions based on distinct decision rules. On the other hand, SVM is a powerful classification algorithm that excels in high-dimensional spaces, often providing robust performance in detecting complex intrusion patterns. By evaluating the effectiveness of these techniques, the research aims to determine which method offers superior accuracy and efficiency in safeguarding cloud environments against unauthorized access and potential threats.

Encryption

- In this step, we can encrypt the predicted data by using RSA and the encrypted data will be stored in cloud.
- The **RSA** algorithm is an asymmetric cryptography algorithm; this means that it uses a public key and a private key (i.e two different, mathematically linked keys).
- As their names suggest, a public key is shared publicly, while a private key is secret and must not be shared with anyone.

Decryption

Decrypting RSA (Rivest-Shamir-Adleman) encrypted data involves using the private key corresponding to the public key used for encryption. Here are the steps involved in RSA decryption:[18] RSA involves generating a public-private key pair**:**

Public Key: Comprises a modulus n and a public exponent e.

Private Key: Comprises the same modulus n and a private exponent d.

Result Generation:

The Final Result will get generated based on the overall classification and prediction. The performance of this proposed approach is evaluated using some measures like,

Accuracy

Accuracy of classifier refers to the ability of classifier. It predicts the class label correctly and the accuracy of the predictor refers to how well a given predictor can guess the value of predicted attribute for a new data.

Accuracy = (TP+TN)/ (TP+TN+FP+FN)

Precision

Precision is defined as the number of true positives divided by the number of true positives plus the number of false positives.

Precision = TP/ (TP+FP)

Recall

Recall is the number of correct results divided by the number of results that should have been returned. In binary classification, recall is called sensitivity. It can be viewed as the probability that a relevant document is retrieved by the query.

Recall = TP/ (TP+FN)

IV SIMULATION RESULT

This Simualtion developing a cloud intrusion detection system using Decision Tree and SVM techniques, where all processes will follow the outlined methodology. The provided dataset will be used for analysis, and a graphical user interface (GUI) will be available to interact with the system, and the entire implementation will be carried out in Python using the Anaconda Navigator – Spyder environment.

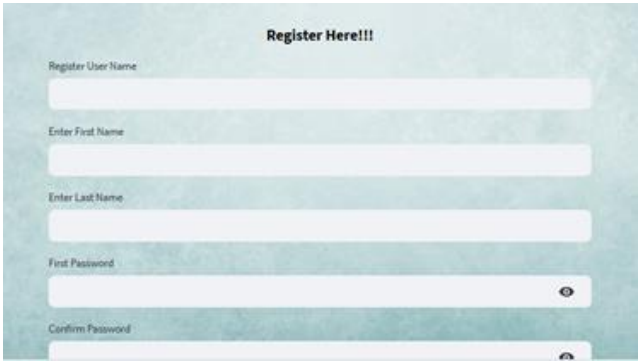

Fig. 2 registration phase

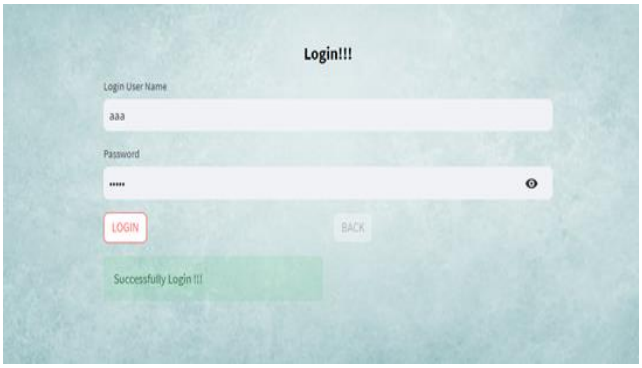

Fig. 3 login phase

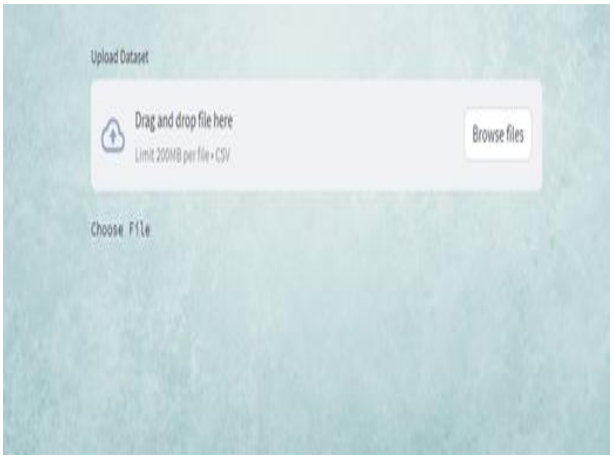

Fig.4 upload file

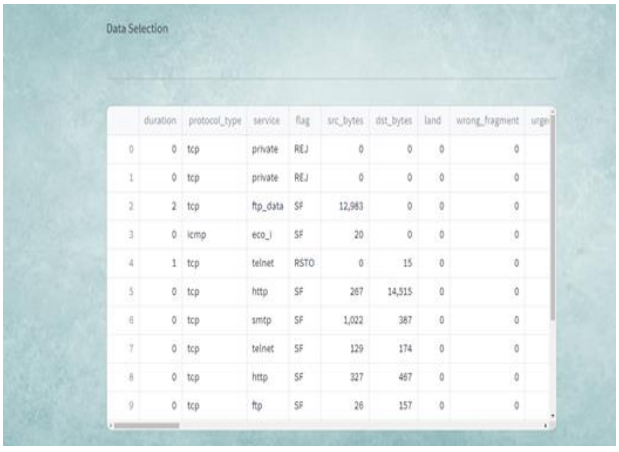

Fig. 5 data selection

Industrial Engineering Journal
ISSN: 0970-2555
Volume : 53, Issue 12, No.2, December : 2024

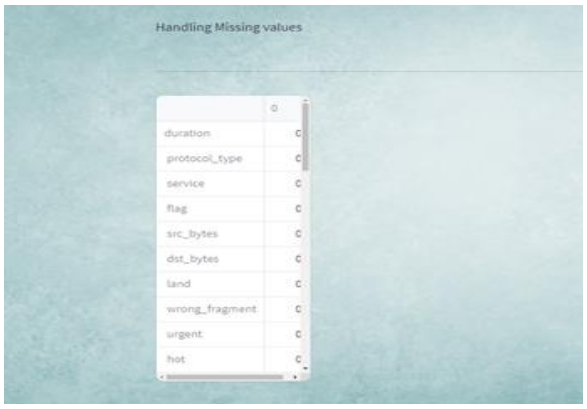

Fig. 5.6 handling missing values

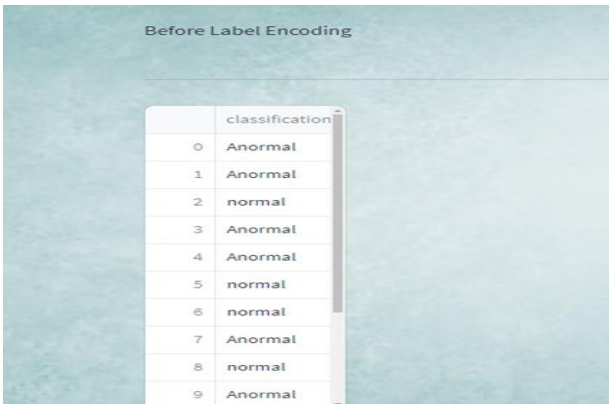

Fig. 5.7 dataset before labeling

Fig.6 label encoding

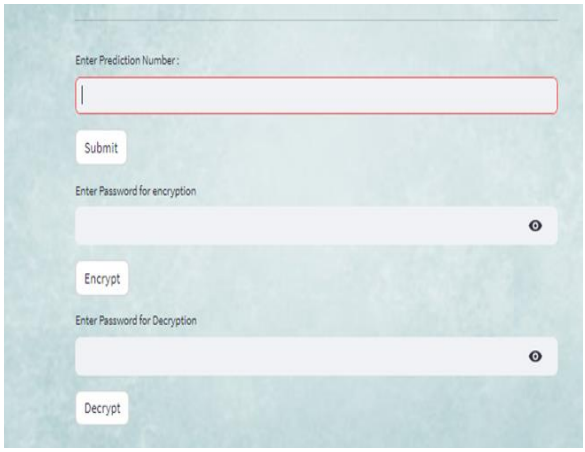

Fig. 7 prediction number and encrypt and decrypt data

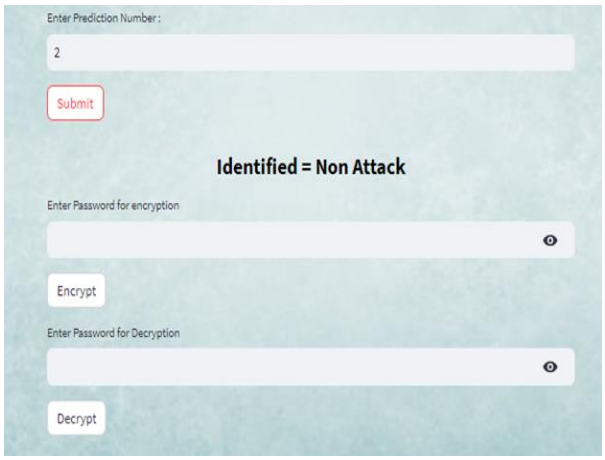

Fig. 8 Attack Detection

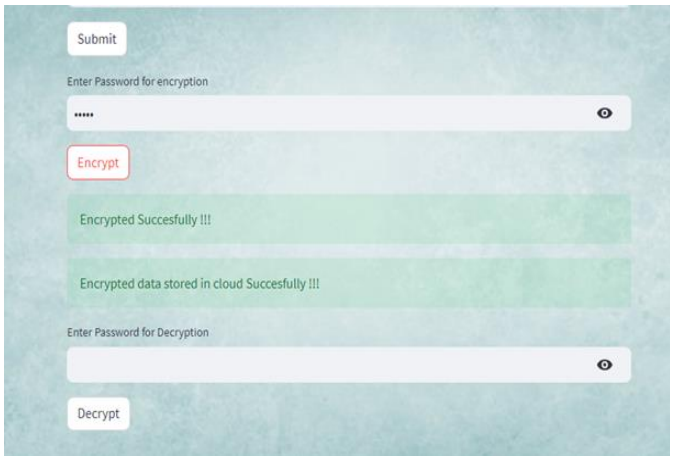

Fig. 9 encrypted data on cloud

Industrial Engineering Journal
ISSN: 0970-2555
Volume : 53, Issue 12, No.2, December : 2024

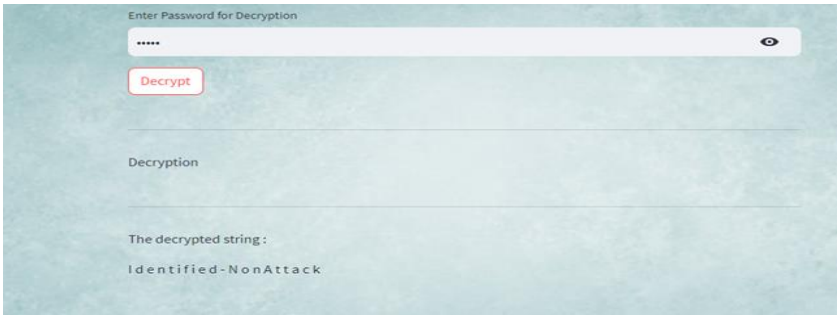

Fig. 10 decrypted data on cloud

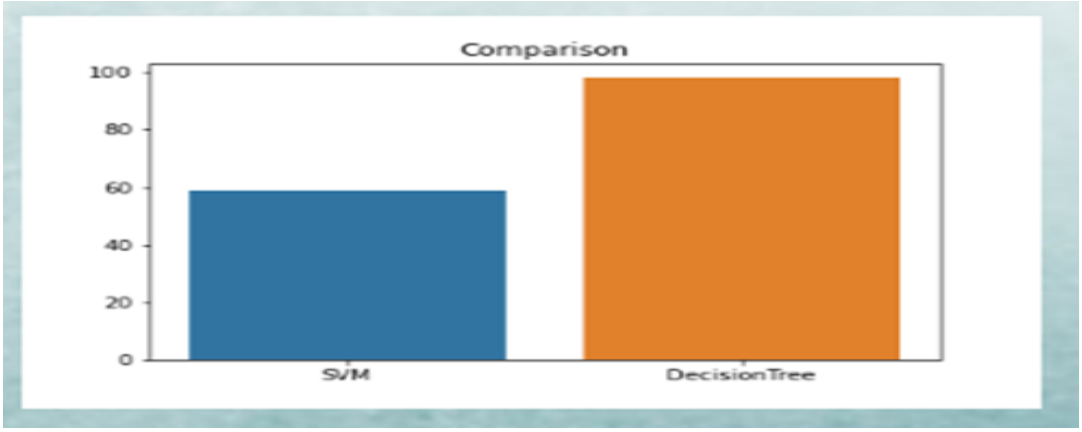

Fig. 11 comparison proposed techniques

Table 1 comparison with existing work

Technique	Accuracy(%)
Radial Basis Function Neural Network (RBFNN)	90.49
Decision tree	98.16

Table 1 presents a comparison of the accuracy achieved by two different machine learning techniques in glaucoma detection. The Radial Basis Function Neural Network (RBFNN) technique yields an accuracy of 90.49%, while the Decision Tree technique demonstrates a higher accuracy of 98.16%. This comparison highlights the superior performance of the Decision Tree approach over the RBFNN in the context of glaucoma detection.

V CONCLUSION

The development of an advanced intrusion detection and data security system addresses the critical need for enhanced cyber security measures in today's digital landscape. By leveraging the NSL-KDD dataset, this project successfully implements robust data preprocessing, feature extraction, and cutting-edge machine learning algorithms to accurately detect and classify network intrusions. Furthermore, the integration of encryption techniques like RSA ensures that sensitive data remains secure during storage and transmission in cloud environments. This comprehensive approach not only improves the accuracy and reliability of intrusion detection but also provides a robust framework for securing data against potential cyber threats.

References

1. Hady, A.A.; Ghubaish, A.; Salman, T.; Unal, D.; Jain, R. Intrusion Detection System for Healthcare Systems Using Medical and Network Data: A Comparison Study. IEEE Access 2020, 8, 106576–106584.
2. Guezzaz, A.; Azrour, M.; Benkirane, S.; Mohy-Eddine, M.; Attou, H.; Douiba, M. A Lightweight Hybrid Intrusion Detection Framework using Machine Learning for Edge-Based IIoT Security. Int.

Arab. J. Inf. Technol. 2022, 19, 822–830.
3. Hazman, C.; Guezzaz, A.; Benkirane, S.; Azrour, M. IDS-SIoEL: Intrusion Detection Framework for IoT-based Smart Environments Security using Ensemble Learning. Clust. Comput. 2022, 1–15.
4. Douiba, M.; Benkirane, S.; Guezzaz, A.; Azrour, M. An improved anomaly detection model for IoT security using decision tree and gradient boosting. J. Supercomput. 2022, 79, 3392–3411.
5. Alshammari, A. Aldribi, A. Apply machine learning techniques to detect malicious network traffic in cloud computing. J. Big Data 2021, 8, 90.
6. Mohy-Eddine, M.Guezzaz, A.Benkirane, S.; Azrour, M. An efficient network intrusion detection model for IoT security usingK-NN classifier and feature selection. Multimed. Tools Appl. 2023, 82, 23615–23633.
7. Jiang, F. Fu, Y. Gupta, B.B. Liang, Y.; Rho, S. Lou, F. Meng, F. Tian, Z. Deep Learning Based Multi-Channel Intelligent Attack Detection for Data Security. IEEE Trans. Sustain. Comput. 2018, 5, 204–212.
8. Burhan, F. Mustafa, G. Nawaz, A. Kiani, A. Ali, T. Securing Cloud Data: A Machine Learning based Data Categorization Approach for Cloud Computing. Res. Sq. 2022.
9. Mubarakali, A. Srinivasan, K. Mukhalid, R. Jaganathan, S.C.B.Marina, N. Security challenges in internet of things: Distributed denial of service attack detection using support vector machine-based expert systems. Comput. Intell. 2020, 36, 1580–1592.
10. Mishra, A. Gupta, B.B. Perakovic, D. Penalvo, F.J.G.; Hsu, C.-H. Classification Based Machine Learning for Detection of DDoS attack in Cloud Computing. In Proceedings of the International Conference on Consumer Electronics, Las Vegas, NV, USA, 10–12 January 2021.
11. Singh, P.; Ranga, V. Attack and intrusion detection in cloud computing using an ensemble learning approach. Int. J. Inf. Technol. 2021, 13, 565–571.
12. Verma, A.; Ranga, V. Machine Learning Based Intrusion Detection Systems for IoT Applications. Wirel. Pers. Commun. 2020, 111, 2287–2310.
13. Mohy-Eddine, M. Guezzaz, A.; Benkirane, S.Azrour, M. An effective intrusion detection approach based on ensemble learning for IIoT edge computing. J. Comput. Virol. Hacking Tech. 2022, 1–13.
14. Liu, Z. Shi, Y. A Hybrid IDS Using GA-Based Feature Selection Method and Random Forest. Int. J. Mach. Learn. Comput. 2022, 12, 43–50
15. . Chaabouni, N.; Mosbah, M.; Zemmari, A. Sauvignac, C. A OneM2M Intrusion Detection and Prevention System based on Edge Machine Learning. In Proceedings of the IEEE/IFIP Network Operations and Management Symposium, Budapest, Hungary, 20–24 April 2020.
16. Rajeev Kumar Tiwari;S. Murugappan(2023)“Enhancing Security in Cloud Computing and Protocols Using Harris Hawks Optimizer with Deep Learning for Intrusion Detection”2023 International Conference on Sustainable Communication Networks and Application (ICSCNA) 2023
17. Govinda Rajulu. G;R. Santhoshkumar;D. Venkatesan;Kalvikkarasi. S;P. Santosh Kumar Patra(2022)“Intrusion Detection in Cloud Architecture Using Machine Learning”2022 4th International Conference on Advances in Computing, Communication Control and Networking (ICAC3N)2022
18. Hanaa Attou, Mouaad Mohy-eddine ,Azidine Guezzaz, Said Benkirane , Mourade Azrour , Abdulatif Alabdultif ,and Naif Almusallam Towards an Intelligent Intrusion Detection System to Detect Malicious Activities in Cloud Computing 4 Appl. Sci. 2023, 13, 9588. https://doi.org/10.3390/app13179588