ENHANCING THE DATA TRANSMISSION SECURITY IN CLOUD USING MACHINE LEARNING

[1]CHITTALA MOHANA SUBHA SREE, [2]K.RAJA RAJESWARI

[1]Students, Department of MCA, B V Raju College, Bhimavaram Ap

[2]Assistant Professor, Department of MCA, B V Raju College, Bhimavaram Ap

ABSTRACT

Cloud computing has emerged as a cost-effective and scalable solution for deploying modern applications by offering high computational power and storage capabilities. However, its reliance on network connectivity exposes it to various cyber threats such as Distributed Denial of Service (DDoS) and Man-in-the-Middle (MITM) attacks, which compromise data transmission security. Traditional cryptographic techniques, while effective, often involve high computational overhead and remain vulnerable if encryption keys are compromised. To address these limitations, this work proposes a machine learning-based framework for enhancing data transmission security in cloud environments. The system utilizes K-Nearest Neighbors (KNN) and Artificial Neural Networks (ANN) for anomaly detection, with further performance enhancement achieved through the integration of the Artificial Bee Colony (ABC) optimization algorithm. The ABC algorithm optimizes ANN by selecting the best feature subset and tuning hyperparameters such as learning rate, number of neurons, and epochs, inspired by the foraging behavior of honeybees. The model is trained and evaluated using the NSL-KDD dataset, which consists of 41 features related to network traffic. After preprocessing and feature optimization, 32 significant features were selected. Experimental results demonstrate that KNN achieved 96% accuracy, ANN achieved 97%, while the optimized ANN-ABC model achieved a superior accuracy of 99%, along with improved precision, recall, and F1-score. The implementation is carried out using Jupyter Notebook for model training and Flask framework for real-time prediction through a web interface. The results confirm that the proposed ANN-ABC model significantly enhances intrusion detection performance, making it a reliable solution for securing cloud data transmission.

Keywords : ***Cloud Security, Machine Learning, KNN, ANN, Artificial Bee Colony (ABC), Intrusion Detection, NSL-KDD Dataset, Data Transmission Security, Cyber Attacks, Anomaly Detection.***

I.INTRODUCTION

Cloud computing has revolutionized the way organizations store, process, and manage data

by offering scalable infrastructure, high computational power, and cost-effective services. It enables both small-scale and large-scale applications to be deployed efficiently without the need for extensive physical hardware. Despite its advantages, cloud computing environments are highly dependent on network connectivity, which makes them vulnerable to various cyber threats. During data transmission between users and cloud servers, attackers can exploit vulnerabilities to launch attacks such as Distributed Denial of Service (DDoS), Man-in-the-Middle (MITM), and unauthorized access attempts. These attacks can disrupt services, compromise sensitive data, and reduce user trust. As cloud adoption continues to grow, ensuring secure data transmission has become a critical concern for researchers and organizations. Traditional security measures, including firewalls and encryption techniques, provide a basic level of protection, but they are often insufficient against evolving and sophisticated cyber threats. Therefore, there is a need for intelligent and adaptive security mechanisms that can detect and prevent attacks in real time while maintaining system performance and scalability.

In recent years, machine learning has emerged as a powerful tool for enhancing cybersecurity due to its ability to analyze large volumes of data and identify hidden patterns. Unlike traditional rule-based systems, machine learning algorithms can learn from historical data and adapt to new and unknown threats. Techniques such as K-Nearest Neighbors (KNN) and Artificial Neural Networks (ANN) are widely used for intrusion detection and anomaly detection in network security. KNN is a simple yet effective algorithm that classifies data based on similarity measures, while ANN mimics the human brain's neural structure to identify complex patterns in data. However, the performance of these algorithms depends heavily on parameter tuning and feature selection. Improper configuration can lead to reduced accuracy and increased false positives. To overcome these challenges, optimization techniques are often integrated with machine learning models to improve their efficiency and predictive capabilities. This integration helps in selecting the most relevant features and fine-tuning model parameters, leading to better detection of cyber threats in cloud environments.

This work proposes an enhanced security framework that combines machine learning algorithms with an optimization technique to improve data transmission security in cloud computing. Specifically, the Artificial Neural Network (ANN) model is optimized using the Artificial Bee Colony (ABC) algorithm, which is inspired by the foraging behavior of honeybees. The ABC algorithm efficiently searches for optimal solutions by evaluating fitness values and iteratively improving the

model's performance. The proposed system is trained and tested using the NSL-KDD dataset, which contains a wide range of network traffic features and attack types. Data preprocessing techniques such as label encoding, normalization, and feature selection are applied to improve model performance. Experimental results show that the optimized ANN-ABC model outperforms traditional KNN and ANN models in terms of accuracy, precision, recall, and F1-score. Furthermore, the system is implemented using Jupyter Notebook for model development and a Flask-based web application for real-time prediction. This approach demonstrates a practical and efficient solution for securing cloud data transmission against modern cyber threats.

II SURVEY OF RESEARCH

[1] The study by Ashish Vaswani et al. (2017) introduced the Transformer architecture, a breakthrough in Natural Language Processing. The model replaces recurrent structures with self-attention mechanisms, allowing parallel processing of input data and capturing long-range dependencies efficiently. This significantly improves performance in tasks such as machine translation and text generation. Experimental results showed superior accuracy and reduced training time compared to traditional models like RNNs and LSTMs. However, the model requires large datasets and high computational resources, making it less suitable for low-resource environments. Despite these limitations, the Transformer architecture has become the backbone of many modern AI systems. In the proposed work, this concept supports intelligent data analysis and pattern recognition, which is essential for detecting anomalies in cloud environments and improving overall system performance.

[2] The research by Ian Goodfellow et al. (2016) on deep learning provides a comprehensive foundation for neural network-based models such as Artificial Neural Networks (ANN). The study explains how multi-layer neural networks can learn complex patterns from large datasets through forward and backward propagation. These models are widely used in classification and prediction tasks. The results demonstrated that deep learning models outperform traditional machine learning algorithms in tasks involving high-dimensional data. However, deep learning requires significant computational power and large labeled datasets for effective training. Overfitting and long training time are additional challenges. This research is highly relevant to the proposed system, as ANN is used for intrusion detection. By leveraging deep learning principles, the system can accurately identify patterns of cyber-attacks and improve cloud data security through intelligent anomaly detection.

[3] The study by Thomas Cover and Peter Hart (1967) introduced the K-Nearest Neighbors (KNN) algorithm, a simple yet effective supervised learning technique. KNN classifies data points based on the majority class of their nearest neighbors using distance metrics such as Euclidean distance. It is widely used in pattern recognition and intrusion detection systems due to its simplicity and effectiveness. The results showed that KNN performs well for small datasets and low-dimensional data. However, its performance decreases with large datasets due to high computational cost and sensitivity to irrelevant features. Despite these drawbacks, KNN remains a baseline model for comparison in many research works. In the proposed system, KNN is used as a comparative model to evaluate the performance of advanced algorithms like ANN and optimized ANN, highlighting improvements in accuracy and detection capability.

[4] The research by Dervis Karaboga (2005) introduced the Artificial Bee Colony (ABC) optimization algorithm inspired by the foraging behavior of honeybees. The algorithm uses employed bees, onlooker bees, and scout bees to explore and exploit the search space efficiently. It is widely used for optimization problems such as feature selection and parameter tuning in machine learning models. Experimental results showed that ABC can effectively find optimal or near-optimal solutions with less computational complexity compared to other optimization techniques. However, it may require proper parameter tuning for best performance. This research is highly relevant to the proposed work, as ABC is used to optimize the ANN model. By selecting the most relevant features and tuning parameters, ABC enhances the performance of ANN, leading to improved accuracy in detecting cyber threats in cloud environments.

III. WORKING METHODOLOGY

The proposed system follows a structured methodology to enhance data transmission security in cloud environments using machine learning techniques. Initially, the NSL-KDD dataset is used as the primary source for training and testing the models. This dataset contains 41 features representing different network traffic characteristics, including both numeric and non-numeric attributes. In the preprocessing stage, non-numeric values such as protocol types and services are converted into numeric format using label encoding techniques. Missing values are handled to ensure data consistency, and normalization is applied to scale the feature values within a specific range. After preprocessing, the dataset is shuffled and split into training and testing sets, typically using an 80:20 ratio. This step ensures that the model learns effectively from diverse data while maintaining unbiased evaluation. Proper preprocessing improves model accuracy and reduces noise, enabling

better detection of anomalies and cyber-attacks during cloud data transmission.

In the next phase, machine learning models such as K-Nearest Neighbors (KNN) and Artificial Neural Networks (ANN) are implemented for anomaly detection. KNN classifies network traffic based on similarity measures, while ANN learns complex patterns through multiple hidden layers and activation functions. The ANN model is configured with optimized parameters such as learning rate, number of neurons, and epochs to achieve better performance. To further enhance the ANN model, the Artificial Bee Colony (ABC) optimization algorithm is applied. ABC works by simulating the foraging behavior of honeybees, where different bee types explore and exploit possible solutions. It selects the best feature subset and tunes ANN parameters by evaluating fitness values iteratively. This optimization process helps in reducing irrelevant features and improving model efficiency, leading to higher accuracy and better generalization in detecting cyber threats within cloud communication systems.

Finally, the performance of all models is evaluated using standard metrics such as accuracy, precision, recall, and F1-score. The results show that KNN achieves moderate performance, while ANN provides improved accuracy due to its ability to learn complex patterns. However, the ANN optimized with ABC significantly outperforms both models by achieving the highest accuracy and better detection rates. Confusion matrices and graphical visualizations are used to analyze classification performance and error distribution. The system is implemented using Jupyter Notebook for model development and training, while a Flask-based web application is developed for real-time prediction. Users can upload test data through the web interface, and the system predicts whether the data represents normal or malicious activity. This integration of machine learning and web technologies provides a practical and efficient solution for securing cloud data transmission against evolving cyber threats.

IV RESULTS EXPLANATIONS

Cloud servers provides heavy computation resources and storage at cheaper cost which is migrating all small and big network based applications for cloud deployment. Due to network connectivity cloud often fall prey of cyber attackers which can perform different number of attacks during communication between cloud server and user. Attackers can generate many forms of attack like DDOS, MITM and many more.

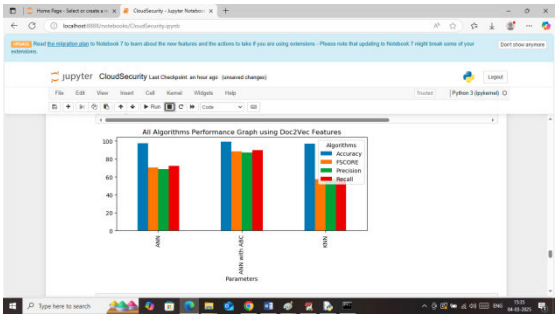
To combat against such attacks many cryptography based algorithms were introduced but its required heavy computation and can leak data if encrypted keys are exposed to attackers.

To overcome from above issues we are employing machine learning based algorithms such as KNN and ANN which are one of the powerful ML algorithms to detect any type of Anomaly activities. To further enhance ANN performance we are tuning ANN with ABC (artificial bee colony) optimization algorithm.

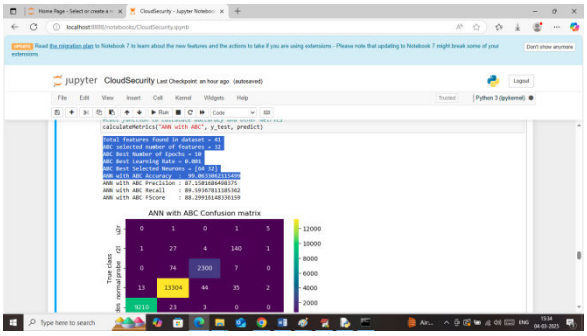
ABC inspired from the foraging behaviour of honeybees to solve optimization problems. ABC will get initialized with number of iterations and populations and then keep optimizing ANN model performance by choosing random population with max given number of iterations. In each iteration ABC will compute fitness values and this process continues till no more fitness can be enhanced.

To train and test above algorithms performance we have used NSL KDD dataset which has total 41 features and after optimization ABC selected 32 features with best learning rate as 0.001, number of epochs as 10 and number of tuning neurons are 32 and 64 for two different ANN layers.

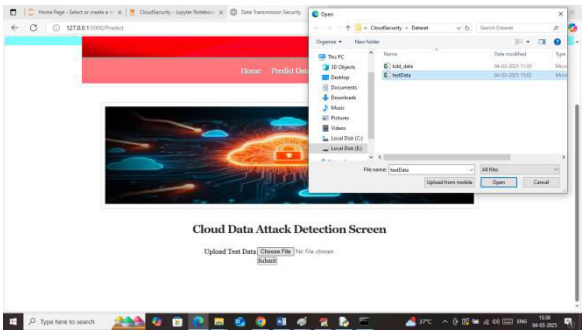
We have compared ANN with ABC performance with KNN and ANN algorithms and each algorithm performance was evaluated in terms of accuracy, precision, recall and FSCORE.



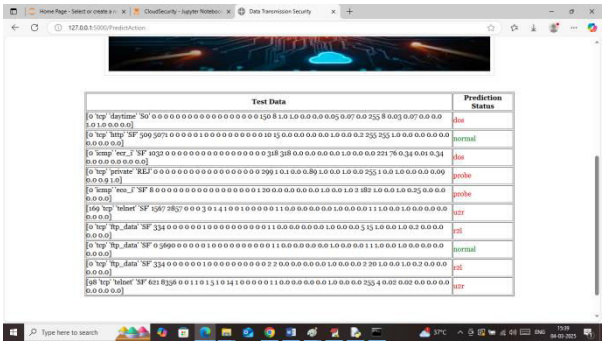
In above screen displaying comparison graph between all algorithms where x-axis represents algorithm names and y-axis represents accuracy and other metrics in different colour bars



In above screen displaying best selected features, neurons and other tuned parameters and then ANN with ABC got 99% accuracy and can see other metrics also



In above screen selecting and uploading ‘test data.csv’ file and then click on buttons to get below page



In above screen in first column can see TEST data values and in second column can see predicted attack types.

V.CONCLUSION

In this work, an effective machine learning-based framework has been proposed to enhance data transmission security in cloud computing environments. Traditional cryptographic methods, although widely used, face limitations such as high computational overhead and vulnerability to key exposure. To overcome these challenges, the proposed system integrates K-Nearest Neighbors (KNN), Artificial Neural Networks (ANN), and an optimized ANN model using the Artificial Bee Colony (ABC) algorithm for efficient anomaly detection. The use of the NSL-KDD dataset, along with proper preprocessing techniques such as label encoding, normalization, and feature selection, significantly improves model performance. Experimental results demonstrate that the ANN optimized with ABC achieves superior accuracy, precision, recall, and F1-score compared to KNN and standard ANN models. The implementation using Jupyter Notebook and Flask further validates the practicality of the system for real-time prediction. Overall, the proposed approach provides a robust, scalable, and intelligent solution for detecting cyber threats and securing cloud data transmission against evolving attacks.

RE.FERENCES

[1] A. Vaswani et al., “Attention is All You Need,” *Advances in Neural Information Processing Systems*, vol. 30, pp. 5998–6008, 2017.

[2] I. Goodfellow, Y. Bengio, and A. Courville, *Deep Learning*. Cambridge, MA, USA: MIT Press, 2016.

[3] T. Cover and P. Hart, “Nearest Neighbor Pattern Classification,” *IEEE Transactions on Information Theory*, vol. 13, no. 1, pp. 21–27, Jan. 1967.

[4] D. Karaboga, “An Idea Based on Honey Bee Swarm for Numerical Optimization,” Technical Report-TR06, Erciyes University, Turkey, 2005.

[5] M. Tavallaee, E. Bagheri, W. Lu, and A. A. Ghorbani, “A Detailed Analysis of the KDD CUP 99 Data Set,” in *Proc. IEEE Symposium on Computational Intelligence for Security and Defense Applications*, 2009, pp. 1–6.

[6] J. Kennedy and R. Eberhart, "Particle Swarm Optimization," in *Proc. IEEE International Conference on Neural Networks*, 1995, pp. 1942–1948.