Identification of Security Issues and Finding their Solution in Cloud Computing

Himanshu Kalra

Salesforce Technical Manager, Technology Department,

Fox Corporation at 1211 6th Ave, New York, NY 10036

Corresponding mail: hkalra.gkg@gmail.com

Abstract

The advent of Cloud Computing has simplified on-demand access to IT services including data storage and administration. In addition, it seeks to secure systems and make them functional. With these benefits, there are significant security constraints for cloud providers. When it comes to cloud computing, one of the biggest obstacles is ensuring the safety of data and services. Considering this, several solutions have been put into place to boost cloud security by keeping an eye on everything from resources to services to networks to identify and stop intrusions as soon as they occur. The term "Intrusion Detection System" (IDS) refers to an improved technique used to regulate network traffic and identify abnormal activity. This paper presents the identification of Security Issues and Finding their Solution in Cloud Computing using machine learning techniques including Support Vector Machine (SVM), Random Forest (RF), K-Nearest Neighbor (KNN), Multi-Layer Protocol (MLP). This model is trained and evaluated using NSL-KDD dataset. The experimental findings show the highest accuracy of 93.5% with the use of SVM model. As a result, the achieved results demonstrate strong performance concerning Accuracy, Precision, Recall, and F1-Score when compared to recent studies.

Keywords: Cloud Computing, Cloud Security, Machine Learning (ML), Intrusion Detection System (IDS).

1. Introduction

A cloud computing or cloud-based environment is an internet-based service that allows the sharing of computer resources and other devices as needed. It serves as a platform that facilitates the sharing of resources on demand. Applications that can save information on a server, data center, or network are just a few examples. That requires little to no work to create. With the use of cloud computing, businesses, and individuals may store their information in remote data centers, which might be placed anywhere from across town to across the globe. The National Institute of Science and Technology (NIST) defines cloud computing as Cloud computing is a framework that empowers ubiquitous and convenient network access on demand to a shared repository of configurable computing resources. These resources include networks, servers, storage, applications, and services, and can be swiftly allocated and de-provisioned with minimal need for management effort or interaction with service providers [1]. Figure 1 illustrates the features of cloud services that aid in the learning and comprehension of cloud computing [2].

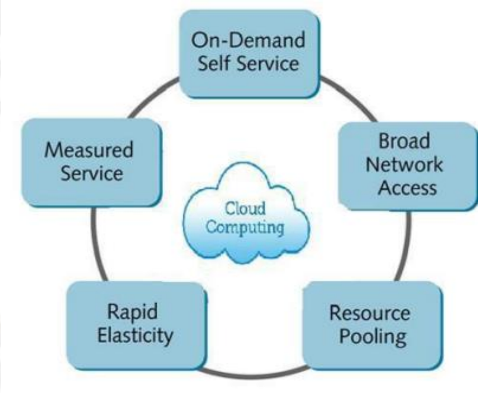


Figure 1: Cloud Computing Characteristics.

The following is a summary of the many ways in which cloud computing differs from conventional computer systems:

- On-Demand Self-Service: The required computational resources (such as network storage or server time) are automatically allocated without requiring human intervention from multiple service providers.
- Broad Network Access: Cloud computing makes resources like servers and databases accessible through the internet and could be used by a wide variety of clients with varying hardware and software requirements.

- Resource Pooling: Multiple users could be supported by a single provider's pooled computer resources (including storage, processor, memory, network bandwidth, and virtual machines) due to the multi-tenant concept.
- Rapid Elasticity: Provides scalable resources on-demand, with the option to add or remove resources as required.
- Measured Service: At some level of abstraction, control, monitoring, and optimization of resource use are performed automatically. It's transparent between the service provider and the user and it matches the service type that was utilized [3].

With its popularity, cloud computing has been plagued by security issues that threaten the confidentiality, integrity, and availability of stored information. The information is sent to a remote server, sometimes known as "the cloud," where it is stored. Data storage breaches include not just data loss but also data abuse and account takeover. Users must take precautions with both the network and the cloud because of the risk of intrusion. The risk of unauthorized people gaining access to sensitive data is high. The data is vulnerable to being altered, misused, corrupted, or stolen because of this. This is the most pressing concern about the security of cloud-stored information. If data security is breached, the database becomes vulnerable to further attacks. This problem statement highlights the need to leverage machine learning to enhance security in cloud computing by providing a foundation for developing effective solutions to address the evolving security challenges in the cloud environment. The following objectives are given below:

- To design and implement a machine learning-based framework for real-time monitoring and detection of security threats in cloud computing.
- To develop a comprehensive taxonomy of security issues specific to cloud computing, including data breaches, insider threats, and DDoS attacks.
- To evaluate the performance and accuracy of the proposed solution in identifying and mitigating security issues in a variety of cloud-based applications and infrastructures.
- To investigate the integration of other emerging technologies, such as blockchain or secure enclaves, with machine learning for enhanced cloud security.

The paper's organization can be summarized as follows. Section II offers an overview of a review of the literature, discussing various methods for dataset normality identification, including prominent data normalization techniques. In Section III, we investigate the specific technique employed to address the security issues. Section IV offers a comprehensive explanation of our proposed data normalization process. Moving forward to Section V, we present a summary of the experiments conducted to assess and validate our proposed approach. Finally, in Section VI, we conclude the paper.

1. Security Issues in Cloud Computing

The safety of data stored online is the primary concern for every cloud service. Everything is stored at the provider's location, which greatly increases the risk of data loss [4]. It's one of the biggest problems with using the cloud for everything. An IDC study of cloud service providers found that privacy and data protection were major issues [5,6]. The results of the study conducted by IDC among 224 IT executives are shown in Figure 2. Service disruptions, data theft, loss of privacy, and data damage are some of the challenges and costs associated with cloud computing. Companies are unable to move forward with cloud computing because of these issues. To get rid of them, just use reliable security measures. IEEEtran.bst files, but the Microsoft Word templates are self-contained. It is important to us that all the templates look the same, therefore they have gone to great lengths to make that happen.

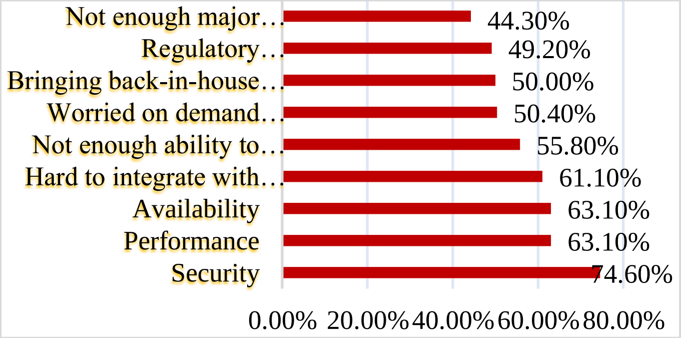


Figure 2: Analysis of major issues of cloud computing [7]

2. Solutions for Security Issues

Cloud computing enables users to store their data on external servers that are not under their ownership. These cloud services are typically offered by a Cloud Provider (CP) [8]. It is crucial that users not lose knowledge of data integrity, and that only the CP has access to data security details. The data must be stored in a cryptographically secure way such that only authorized users have access to the primary data. This is accomplished using encryption and decryption techniques. The confidentiality of information is protected by using the AES encryption algorithm. It generates a digital signature that can be checked using the Digital Signal Algorithm (DSA). This ensures that only the user has access to the data and that any modifications made are encrypted once again. Data backups, network traffic, and file system security are just a few areas where cryptography might be useful. The system uses both symmetric and asymmetric algorithms to function. DES, AES, and 3DES are three important symmetric algorithms [9,10].

2. Literature of Review

In this Section, the authors define the previous studies of several authors built on the Identification of Security Issues and Finding their Solutions in Cloud Computing.

Kassabi H. et al., (2023) [11] offered a framework and a theoretical model to ensure the safety of cloud-based workflow orchestration. To identify and anticipate abnormalities in cloud workflow orchestration, the suggested design focuses on observing cloud resources, workflow tasks, and the data by combining Deep Learning (DL), one-class classification, and clustering. Predictions of anomalies are very accurate in terms of precision, recall, and F1 scores, both during training and in subsequent prediction trials. Despite this, other investigation findings showed that the cloud workflow's execution speed remained consistently high when an adaptation technique was used in response to certain observed abnormalities. The results of the studies show that the suggested design saves time and energy by detecting and predicting anomalies before they occur.

Attou H. et al., (2023) [12] introduced a cloud-centric intrusion detection model built around the principles of the RF algorithm and feature engineering. The RF classifier has been included in the model without causing any disruptions to increase the accuracy. This approach was rigorously tested and verified on two distinct datasets, resulting in an impressive 98.3% ACC for the Bot-IoT dataset and an astounding 99.99% ACC for the NSL-KDD dataset. These findings underline the model's outstanding performance in terms of accuracy, precision, and recall when benchmarked against recent, similar studies.

Abbas Z. et al., (2023) [13] aimed to improve cloud computing security by using ML approaches (including Support Vector Machine, XGBoost, and Artificial Neural Networks). The goals of the ML research have been met with the selection of these 11 characteristics. This study identifies deficiencies in the utilization of Machine Learning (ML) methods within the domain of cloud cybersecurity. Furthermore, the primary objective of this study is to formulate a pragmatic approach for forecasting the adoption of ML within an industrial cloud setting, with a specific focus on concerns related to trust and privacy. The empirical findings demonstrated that the XGBoost model had superior accuracy, precision, recall, F1-score, and ROC-AUC compared to the other methods. This study demonstrates how ML algorithms could be used to better secure cloud-based enterprise applications.

Alhazmi L. et al., (2023) [14] presented an optimization-based DL model to detect cloud-based DoS attacks using a Deep Convolutional Generative Adversarial Network (DCGAN). The suggested model uses DCGAN to understand the geographical and temporal characteristics of network traffic data, which enables the identification of patterns indicative of DoS attacks. In addition, a large amount of network traffic data is used to train the DCGAN, making it more accurate and secure. The results of the tests show that the proposed model outperforms state-of-the-art technological approaches in terms of both accuracy and the number of false positives generated while looking for cloud-based DoS attacks.

Amitha M. et al., (2023) [15] evaluated a model that combines the strengths of Radial Basis Function (RBF) and Long Short-Term Memory (LSTM) networks to better protect cloud-based systems against DDoS attacks. The evaluation of the CICDDoS2019 benchmark dataset reveals the efficacy of the suggested approach in detecting DDoS assaults and decreasing their influence on cloud infrastructure. DDoS assaults have an accuracy of about 99.95%, much surpassing that of KNN and other ML models. When applied to a larger dataset, transfer learning techniques allow for improved performance.

Bingu R. et al., (2023) [16] introduced a DL ensemble method for monitoring cloud and SDN infrastructures for intrusions. In this context, the ensemble model is constructed by amalgamating K-means with DL classifiers, including the Long Short-term Memory (LSTM) network, Convolutional Neural Network (CNN), Recurrent Neural Network (RNN), Gated Recurrent Unit (GRU), and Deep Neural Network (DNN). This model is trained and tested on the CICIDS 2018 and SDN-based DDOS attack datasets, respectively. The suggested method improves on previous methods in terms of the F1 measure, precision, accuracy, and recall, all of which pertain to the efficacy of an Intrusion Detection System (IDS). With the suggested method, the accuracy and precision could be improved to 0.9685 and 0.992, respectively.

Alharbi H. et al., (2023) [17] developed a novel compression, security, and energy-aware task offloading architecture for the ECC system environment to overcome the constrained bandwidth and solve the difficulty of possible security risks. Additionally, an additional layer of protection is offered to safeguard offloaded and sensitive data from various vulnerabilities, based on an Advanced Encryption Standard (AES) cryptographic technology, which helps solve the security problem. In conclusion, simulation findings show

***Article Received:** 10 July 2023 **Revised:** 12 August 2023 **Accepted:** 28 August 2023*

that this approach is scalable and could result in a considerable decrease in energy usage compared to other benchmarks (i.e., local, edge, cloud, and additional benchmark models).

Stergiou C. et al., (2023) [18] addressed IoT-exported Cloud Computing and Big Data, emphasizing associated security and management issues. Additionally, they mix the two technologies to investigate their shared features, uncover novel angles for their integration, and finally arrive at a sustainable ecosystem they call a "Digital Twin." Next, they discuss the benefits of Cloud Computing for IoT-based Big Data, to close a research gap in this area of study. The authors show empirical findings using the encryption methods AES, RC5, and RSA to demonstrate how the suggested approach enhances the state of the art in cloud computing and IoT-based big data by providing a highly unique and scalable service platform on which to realize improved privacy and security.

Ahmad F. et al., (2022) [19] introduced a method for protecting sensitive data in the cloud that relies on machine learning. The suggested system is trained to utilize automated feature extraction utilizing the classifiers RF, Naive Bayes (NB), KNN, and SVM based on three levels of sensitivity (basic, confidential, and extremely confidential). Simulation findings validated the suggested model's predictions within a margin of error of 92%. The conclusion is that RF, NB, and KNN all outperform SVM based on these results. Guidelines for cloud service providers (like Dropbox and Google Drive) and academics are provided for this study.

Table 1 presents the authors' study strategies and provides an overview of the pertinent literature.

Table 1. Summarize the table of reviewed literature.

Authors	Years	Techniques Used	Outcomes
Kassabi H. et al., [11]	**2023**	ML	The experimental results demonstrate that clustering provides slightly better performance in terms of accuracy of 96.43%, precision of 0.94, recall of 0.99, and F1 scores of 0.96 over the one class classification with k-means outperforming other clustering algorithms.
Attou H. et al., [12]	**2023**	RF	According to the results of the experiments, the suggested model achieves a 98.3% accuracy on the NSL-KDD dataset and a 100% accuracy on the BoT-IoT dataset.
Abbas Z. et al., [13]	**2023**	SVM, ANN, and XGBoost	The outcomes indicated that the XGB model exhibited superior performance across all metrics, achieving an accuracy rate of 97.50%, a precision score of 97.60%, a recall value of 97.60%, and an F1 score of 97.50%.
Alhazmi L. et al., [14]	**2023**	DCGAN	According to the findings, the suggested DCGAN method significantly improves Accuracy, Precision, Recall, and F1-score values, with increases of 0.997, 0.988, 0.96, and 0.978, respectively.
Amitha M. et al., [15]	**2023**	RBFNN and LSTM	The suggested model achieved a 99.94% accuracy rate, a 99.94% precision rate, a 99.96% recall rate,

			and an F1 score of 99.95% on the CICDDoS2019 dataset.
Bingu R. et al., [16]	**2023**	Ensemble-based DL	The ensemble-based approach performs better than individual methods, with an increased detection rate of roughly 99.8%.
Alharbi H. et al., [17]	**2023**	AES cryptographic	Findings from the simulations show that the system is scalable and could reduce energy usage by 19%, 18%, 21%, 14.5%, 13.1%, and 12% compared to existing benchmark systems (i.e., local, edge, cloud, and models).
Stergiou C. et al., [18]	**2023**	AES, RSA, RC5	The findings indicate that through the implementation of the proposed model, the authors were able to attain superior and expedited data processing when compared to the established encryption algorithms AES, RC5, and RSA. This is primarily due to its ability to handle larger volumes of data simultaneously.
Ahmad F. et al., [19]	**2022**	RF, NB, KNN, and SVM	The testing results demonstrate that the suggested method has identified data more correctly than SVM (43% accuracy), the KNN algorithm (83% accuracy), NB (72% accuracy), and RF (92% accuracy).

3. Techniques Used

In this section, the authors used some techniques i.e., Random Forest (RF), Support Vector Machine (SVM), K-Nearest Neighbour (KNN), and Ensemble Classifier to solve the security issues in cloud computing based on these approaches.

- **Random Forest (RF)**

It is one of the ensemble approaches that is exclusively utilized to increase the success and accuracy of Machine Learning (ML) algorithms in artificial intelligence. An RF technique could also aid in identifying the relevant independent variables and allowing the system to choose functionality. Several results show its value in picking several options for each shrub in empirical study, and it turns out that this method is also the most accurate in terms of predictions [20]. The data collected from the trees is then utilized to make the most precise predictions. Decision Trees (DTs) are shown in Figure 3, where a classification model and a regression model are used to predict the values of the dependent variables [21]. A single DT has just one conclusion and a restricted range of groups, but a forest of DTs has a wider number of groups and possibilities, guaranteeing a more accurate outcome. The RF algorithm is quick, flexible, and more effective than other DT approaches despite its reliance on several different tree architectures [22]. Table 2 gives the advantages and disadvantages of RF.

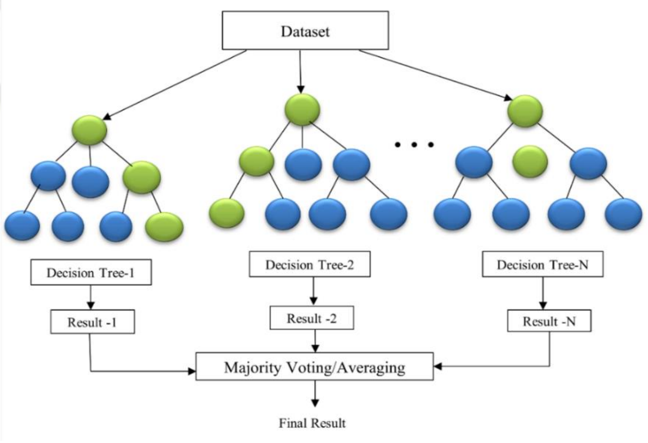


Figure 3: The structure of Random Forest [23-25].

Table 2: Advantages and disadvantages of RF [26,27]

Advantages	Disadvantages	Application
It's useful for a wide variety of tasks because it produces accurate predictions.	RF has a bias for characteristics with more levels when the data set contains categorical variables with varying numbers of levels.	It is used in image categorization and pixel analysis.
It can quantify the significance of each feature with the original training data.	If the input data contains clusters of related traits that are all equally important to the output, then the smaller clusters will be prioritized.	It is used for intricate analyses of biological data in Bioinformatics.
The training data set allows for the measurement of pairwise sample proximity.	DT building is dependent on the following factors: 1. The form of the choice being made at each node. 2. Using a certain class of predictor in each leaf. 3. The goal of optimizing each node's splitting.	It is used in the high-dimensional data process of video segmentation.

- **Support Vector Machine (SVM)**

SVMs depend primarily on structural risk reduction, as opposed to the empirical risk minimization emphasized by other NNs [28]. Vapnik first introduced this method in 1992 to fix bugs and address the issue of binary classification; nowadays, it has been expanded to include nonlinear regression as well [29]. SVMs locate the hyperplane that maximizes the margin between any two classes by mapping the data to an existing extremely high dimensional space using a specific kernel function. The solution to SVM difficulties relies heavily on outlying data points. Such factors are referred to as support vectors [30]. Figure 4 depicts the architecture of SVM.

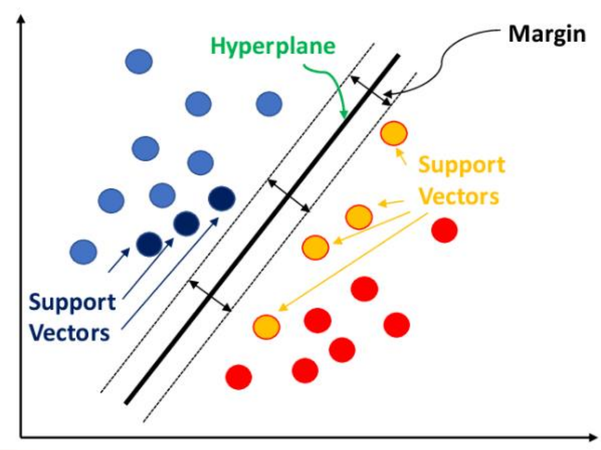


Figure 4: SVM visualization [31].

- **K-Nearest Neighbour (KNN)**

The KNN classifier seems to be a basic, easy-to-implement supervised ML technique that could potentially be used to tackle classification and regression issues. This method is used to find people who have made fraudulent credit card purchases or filed fraudulent vehicle insurance claims [32].

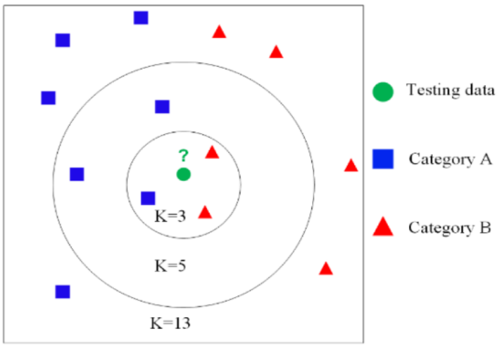


Figure 5: Schematic diagram of KNN [33,34]

The test data is represented by the green dot in Figure 5, while category A of the training set is represented by the blue squares and category B by the red triangles. If k is set to 3, there will be 2 more red triangles than blue squares (of which there are only 1). The red triangle falls under the category of green dot. However, when k approaches 5, the category of green dot shifts to the blue square because there are more blue squares than red triangles. The closest neighbor algorithm is a particular example of the k-nearest neighbor method with k = 1. The benefits and drawbacks of KNN are shown in Table 3.

Table 3: Advantages and Disadvantages of KNN

Advantages	**Disadvantages**	**Application**
Can be used with data from any distribution; therefore, they don't necessarily need to be able to be separated by a linear border.	Choosing k may be difficult.	Used in classification
Very easy to understand and use.	The test step is computationally costly.	To get data while some is missing
If there are enough samples, the categorization is quite accurate.	There is no learning phase since everything is accomplished in the testing phase.	Used in pattern recognition software
This is the exact opposite of what it desires. They can usually afford a lengthy training phase, but they want a quick test phase.		Used for the expression of genes. Used for protein-protein interaction prediction and obtaining protein 3D structures.

- **Multi-Layered Perceptron (MLP)**

MLP deals with data that cannot be linearly divided. A feed-forward ANN and the MLPs propagate the weights of each unit to the next layer. The layer might be a hidden layer, an input layer, or an output layer, depending on their needs and preferences. As an additional feature, a summing function and an activation function are available for mapping weighed inputs to weighted outputs. MLP is a FFNN that processes inputs into outputs. As depicted in Figure 6, a typical network comprises three layers: a hidden layer or layers, an output layer, and an input layer.

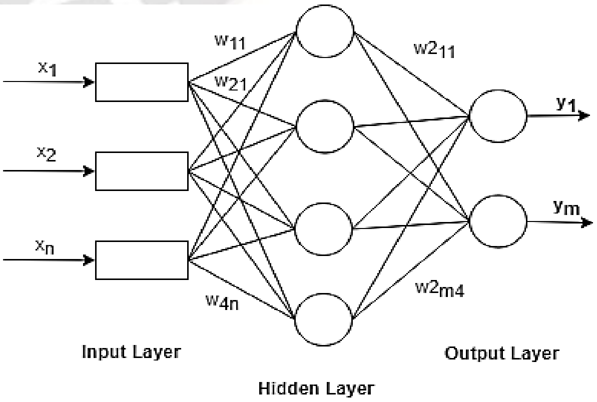


Figure 6: The structure of MLP [35].

Where x represents a node as well as the input layer, y is the output layer, and $w_{i,j}$ Represents the weight, and $i, j = 1,2,3$.

4. Proposed Methodology

The methodology encompasses data collection, evaluation, and the architecture for machine learning. This includes tasks such as model development, platform selection, data transfer, and an assessment of cloud platforms. A visual representation of the flowchart of the proposed methodology as shown in Figure 7 below.

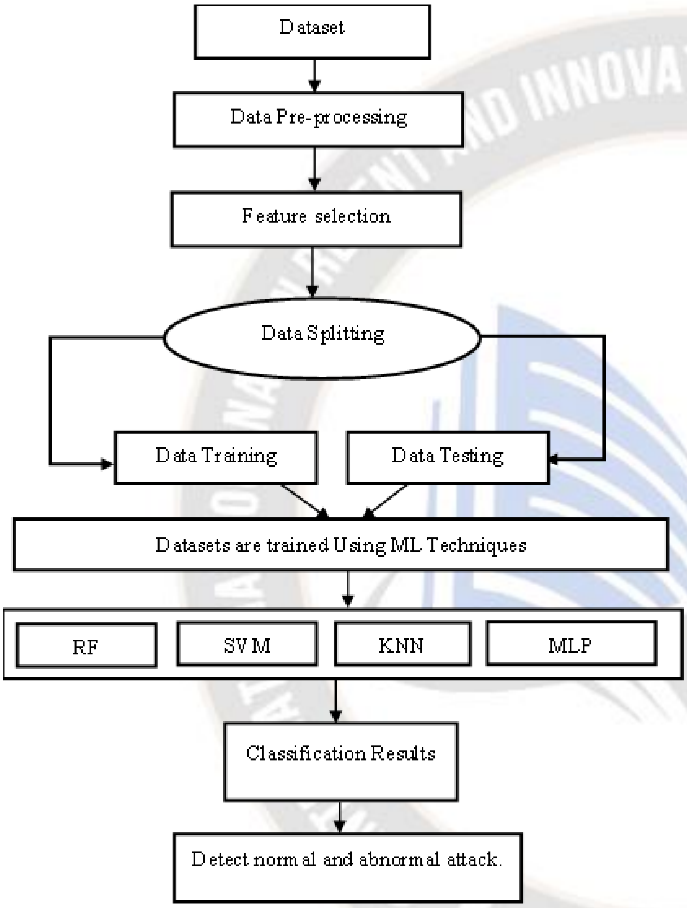


Figure 7: Proposed Methodology

The following are steps of the proposed methodology are explained in given below:

Step 1: In the first step, it involves collecting data from various sources, such as sensors, databases, and APIs.

Step 2: After that, this involves cleaning and preparing the data for analysis. This may include removing outliers, handling missing values, and normalizing the data.

Step 3: Then, selecting the most relevant features from the dataset. It could be done using techniques such as principal component analysis (PCA) and correlation analysis.

Step 4: In this step, dividing the dataset into a training set and a test set. The training set is used to train the machine learning model, while the test set is used to evaluate the model's performance.

Step 5: After that training the machine learning model on the training set. The model learns to map the input features to the output labels. This involves evaluating the model's performance on the test set. This is done by measuring the model's accuracy, precision, recall, and F1-score.

Step 6: In this step, the authors using machine learning techniques to train the datasets such as, RF, SVM, KNN, and MLP.

Step 7: This step refers to the results of the classification task. The model will output a classification for each data point.

Step 8: In the last step, this indicates to the task of detecting normal and abnormal network traffic. The model can be used to identify malicious traffic and take appropriate action.

5. Implementation and Results

This research section details the implementation using the suggested technique, and the implementation tools and dataset are provided below. The authors used the Matrix Laboratory (MATLAB) tool to obtain the results of this research. MathWorks created MATLAB, a commercial programming language and numerical computing environment supporting many examples. Matrix processes, charting of functions and information, procedure development, user interface design, and linking with other programming languages are all probable with MATLAB. The findings provided to support the suggested effort stated below are as follows.

5.1.Dataset Description

The NSL-KDD (NSL-KDD: KDD Cup 1999 dataset) is a widely used and benchmark dataset in the field of network intrusion detection and cybersecurity. It is an improved version of the original KDD Cup 1999 dataset, designed to address some of its limitations and provide a more realistic and challenging environment for evaluating intrusion detection systems. The NSL-KDD dataset consists of network traffic data captured in a controlled environment, simulating various types of attacks and normal network activities. It includes features extracted from network packets, such as connection duration, protocol types, service types, and flag information. The dataset is categorized into multiple classes of attacks, including DoS (Denial of

Service), Probe, R2L (Unauthorized access from a remote machine), and U2R (Unauthorized root access). Researchers and practitioners use the NSL-KDD dataset to develop and test intrusion detection algorithms and assess the performance of security systems in identifying and mitigating network-based threats.

5.2.Performance Evaluation

They considered metrics such as Accuracy, Precision, Recall, and F1-Score to evaluate and analyse the ML models' performances.

- **Accuracy**

The ratio of the total number of values to the sum of the true positive and true negative.

$$\text{Accuracy} = \frac{\text{TruePositive+TrueNegative}}{\text{TruePositive+TrueNegative+False Positive+FalsePositive}} \qquad (1)$$

True Positive (TP): Detected the altered images without error.

False Positive (FP): Images mistakenly recognized as genuine or manipulated.

True Negative (TN): Validated as authentic on visual inspection.

False Negative (FN): Falsely recognized manipulated images or images mistakenly thought genuine.

- **Precision**

The word precision is used to describe the unavoidable variation in measuring results. Thermal effects probably cause a random fluctuation in the observed value. It can be calculated as:

$$\frac{\text{TP}}{\text{TP+FN}} \qquad (2)$$

- **Recall**

One of the other most crucial parameters for testing an ML model is Recall. The formula for determining the Recall is:

$$\text{Recall} = \text{TP}/(\text{TP} + \text{FN}) \qquad (3)$$

- **F1-Score**

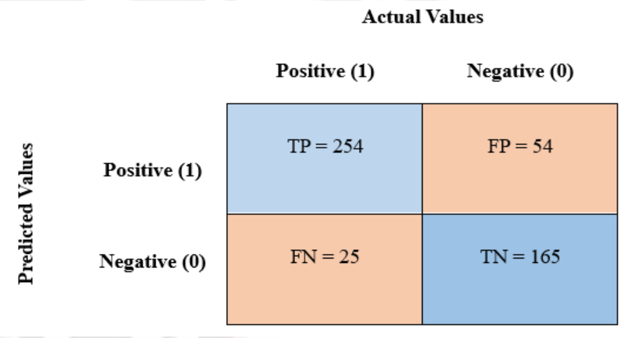
F1-score is a single metric that combines a model's precision and Recall, providing a balanced assessment of its performance in binary classification tasks.

$$\text{F1score} = \frac{2(\text{Precision}*\text{Recall})}{\text{Precision+Recall}} \qquad (4)$$

5.3.Performance Analysis

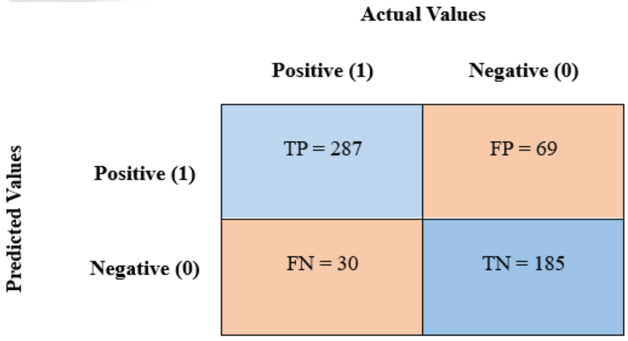
5.3.1. Testing Confusion Matrix

Figure 8 depicts the confusion metrics for evaluating the ML models are reported as TP, FP, TN, and FN. Figure 8 (a) depicts that the SVM model correctly classified 165 instances as Negative (0) and 254 instances as Positive (1). However, it misclassified 54 instances as Positive when they were Negative FP and 25 instances as Negative when they were Positive FN. This indicates that the model had a relatively higher rate of FP compared to FN. Figure 8 (b) shows for RF model the model classified 185 instances as Negative (0) correctly TN, and 287 instances as Positive (1) correctly TP. However, it misclassified 69 instances as Positive when they were Negative FP and 30 instances as Negative when they were Positive FN. The model demonstrated a good performance in correctly identifying Negative and Positive instances, but with some misclassifications. the model classified 271 instances as negative (0) correctly TN and 172 instances as positive (1) correctly TP. However, it misclassified 16 instances as positive when they were negative FP and 48 instances as negative when they were positive FN as shown in Figure 8 (c). In Figure 8 (d) It was observed that the KNN method, there are 193 instances correctly classified as positive TP and 127 correctly classified as negative TN, demonstrating the model's effectiveness. However, there are 35 instances falsely classified as positive FP and 17 instances incorrectly labelled as negative FN, indicating areas for potential improvement in the model's accuracy, especially in reducing FP predictions.



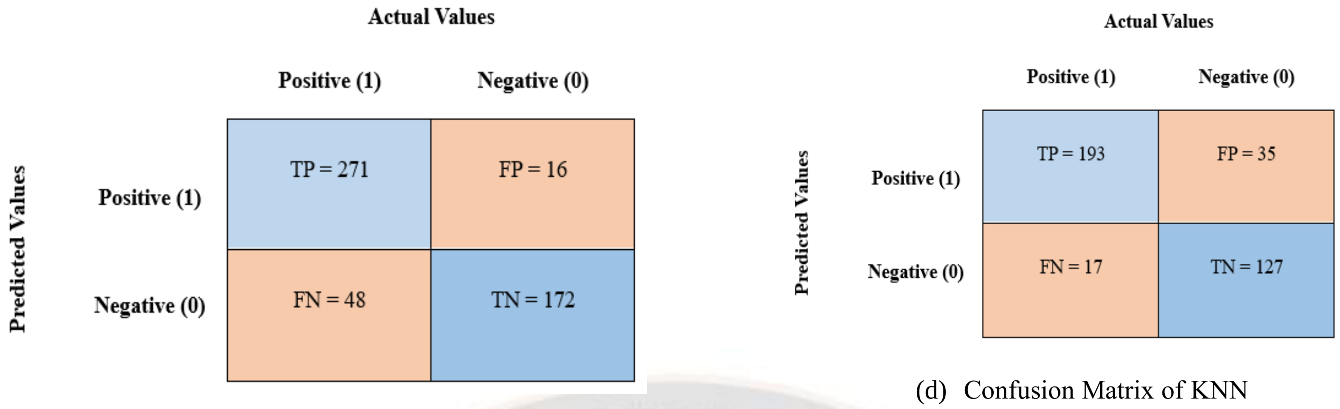
		Actual Values	
		Positive (1)	Negative (0)
Predicted Values	Positive (1)	TP = 254	FP = 54
	Negative (0)	FN = 25	TN = 165

(a) Confusion Matrix of SVM



		Actual Values	
		Positive (1)	Negative (0)
Predicted Values	Positive (1)	TP = 287	FP = 69
	Negative (0)	FN = 30	TN = 185

(b) Confusion Matrix of RF



		Actual Values	
		Positive (1)	Negative (0)
Predicted Values	Positive (1)	TP = 271	FP = 16
	Negative (0)	FN = 48	TN = 172

(c) Confusion Matrix of MLP

		Actual Values	
		Positive (1)	Negative (0)
Predicted Values	Positive (1)	TP = 193	FP = 35
	Negative (0)	FN = 17	TN = 127

(d) Confusion Matrix of KNN

Figure 8: The suggested model's Confusion Matrix, which shows the TP, TN, FP, and FN ratio in the validation dataset of the (a) CNN, (b) RF, and (c) MLP, (d) KNN.

5.3.2. Classification of Model Performance

Tables 4, 5, 6, and 7 describe the Precision, Recall, and F-measure of the four classes generated with SVM, RF, MLP and KNN models, respectively.

Table 4. Precision, Recall, and F-measure of the SVM model

Models	**Precision**	**Recall**	**F1-Score**
Normal	82.4%	91%	93.5%
Attacks	99.7%	89.5%	92.6%
Weighted Average	98	98	98
Accuracy			93.5%

Table 5. Precision, Recall, and F-Measure of RF

Models	**Precision**	**Recall**	**F1-Score**
Normal	80.5%	90.4%	82.6%
Attacks	79.8%	89.9%	81.5%
Weighted Average	98	98	98
Accuracy			82%

Table 6. Precision, Recall, and F-Measure of MLP

Models	**Precision**	**Recall**	**F1-Score**
Normal	94.5%	84.9%	89.5%
Attacks	93.6%	83.2%	88%
Weighted Average	98	98	98

***Article Received:** 10 July 2023 **Revised:** 12 August 2023 **Accepted:** 28 August 2023*

Accuracy			87.4%

Table 7. Precision, Recall, and F-Measure of KNN

Models	**Precision**	**Recall**	**F1-Score**
Normal	84.6%	91.9%	88.1%
Attacks	83.5%	90.7%	87.8%
Weighted Average	98	98	98
Accuracy			86%

5.3.3. Comparison Analysis

Table 8 illustrates a comparative analysis between the previous methods, revealing that the proposed method outperforms the others with an accuracy of 93.5% when evaluated on the NSL-KDD dataset. Figure 9 displays a comparative graph featuring state-of-the-art methods.

Table 8. Performance comparison of the various methods.

Authors [Reference]	**Datasets**	**Models**	**Accuracy**
Mohammad A. et al., (2021) [36]	-	Machine Learning-Assisted Cloud Computing Model (ML-CCM)	91.7%
Kushwah G. et al., (2021) [37]	NSL-KDD	Self-Adaptive Evolutionary Extreme Learning Machine (SaE-ELM)	86.80%
Kushwah G. et al., (2020) [38]	ISCX	Voting Extreme Learning Machine (V-ELM)	92.11%
Our Work	NSL-KDD	Proposed	93.5%

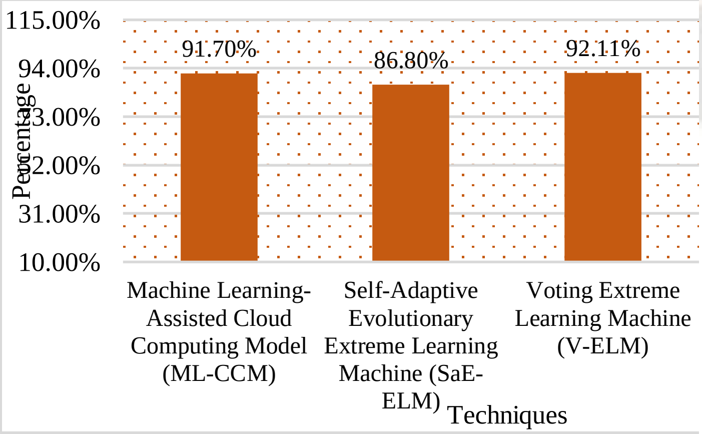


Figure 9: Comparison graph of the previous works

6. Conclusion and Future Scope

The use of cloud computing, a relatively new and internet-based technology, has become more common in today's society, particularly in the computing and IT sectors. Cloud computing, which makes use of a network to provide access to a shared pool of resources, is rapidly growing in popularity because of its low cost, high availability, and high productivity. Cloud computing has many advantages, but it also introduces new problems with data privacy, data protection, authorized access, etc. These challenges are making widespread use of cloud computing more challenging in the present day. This paper illustrates the Identification of Security Issues and Finding their Solution in Cloud Computing using ML approaches such as SVM, RF, KNN, and MLP. The validation was performed using the recently produced and frequently used benchmarking dataset NSL-KDD. Next, they check these datasets for their F1-score, accuracy, precision, and recall. By comparing the obtained values to those of contemporary security models, it is shown that the proposed system is successful at detection. The results of the performance study show that the accuracy of 93.5% with the KNN method outperforms the other security models. In future endeavours, they would direct the attention to this aspect, leveraging DL and Ensemble Learning techniques to enhance the performance of the model.

References

1. Cloud Computing Definition. 2011; Available from: https://www.nist.gov/newsevents/news/2011/10/final-version-nist-cloudcomputing-definition-published.
2. Sabir, Sabiyyah. "Security issues in cloud computing and their solutions: a review." International Journal of Advanced Computer Science and Applications 9, no. 11 (2018).
3. El Kafhali, Said, Iman El Mir, and Mohamed Hanini. "Security threats, defense mechanisms, challenges, and future directions in cloud computing." Archives of Computational Methods in Engineering 29, no. 1 (2022): 223-246.
4. Architecture for Managing Clouds White Paper (DSP-IS0102), http://www.dmtf.org/sites/default/files/standards/documents/DSP -IS0102_1.0.0.pdf.
5. Special Publication 800-53, Recommended Security Controls for Federal Information Systems, (2006) December.
6. Special Publication 800-125, "Guide to Security for Full Virtualization Technologies".
7. https://www.ijert.org/research/an-empirical-study-on-security-issues-in-cloud-computing-environments-IJERTCONV9IS05096.pdf.
8. G. B. Prasanth SP, "AES and DES Using Secure and Dynamic Data Storage in Cloud," IJCSMC, vol. III, no. 1, p. 401 – 407, 2014.
9. M. S. S. D. Miss. Shakeeba S. Khan, "Security in Cloud Computing Using Cryptographic Algorithms," International Journal of Computer Science and Mobile Computing, vol. III, no. 9, pp. 517-525, 2014.
10. Sabir, Sabiyyah. "Security issues in cloud computing and their solutions: a review." International Journal of Advanced Computer Science and Applications 9, no. 11 (2018).
11. El-Kassabi, Hadeel T., Mohamed Adel Serhani, Mohammad M. Masud, Khaled Shuaib, and Khaled Khalil. "Deep learning approach to security enforcement in cloud workflow orchestration." Journal of Cloud Computing 12, no. 1 (2023): 10.
12. Attou, Hanaa, Azidine Guezzaz, Said Benkirane, Mourade Azrour, and Yousef Farhaoui. "Cloud-Based Intrusion Detection Approach Using Machine Learning Techniques." Big Data Mining and Analytics 6, no. 3 (2023): 311-320.
13. Abbas, Zaheer, and Seunghwan Myeong. "Enhancing Industrial Cyber Security, Focusing on Formulating a Practical Strategy for Making Predictions Through Machine Learning Tools in Cloud Computing Environment." Electronics 12, no. 12 (2023): 2650.
14. Alhazmi, Lamia. "Enhancing Cloud Security: An Optimization-based Deep Learning Model for Detecting Denial-of-Service Attacks." International Journal of Advanced Computer Science and Applications 14, no. 7 (2023).
15. Amitha, Marram, and Muktevi Srivenkatesh. "DDoS Attack Detection in Cloud Computing Using Deep Learning Algorithms." International Journal of Intelligent Systems and Applications in Engineering 11, no. 4 (2023): 82-90.
16. Bingu, Rajesh, and S. Jothilakshmi. "Design of Intrusion Detection System using Ensemble Learning Technique in Cloud Computing Environment." International Journal of Advanced Computer Science and Applications 14, no. 5 (2023).
17. Alharbi, Hatem A., Mohammad Aldossary, Jaber Almutairi, and Ibrahim A. Elgendy. "Energy-Aware and Secure Task Offloading for Multi-Tier Edge-Cloud Computing Systems." Sensors 23, no. 6 (2023): 3254.
18. Stergiou, Christos L., Elisavet Bompoli, and Konstantinos E. Psannis. "Security and Privacy Issues in IoT-Based Big Data Cloud Systems in a Digital Twin Scenario." Applied Sciences 13, no. 2 (2023): 758.
19. Ahmad, Fahad Burhan, Asif Nawaz, Tariq Ali, Azaz Ahmed Kiani, and Ghulam Mustafa. "Securing cloud data: a machine learning based data categorization approach for cloud computing." (2022).
20. Patil, Suraj, Varsha Nemade, and Piyush Kumar Soni. "Predictive modelling for credit card fraud detection

using data analytics." Procedia computer science 132 (2018): 385-395.

21. Fu, Yijie. "Combination of random forests and neural networks in social lending." Journal of Financial Risk Management 6, no. 4 (2017): 418-426.
22. Rodriguez-Galiano, Victor Francisco, Bardan Ghimire, John Rogan, Mario Chica-Olmo, and Juan Pedro Rigol-Sanchez. "An assessment of the effectiveness of a random forest classifier for land-cover classification." ISPRS journal of photogrammetry and remote sensing 67 (2012): 93-104.
23. Mbaabu, Onesmus. "Introduction to random forest in machine learning." Engineering Education (EngEd) Program| Section (2020).
24. Dimitriadis, Stavros I., Dimitris Liparas, and Alzheimer's Disease Neuroimaging Initiative. "How random is the random forest? Random forest algorithm on the service of structural imaging biomarkers for Alzheimer's disease: from Alzheimer's disease neuroimaging initiative (ADNI) database." Neural regeneration research 13, no. 6 (2018): 962.
25. Çavuşoğlu, Ünal. "A new hybrid approach for intrusion detection using machine learning methods." Applied Intelligence 49 (2019): 2735-2761.
26. Learn Random Forest Using Excel," 27 12 2017. [Online]. Available: https://www.newtechdojo.com/learnrandom-forest-using-excel/. (12/8/2019)
27. Akkaya, Berke, and Nurdan Çolakoğlu. "Comparison of multi-class classification algorithms on early diagnosis of heart diseases." (2019).
28. Li, Xurui, Wei Yu, Tianyu Luwang, Jianbin Zheng, Xuetao Qiu, Jintao Zhao, Lei Xia, and Yujiao Li. "Transaction fraud detection using gru-centered sandwich-structured model." In 2018 IEEE 22nd International Conference on Computer Supported Cooperative Work in Design ((CSCWD)), pp. 467-472. IEEE, 2018.
29. Abakarim, Youness, Mohamed Lahby, and Abdelbaki Attioui. "An efficient real time model for credit card fraud detection based on deep learning." In Proceedings of the 12th international conference on intelligent systems: theories and applications, pp. 1-7. 2018.
30. Madhurya, M. J., H. L. Gururaj, B. C. Soundarya, K. P. Vidyashree, and A. B. Rajendra. "Exploratory analysis of credit card fraud detection using machine learning techniques." Global Transitions Proceedings 3, no. 1 (2022): 31-37.
31. Manjrekar, Onkar N., and Milorad P. Dudukovic. "Identification of flow regime in a bubble column reactor with a combination of optical probe data and machine learning technique." Chemical Engineering Science: X 2 (2019): 100023.
32. Wei, Jinlong, Lilin Yi, Elias Giacoumidis, Qixiang Cheng, and Alan Pak Tao Lau. "Special Issue on “Optics for AI and AI for Optics”." Applied Sciences 10, no. 9 (2020): 3262.
33. Zhang, Wei, Xiaohui Chen, Yueqi Liu, and Qian Xi. "A distributed storage and computation k-nearest neighbor algorithm-based cloud-edge computing for cyber-physical-social systems." IEEE Access 8 (2020): 50118-50130.
34. Zhang, Qianwu, Hai Zhou, Yuntong Jiang, Bingyao Cao, Yingchun Li, Yingxiong Song, Jian Chen, Junjie Zhang, and Min Wang. "A simple joint modulation format identification and OSNR monitoring scheme for IMDD OOFDM transceivers using K-nearest neighbor algorithm." Applied Sciences 9, no. 18 (2019): 3892.
35. Alboaneen, Dabiah Ahmed, Huaglory Tianfield, and Yan Zhang. "Glowworm swarm optimisation for training multi-layer perceptrons." In Proceedings of the Fourth IEEE/ACM International Conference on Big Data Computing, Applications and Technologies, pp. 131-138. 2017.
36. Mohammad, Abdul Salam, and Manas Ranjan Pradhan. "Machine learning with big data analytics for cloud security." Computers & Electrical Engineering 96 (2021): 107527.
37. Kushwah, Gopal Singh, and Virender Ranga. "Optimized extreme learning machine for detecting DDoS attacks in cloud computing." Computers & Security 105 (2021): 102260.
38. Kushwah, Gopal Singh, and Virender Ranga. "Voting extreme learning machine based distributed denial of service attack detection in cloud computing." Journal of Information Security and Applications 53 (2020): 102532.