Spectrum of Engineering Sciences
ISSN (e) 3007-3138 (p) 3007-312X Volume 4, Issue 4, 2026

EXPLAINABLE AI-BASED ANOMALY DETECTION FOR CYBERSECURITY IN CLOUD COMPUTING ENVIRONMENTS

***[1]Sohail Basheer, [2]Imran Hayat, [3]Asif Ali Leghari, [4]Zuhaib Phul, [5]Syed Kashif Ali Quadri**

**[1]Department of Computer Science, Sindh Madressatul Islam University, Karachi Pakistan*
[2]Federal Government Educational Institutions (C/G) Rawalpindi Pakistan
[3]Department of Computer Science, Sindh Madressatul Islam University, Karachi Pakistan
[4]Department of Computer Science, Sindh Madressatul Islam University, Karachi Pakistan
[5]Department of Computer Science, Sindh Madressatul Islam University, Karachi Pakistan
**[1]sohail1988@gmail.com, [2]imranhayat974@gmail.com, [3]asifalileghari@smiu.edu.pk, [4]zohaibphul53@gmail.com, [5]syedkashifaliofficial@gmail.com*

DOI:- https://doi.org/10.5281/zenodo.20103419

Keywords
Cyber Security, Cloud Computing, Machine Learning, Phishing Detection & Prevention

Article History
Received: 06 Jan 2026
Accepted: 04 Mar 2026
Published: 06 Apr2026

Corresponding Author: *
Sohail Basheer

Abstract
An increase in dependence on technology and internet based service providers has led to growth of cloud computing infrastructure across multiple sectors. Cloud-based systems have grown in popularity as organizations migrate their applications and data to cloud platforms because they offer scalable and flexible solutions. In addition, migration to cloud platforms has created additional risk areas for cyber attackers to exploit. A research study will propose the use of Explainable Artificial Intelligence (XAI) based anomaly detection system in order to improve cybersecurity in cloud computing. Machine Learning (ML), specifically traditional approaches, operate like "black box" models. Black box models limit ML's application in security critical systems that require interpretable information. The proposed method utilizes explainability techniques (SHAP & LIME) along with Deep Learning Models to provide interpretability. Results from experiments utilizing benchmark datasets (CICIDS2017) indicate that the proposed method can identify anomalous network behavior effectively while producing transparent decisions. These results show that using XAI improves trust, interpretability, and overall performance in cloud-based IDS systems.

1. Introduction

Over the last few years many different types of intrusion detection systems (IDS) have been created as a way to improve the strength of a computer networks' defense against malicious attacks. Although the CICIDS2017 dataset,

which was produced by the Canadian Institute of Cybersecurity, has a wealth of new data about how attackers are using modern techniques with much richer features than before and therefore is commonly used to evaluate IDSs; however, some major flaws have appeared and could produce bias into an IDS's ability to reliably detect threats. If an IDS cannot generalize well, it will be unable to accurately or consistently perform at the level of intrusion detection needed [1]. The interest in anomaly detection is rising as a way to address many of the shortcomings with signature-based intrusion detection. One such example of this type of data is the KDD Cup '99 dataset which has become widely used for testing purposes. However, there are significant issues (both statistically redundant and biased) contained within the KDD Cup '99 dataset which can significantly limit its use as an effective test of how well various types of anomaly detection techniques perform. In order to provide a more stable and representative platform on which to evaluate both traditional and innovative approaches to anomaly detection, the NSL-KDD dataset was developed [2]. As a result of the growing number of sophisticated cyber threats against networks, there is increased interest in using Machine Learning (ML) technologies to protect organizational IT assets. Recent advances in AI technology have provided an opportunity for Deep Learning (DL) to be used effectively to enhance the capabilities of Intrusion Detection Systems (IDS). DL technologies are being employed by researchers to develop new methods that can significantly increase the accuracy of traditional intrusion detection techniques. This includes employing advanced neural network architectures to detect sophisticated patterns of malicious activity in network traffic. An overview of these emerging areas provides insight into what is currently working with respect to enhancing the capability of IDS and what opportunities exist for future research and development [3]. The use of machine learning has become widespread in intrusion detection systems (IDS); nonetheless, the two most serious problems associated with IDS are its time complexity and the problem of classifying data. The results from a number of studies using a systematic review methodology show that four of the most popular types of models used in IDS include Random Forest, Support Vector Machines, Decision Trees, and K-Nearest Neighbor. All these models are evaluated based upon their ability to perform well by measuring them against various different metrics, including accuracy, precision, recall and F1-Score. Benchmark datasets used to train and evaluate IDS include several datasets like KDD-Cup '99, NSL-KDD and CICIDS2017. While each type of IDS provides excellent detection capability there continues to be a couple of very serious factors affecting an IDS's total performance – specifically data imbalances and the effect of large dimensional data [4]. Trust is at the center of cloud computing's ability to offer scalable and flexible services. Although there is growing interest in this area, trust remains one of the greatest barriers to cloud computing's widespread adoption. Many have reported various security-related issues that arise when using cloud-based services; however, many of these issues are being addressed through partial measures such as implementing

some level of trust based on how the user chooses to implement their cloud environment [5]. Understanding why a prediction was made is important in all applications where we want both strong performance from our model and strong understanding of that performance. While complex models like ensemble methods or deep learning can give us stronger predictive results than simpler models, they are less interpretable. Therefore, there is an ongoing conflict between having a highly accurate model and being able to understand how it makes decisions. As one method to help balance these competing objectives, SHAP (SHapley Additive exPlanations) has provided a unifying methodology to explain model predictions through feature attribution [6]. There has been rapid development of artificial intelligence over the past few years which provides numerous potential uses in many areas. However, the primary challenge with complex machine learning models (such as those using ensemble techniques or deep learning) is that they are often difficult to understand due to the lack of explanation. Explainable Artificial Intelligence was developed to overcome these challenges by increasing both the transparency and the interpretability of complex machine learning models. In order to create a common understanding of how explainability can be defined and categorized when applied to various types of machine learning models there have been a number of different approaches developed. This work supports the development of responsible and trustworthy artificial intelligence for use in real world applications [7]. In an effort to be able to keep up with all that is being done in Explainable AI; due to the speed at which new methods are emerging; the need for taxonomy-based reviews (which would provide a structured overview) of the different types of explainability techniques (XAI), has become apparent. In addition to the use of taxonomy-based reviews, multiple methodologies have also developed for creating such taxonomies. Each methodology has its own advantages as well as disadvantages. These differences point out some of the problems that exist in establishing a single way of viewing or representing the entire field of Explainable AI. As such, comprehensive frameworks and systematic methods will help to better understand how to effectively apply these XAI techniques [8]. With the rapid increase in use of the Internet, there is a growing need for securing user data as it relates to modern networks. To help protect against malicious activity on computer networks an intrusion detection system (IDS) may be used. However, IDS's often have problems with false alarms and false negatives. False alarms and false negatives can cause disruptions to legitimate users and/or allow unauthorized access into the systems which will affect overall system reliability. Real time detection systems using machine learning techniques can improve detection rates for identifying malicious activity by examining more complex patterns of network traffic than traditional IDS. Additionally, distributed and parallel processing techniques can also enhance performance of these systems allowing them to detect malicious activity much quicker and with greater accuracy [9]. Anomaly-based intrusion detection model using a decision tree approach on the cic ids 2017 dataset was developed. The categorical features were encoded and relevant

attributes selected to improve performance of model using RFE (Recursive Feature Elimination). Training and testing sets were created from the data set where network traffic was classified as malicious or benign. Results of experiments indicated high accuracy with strong true positive rates and low false positives. Improved effectiveness over traditional datasets was demonstrated by this methodology; big data analytics can be used for enhanced detection [10].

2. Literature Review

The use of deep learning in Cybersecurity has greatly increased the ability to detect sophisticated threats including malware and vulnerabilities. However, many of these models are "black boxes" and do not provide insight into how they reached a particular conclusion; this has been identified by researchers as an opportunity for Explainable Artificial Intelligence (XAI) to increase the transparency and reliability of decision making based on AI. As it relates to Cybersecurity, the integration of XAI will help analysts better interpret alerts generated by AI-based tools and reduce false positive alarm [11]. The increasing complexity of cyber threats is driving the growth of new generation intrusion detection systems that provide a means of continuous monitoring of network security. Traditional Machine Learning based Intrusion Detection Systems (IDS) have shown very good performance levels regarding accuracy; however, they are lacking in providing transparent reasoning as to why certain intrusions were detected. Explainable AI can be applied to traditional IDS's to make them more understandable, to improve user's confidence in the decision-making process, and to optimize the detection capabilities of the system. In recent years several research studies have indicated the necessity of developing models that are both explainable and scalable in order to increase dependability in Cyber Security Applications [12]. XAI has become increasingly important in the area of cybersecurity because of its reliance on a large number of machine learning (ML) and deep learning (DL) based threat detection systems. ML and DL can be very accurate compared to many legacy systems used in threat detection. However, their inability to provide transparent explanations results in an inability for users to rely upon them and limits their ability to make informed decisions during a cyber event. In addition to enhancing the understanding of AI based security systems by providing transparent explanations, integrating XAI into security systems also enhance the reliability of such systems [13]. There is a significant concern about the increasing number of AI's used within highly sensitive and important areas such as Cyber Security that are currently opaque. Therefore, Explanatory Artificial Intelligence (XAI) was developed to provide AI with an ability to produce intelligible or interpretable results. In addition to improving threat analysis in Cyber Security, Explanatory AI will also improve defensive system designs. A variety of approaches have been explored to increase the ability to understand and analyze results generated by AI systems while still maintaining detection capabilities. Research today indicates there is a growing need for more efficient and scalable methods of achieving Explanatory AI in secure systems [14]. Supervised Machine Learning (SML) has become one of the most

used tools in classifying system operating states. SML uses simulated systems to create a model that will be able to classify system conditions as either stable or unstable. The major challenge associated with this type of model generation is to ensure that the decision-making process is interpretable. This challenge can be even greater when applying SML to complex and safety-critical systems. An example of a tool that could help address this problem is decision trees which provide clear, easy-to-understand classification rules. There are recent developments that aim to find a balance between interpretability and predictive capability so that these types of models are useful and efficient [15]. Artificial intelligence (AI) as an area of research has grown rapidly and significantly influenced contemporary technological advancements. However, issues related to unpredictable behavior and transparency are emerging. Because of their "black box" nature, there exists limited comprehension with respect to many current artificial intelligence systems; thus, they create significant barriers to both trust and the eventual use of such systems. Trustworthy AI addresses the need to improve the dependability, robustness, and explainability of intelligent systems. Thus, it is critical that trustworthy properties be developed so as to ensure the security and dependability of future applications based upon AI [16]. Deep Learning Anomaly Detection Methods are used for detecting abnormality in Cloud Environments through System & Network Metrics. Traditional Techniques have lower Detection Performance and higher False Alarm Rates than Deep Learning Based Approaches. In order to effectively address the Security Concerns associated with the Increased Data Processing and Storage due to increasing use of Cloud Services, it is necessary to develop Effective Mechanisms for Detecting Abnormal Behavior as well as Maintaining System Reliability [17]. Deep learning has grown as an area for enhancing cyber security in IoT. Techniques based on this area are being used to improve the ability of intrusion detection systems within IoT. Anomalies can be defined as behaviors or patterns that differ from what is expected (i.e., normal). Anomaly-based intrusion detection systems identify anomalies as potential threats by comparing network traffic against baseline data. While anomaly-based intrusion detection may identify some types of unknown or zero day attacks better than signature-based intrusion detection, it also produces false positives [18]. Software-defined networking (SDN) has introduced a number of significant security risks due to a centralization of network management. Centralized architectures for SDNs introduce vulnerabilities that are significant enough to disrupt the reliability and availability of networks. While several types of attacks (i.e., denial-of-service [DDoS], unauthorized access) can be detrimental to a network's operation; a variety of machine-learning based solutions have been proposed to detect and classify different types of malicious traffic and improve detection [19]. Cyber-attacks on important infrastructure systems have become one of the largest challenges in contemporary cybersecurity. The use of intrusion detection systems (IDS) is wide-spread for identifying and preventing abnormal events from occurring throughout various types of

sensitive infrastructures. This paper proposes an autoencoder-based intrusion detection model utilizing the UNSW-NB15 data set that contains complex attack scenarios; training/testing were used to categorize network traffic as either "normal" or "attack" to perform anomaly detection. The experimental results support that the proposed model provides high-quality detection of attacks and effectively identifies threats to security [20]. Advances in technology to use web-based networks have increased our dependency upon these systems. Therefore, there is a growing need to provide reliable network protection. As a result of its capabilities to learn from time-series data, deep-learning-based intrusion detection systems are being used more often. In particular, the Long Short-Term Memory (LSTM) model has shown promise to detect intrusions within a network based upon long-term trends in network activity. Improved performance was demonstrated by using a hybrid model which combined Random Forest as the feature selector for the data with the LSTM-based model. This resulted in improvements over previously published approaches and improved performance relative to other recently developed approaches. For example, this model provided very good testing accuracy of 99.66% with a very low loss of 0.12%. [21]. Cloud computing is now an increasingly common choice for many organizations because of its flexibility in delivering services at low costs. While this offers benefits, it also presents new opportunities for security breaches and loss of private information in a distributed network. The ability to detect intrusions or malicious activity (such as unauthorized access) by means of intrusion detection systems is critical to protecting cloud-based networks. This paper will compare how existing cloud-based detection methods have been deployed with respect to attack type detection mechanism or methodology (i.e., what type of attacks can be detected), as well as the strategy they were designed to employ. It emphasizes multi-detection methodologies which can provide better protection against various cyber threats [22]. Rapid expansion in Cloud Computing increases the operational efficiency of infrastructure while at the same time creates numerous potential security risks. As a result of these new security challenges, intrusion detection systems (IDS) have become a popular tool for protecting cloud based applications from all manner of cyber-attacks. Previous research has identified IDS methods as being either Hypervisor-Based IDS, Network-Based IDS, Machine Learning-Based IDS or Hybrid IDS. A comparative review of existing research provides evidence that common evaluation criteria for IDS include Performance Criteria which include Accuracy, Response Time and Overhead. Additionally, there is still much room for improvement when it comes to the use of IDS regarding Cost Sensitive and Attack Tolerant criteria [23]. The rapid growth of Cloud Computing is attributed to the advancement in Information & Communication Technologies (ICT) as well as the cost-efficient Service Model offered. Nevertheless, out-sourcing your Data and Applications generates numerous Security and Privacy issues which may impede the rate of adoption. Several Studies have proposed a

number of Solutions; however, most are inflexible with respect to offering solutions for multiple and continuously changing threats. A number of other methods identify problems but do not provide adequate Mitigation Strategies or provide no explanation of underlying Security Risks. Therefore, Reliable and Resilient Cloud Environments require adaptive and complete Security Mechanisms [24]. Network Traffic Monitoring and Suspicious Activity Identification within Cloud Environments is accomplished using Intrusion Detection Systems. Traditional methods of intrusion detection do not have the capability to detect new types of intrusions as well as handle the complexities associated with a virtualized environment. This is improved through use of deep-learning algorithms that allow for automatic extraction of meaningful features in network data. Combining Sparse Auto-Encoder, Stacked Contractive Auto-Encoder and Bi-LSTM with Attention into a Hybrid Model improves the classification. The proposed system provides an impressive precision rate of 99% as well as an impressive Recall rate of 98% and Accuracy greater than 98%. Thus, it demonstrates its effectiveness in performing intrusion detection [25]. High dimensionality of features used in intrusion detection systems may lead to increased complexity during computation and slow down the time needed to classify attacks. By applying principal component analysis (PCA), it is possible to decrease the number of features from 81 to 10 while maintaining important information in the feature space; therefore, PCA reduces the amount of computation required when evaluating different types of attacks. A variety of classifiers are evaluated in this study including a random forest classifier, a bayesian network classifier, linear discriminant analysis (LDA) classifier, and quadratic discriminant analysis (QDA) classifier. To balance the classes within the CICIDS 2017 dataset, we have developed a technique that addresses the problem of class imbalance. Using our proposed method allows us to achieve a high level of accuracy (approximately 99.6%) as well as an improvement in detection rates and an improvement in the false alarm rate [26]. The development of network intrusion detection systems is crucial in order to detect unauthorized activity and provide time-sensitive security responses. Organizations face significant challenges when attempting to develop a quality detection model due to limited access to large quantities of high-quality attack data. This challenge can be addressed through the use of federated learning which enables the collaborative training of a model with all parties maintaining their respective data privacy. DAFL (Dynamic Weighted Aggregation Federated Learning) enhances detection performance by implementing an adaptive strategy for both filtering and weighting communications. The experimental results show that DAFL significantly improved the accuracy of detection while also reducing the amount of communication required and improving privacy protection [27]. As the majority of systems become increasingly dependent upon data networks, so too will the importance of network security increase. Network security is of paramount importance as an increasing number of threats exist to digitally protected infrastructure. As such, intrusion

detection systems (IDS) play an important role in ensuring system security. Traditionally, IDS approaches have utilized signature based methods to identify malicious traffic. However, the incorporation of machine learning into IDS systems has enabled a more effective method of detecting intrusions. Although current machine learning based IDS systems can detect many types of intrusion attempts they are still susceptible to adversarial attacks. This highlights the necessity to develop and deploy secure and reliable detection schemes for intrusion detection systems [28]. Advances in the area of cybersecurity have generated an increased interest in graph based technologies that can be used to identify complex cyber threats. Technologies such as Graph Neural Networks (GNN), Behavioural Graphs and Botnet Detection Models are examples of how graph based technology can be effectively utilized to identify anomalies in network communication. These types of technologies look at the relationships between nodes and edges in a graph to determine when there is a deviation from expected behaviour. As a result, Anomaly Detection using Graph Technology provides high levels of accuracy in identifying malicious activity such as Malware and Network Attacks. Additionally these systems improve overall security analysis capabilities with respect to workloads and reduce unnecessary alerts and false positives via efficient alert filtering [29]. IoT is becoming increasingly large and that is making it difficult for Distributed IDS (Intrusion Detection Systems) to address the issues of scalability, security, and privacy. Federated Learning methods are being developed as a way to train models collaboratively, while keeping users' data private. The proposed F-NIDS method combines Federated AI with Asynchronous Communication to facilitate scalable implementations on Cloud and Fog Environments. Additionally, Differential Privacy techniques such as adding Gaussian Noise can protect users' data from Inference Attacks [30]. Cloud Computing offers scalable and effective cloud services; however it creates a variety of Cybersecurity issues that include Data Breaches, Insider Threats, Advanced Persistent Threats. The traditional security measures for many organizations fail to address the ever changing cyber threats in a sufficient manner which has led to the use of Artificial Intelligence (AI) technologies. There are four machine learning techniques used to improve the ability to detect and prevent cyber threats: Supervised Learning Models can be trained to recognize specific types of threats by labeling them. Unsupervised Learning Models find patterns or "anomalies" within a large amount of data. Reinforcement Learning is a technique where a model learns based on trial and error. Hybrid models combine two or more machine learning techniques together. Together these different machine learning techniques have been shown to greatly increase an organization's ability to identify and block both known and unknown cyber threats. Although they offer tremendous benefits, there are still some challenges that need to be addressed before machine learning will become widely accepted. Challenges associated with machine learning models include: Data Privacy; Adversary Attacks on machine learning models; Limited Interpretation of results [31]. The wide spread

use of cloud computing has made better use of data management, while at the same time creating a multitude of new security issues with respect to anomaly detection. The traditional approaches have generally failed because they cannot keep up with the constantly changing and highly dynamic characteristics of modern cloud based systems. AI technologies (machine learning and deep learning) offer an alternative method for detecting anomalous behavior that is both very complex and unknown. Supervised, unsupervised and combinations of different models improve detection capabilities and overall system performance. Although this provides many benefits there are still several major concerns regarding data quality, scalability and computational cost [32].

3. Research Gap

Existing solutions to improve detection accuracy of intrusion detection and cloud security models are generally limited to improving detection accuracy without much attention towards the interpretability and transparency of a model. Most of today's anomaly detection models do not provide sufficient insight into how the underlying machine learning or deep learning algorithm makes its classification decision in an anomaly detection system. Many anomaly detection studies have not included the integration of explainable AI (XAI) methods that will allow users to establish trust and understand the reasoning behind their results. In addition to these general challenges to anomaly detection there exist other specific challenges in detecting anomalies in cloud environments including but not limited to, data imbalances, scalability challenges as well as the need to implement high speed, real time anomaly detection mechanisms.

4. Research Objective

The purpose of this study was to create a robust (efficient) yet explainable anomaly detection system that can be used for cybersecurity in the Cloud. It applied Machine Learning algorithms to achieve high-quality Threat Detection incorporated Explainable AI to increase the models' Interpretability, and applied Standard Metrics including Accuracy, Precision, Recall, and F1-Score for Assessing the Reliability of Security Assessment

5. Methodology

Despite advancements made to intrusion detection and cloud security technologies, there have been very few attempts to make machine learning and deep learning technology more interpretable. Most current intrusion detection and cloud security technologies use non-transparent "black box" type machine learning and deep learning models that do not allow end-users or developers to understand the reasoning behind decisions made by these models. The lack of transparency in AI based technologies contributes to reduced trust and reliability in many cybersecurity products. Current anomaly detection research has provided minimal methods of integrating Explainable Artificial Intelligence (XAI) into their methodologies. Therefore, understanding how an anomaly was classified continues to be one of the most challenging areas of anomaly detection. Additionally, detecting anomalies in cloud-based environments can present two major challenges: data imbalance and scalability. Efficient and adaptable anomaly

detection frameworks will need to be developed to handle the scalability of large-scale cloud environments. Finally, real-time anomaly detection remains an area where current anomaly detection methodologies remain insufficient. These shortcomings continue to limit the practical application of current cybersecurity models within highly dynamic cloud environments. Thus, more robust, scalable, and explainable anomaly detection techniques will be necessary to create reliable cloud computing systems as shown in Fig 1.

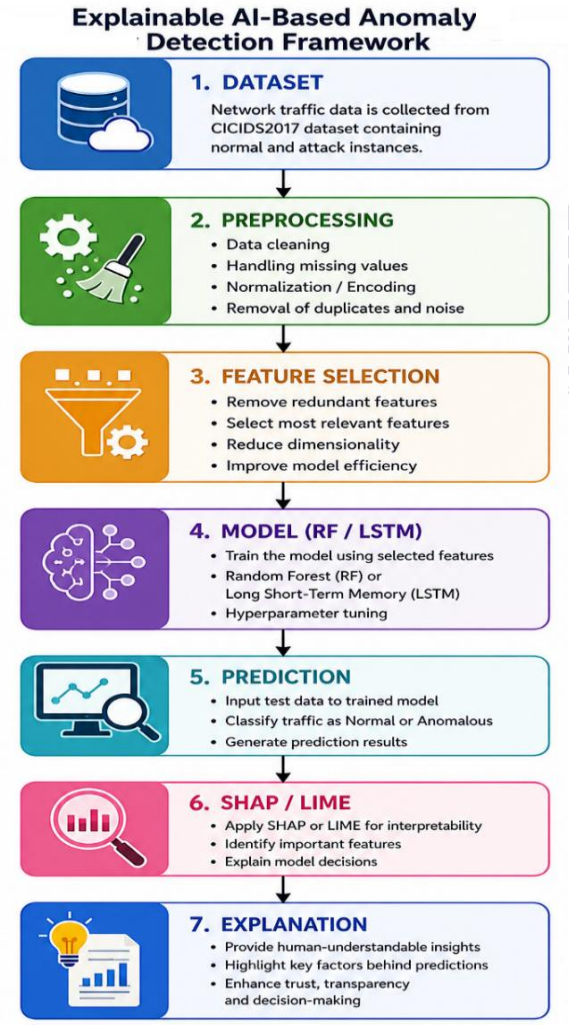

Figure 1

a. Experimental Setup

This area discusses the specifics of how the authors implemented their proposed anomaly detection technique as well as the experimental settings they utilized to test it. Specifically, this section provides information about the computing environment used by the authors (including tools), data set characteristics, pre-processing that was done to the data prior to using it for testing (and/or training) and other items the authors felt would be important for providing reliable and replicable results. The authors' experimental design has been intended to provide an accurate representation of an intrusion detection scenario within a cloud computing environment with all aspects being consistent from the training and testing phases.

b. Experimental Environment and Tools

The overall research setting used in the current study is presented in Table 1. Due to the wide variety of support that it offers both for machine learning and for analyzing data, the programming language Python was chosen to implement the methods. A Jupyter notebook was used as the development tool to allow for easy testing and visualizing different combinations of parameters. For developing traditional models (like random forests), libraries such as scikit-learn were used; for developing deep models (such as LSTMs) that use neural networks, either TensorFlow or Keras was selected. To perform numeric computations and manipulate the data, pandas and numpy were employed. Visualizations of model performance, including the generation of graphs, were performed by matplotlib and seaborn. In order to be able to train and test

models efficiently, the system had sufficient computing capabilities to run these types of experiments.

Component	Description
Programming Language	Python
Environment	Jupyter Notebook
ML Library	Scikit-learn
DL Library	TensorFlow / Keras
Data Handling	Pandas, NumPy
Visualization	Matplotlib, Seaborn
Hardware	Standard CPU System

Table 1

c. Dataset Description

The data used in this research is presented in Table 2. The CIC IDS 2017 is utilized because it presents an extensive view of how benign and malicious internet traffic can be represented within a single set of data. In addition to these two categories of traffic, the CIC IDS 2017 also represents other forms of cyber attacks, i.e., Distributed Denial-of-Service (DDoS), Brute Force attacks, Botnets, and Infiltration. This dataset provides a broad array of features from which Network Flow information may be derived at either a Packet level or Behavioral Level; therefore, it will provide sufficient features for the evaluation of Intrusion Detection Systems. The large number of types of network flows that are contained within the CIC IDS 2017 dataset improve the validity and performance of machine learning algorithm evaluation and development using this data set to detect anomalies in network traffic. In addition, having a variety of traffic patterns allows for better experimentation in analyzing the results and increases the ability of the model used to identify anomalies to generalize well among various cyber security applications.

Feature	Description
Dataset Name	CICIDS2017
Total Records	Approximately 2.8 Million
Number of Features	80+
Attack Types	DDoS, Brute Force, Botnet, Infiltration
Data Type	Network Traffic Flows
Format	CSV

Table 2

d. Data Processing Steps

The data is preprocessed several times before a final version can be used for training a machine learning model that will perform well and accurately detect anomalies in a cloud environment. The preprocessing allow the dataset to have higher quality and consistency than it would otherwise. They also allow the various noise reduction techniques, missing value management and imbalanced data management, as well as improvements on how features are represented to contribute to the overall improvement of the effectiveness and dependability of the anomaly detection framework in cloud environments

Data Cleaning: The goal of Data Preprocessing is to remove noise from data by removing missing values, duplicate records and inconsistency so that the bias for data does not occur.

Encoding: For the purpose of the analysis, the Categorical attributes are converted into numerical form for using label encoding

technique, which enable compatibility to work with Machine-Learning algorithms.

Normalization: Numerical features are normalized using min-max scaling in order to have an equal weight of each feature when building models through supervised or unsupervised machine learning techniques. The use of this form of normalizing data for machine learning is to ensure that none of the data values for a particular attribute will be greater than those for other attributes, thereby creating a balance across the input attributes which can lead to better convergence of the model, as well as improve the quality of predictions.

Feature Selection: In order to select the relevant features that are used in identifying anomalies, data scientists use a variety of methodologies including correlation analyses (to measure how strongly each feature correlates with the target variable) and recursive feature elimination (to determine which combination of variables will provide the best fit to the given problem). In addition to helping improve the predictive power of models, feature selection helps reduce the number of features needed to analyze (i.e., "reduces dimensionality") and eliminates redundant or correlated features from the dataset; all three help make it easier and faster to train those models.

Data Splitting: Training and Testing Data Sets. The data sets are split into two groups (training and testing) to assess how well the Machine Learning Model can generalize from the training set to new, unseen data in the test set. In addition, Min-Max normalization was used to normalize all numeric columns so they fall within a common interval; this should improve both the training process, as well as the predictive quality of the models:

$$x' = \frac{x - x_{min}}{x_{max} - x_{min}}$$

where x represents the original feature value, x_{min} and x_{max} shows the minimum and the maximum values of the feature. x' represent the normalized value. These pre-processing steps will enhance the model efficiency, reduce overfitting, improve accuracy and reliability of intrusion detection system in the cloud environments.

e. Model Training and Configuration

The configurations of the proposed models are outlined in Table 3. A training set is created from the preprocessed dataset and used to create predictive models for detecting anomalous behavior on networks. The Random Forest classifier uses an ensemble of decision trees to produce decisions that are combined to increase the reliability of predictions while reducing the risk of over-fitting. Important parameters of this method, such as the number of trees, maximum tree depth, and split criteria have been optimized in order to obtain optimal performance. In addition, the long short-term memory (LSTM) type deep learning model has also been implemented to detect time-based trends within network traffic data. The LSTM model includes three layers including input layer, hidden layers and output layer. Activation functions were added to each layer to allow for the identification of various relationships in the data. Back propagation was utilized during training with the Adam optimizer to optimize convergence speed. In addition, explainable

artificial intelligence (XAI) technologies such as SHAP and LIME were included to provide an explanation of how the models made their predictions by assigning weights or values to the input variables. By providing explanations of how these models make predictions, they provide additional transparency and build trust in the anomaly detection systems.

Parameter	Value / Description
Model Type	Random Forest / LSTM
Train-Test Split	80:20
Number of Trees (RF)	100
Max Depth (RF)	Optimized
Epochs (LSTM)	20–50
Batch Size	32
Optimizer	Adam
Activation Function	ReLU / Sigmoid

Table 3

f. **Evaluation Metrics**

The anomaly detection model proposed is evaluated using standard classification metrics. They evaluate a comprehensive assessment of the models ability to identify both normal and malicious network activities correctly. Accuracy: It evaluates the overall accuracy of the model by comparing the number of correct classifications against all classified instances.

Precise: Provides a comparison of the proportion of actual attacks among all predicted attack instances.

Detection Rate (Recall): Provides a comparison of the model's capability to detect the correct amount of true attack activity.

F1 Score: Represents the harmonic mean of precision & recall which offers a balanced evaluation of the model's performance.

Each metric is calculated based on the following parameters:

- True Positive (TP) = Correctly Identified Attack
- True Negative (TN) = Correctly Identified Non-Attack
- False Positive (FP) = Non-Attack Misclassified as an attack
- False Negative (FN) = An Attack Misclassified as a Non-Attack.

The above defined metrics are mathematically represented as follows:

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN}$$

$$Precision = \frac{TP}{TP+FP}$$

$$Recall = \frac{TP}{TP+FN}$$

$$F1 = 2 \times \frac{Precision \times Recall}{Precision+Recall}$$

Performance definition metrics are summarized in Table 4.

Metric	Description
Accuracy	Overall correctness of the model
Precision	Correct predictions among predicted attacks
Recall	Detection rate of actual attacks
F1-Score	Balance between precision and recall

Table 4

6. Result and Discussion

The proposed anomaly detection model is assessed by means of common classification metrics such as accuracy, precision, recall and F1-score. In order to compare the performance of the developed model, experimental analysis results are presented graphically in Fig. 2 and

also in summary form (Table 5) with respect to their efficacy.

Metric	Value (%)
Accuracy	99.2
Precision	98.9
Recall	98.7
F1-Score	98.8

Table 5

The experimental data shows that the developed anomaly detection method is able to obtain good accuracy and be very effective at identifying anomalies as well as attacks from inside the cloud. High levels of precision result in fewer false positives being produced by the classifier, which means less time wasted investigating non-attacks. Strongly performing recalls show that the system can find most or all of the attacks/anomalies on each case.

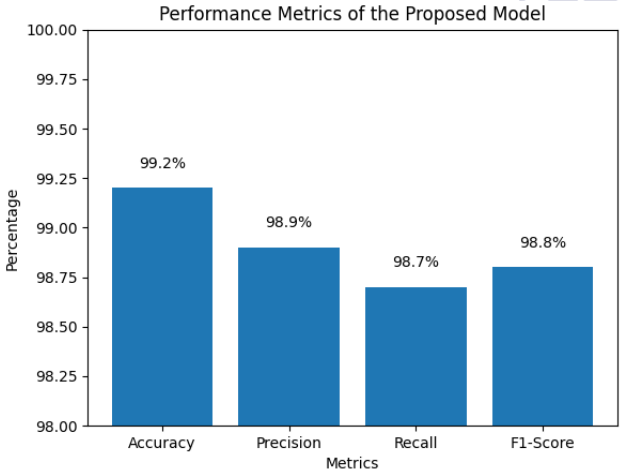

Figure 2

The graphical comparison shown in Fig. 2 illustrates how all three of the evaluation metrics (F1-Score, Recall and Precision) provide evidence to support that the proposed Anomaly Detection Model is consistent, reliable and stable. The balanced F1 score also supports that the Anomaly Detection Model maintains an appropriate balance between accuracy and dependability for classifying anomalies in cloud environments.

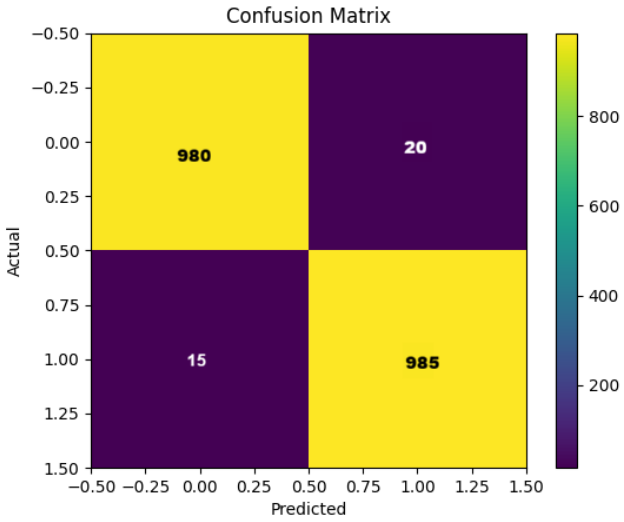

Figure 3

The confusion matrix in Figure 3 shows the level of detail with which we can examine the classification performance of the anomaly detection model developed here. Correctly classified data greatly outnumbers incorrectly classified data for all the different types of attacks and normal traffic. The results clearly show how well the proposed technique performs at reducing both false positives and false negatives when performing an anomaly detection. Additionally, to better assess the quality of the models' performance on these anomalies, a Receiver Operator Curve (ROC) analysis was conducted. An ROC curve is used to illustrate the balance between true positives and false positives that are being made based upon the threshold of classification. Typically, models located near the upper left hand side of the ROC plot will be considered to have better classification capabilities. Therefore, it should come as no surprise that our ROC analysis

validated the reliability, consistency, and ability to maintain its strength within a wide range of anomalies from the proposed anomaly detection system in cloud security systems.

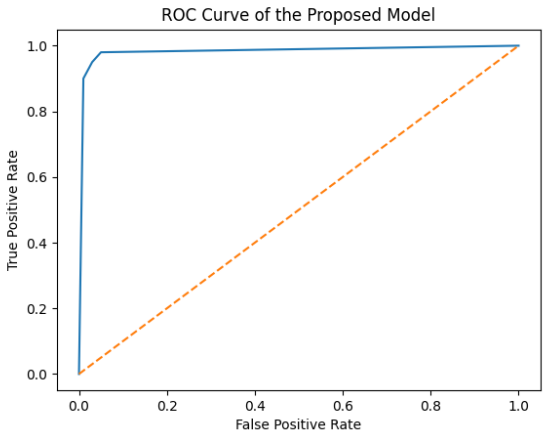

Figure 4

In Fig 4, the ROC Curve shows that this model can be at an extremely high True Positive Rate while having a False Positive Rate of very little. Also, the ROC Curve is very close to the optimal area; which means it has good Classification Performance and does a great job of distinguishing Normal Traffic from Anomalous Network Traffic. The results show that the proposed method is reliable and will work well for detection of Cyber Threats in Cloud Environments.

7. Conclusion

This research has been a presentation of a methodology which is used to develop a transparent anomaly detection process in order to enhance cybersecurity in Cloud Computing Environments. The developed methodology has combined Machine Learning (ML) methods with Explainable Artificial Intelligence (XAI), to increase the detection capability of ML models and at the same time their interpretability. Results from experiments have shown that the model was able to achieve very good detection rates in terms of Accuracy, Precision, Recall, and F1-Score, demonstrating its ability to detect anomalies in Network Activities. In addition, XAI will provide additional transparency into the decision making capabilities of the model providing greater clarity into how the model makes decisions. However, there are also several limitations to the developed model. Firstly, it is imperative that the dataset used to train the model be of high quality and diverse; otherwise, the performance of the model will suffer. Secondly, due to the nature of real-time data (the larger the volume of data the greater the potential for computational overhead); and finally, as new types of attacks occur, the model will need to continually adapt to accommodate this changing threat landscape.

8. Future Work

Future research is to extend the proposed framework to real-time cloud environments and to combine new techniques with existing ones (such as Federated Learning) in order to protect customer data through privacy-preserving intrusion detection. Additional enhancements may also be realized by developing Hybrid Deep Learning Models and selecting optimal features. In addition, Adaptive & Self-Learning Mechanisms will improve both the scalability and reliability of a system for a wide range of dynamic Cybersecurity Scenarios.

9. Figures & Tables

Fig. 1. Proposed Explainable AI-Based Anomaly Detection Framework

Fig. 2. Performance Metrics of the Proposed Model

Fig. 3. Confusion Matrix

Fig. 4. ROC Curve of the Proposed Model

Table 1: Experimental Environment

Table 2: Dataset Summary (CICIDS2017)

Table 3. Model Configuration

Table 4. Evaluation Metrics

Table 5. Model Performance

10. References

[1] Panigrahi R, Borah S. A detailed analysis of CICIDS2017 dataset for designing Intrusion Detection Systems. vol. 7. 2018.

[2] Tavallaee M, Bagheri E, Lu W, Ghorbani AA. A detailed analysis of the KDD CUP 99 data set. IEEE Symposium on Computational Intelligence for Security and Defense Applications, CISDA 2009 2009. https://doi.org/10.1109/CISDA.2009.5356528.

[3] Lansky J, Ali S, Mohammadi M, Majeed MK, Karim SHT, Rashidi S, et al. Deep Learning-Based Intrusion Detection Systems: A Systematic Review. IEEE Access 2021;9:101574–99. https://doi.org/10.1109/ACCESS.2021.3097247.

[4] Dahunsi N:, Adeyemi S, Adeyemi DS. Effectiveness of Machine Learning Models in Intrusion Detection Systems: A Systematic Review. Communication In Physical Sciences 2024;11:1060–88.

[5] Shaikh R. Security Issues in Cloud Computing: A survey. 2012.

[6] Lundberg SM, Allen PG, Lee S-I. A Unified Approach to Interpreting Model Predictions. Adv Neural Inf Process Syst 2017;30.

[7] Barredo Arrieta A, Díaz-Rodríguez N, Del Ser J, Bennetot A, Tabik S, Barbado A, et al. Explainable Artificial Intelligence (XAI): Concepts, taxonomies, opportunities and challenges toward responsible AI. Information Fusion 2020;58:82–115. https://doi.org/10.1016/J.INFFUS.2019.12.012.

[8] Speith T. A Review of Taxonomies of Explainable Artificial Intelligence (XAI) Methods. ACM International Conference Proceeding Series 2022:2239–50. https://doi.org/10.1145/3531146.3534639;PAGEGROUP:STRING:PUBLICATION.

[9] Jadhav AD, Pellakuri V. Intrusion detection system using machine learning techniques for increasing accuracy and distributed and parallel approach for increasing efficiency. Proceedings - 2019 5th International Conference on Computing, Communication Control and Automation, ICCUBEA 2019 2019. https://doi.org/10.1109/ICCUBEA47591.2019.9128620.

[10] Jamadar RA. Network Intrusion Detection System Using Machine Learning. Indian J Sci Technol 2018;11:1–6. https://doi.org/10.17485/ijst/2018/v11i48/139802.

[11] Charmet F, Tanuwidjaja HC, Ayoubi S, Gimenez PF, Han Y, Jmila H, et al. Explainable artificial intelligence for cybersecurity: a literature survey. Annals of Telecommunications 2022 77:11 2022;77:789–812. https://doi.org/10.1007/S12243-022-00926-7.

[12] Mohale VZ, Obagbuwa IC. A systematic review on the integration of explainable

artificial intelligence in intrusion detection systems to enhancing transparency and interpretability in cybersecurity. Front Artif Intell 2025;8:1526221. https://doi.org/10.3389/FRAI.2025.1526221/TEXT.

[13] Zhang Z, Hamadi H Al, Damiani E, Yeun CY, Taher F. Explainable Artificial Intelligence Applications in Cyber Security: State-of-the-Art in Research. IEEE Access 2022;10:93104–39. https://doi.org/10.1109/ACCESS.2022.3204051.

[14] Rjoub G, Bentahar J, Abdel Wahab O, Mizouni R, Song A, Cohen R, et al. A Survey on Explainable Artificial Intelligence for Cybersecurity. IEEE Transactions on Network and Service Management 2023;20:5115–40. https://doi.org/10.1109/TNSM.2023.3282740.

[15] Cremer JL, Konstantelos I, Strbac G. From Optimization-Based Machine Learning to Interpretable Security Rules for Operation. IEEE Transactions on Power Systems 2019;34:3826–36. https://doi.org/10.1109/TPWRS.2019.2911598.

[16] Chander B, John C, Warrier L, Gopalakrishnan K. Toward Trustworthy Artificial Intelligence (TAI) in the Context of Explainability and Robustness. ACM Comput Surv 2025;57. https://doi.org/10.1145/3675392;SERIALTOPIC:TOPIC:ACM-PUBTYPE.

[17] Sreenivasa Chakravarthi S, Jagadeesh Kannan R, Anantha Natarajan V, Gao XZ. Deep Learning Based Intrusion Detection in Cloud Services for Resilience Management. Computers, Materials and Continua 2022;71:5117–33. https://doi.org/10.32604/cmc.2022.022351.

[18] Alsoufi MA, Razak S, Siraj MM, Nafea I, Ghaleb FA, Saeed F, et al. Anomaly-Based Intrusion Detection Systems in IoT Using Deep Learning: A Systematic Literature Review. Applied Sciences 2021, Vol 11, Page 8383 2021;11:8383. https://doi.org/10.3390/APP11188383.

[19] Ben Said R, Sabir Z, Askerzade I. CNN-BiLSTM: A Hybrid Deep Learning Approach for Network Intrusion Detection System in Software-Defined Networking With Hybrid Feature Selection. IEEE Access 2023;11:138732–47. https://doi.org/10.1109/ACCESS.2023.3340142.

[20] Altunay HC, Albayrak Z, Çakmak M. AUTOENCODER-BASED INTRUSION DETECTION IN CRITICAL INFRASTRUCTURES. Current Trends in Computing 2024;2:1–12.

[21] Vibhute AD, Khan M, Kanade A, Patil CH, Gaikwad S V., Patel KK, et al. An LSTM-based novel near-real-time multiclass network intrusion detection system for complex cloud environments. Concurr Comput 2024;36:e8024. https://doi.org/10.1002/CPE.8024;CTYPE:STRING:JOURNAL.

[22] Mehmood Y, Shibli MA, Habiba U, Masood R. Intrusion detection system in cloud computing: Challenges and opportunities. Conference Proceedings - 2013 2nd National Conference on Information Assurance, NCIA 2013 2013:59–66.

https://doi.org/10.1109/NCIA.2013.6725325.

[23] Liu Z, Xu B, Cheng B, Hu X, Darbandi M. Intrusion detection systems in the cloud computing: A comprehensive and deep literature review. Concurr Comput 2022;34:e6646. https://doi.org/10.1002/CPE.6646;WGROUP:STRING:PUBLICATION.

[24] Abdulsalam YS, Hedabou M. Security and Privacy in Cloud Computing: Technical Review. Future Internet 2022, Vol 14, Page 11 2021;14:11. https://doi.org/10.3390/FI14010011.

[25] Sharon A, Mohanraj P, Abraham TE, Sundan B, Thangasamy A. An Intelligent Intrusion Detection System Using Hybrid Deep Learning Approaches in Cloud Environment. IFIP Adv Inf Commun Technol 2022;651 IFIP:281–98. https://doi.org/10.1007/978-3-031-11633-9_20/SAVE-RESEARCH.

[26] Abdulhammed R, Faezipour M, Musafer H, Abuzneid A. Efficient network intrusion detection using PCA-based dimensionality reduction of features. 2019 International Symposium on Networks, Computers and Communications, ISNCC 2019 2019. https://doi.org/10.1109/ISNCC.2019.8909140.

[27] Li J, Tong X, Liu J, Cheng L. An Efficient Federated Learning System for Network Intrusion Detection. IEEE Syst J 2023;17:2455–64. https://doi.org/10.1109/JSYST.2023.3236995.

[28] Moisejevs I. Adversarial Attacks and Defenses in Intrusion Detection Systems: A Survey. n.d.

[29] Sozol MS, Saki GM, Rahman MM. Anomaly Detection in Cybersecurity with Graph-Based Approaches. INTERANTIONAL JOURNAL OF SCIENTIFIC RESEARCH IN ENGINEERING AND MANAGEMENT 2024;08:1–5. https://doi.org/10.55041/IJSREM37061.

[30] de Oliveira JA, Gonçalves VP, Meneguette RI, de Sousa RT, Guidoni DL, Oliveira JCM, et al. F-NIDS — A Network Intrusion Detection System based on federated learning. Computer Networks 2023;236:110010. https://doi.org/10.1016/J.COMNET.2023.110010.

[31] Eleweke I, Umakor MF, Ndubuisi CW, Amomo CG, Adeniji S, Temidayo M. Ai-Driven Threat Detection and Prevention in Cloud Computing Environments. American Journal of Innovation in Science and Engineering 2025;4:49–56. https://doi.org/10.54536/ajise.v4i3.5041.

[32] Chukwuemeka Nwachukwu, Kehinde Durodola-Tunde, Chukwuebuka Akwiwu-Uzoma. AI-driven anomaly detection in cloud computing environments. International Journal of Science and Research Archive 2024;13:692–710. https://doi.org/10.30574/ijsra.2024.13.2.2184.