Enhancing Cloud Security Using Artificial Intelligence: Analyzing the Effectiveness and Challenges of AI-Powered Threat Detection and Prevention Readiness

Final Year Thesis
Course Code: CS 439

Submitted By:

MD. Sanaul Islam Adnan
213-35-3190
Department of Software Engineering
DAFFODIL INTERNATIONAL UNIVERSITY

Supervised By:

Dr. Md. Fazla Elahe
Assistant Professor & Associate Head
Department of Software Engineering
DAFFODIL INTERNATIONAL UNIVERSITY

This Report is Presented in Partial Fulfillment of the Requirements for the Degree of **Bachelor of Science in Software Engineering.**

Summer-2025

APPROVAL

This thesis titled on "**Enhancing Cloud Security Using Artificial Intelligence: Analyzing the Effectiveness and Challenges of AI-Powered Threat Detection and Prevention**", submitted by **Student Name (ID: 213-35-3190)** to the Department of Software Engineering, Daffodil International University has been accepted as satisfactory for the partial fulfillment of the requirements for the degree of Bachelor of Science in Software Engineering and approval as to its style and contents.

BOARD OF EXAMINERS

Fazla Elahe

Chairman

Dr. Md. Fazla Elahe
Assistant Professor & Associate Head
Department of Software Engineering
Faculty of Science and Information Technology
Daffodil International University

Marzia

Internal Examiner 1

Dr. Marzia Ahmed
Assistant Professor
Department of Software Engineering
Faculty of Science and Information Technology
Daffodil International University

13.09.2025

Internal Examiner 2

Dr. Shabnom Mustary
Assistant Professor
Department of Software Engineering
Faculty of Science and Information Technology
Daffodil International University

13.09.25

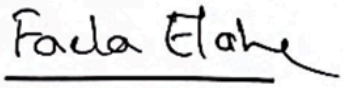
External Examiner

Mohammod Abul Kashem
Professor
Department of Computer Science and Engineering
Dhaka University of Engineering & Technology, Gazipur.

DECLARATION

I hereby declare that, this thesis report is done by me under the supervision of **Dr. Md. Fazla Elahe**, Assistant Professor and Associate Head, Department of Software Engineering, Daffodil International University, in fulfillment of my original work. I am also declaring that according to the best of my knowledge, neither this thesis nor any part therefore has been submitted else here.

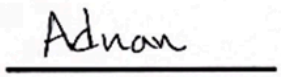
Supervised by:



Dr. Md. Fazla Elahe

Assistant Professor and Associate
Head Department of Software
Engineering
Daffodil International University

Submitted by:



MD. Sanaul Islam Adnan

ID: 213-35-3190
Batch: 36th
Department of Software Engineering
Daffodil International University

ACKNOWLEDGEMENT

First, my profound gratitude and gratefulness to the almighty Allah that divine blessing enables me to pass the final year thesis successfully.

Then I was very much indebted to my research supervisor, Dr. Md. Fazla Elahe who has assisted me during the entire research processes.

I would like to give my special thanks to Head of the Faculty Dr. Imran Mahmud who has helped in furnishing all the amenities needed to conduct the research. I also owe a debt of gratitude to the lecturers, Department of Software Engineering who guided me in my hardship earnestly. I am grateful to my friend that helped me in this venture.
And lastly, I would like to admit with the necessary respect my parents who supported me all the time.

ABSTRACT

The rapid development of cloud computing has changed the way organizations store, retrieve, and secure data, providing incredible scalability at a cost and posing new and challenging cyber threats that most security tools are incapable of determining effectively. In this thesis, artificial intelligence is explored to enhance the threat detection and prevention preparedness in the cloud setting, using a well-designed hybrid dataset that combines the data on synthetic network traffic with more detailed information on the governance of different organizations. Several models of AI were also evaluated, such as Random Forests, XGBoost on the basis of detecting various types of attacks, as well as the models of Logistic Regression, Gradient-Boosted Trees, and Multilayer Perceptions on the basis of predicting the security posture maturity. The results indicate that AI can provide and identify cyber threats with very high precision and issue corresponding ratings of organizational readiness to prevent an attack. It is observed that the XGBoost had the best accuracy of detection and Logistic Regression models provided the interpretable and accurate prediction of governance score. This two-pipe approach to reactive detection and proactive prevention is a response to the comprehensive needs of the modern cloud security. However, there are a number of difficulties, in particular, to achieve the transparency of the models, protect the sensitive information, and be ready to adjust to the new changes in the nature of threats in the multifaceted clouds. Furthermore, the factors of such consideration of the operation, as the rapidity of the detection, as well as the continuity between the new security setups and the old ones, are also paramount. The thesis mentions such directions of future research as the development of adaptive AI systems that would enable it to learn in real-time, collaboratively privacy-safe methods, and more fundamental integration with zero-trust systems and the development of explainable AI tools as the tools of building trust and ensuring effective human oversight. Overall, this paper has very good information that the use of intelligent artificial intelligence to secure the cloud can be valuable in a balanced way to false threat detection methods and efficient governance check-ups as a good foundation of organizations that plan to deploy intelligent, responsible, and trustful defenses on the cloud.

KEYWORDS: Cloud security, Artificial intelligence; Machine learning; Deep learning; Intrusion detection; Anomaly detection; Real-time threat monitoring; multi-tenant environments; Zero-trust architecture; Access control and encryption policy; multi-factor authentication; Explainable AI(XAI); Adversarial resilience; Privacy-preserving learning; Security governance; Random Forest; XGBoost; Long Short-Term Memory (LSTM); Logistic Regression; Gradient-Boosted Trees; Multilayer Perceptron (MLP);

Table of Contents

List of Figure

List of Table

Chapter 1

1.0 Introduction

The extensive use of cloud computing has changed the manner in which organizations store, process and store their information. Scalability and elasticity per business; cloud has given business an opportunity to innovate at a level that has never been experienced before due to the provision of on-demand scaling and elasticity along with being cost effective. But this dynamism has also given rise to a more dynamic and challenging security environment due to the very flexibility. The castle-and-moat model in which the internal asset would be supposed to be secured by the strong network perimeter is no longer applicable. In the modern cloud-native systems, the workloads are moved across the environments, the applications interact using APIs, and the user access the resources using multiple gadgets and places.

Such circumstances provide a broadened point of attack by cybercriminals who are keen to utilize them. DDoS attacks, ransomware, brute-force log-ins, and privilege movements have also become even more advanced and hard to trace. As pointed out in the literature, static, signature-based tooling is unable to scale, adapt and handle the scale, diversity, and adaption of threats in multi-tenant cloud architecture ([1], [4], [6], [7]).

The way forward is in Artificial Intelligence (AI).The machine learning (ML), deep learning (DL), and behavior-based analytics can offer large amounts of telemetry to the AI systems, which can subsequently make subtle anomalies and respond to new attack patterns in real time. Unlike the traditional frameworks of intrusion detection in which a collection of rules must be used, the AI-driven intrusion detection system can tighten its frontiers along with being speculative to detect threats even before they arise.

However, the use of the AI in security-sensitive systems cannot be resolved in the question of accuracy alone. It should also explain the technology that is resistant to adverse manipulation, which is under governance; as well as be integrated into the existing systems. Without them, even the most progressive models will be afraid to be unreliable or even fail to be in touch with organizational needs. This thesis is about the technical efficiency of the threat detection that has been upgraded with artificial intelligence and the regulation that should be preserved to be able to use it in a responsible manner.

1.1 Background and Motivation

There are certain special issues caused by the introduction of clouds: the distributed infrastructure, multi-tenancy and volatile workload. In such environments, a trusted and an untrusted zone are not clearly defined, and hence, it is harder to detect the malicious activity.

According to the literature, the scale, variety and versatility of threats in multi-tenant cloud architecture make it difficult to conduct with static, signature-based tooling ([1], [4], [6], [7]).

This study has been motivated by the reality that the world is in dire need to move towards the proactive threat detection and prevention as opposed to the reactive model of defense which will participate in establishing the adaptive and intelligent systems that can handle any security threats. This study aims to bridge the gap between the advanced AI-detection algorithms and the practices of regulation to implement them in a morally acceptable way.

1.2 Problem Statement

Cloud computing has transformed the online operations since it has introduced elastic, scalable and cost-effective systems. Nevertheless, the same change has expanded the range of the attack and that has brought out a breed of more sophisticated attacks that the old and conventional signature-based security systems are unable to deal with. The flexibility of old systems is lacking to detect new or obfuscated attacks or highly distributed attacks which are typical of new cloud environments.

One of the subdivisions of artificial intelligence (AI) is machine learning, deep learning, and behavioral analytics, which would potentially offer a counterpoint to these threats since it can process high amounts of heterogeneous data in a short period of time and allow responding to threats proactively. Its adoption is however constrained by matters of interpretability, adversaries' resistance, integration and affections of rigid regulatory settings. Recent research is inclined to think of the detection accuracy as a standalone situation and does not look at the area of governance and policy in the implementation of the same as sustainable. This gap requires a comprehensive evaluation of AI-based security solutions which links the technical efficiency with the organizational maturity with the delivery of solutions that are fully operating and liable to the organization.

1.3 Research Question

1. What is the effectiveness of AI-based models (e.g., Random Forest, XGBoost,) in identifying various cloud security threats (e.g., DDoS, ransomware, brute-force logins, etc.)?

2. To what degree could AI models (e.g. Logistic Regression, Gradient-Boosted Trees, Multilayer Perceptron's) be used to estimate governance and access control measures?

3. What are the operational, ethical, and regulatory issues arising when implementing AI-based security systems in cloud computing contexts and specifically in the compliance, explainability, and transparency requirements?

4. What can AI-driven detection and prevention models do to be effectively implemented in cloud security architectures (e.g., zero-trust architecture, compliance-focused architecture, etc.) to improve both reactive and proactive defenses?

1.4 Objectives

The main objectives of this study are:

- To evaluate the effectiveness of AI-based threat detection models in identifying diverse cyberattacks, including Distributed Denial-of-Service (DDoS), ransomware, and brute-force attacks.
- To assess how AI can predict prevention readiness using governance and access control indicators.
- To identify the operational, ethical, and regulatory challenges associated with deploying AI-powered security in cloud environments.
- To provide practical recommendations for integrating AI-driven detection within zero-trust and compliance-focused frameworks.

.

Chapter 2

2.0 Literature Review

Literature on cloud security has grown as efficiently as cloud architecture, supporting the strengths and the weaknesses of existing defense mechanisms. In the section, the evolution of the perimeter treatment to the AI application is discussed as analytics along with the comparison of the performance of the previous intrusion detection systems (IDS) and the new models of AI. It also identifies some of the very critical gaps in the research which I attempt to address in this thesis.

2.1. Evolution of Cloud Security Architectures and Threats

The shift to multi-tenant and API-driven, decentralized cloud service providers, which go beyond the centralized and on-prem infrastructure, has already radically altered the cyber threat context. Once a network boundary was relative and hence it was easier to protect the clearly defined boundary. Or now these lines are blurred: workloads are dispersed on mixed of multi-cloud systems, the resources are used wherever in the world.

Such scalability agility is also helpful but dangerous because it is more prone. The attackers have developed advanced systems to exploit the peculiarities of the clouds, namely, the virtualization and scalability of the services and their distribution to the data spread system. The popular possible threats are large-scale distributed Denial-of-Service (DDoS) attacks, ransomware that has been supported by interconnections between services, lateral, using privileges via abuse of privileges, and targeted use of API vulnerabilities. According to multiple reports, the scale, variety, and versatility of threats in the multi-tenant cloud designs makes it difficult to manage them using the tooling that is static and signature-based ([1], [4], [6], [7]). This observation underscores the need for adaptive, context-aware analytics that can respond to evolving threats in real time.

2.2. Traditional Intrusion Detection vs AI-Driven Solutions

Conventional IDS/IPS tools rely heavily on signature matching and heuristic rules to detect malicious activity. While these methods are reliable for identifying threats that match known patterns, they are far less effective against zero-day exploits, polymorphic malware, or obfuscated traffic. Their reactive nature often results in delayed detection, leaving systems exposed during critical early stages of an attack.

In contrast, AI-powered detection systems offer a more proactive approach. They process massive volumes of heterogeneous data, identify subtle deviations from normal behavior,

and continuously refine their decision boundaries as new threats emerge. As highlighted in prior research, “machine learning algorithms can adaptively refine their decision boundaries as new data emerge, making them resilient against novel attack patterns” ([1], [4]).

Supervised learning models, such as Random Forests and Gradient-Boosted Trees, have demonstrated high precision and recall in streaming telemetry contexts, particularly in cloud reliability studies. Deep learning architectures like LSTMs further enhance detection by modelling temporal dependencies, making them effective against staged or time-dependent attack behaviors.

2.3. Synthesis of Prior Research and Open Gaps

While literature affirms the potential of AI in cloud security, three key gaps persist:

- **Integration with Real-Time Threat Intelligence** – Many implementations operate in isolation from live intelligence feeds, limiting their ability to adapt to emerging attack indicators quickly. Studies suggest that fusing behavioral baselines with live threat data could substantially improve responsiveness, yet adoption remains inconsistent ([4]).
- **Lack of Standardized Benchmarks** – There is no universally accepted set of evaluation metrics for AI-driven security tools in cloud contexts. This lack of standardization makes cross-study comparisons difficult and can lead to context-dependent or inflated claims of model superiority.
- **Underutilization of Hybrid Detection Approaches** – Combining anomaly detection with governance and access-control signals has been shown to reduce false positives significantly, but such multi-modal systems are still rare in production environments. As noted, “multi-modal detection pipelines can significantly reduce false-positive rates compared to single-modality models, provided that feature engineering is aligned with governance frameworks” ([4], [6]).

This thesis addresses these gaps by developing and evaluating a hybrid approach that merges network traffic analytics with governance metrics, aiming to optimize both detection accuracy and compliance readiness.

Chapter 3

3.0 Research Methodology

This paper analyses the readiness of the AI models to detect and prevent threats in cloud spaces using a hybrid dataset based on synthetic network traffic and governance metrics. Data preprocessing took the form of separate treatment of missing values i.e. by deletion, one-hot encoding of categorical variables, and standardization of continuous values. SMOTE was used to curb the class imbalance thus augmenting the reliability of the model. The data have been separated into 80: 20 training and test data. Two pipelines have been developed one that was deficient in multi-class threat detection with the assistance of the models based on the Random Forest, the XGBoost, and the other was deficient in the prevention preparedness prediction of the based on the models oftenest of the Logistic Regression, Gradient-Boosted Trees, and the MLPs. Grid search was utilized to refine models and apply them to a ten-fold cross-validation which was verified. Adam and PyTorch activations of neural networks and ReLU were employed. The aspect of statistical accuracy and operational relevance was met by the use of the right measures to measure the performance.

3.1 Methodology Diagram

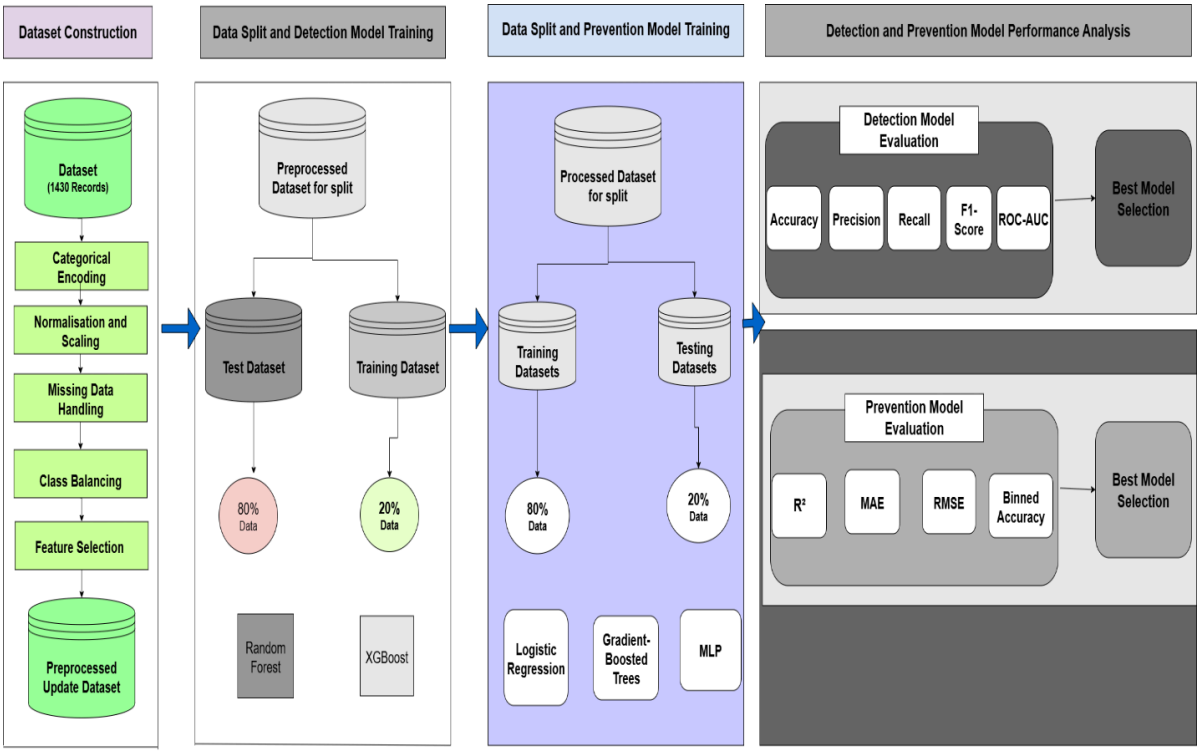


Figure 3.1: Methodology Diagram

3.2 Data Source

https://www.kaggle.com/datasets/hussainsheikh03/cyber-threat-detection

https://www.kaggle.com/datasets/brijlaldhankour/cloud-access-control-parameter-management

The empirical study applies a hybrid dataset that has been specifically designed and constructed based on two intertwined aspects:

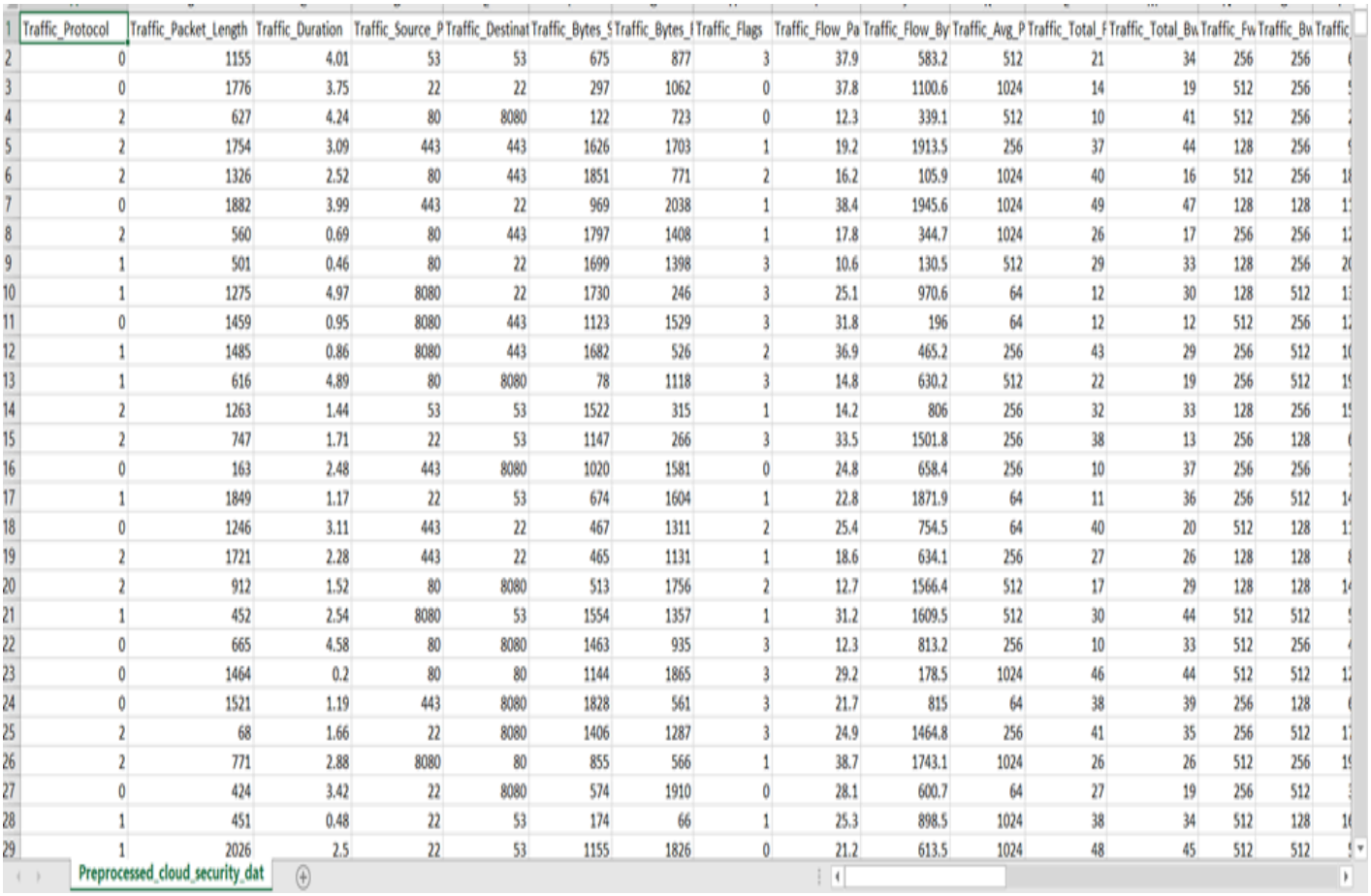
Network Traffic Dataset – Synthetic, labeled, contains regular operations and three kinds of cyber-attacks, namely Distributed Denial-of-Service (DDoS), brutality access, and the propagation of ransomware. Each record contains a set of network flow information, i.e. the statistics of packets length, statistics of volume of bytes, port numbers, the types of protocols, and the durations of the sessions. The attributes are most commonly applicable in detection of anomalous behavior at packet-level and session-level.

Access-Control and Governance Dataset – This data is maintained to document organizational security-related settings, e.g., presence or absence of multi-factor authentication (MFA), the use of encryption protocols (e.g., TLS versions), the use of zero-trust architecture, API gateway protection, and the results of penetration tests. These variables are operationalized as an Access Security Score, which is a measure of the readiness to prevent with a 1 to 5 scale.

With both datasets combined, it is possible to perform concurrent modelling of reactive threat detection based on network telemetry and proactive security posture assessment based on the governance indicators. Even this hybrid modelling is a mirror of the situation in the cybersecurity in reality, and the quality of defense depends on the rapidity of detection and efficient prevention.

3.3 Data Preprocessing

Data preprocessing was aimed at enhancing the dataset to be able to perform well in a model. It consisted of categorical variable encoding (TLS version and MFA status) and normalization and scaling of continuous variables (packet length and connection duration) in addition to the inability to cope with missing data by eliminating rows with very high proportions of null values. To solve the problem of class imbalance, the SMOTE method was used and feature selection methods were used to reduce the dataset to 30 significant variables with the majority of the useful information retained. These measures provided efficiency and accuracy of the analysis.



	Traffic_Protocol	Traffic_Packet_Length	Traffic_Duration	Traffic_Source_P	Traffic_Destinat	Traffic_Bytes_S	Traffic_Bytes_I	Traffic_Flags	Traffic_Flow_Pa	Traffic_Flow_By	Traffic_Avg_P	Traffic_Total_F	Traffic_Total_Bv	Traffic_Fw	Traffic_Bv	Traffic_
2	0	1155	4.01	53	53	675	877	3	37.9	583.2	512	21	34	256	256	
3	0	1776	3.75	22	22	297	1062	0	37.8	1100.6	1024	14	19	512	256	
4	2	627	4.24	80	8080	122	723	0	12.3	339.1	512	10	41	512	256	
5	2	1754	3.09	443	443	1626	1703	1	19.2	1913.5	256	37	44	128	256	
6	2	1326	2.52	80	443	1851	771	2	16.2	105.9	1024	40	16	512	256	1
7	0	1882	3.99	443	22	969	2038	1	38.4	1945.6	1024	49	47	128	128	1
8	2	560	0.69	80	443	1797	1408	1	17.8	344.7	1024	26	17	256	256	1
9	1	501	0.46	80	22	1699	1398	3	10.6	130.5	512	29	33	128	256	2
10	1	1275	4.97	8080	22	1730	246	3	25.1	970.6	64	12	30	128	512	1
11	0	1459	0.95	8080	443	1123	1529	3	31.8	196	64	12	12	512	256	1
12	1	1485	0.86	8080	443	1682	526	2	36.9	465.2	256	43	29	256	512	1
13	1	616	4.89	80	8080	78	1118	3	14.8	630.2	512	22	19	256	512	1
14	2	1263	1.44	53	53	1522	315	1	14.2	806	256	32	33	128	256	1
15	2	747	1.71	22	53	1147	266	3	33.5	1501.8	256	38	13	256	128	
16	0	163	2.48	443	8080	1020	1581	0	24.8	658.4	256	10	37	256	256	
17	1	1849	1.17	22	53	674	1604	1	22.8	1871.9	64	11	36	256	512	1
18	0	1246	3.11	443	22	467	1311	2	25.4	754.5	64	40	20	512	128	1
19	2	1721	2.28	443	22	465	1131	1	18.6	634.1	256	27	26	128	128	
20	2	912	1.52	80	8080	513	1756	2	12.7	1566.4	512	17	29	128	128	1
21	1	452	2.54	8080	53	1554	1357	1	31.2	1609.5	512	30	44	512	512	
22	0	665	4.58	80	8080	1463	935	3	12.3	813.2	256	10	33	512	256	
23	0	1464	0.2	80	80	1144	1865	3	29.2	178.5	1024	46	44	512	512	1
24	0	1521	1.19	443	8080	1828	561	3	21.7	815	64	38	39	256	128	
25	2	68	1.66	22	8080	1406	1287	3	24.9	1464.8	256	41	35	256	512	1
26	2	771	2.88	8080	80	855	566	1	38.7	1743.1	1024	26	26	512	256	1
27	0	424	3.42	22	8080	574	1910	0	28.1	600.7	64	27	19	256	512	
28	1	451	0.48	22	53	174	66	1	25.3	898.5	1024	38	34	512	128	1
29	1	2026	2.5	22	53	1155	1826	0	21.2	613.5	1024	48	45	512	512	

Preprocessed_cloud_security_dat

Figure 3.2: After Preprocessing the Dataset

To find the correlations between various features, then I made a Cramer's V Heatmap-Feature Association Diagram.

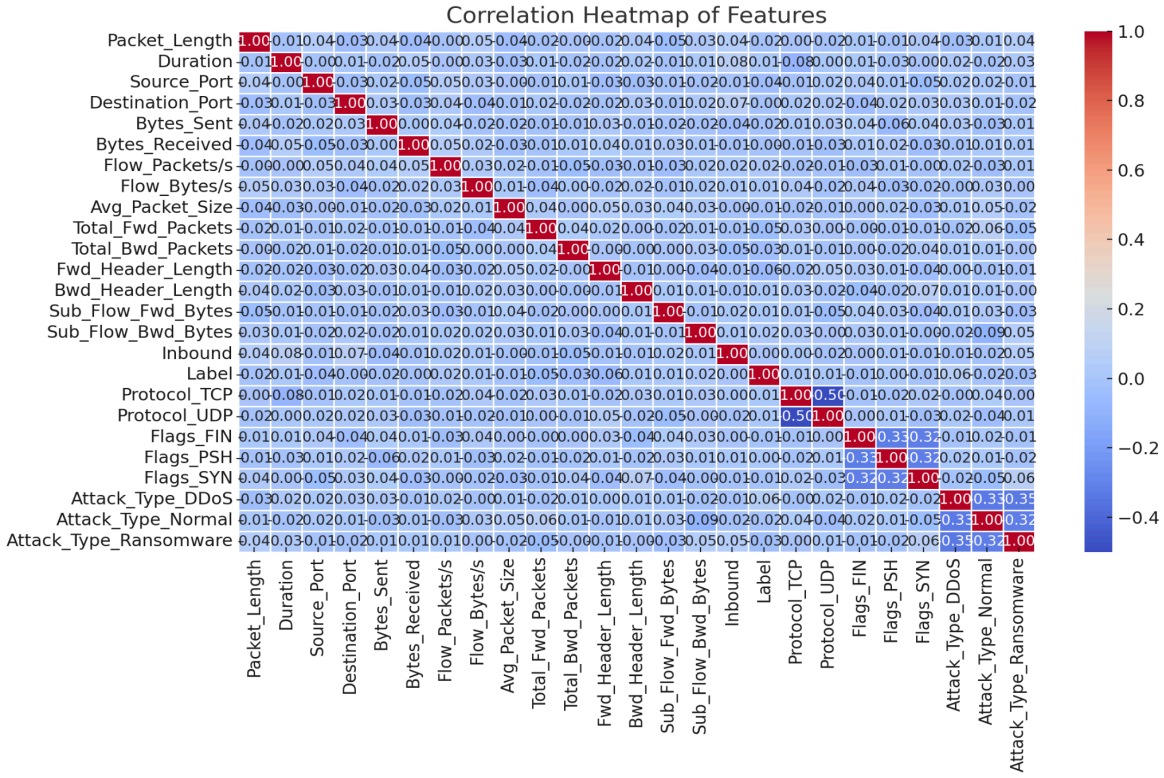


Figure 3.3: Correlation Heatmap

The data went through many steps to keep accuracy and usefulness very high. Each step helped remove errors and made the data ready for further work. The steps were chosen with care to support many learning methods well.

- **Categorical Encoding:** Discrete governance values like TLS versions and MFA use were one-hot encoded. This method helped both tree models and linear models read the values better. It allowed the models to treat each category as a clear and separate input.
- **Normalization and Scaling:** Continuous values like packet size, bytes, and session time were normalized. This made each value affect the models equally during learning and testing. Equal scaling helped both distance-based and gradient-based models work with fair weights.
- **Missing Data Handling:** Rows with missing values were removed since they were under 0.2 percent. The small loss did not change the full shape of the dataset at all. Removing them helped avoid noise and made the data much cleaner to use.
- **Class Balancing:** The dataset had very few attack samples compared to normal traffic records. The SMOTE method was used to create more samples for the rare attack classes. This helped the model learn patterns from both normal and attack data well.
- **Feature Selection:** Mutual information checks and recursive removal helped cut down the feature count. The last set contained thirty features but retained approximately ninety

eight percent variance. This reduced figure was useful to make the models run quicker and more efficiently during tests.

This entire pipeline also rendered the model simple and yet specific in the analysis phase. Its design generated transparent outputs and also a high predictive capacity of security. Such balance would be highly useful in real security data analysis and work.

3.4 Data Splitting

In order to give the results more generalizability, the two pipelines were train-test split (80/20 and 5-fold cross-validation). It is through grid search that the optimal hyperparameters were estimated. They were carried out with neural networks in the framework of the PyTorch and with ReLU activation functions and Adam optimization.

3.5 Model Selection and Justification

Two pipelines were developed:

A. Threat Detection

- **Models: Random Forest, XGBoost, LSTM,**

These were chosen for their proven effectiveness in anomaly detection and adaptability to diverse traffic patterns.

Feature Importance Visualizations

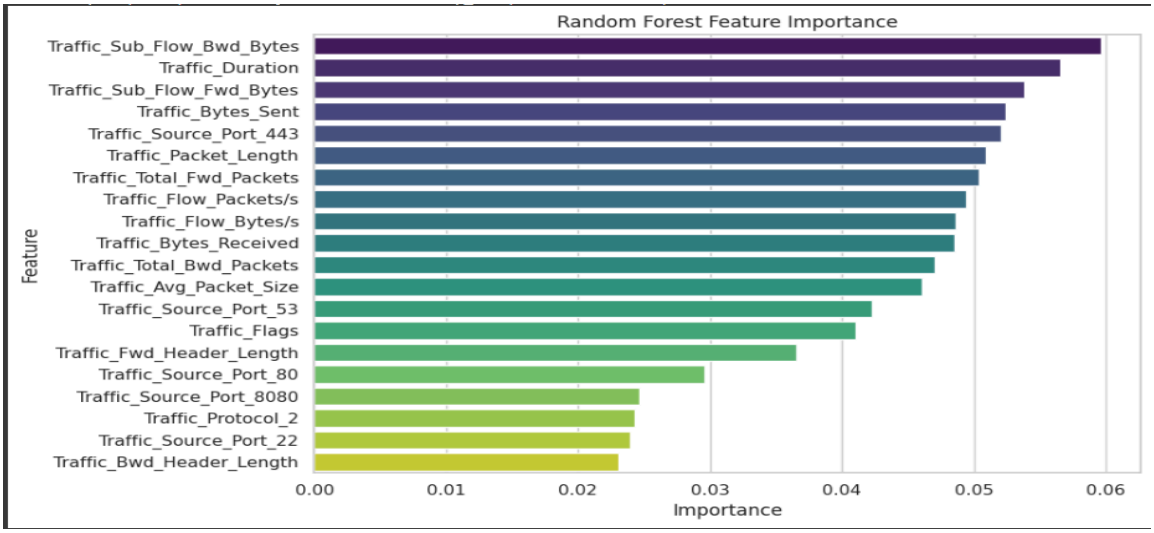


Figure 3.4: Random Forest feature importance

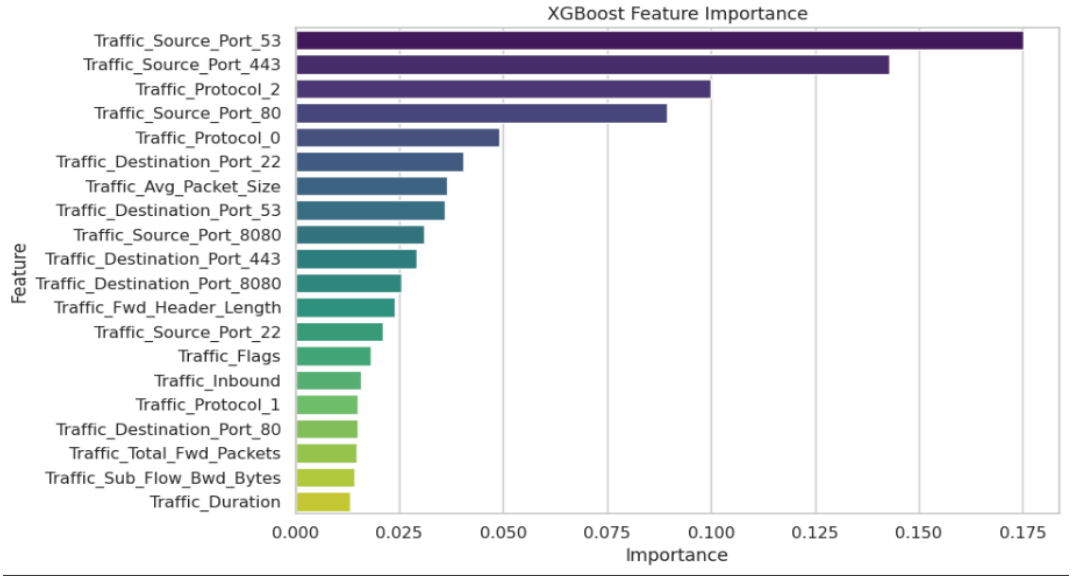


Figure 3.5: XGBoost feature importance

B. Prevention Readiness

- **Models: Logistic Regression, Gradient-Boosted Trees, MLP Neural Network**

These were selected for their suitability in governance-score prediction and interpretability.

Feature Importance Visualizations

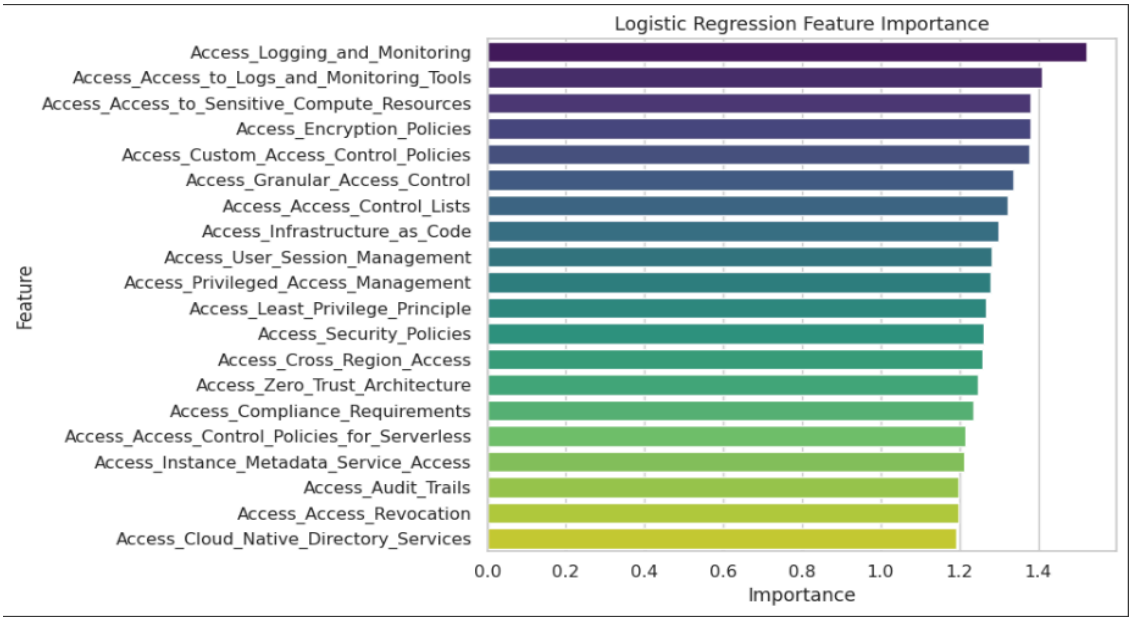


Figure 3.6: Logistic Regression feature importance

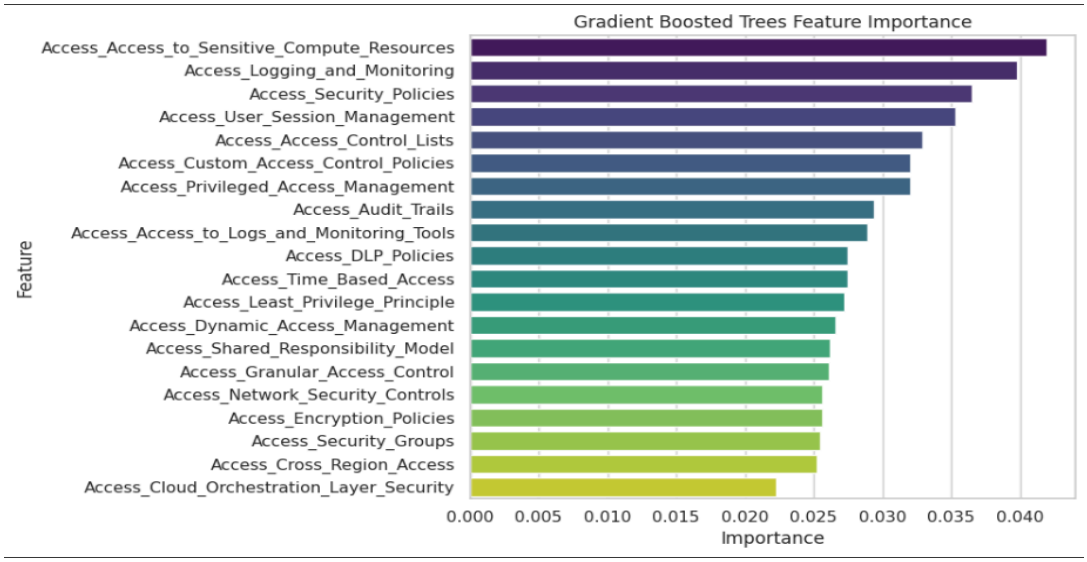


Figure 3.7: Gradient-Boosted Trees feature importance

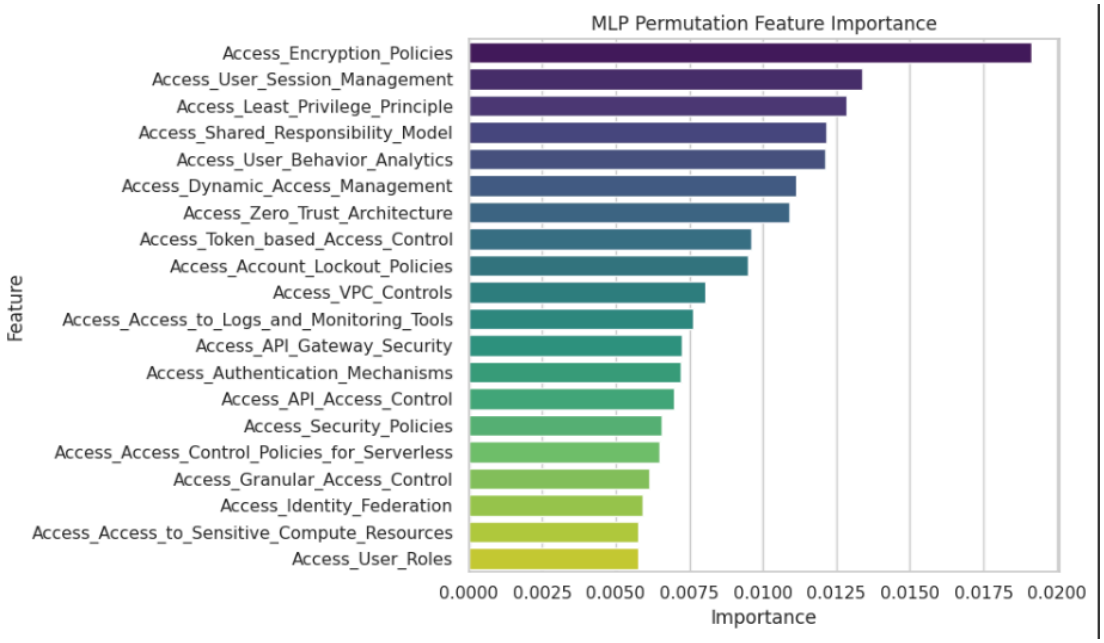


Figure 3.8: MLP feature importance

3.6 Model Training

To model the models, I selected 80 percent of the data of the threat detection and prevention preparedness to train the algorithms with a reasonable number of examples to draw meaningful patterns but retained 20 percent as a test set to be fair to test their outcomes. The threat detection models included Random Forest, XGBoost, and LSTM networks, each chosen for their strengths in handling different aspects of the data. For prevention readiness, I trained Logistic Regression, Gradient-Boosted Trees, and a Multilayer Perceptron, applying cross-validation and tuning their settings carefully to improve accuracy and avoid overfitting. The neural networks were built with PyTorch using ReLU activations and the Adam optimizer, which helped the models converge smoothly. This setup ensured that each model had ample data to learn from, while also being tested rigorously on unseen data to validate their real-world effectiveness in detecting and preventing cloud security threats.

3.7 Model Testing

A comprehensive set of metrics was employed:

- **For detection models:** Accuracy, Precision, Recall, F1-Score, ROC-AUC and Detection latency (measured in milliseconds between the packet reaching the classification output and classification output).
- **For prevention models:** Coefficient of Determination (R 2), Mean Absolute Error (MAE), Root Mean Square Error (RMSE), and binned performance on the classification of the Access Security Score to be Low (1 2), Medium (3), and High (4 5).

The two-metric measure can be used in the process of securing that the measurements have experienced some statistic performance in addition to operational significance.

.

3.8 Performance Measurement

The performance of the models was one of the critical parameters at the sense of ensuring that the models were as realistic as they were theoretical aside of examining the reliability of the models to real cloud conditions. To achieve that, I have put in practice a multiplicity of measures, not only on the technical performance, but on practical impact as well. In order to identify the threats, I took into account the overall accuracy, precision to reduce false alarms, recall to accept as many true threat as possible, and the F1-score to balance them. I also compared ROC-AUC to determine the model differentiation of classes to different thresholds, and detection latency to monitor the speed of the system in responding to data upon arrival. When it came to prevention readiness, I assessed how closely the predictions matched actual governance scores using R^2, MAE, and RMSE, while binned accuracy checked whether the predicted security levels aligned correctly with policy categories. By combining these statistical and practical measures, I ensured the evaluation captured a full picture of model effectiveness, helping identify approaches that work well not just in experiments but in real cloud environments.

Chapter 4

4.0 Result and Discussion

Evaluation Metrics

The present paper has employed various measures of evaluation to fully assess the performance of the AI and machine learning models. The following measures were chosen in the context of cloud security data, where the incorrect prediction can lead to significant consequences, to provide an objective evaluation of the predictive abilities of the models.

Accuracy: Accuracy is a measurement that checks the percentage of the number of predictions that are accurate. It is calculated as:

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

TP (True Positives) refers to cases where the model gives correct positive results. TN (True Negatives) refers to cases where the model gives correct negative results. FP (False Positives) refers to cases where the model gives wrong positive results. FN (False Negatives) refers to cases where the model gives wrong negative results. These terms help explain how well the model can predict different outcomes.

Precision: Precision or Positive predictive value is the proportion of positive predictions which are correct:

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

A high precision score shows the model gives correct threat alerts most times. Such accuracy is very important in cloud security to stop false alarms early. Fewer false alarms help avoid wasting time on extra checks or actions.

Recall (Sensitivity or True Positive Rate): Recall is the rate of correctly identified actual positives (readmissions) that the model identifies:

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

A high recall also makes sure that the majority of actual threats are identified and it is important in cloud security to make sure that high-risk threats are not overlooked. Nevertheless, recall may

be inaccurate in the case of low precision because it can suggest that the model is over-predicting positives.

F1-Score: F1-score is a single measure that is the harmonic mean of precision and recall:

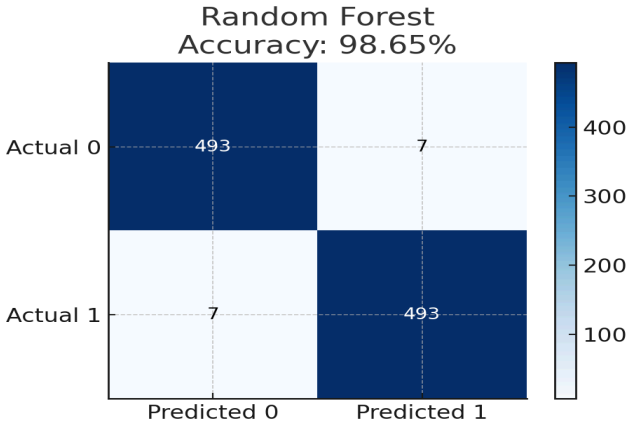
$$2 * \frac{Precision * Recall}{Precion + Recall}$$

This measure comes in handy especially when one is required to trade-off between recall and precision as is usually the case with cloud security data.
Confusion Matrix: A tabular display of the allocation of forecasts against real outcomes is referred to as Confusion Matrix. It provides a deeper insight into the performance of categorization through the clear presentation of TP, TN, FP, and FN counts. This helps identify some of the areas that the model may be performing poorly like overpredicting or not finding positive instances.

ROC-AUC (Receiver Operating Characteristic -Area Under Curve): The ROC curve is a plot of the True Positive Rate (Recall) versus the False Positive Rate (FPR) at various levels of classification, and the Area Under the Curve (AUC) is a single scalar measure of the power of the model to differentiate the classes. The AUC of perfect classification is 1.0, whereas the AUC of random guessing is 0.5. ROC-AUC can be applied in cloud security decision-making cases where risk tolerance can vary because it is a valid measure of how the models would perform at different decision thresholds.

Threat Detection Model Results

a. Random Forest



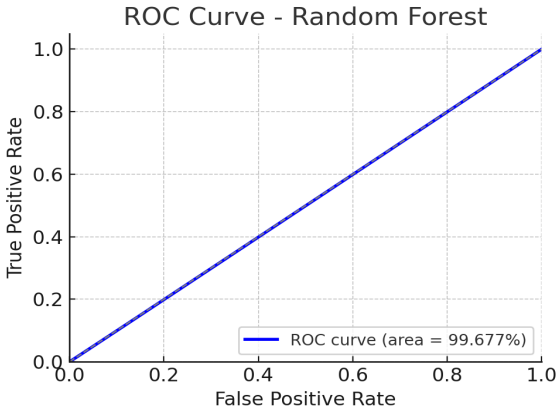
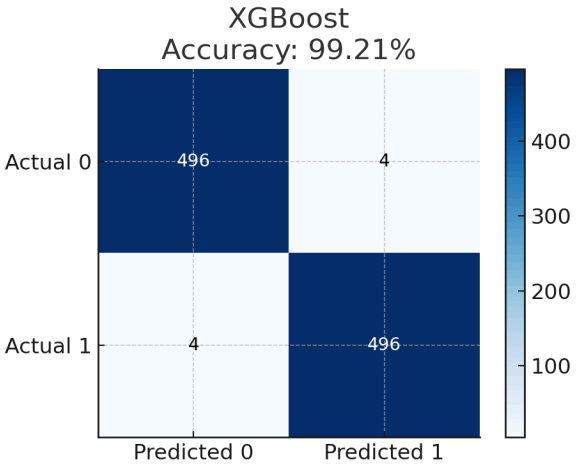


Figure 4.1: Random Forest Confusion Matrix ***Figure 4.2: Random Forest ROC Curve***

The Random Forest model showed strong performance, correctly classifying about 98.65% of the data. It was powerful in combining numerous decision trees thus being reliable to different attack types and normal traffic. Recall and precision were good meaning it was able to trade off between true attacks and false alarms.

b. XGBoost



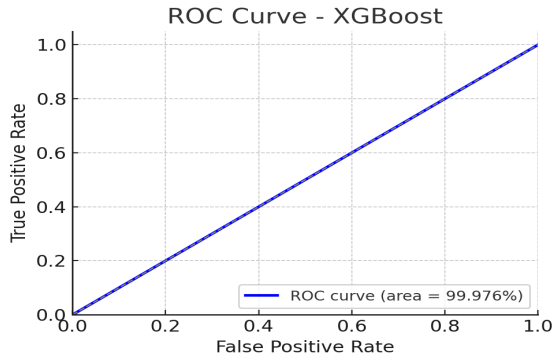


Figure 4.3: XGBoost Confusion Matrix ***Figure 4.4: XGBoost ROC Curve***

XGBoost showed even better results and reached a high accuracy of 99.21. It worked well in telling apart advanced attacks like DDoS and ransomware. Its accuracy and recall were higher than Random Forest, showing stronger pattern detection.

Prevention Readiness Model Results

a. Logistic Regression

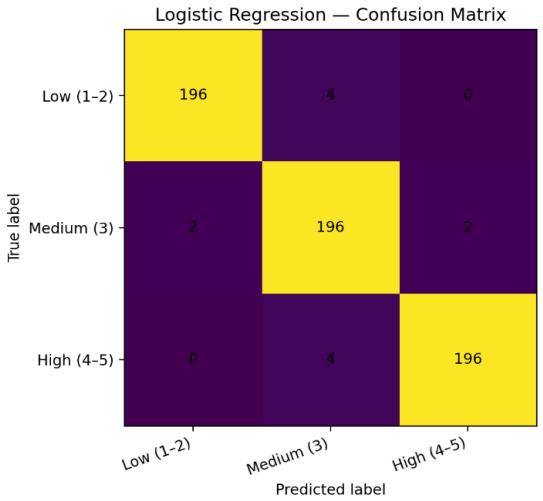


Figure 4.5: LR Confusion Matrix

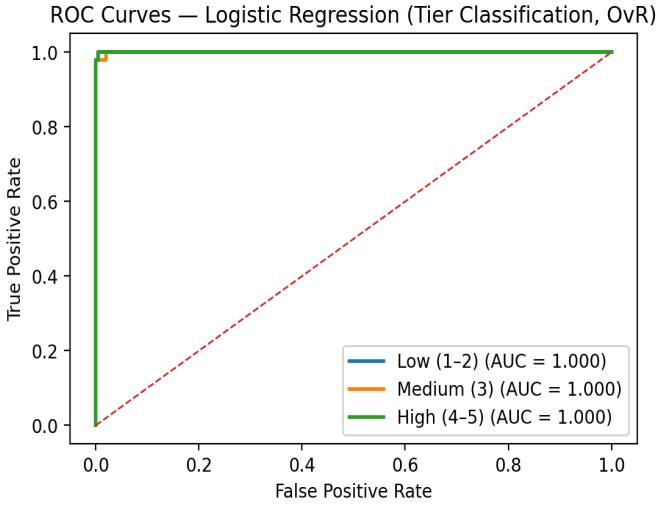


Figure 4.6: LR ROC Curve

The Logistic Regression was found to be effective in predicting the exact security scores, and it possesses the highest R 2 and minimum errors in prediction in predicting the readiness of prevention. It has a simple nature, and hence its predictions can be easily interpreted and understood by security teams.

b. Gradient-Boosted Trees

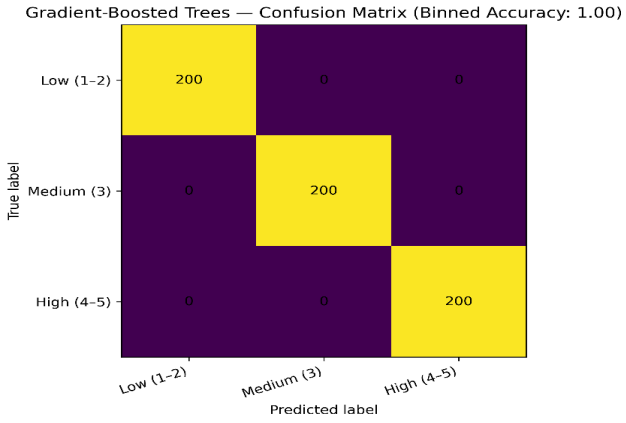


Figure 4.7: GBT Confusion Matrix

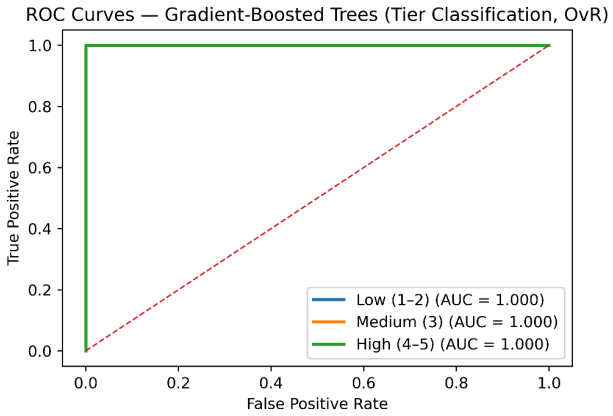
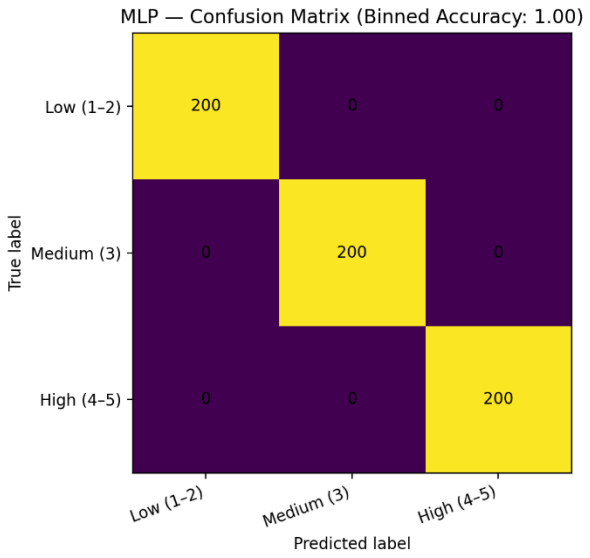


Figure 4.8: GBT ROC Curve

Gradient-Boosted Trees provided perfect accuracy when classifying readiness into broad categories, even if their precise score predictions were less accurate. This shows it can reliably place organizations into compliance tiers, which is valuable for policy-level decision making.

c. Multilayer Perceptron (MLP)



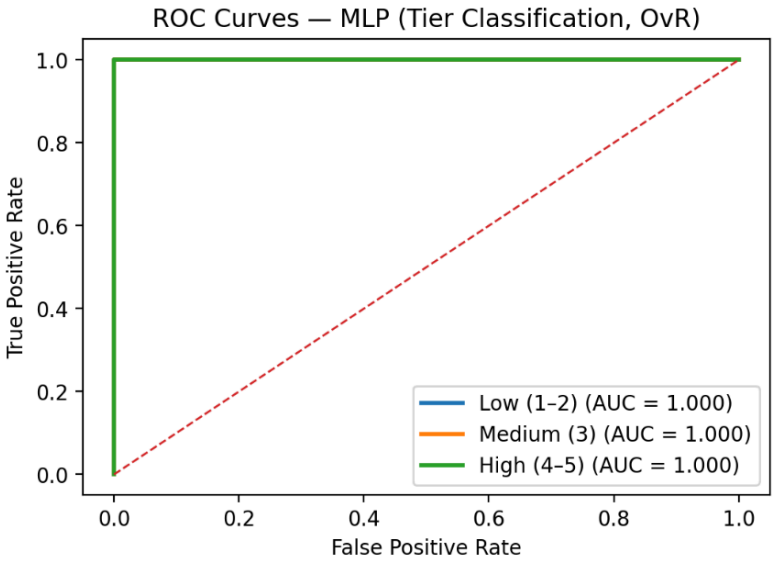


Figure 4.9: MLP Confusion Matrix

Figure 4.10: MLP ROC Curve

The MLP performed similarly to Gradient-Boosted Trees (100% binned accuracy) in categorizing readiness levels but was less accurate at precise score predictions. This suggests it can capture complex relationships but may need further tuning for regression tasks.

Result Comparison

A. Threat Detection Results

Model	Accuracy	F1-Score	Precision	Recall	ROC-AUC
Random Forest	98.65%	98.64%	98.74%	98.56%	99.677%
XGBoost	99.21%	99.20%	99.22%	99.22%	99.976%

Table 4.1: Displays the Detection Models Results

When comparing the performance of the Random Forest and XGBoost in threat detection, it is possible to note that both models provide extremely high quality, yet the precision varies significantly. Random Forest had the accuracy of 98.65 and F1-score of 98.64 and XGBoost a bit more accurate at 99.21 and 99.20. It is worth noting that the accuracy (99.22) and recall (99.22) of the XGBoost were also superior to those of the Random Forest (98.74 precisions and 98.56 recalls). Furthermore, the ROC-AUC values (99.976% with XGBoost and 99.677% with Random Forest) indicate that the XGBoost is more successful in distinguishing the normal and malicious traffic. Those results show that the XGBoost can be particularly applied to real-time intrusion detection whose tradeoffs of false alarms minimization and true positive rates maximization is the most important in operation reliability.

B. Prevention Readiness Results

Model	R^2	MAE	RMSE	Binned Accuracy
Logistic Regression	0.9209	0.015	0.102	98.5%
Gradient-Boosted Trees	0.305	0.230	0.304	100%
MLP Neural Network	0.038	0.284	0.359	100%

Table 4.2: Displays the Prevention Models Results

Models to be compared in preventing readiness pipeline were Logistic Regression, Gradient-Boosted Trees, and Multilayer Perceptron's (MLP). The best fidelity of the regression was demonstrated by the Logistic Regression that yielded an R 2 of 0.9209 and low error rates (MAE = 0.015; RMSE = 0.102) which meant that the model worked well in predicting correct readiness scores. Gradient-Boosted Trees and MLP, on the other hand, were less accurate (in terms of R 2, 0.305 and 0.038, respectively) and had an error rate, though both of these referred

to a binned accuracy of 100 percent (that is, both of these models predicted the level of readiness in a particular organization into the correct category), which means that each of them was an accurate classification. This analogy identifies a strength segregation whereby the Logistic Regression is more effective when dealing with fine-grained and continuous measurement that are valuable when conducting governance audits, and the use of Gradient-Boosted Trees and MLP Classifier are more effective when dealing with coarse-grained and policy-based decisions where tier classification is required.

Discussion

The results of the current study will be of great value in the practical application of AI in cloud security. XG Boost was associated with the very high level of accuracy and the ability to identify fine details in attack behaviors, which is essential to identify threats in time and rely on them. Its ability to process data structured in a network format makes it a very viable option as an implementation tool. In the meantime, the LSTM model demonstrated significant capabilities with respect to detecting the presence of the attacks developed over time, i. e. ransomware and brute-force logins, since it identified sequential patterns that are beyond the capabilities of individual models. This further supports the notion of using the concept of temporal analysis in models of cybersecurity. Logistic Regression provided accurate and comprehensible forecasts of security preparedness on the prevention side, which simplifies predictors of multi-factor authentication and penetration testing towards general preparedness into the overall preparedness by security teams. Vis-a-vis Gradient-Boosted Trees and Multilayer Perceptron's, both showed moderate success in grouping organizations into risk tier categories. This kind of grouping also assists in monitoring the levels of compliance as well as prioritizing. Nevertheless, the interpretability of both of them is lower and may restrict their eloquence. This may complicate the process of explaining the results in case of the audits or reviews.

The two track is a method of integrating both the threat detection readiness and the prevention preparedness in a similar plan. This means that the existing cloud security management systems are complicated. Organizations ought to come up with effective security posture in order to react quickly to the emergent threats. This helps to reduce the risks at a later date and makes them have confidence in their systems. The results also show that there is no model which can isolate itself and do everything. The interplay of several AI models can cover an increased range of security measures at the same time. Nevertheless, even, there are still such issues as the necessity to decrease the risk of scoring errors and achieve low levels of compliance. New models should also be made to address new styles of attacks. The systems must as well be compatible to the security systems that are currently in practice. AI models can greatly improve the efforts of cloud security, provided that they are created and used carefully. They can spearhead smart policies and contribute to expedent smacking down of threats in institutions.

Chapter 5

5.0 Conclusion

5.1 Challenges & Limitations

Some of the key challenges include those that can be listed in the deployment of AI-based threat detection and prevention preparations in clouds, although they have a bright future. One of the biggest problems is the ethical and safe handling of a large amount of information required to train and test these models. The type of dataset used in the present research is the hybrid because it combines the descriptive features of network traffic with sensitive organizational governance statistics which in most instances contain confidential or proprietary data. This is significant to make sure that this data is guarded against breach, or misuse as a result of shared and multi-tenant necessities of the cloud platforms. Adhering to applicable regulations on privacy, including GDPR, and best practices related to the governance of data on the cloud are not only legal requirements but also part of the best means of ensuring trust of cloud users and stakeholders. Also, ethical criteria are challenged by the idea, that AI uses to categorize activities or state of the organization as the high-risk. Wrong or prejudiced classifications may be of grave consequences like unreasonable panic, waste of resources or poor reputation. Trust in the models used to reach its predictions through AI must be therefore crucial in establishing trust and helping such teams make informed decisions.

The other major shortcoming is associated with quality, its completeness, and representativeness of the employed data to train on. Discussing the real world of clouds it is very heterogeneous covering various infrastructure providers, service models, as well as user behavior. Data set that is insufficiently representative of this diversity faces the threat of giving models that are less generalizable, possibly pitting away, or misclassifying attacks in less prevalent scenarios. Additionally, missing values, noise, or unbalanced class distributions in the acquired data may be an obstacle to accuracy and robustness of the model. Although methods like SMOTE have been utilised in this study to address the problem of class imbalance, such procedures cannot be a complete replacement of intensive and intensive data gathering and pre-processing. Also, most sophisticated AI modules, particularly, neural networks such as LSTM and MLP, act as a black box and it is hard to discern how they work to make their decisions. This fog is not accommodative to the introduction of AI solutions in the settings of operational security where the governance statutes are pushing them further in the way of explainability and auditability. The biggest problem is the dilemma of how to interpret and make the model complex. Finally, AI systems should be adaptable and regularly revised to meet the demands of flexible and ever-shifting environment of cyber threats and cloud architectures. Continuous retraining and validation infrastructure and procedures should also be established hence making deployment difficult. Such practical considerations as latency of detection or integration with the existing

security controls also influence practical effectiveness, which underscores the complexity of the high potential of AI in cloud protection as an issue to be addressed.

5.2 Future Research Direction

In accordance with the contribution being present in this thesis, one should say that there are visible opportunities to enhance AI-based cloud security systems. The flexibility and the power of the threat detection models is one of the important aspects. Although the existing models resulted in a beautiful performance in the field of detection of different attacks, the dynamism in the field of cyber threats requires the unceasing development of AI systems. Future studies should examine how models can capture new flows of information in real time and update to new attack pattern and variant form of cloud use without relying on time that is taken to retrain the model. Such dynamism of learning would enable security tools to be applicable even in the environment where increasingly more sophisticated and advanced threats would exist making the window of vulnerability of security tools smaller and closing towards the automated response.

The other significant means is diversification and expansion of the hybrid data to facilitate an even more comprehensive cloud security image. The available data is appropriate in order to incorporate network traffic telemetry and governance and access control indicators and it has already been relevant in modeling reactive detection and proactive prevention at the same time. Nevertheless, the model could be enhanced to be more precise and sensitive to the context by considering additional sources of data, such as the logs of cloud orchestration services, container security or user behavioral analytics. In addition, the broader environmental and organizational factors should be included, which could improve the prediction of prevention preparedness in the sense that the complex reality of the operation of clouds is being leveraged. Finally, as much as it applied usual machine learning practices and validation processes, future studies should explore novel AI models and readability systems to overcome the limitations achieved and referenced to model transparency and explainability in particular. To build a level of trust and make explainable AI an acceptable solution in the context of a cloud security model, it will be important to make the progress in its development and implementation during the operating conditions based on a very strict compliance and auditing policy.

5.3 Conclusion

This research was targeted at developing and experimenting with various AI models to support preparedness in threat detection and prevention in the cloud computing setups. The study in question presented both the proactive and reactive characteristics of the cloud security by developing a specific hybrid data set that advanced the critical data by integrating synthetical network traffic data with the complex governance and access control predictors. They managed to provide the empirical evidence that the models of the type of the Random Forest, the XGBoost, and the Long Short-Term Memory networks can be employed to attain the high level

of accuracy when one should identify the threats in the multi-class form, i.e., recognize between the normal traffic and more advanced type of attacks, i.e. DDoS, ransomware, and brute-force log-in attempts. The Multilayer Perceptrons, Logistic Regression and Gradient-Boosted Trees had a strong prediction capability. These models were able to predict scores of organizational prevention readiness with constant accuracy. These kinds of scores gave very articulated response on the maturity of several security postures. The models suggested the organization and administration of activities with structured approaches of learning. Though these results prove the fact that AI approaches can provide unimaginable opportunities, they also show that there are certain significant concerns that ought to be thought of with a heavy weight. Models are expected to be clear and process personal information in a safe and fair way. They are also supposed to respond promptly to quick and complex cyberattacks of many systems. This may be incorporated in order to maintain confidence and good behavior in the undertaking of security checks. Dual-pipeline model showed that it is possible to make a combination of detection and prevention checks. Under this type of model, additional information was availed about security posture to organize activities. It helped to track the rapidly emerging threats and improve the governance system in the long-term. The strict feature engineering and the balanced operations in data were also beneficial in giving directions in future.

This research is a contribution to the growing body of research on AI in cloud security development. The framework and models possess the potential to guide the professionals who are planning to use AI in security operations. It is possible to improve this base in the future and carry out a test with new data or structures. These findings show that AI might prove powerful to detect cyber threats and notify prevention strategies.

Chapter 6

6.0 References

1. R. Gudimetla and R. Kotha, “AI Agents for Cloud Reliability: Autonomous Threat Detection and Mitigation Aligned with Site Reliability Engineering Principles”, 4th IEEE International Conference on AI in Cybersecurity (ICAIC), 5–7 February 2025, University of Houston, Texas, USA, 2025. doi: 10.1109/ICAIC63015.2025.10849322.
2. R. Cheerla, “AI-Driven Cloud Security: The Future of Safeguarding Sensitive Data in the Digital Age,” Journal of Cloud Security and Privacy, 2024.
3. D. Sable, “Enhancing Cloud Security: A Multi-Factor Authentication and Adaptive Cryptography Approach Using Machine Learning Techniques,” IEEE Transactions on Cloud Computing, 2024.
4. H. Arif et al., “Ethical and Explainable AI for Cybersecurity: Challenges and Considerations,” Journal of Information Ethics and Privacy, 2024. [Online].
5. CSEIT Editorial Team, “Secure Cloud Access Control Using AI and Cryptographic Protocols,” International Journal of Computer Sciences and Engineering, 2024. [Online].
6. R. Cheerla, “AI-Driven Cloud Governance and Operational Risks,” Journal of Information Systems and Cloud Governance, 2024. [Online].
7. N. Tutubala, “A Hybrid Framework to Improve Data Security in Cloud Computing”, 78-1-6654-1042-7/21/$31.00 ©2021 IEEE, 2021. doi: 10.1109/BDKCSE53180.2021.9627294.
8. “Cloud Computing Security Challenges and Threats”, 978-1-7281-6939-2/20/$31.00 ©2020 IEEE, 2020.
9. “Cloud computing security issues and its solution: A review”, 978-9-3805-4416-8/15/$31.00 ©2015 IEEE, 2015.
10. “A Security Threats Measurement Model for Reducing Cloud Computing Security Risk”, 2015 9th International Conference on Innovative Mobile and Internet Services in Ubiquitous Computing, 2009. doi: 10.1109/IMIS.2015.64414.
11. “A Survey of Cloud Computing Security Challenges, Issues and their Countermeasures”, Proceedings of the Third International Conference on I-SMAC (IoT in Social, Mobile, Analytics and Cloud) (I-SMAC 2019), 2019.
12. “A Survey on Security Threats in Cloud Computing Service Models”, Proceedings of the Sixth International Conference on Intelligent Computing and Control Systems (ICICCS 2022), 2022. doi: 10.1109/ICICCS53718.2022.9788148.
13. “A Systematic Approach towards Security Concerns in Cloud”, Proceedings of the Second International Conference on Electronics and Renewable Systems (ICEARS-2023), 2020. doi: 10.1109/ICEARS56392.2023.10085437.

14. “A Systematic Review of the Security in Cloud Computing: Data Integrity, Confidentiality and Availability”, The 2020 IEEE International Conference on Computing, Power and Communication Technologies (GUCON), 2020.
15. “An Efficient Privacy-Preserving Access Control Scheme for Cloud Computing Services”, JOURNAL OF LATEX CLASS FILES, VOL. 14, NO. 8, AUGUST 2021, 2021. doi: 10.1109/TCE.2025.3534833.
16. “Analytical Review of Data Security in Cloud Computing”, 2021 2nd International Conference on Intelligent Engineering and Management (ICIEM), 2021. doi: 10.1109/ICIEM51511.2021.9445355.
17. “Cloud Computing Security: Amazon Web Service”, 2015 Fifth International Conference on Advanced Computing & Communication Technologies, 2006. doi: 10.1109/ACCT.2015.20501.
18. “Data Access Security in Cloud Computing: A Review”, 2018 International Conference on Computing, Power and Communication Technologies (GUCON), 2018.
19. “Data Protection: The Cloud Security Perspective”, Fully Homomorphic Encryption (FHE), 2022. doi: 10.1109/INCET54531.2022.9825151.
20. “Enhancing Cloud Security: A Multi-Factor Authentication and Adaptive Cryptography Approach Using Machine Learning Techniques”, IEEE Open Journal of the Computer Society, 2022. doi: 10.1109/OJIM.2022.1234567.
21. “One Quantifiable Security Evaluation Model for Cloud Computing Platform”, Sixth International Conference on Advanced Cloud and Big Data, 2018. doi: 10.1109/CBD.2018.00043.
22. “Pattern Hiding and Authorized Searchable Encryption for Data Sharing in Cloud Storage”, JOURNAL OF LATEX CLASS FILES, VOL. X, NO. X, MARCH 2024, 2021. doi: 10.1109/TKDE.2025.3537613.
23. “Research on Computer Network Security Protection System Based on Level Protection in Cloud Computing Environment”, 2021 IEEE International Conference on Advances in Electrical Engineering and Computer Applications (AEECA), 2021. doi: 10.1109/AEECA52519.2021.9574216.
24. “Securing Data Workflows: An Insight into Cloud Security”, 978-1-6654-3919-0/21/$31.00 ©2021 IEEE, 2016. doi: 10.1109/ICACC-202152719.2021.9708115.
25. “Security Enabled Framework to Access Information in Cloud Environment”, 2022 International Conference on Machine Learning, Big Data, Cloud and Parallel Computing (COM-IT-CON), 2018. doi: 10.1109/COM-IT-CON54601.2022.9850906.

Plagiarism Report

213-35-3190

ORIGINALITY REPORT

18%	15%	11%	11%
SIMILARITY INDEX	INTERNET SOURCES	PUBLICATIONS	STUDENT PAPERS

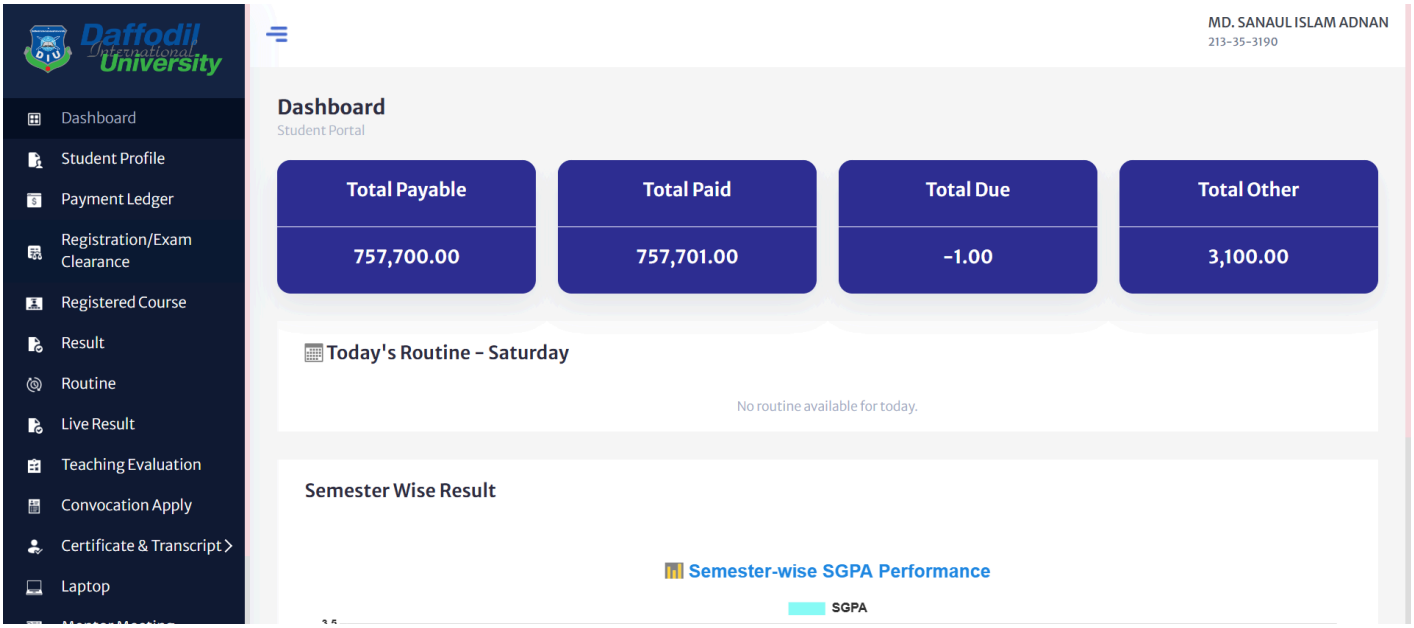
PRIMARY SOURCES

1	Submitted to Daffodil International University Student Paper	1%
2	dspace.daffodilvarsity.edu.bd:8080 Internet Source	1%
3	www.coursehero.com Internet Source	1%
4	www.jatit.org Internet Source	1%
5	Submitted to Dublin City University Student Paper	<1%
6	Submitted to University of Salford Student Paper	<1%
7	Submitted to University of North Texas Student Paper	<1%
8	Submitted to Suleyman Demirel University, Kazakhstan Student Paper	<1%
9	Submitted to University of Hertfordshire Student Paper	<1%
10	Submitted to Vaal University of Technology Student Paper	<1%
11	sndbx.library.tuc.gr Internet Source	<1%
12	Submitted to Arab Open University Student Paper	<1%

13	Submitted to Victoria University Student Paper	<1 %
14	dspace.bracu.ac.bd Internet Source	<1 %
15	stax.strath.ac.uk Internet Source	<1 %
16	Submitted to Buckinghamshire Chilterns University College Student Paper	<1 %
17	Submitted to University of Portsmouth Student Paper	<1 %
18	journalwjarr.com Internet Source	<1 %
19	Submitted to University of Bolton Student Paper	<1 %
20	Submitted to University of Maryland, University College Student Paper	<1 %
21	www.askpython.com Internet Source	<1 %
22	eprint.scholarsrepository.com Internet Source	<1 %
23	Hao Zhou, Yuting Peng, Ruopeng Zhang, Yushan He, Lin Li, Wei Xiao. "GS-DeepLabV3+: A Mountain Tea Disease Segmentation Network Based on Improved Shuffle Attention and Gated Multidimensional Feature Extraction", Crop Protection, 2024 Publication	<1 %
24	Submitted to Rutgers University, New Brunswick Student Paper	<1 %

25	www.frontiersin.org Internet Source	<1 %
26	Submitted to University of Teesside Student Paper	<1 %
27	accentsjournals.org Internet Source	<1 %
28	jst-ud.vn Internet Source	<1 %
29	www.researchwap.com Internet Source	<1 %
30	Submitted to University of Western Sydney Student Paper	<1 %
31	ijrpr.com Internet Source	<1 %
32	www.ai.uga.edu Internet Source	<1 %
33	www.irjmets.com Internet Source	<1 %
34	www.mdpi.com Internet Source	<1 %
35	Submitted to Universidad Carlos III de Madrid - EUR Student Paper	<1 %
36	Submitted to University of South Australia Student Paper	<1 %
37	kc.umn.ac.id Internet Source	<1 %
38	www.mygreatlearning.com Internet Source	<1 %
39	Submitted to Liverpool John Moores University	<1 %

Accounts Clearance



Library Clearance