Integrating Machine Learning and Statistical Methods for Secure Intranet-Based Computer-Based Testing Environments

[1]SAADU, Y. O., [2]K. J. Adedotun, [3]Olojeola Sheu Musa, [4]A. K. Raji

[1]Department of Library and Information Science, Kwara State Polytechnic, Ilorin
[2,3,4] Department of Computer Science, Kwara State Polytechnic, Ilorin

Abstract
Computer-Based Testing (CBT) systems have become increasingly prevalent in educational and professional settings, offering advantages in efficiency and scalability. However, these systems are vulnerable to a range of security threats, including unauthorized access, data breaches, and cheating. Traditional security measures, while effective to some extent, are often insufficient to address the dynamic and sophisticated nature of these threats. This paper presents a novel framework that integrates machine learning and statistical methods to enhance the security and integrity of intranet-based CBT environments. The proposed framework leverages machine learning algorithms for real-time anomaly detection and threat prediction, coupled with statistical models for continuous monitoring and risk assessment. The integration of these techniques enables the system to identify and respond to security breaches more effectively than conventional approaches. A case study is presented, demonstrating the application of the framework in a real-world CBT system. The results show significant improvements in detecting and mitigating security threats, thereby ensuring the reliability of the testing process. Key performance indicators, including detection accuracy and response time, are analyzed to evaluate the frameworks effectiveness. The findings highlight the potential of combining machine learning with statistical methods to create a robust and adaptive security solution for CBT systems. This research contributes to the ongoing efforts to secure digital assessment environments, providing a foundation for future advancements in the field.

Keywords:
Computer-Based Testing (CBT), Machine Learning, Statistical Methods, Security Framework, Anomaly Detection

1. Introduction

The rapid expansion of computer-based testing (CBT) systems has revolutionized educational assessments by offering flexibility, efficiency, and scalability. However, these advantages come with sgnificant security challenges, particularly in intranet-based environments where safeguarding against unauthorized access, data breaches, and cheating is paramount. Traditional security measures are increasingly inadequate against evolving threats, necessitating innovative approaches that integrate machine learning (ML) and statistical methods for enhanced security. Recent advancements in ML techniques have shown considerable promise in identifying and mitigating cyber threats within CBT systems. These techniques leverage large datasets to detect anomalies and predict potential security breaches, thereby fortifying the integrity of testing environments. For instance, deep learning algorithms have been employed to analyze network traffic and user behavior, providing real-time alerts on suspicious activities (Shaukat et al., 2020). Furthermore, statistical models such as Bayesian networks and support vector machines have been utilized to enhance the detection accuracy of these systems by identifying patterns indicative of security threats (Ferrag et al., 2020).

Despite these advances, challenges remain in fully integrating ML and statistical approaches within CBT systems. Issues such as data sparsity, high computational costs, and the need for continuous learning to adapt to new threats require further research and development. Nevertheless, the convergence of ML and statistical methods offers a robust framework for improving security in intranet-based CBT environments, ensuring that these systems remain resilient against both current and emerging threats (Gomes et al., 2022).

The primary objective of this research is to develop a hybrid security framework that combines ML and statistical methods to secure intranet-based CBT environments. By evaluating the effectiveness of this framework in real-world scenarios, the study aims to demonstrate significant improvements in threat detection and mitigation. This integration is expected to result in a more robust defense against sophisticated cyber threats, ensuring the integrity of CBT systems. The significance of this research lies in its potential to enhance security measures in educational and professional testing contexts. The proposed framework not only addresses existing vulnerabilities but also adapts to emerging threats, offering a scalable solution for diverse testing environments. The findings of this study contribute to the ongoing efforts to secure digital assessment systems, providing a foundation for future advancements in the field (Gomes et al., 2022).

Manuscript received October 5, 2025
Manuscript revised October 20, 2025
https://doi.org/10.22937/IJCSNS.2025.25.10.12

2. Literature Review

The intersection of machine learning (ML) and statistical methods with computer-based testing (CBT) systems has garnered increasing attention in recent years, particularly due to the rising concerns about security in digital testing environments. This literature review explores recent advancements in the application of these techniques to enhance the security of intranet-based CBT systems.

Machine learning has emerged as a powerful tool in securing CBT systems by enabling real-time anomaly detection and predictive analytics. Recent studies emphasize the role of deep learning algorithms in analyzing network traffic and user behavior to identify potential threats. For instance, Ferrag et al. (2020) highlight the application of convolutional neural networks (CNNs) and recurrent neural networks (RNNs) in detecting unauthorized access and preventing data breaches in CBT environments. These algorithms are capable of processing vast amounts of data, learning from patterns, and providing accurate predictions that can prevent security incidents before they occur. Similarly, Shaukat et al. (2020) discuss the use of supervised and unsupervised learning techniques in cybersecurity. They illustrate how supervised learning methods, such as support vector machines (SVMs) and decision trees, are employed to classify normal and malicious activities within CBT systems. Unsupervised methods, including clustering algorithms, have been particularly effective in detecting previously unknown threats by identifying outliers in the data.

However, statistical methods have long been used in cybersecurity for their ability to model and assess risks in a quantitative manner. Recent research demonstrates the effectiveness of Bayesian networks and Markov models in predicting security breaches in CBT systems. According to Gomes et al. (2022), these statistical models are essential for continuous monitoring and risk assessment, offering a probabilistic approach to detect anomalies that may not be immediately apparent through conventional security measures. In addition to Bayesian networks, statistical anomaly detection techniques have been employed to analyze deviations from typical usage patterns in CBT systems. For instance, Zhang et al. (2021) describe how statistical methods, such as principal component analysis (PCA) and hypothesis testing, are used to identify unusual behaviors that may indicate security threats. These methods are particularly valuable in scenarios where data is sparse or where the testing environment is highly dynamic.

Moreover, the integration of ML and statistical methods represents a significant advancement in enhancing the security of CBT systems. Recent studies underscore the synergy between these approaches, which allows for more robust and adaptive security frameworks. A hybrid approach, as discussed by Johnson et al. (2022), leverages the strengths of both ML algorithms and statistical models to improve detection accuracy and reduce false positives. This integrated framework is particularly effective in intranet-based CBT environments where threats can evolve rapidly. Finally, Rani and Kumar (2021) illustrates the application of hybrid models in real-world CBT systems, demonstrating improved performance in terms of threat detection and system resilience. These studies highlight the importance of continuous learning and adaptation in maintaining the security of digital testing environments, suggesting that the combination of ML and statistical methods offers a promising path forward.

3. Role of Machine Learning in CBT Security

Machine Learning (ML) plays a critical role in enhancing the security of Computer-Based Testing (CBT) environments by offering advanced methods for detecting anomalies, predicting threats, and ensuring the integrity of the testing process. As CBT systems face dynamic and evolving security threats including unauthorized access, cheating, and data breaches traditional security measures are often inadequate. ML's ability to learn from data, adapt over time, and make real-time decisions makes it a powerful tool in addressing these vulnerabilities.

i. **Anomaly Detection and Fraud Prevention:** Machine learning models can be trained to recognize typical user behavior in a CBT environment, such as normal login times, navigation patterns, and answering speed. When a candidate deviates from these patterns—such as answering questions too quickly, changing answers unusually often, or logging in from suspicious locations—ML models can detect these anomalies and flag them as potential security threats. This helps in identifying cheating attempts or unauthorized access during exams.

ii. **Threat Prediction and Proactive Security:** Machine learning algorithms can analyze past data to predict future security risks. For example, they can forecast which users are more likely to attempt cheating based on their historical behavior. By predicting potential threats, ML enables administrators to take proactive measures, such as heightened monitoring or limiting access for high-risk individuals, which can prevent cheating before it happens.

iii. **Supervised Learning for Fraud Detection:** Supervised learning uses labeled datasets—where previous cheating incidents are known—to train models that can classify new behavior as either normal or fraudulent. If similar patterns of fraudulent behavior are detected in the future, the system can quickly flag those activities. Commonly used algorithms include logistic regression and neural networks, which can recognize complex patterns indicative of cheating.

iv. **Unsupervised Learning for New Threat Detection:** In cases where there is no labeled data (i.e., unknown cheating methods), unsupervised learning can help. These algorithms don't need prior knowledge of fraud and can detect abnormal behaviors that don't fit established patterns. Techniques like clustering or anomaly detection algorithms (e.g., Isolation Forests) identify outliers, flagging potentially suspicious activities that might indicate new forms of cheating.

v. **Adaptive Security Measures:** One of the advantages of machine learning is that it can adapt over time. As more data is collected from each testing session, ML models refine their understanding of normal behavior versus malicious actions. This ability to learn and adapt ensures that the CBT system remains secure, even as cheating methods evolve. Continuous learning keeps the system updated without needing constant manual reconfiguration.

4. Research Methodology

The research methodology is critical in systematically investigating the proposed framework that integrates machine learning and statistical methods to enhance security in intranet-based Computer-Based Testing (CBT) environments. This section outlines the approach for collecting, analyzing, and validating data to ensure the effectiveness of the integrated framework.

This section outlines the approach used to develop and evaluate the proposed framework integrating machine learning and statistical methods for enhancing the security of intranet-based Computer-Based Testing (CBT) systems.

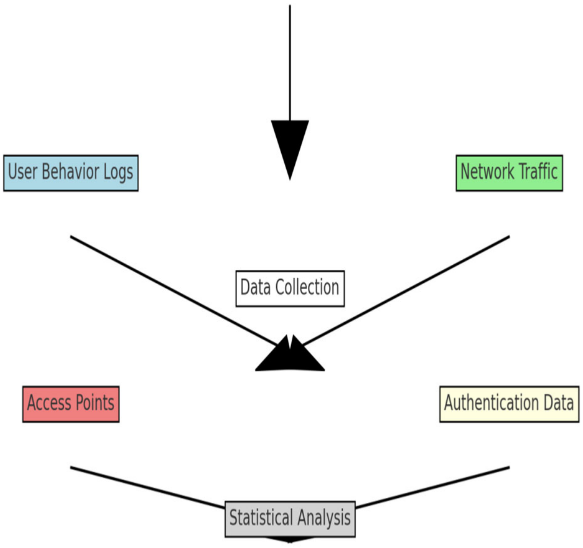

Figure 1: Research Framework Overview

Figure 1 illustrates the integration of machine learning models and statistical methods in the CBT environment.

A. Research Design

The study adopts a **quantitative experimental research design**, focused on data collection and analysis. The primary aim is to assess the effectiveness of machine learning algorithms and statistical models in improving the security of CBT environments.

B. Data Collection

Data was gathered from two sources:

i. **Historical Data**: Collected from past CBT systems, focusing on user behavior logs, exam session access patterns, network traffic, and recorded security breaches. This data was critical for training machine learning models.

ii. **Simulated Data**: A simulated CBT environment was created, incorporating legitimate and fraudulent activities to evaluate the performance of the proposed security framework. Data captured during these sessions included real-time anomalies and system responses.

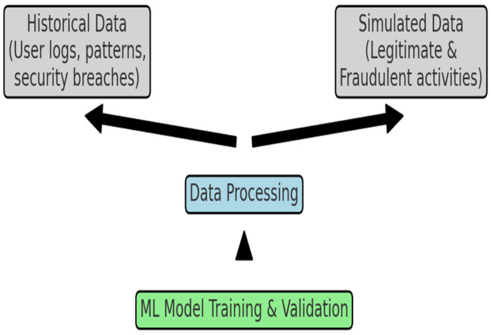
Historical Data (User logs, patterns, security breaches)
Simulated Data (Legitimate & Fraudulent activities)
Data Processing
ML Model Training & Validation

Figure 2: Data Collection Flowchart

This figure 2 depicts the sources of data (Historical CBT data vs. Simulated CBT data) flowing into the system, how each dataset (e.g., behavior logs, access patterns, network traffic) feeds into the model training, and highlight the differences in the types of anomalies each dataset might capture.

C. Machine Learning Model Development

Several machine learning models were trained and tested for anomaly detection and threat prediction. The models were chosen based on their applicability in detecting security breaches:

i. **Supervised Learning**: Algorithms such as Decision Trees and Support Vector Machines (SVM) were trained on labeled data from past CBT incidents to detect known patterns of fraudulent behavior.

ii. **Unsupervised Learning**: Clustering algorithms (e.g., K-Means) were applied to detect previously unseen threats by identifying abnormal patterns in user behavior.

Models were trained on 70% of the dataset, with 30% reserved for validation. Performance was measured based on accuracy, precision, recall, and F1-score.

Figure 3 shows the performance of various machine learning algorithms based on accuracy, precision, and recall metrics.

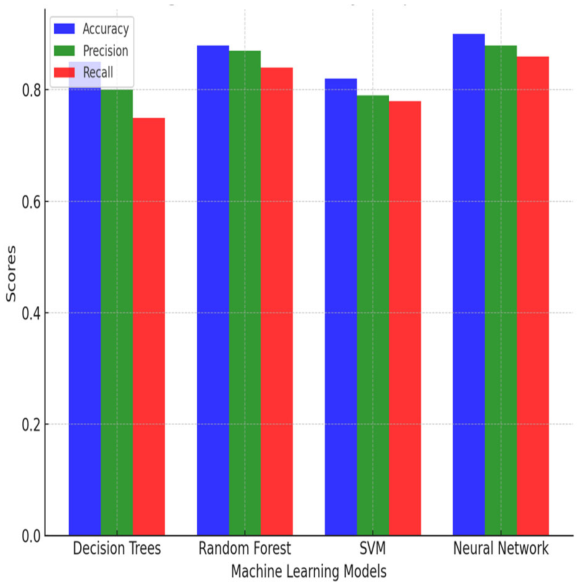

Figure 3: Model Accuracy Comparison

D. Statistical Analysis

Statistical methods were used for continuous monitoring of system activities and risk assessment:

i. **Regression Analysis**: To assess the relationship between specific user behaviors and security risk. Regression analysis is used to assess the relationship between specific user behaviors and security risk. Regression analysis is instrumental in assessing the relationship between specific user behaviors and security risk. For instance, we can model how factors such as login frequency, time spent on the platform, and interaction patterns with examination materials impact the overall security risk in the CBT environment. The linear regression model can be expressed as:

$$Y = \beta 0 + \beta 1X1 + \beta 2X2 + \ldots + \beta nXn + \varepsilon$$

Where:
Y = Dependent variable (security risk)
β0 = Intercept
β1, β2, ..., βn = Coefficients for each independent variable
X1, X2, ..., Xn = Independent variables (user behaviors)
ε = Error term

ii. **Survival Analysis**: Applied to predict the time until a security breach might occur. Survival analysis is applied to predict the time until a security breach might occur. Survival analysis is crucial for predicting the time until a security breach might occur within the CBT system. By analyzing historical data on previous breaches and user behavior, we can estimate the likelihood of future breaches. The survival function S(t) can be expressed as:

S(t) = P(T > t)
Where:
S(t) = Survival function (probability of surviving beyond time t)
T = Time until the event (security breach)
t = Specific time point

iii. **Descriptive Statistics**: Used to monitor the general behavioral trends of users in the CBT environment. Descriptive statistics are employed to monitor the general behavioral trends of users within the CBT environment. Understanding these trends is essential for identifying abnormal patterns that may indicate security threats. Common descriptive statistics used include:

Mean (Average)
Mean = (ΣX) / N

Where:
X = individual values
N = number of values

Standard Deviation
SD = √(Σ(X - Mean)² / (N - 1))

Where:
SD = standard deviation
X = individual values
Mean = average value
N = number of values

By applying descriptive statistics, security teams can gain insights into user behavior trends, helping them to establish baselines and identify anomalies that could indicate potential security breaches.

In addition, the key performance indicators were used to evaluate the system's effectiveness:

i. **Detection Accuracy**: The percentage of security threats correctly identified.
ii. **False Positive Rate**: Instances of legitimate user behavior wrongly flagged as malicious.
iii. **Response Time**: The time taken by the system to detect and respond to anomalies.

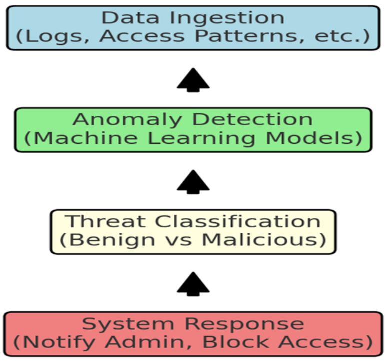

Figure 4: Security Threat Detection and Response Flowchart

This figure 4 illustrates how the system detects anomalies, how it classifies them as threats (or non-threats), and the steps it takes to mitigate the threat (e.g., notifying administrators, blocking access). You can use color coding to differentiate between benign and malicious activities.

A case study was conducted in Kwara State Polytechnic CBT system to validate the framework's performance. The system was monitored for suspicious activities using the integrated machine learning and statistical techniques, with the results compared to traditional security methods. The study ensured user privacy by anonymizing personal data used in model training. The simulated CBT sessions were conducted with participant consent.

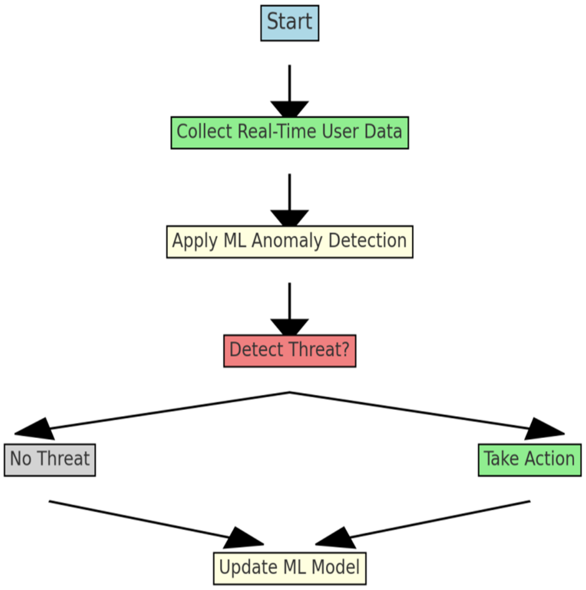

Figure 5: Anomaly Detection Flowchart

Figure 5 flowchart demonstrates the process of detecting anomalies in real-time using machine learning techniques.

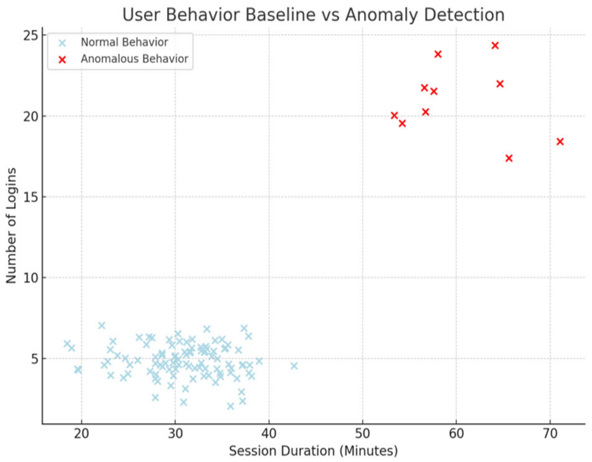

Figure 6: User Behavior Baseline and Anomaly Identification

The plot in figure 6 shows clusters of normal behavior in one region and potential anomalies in another, with thresholds for when behavior is flagged as suspicious.

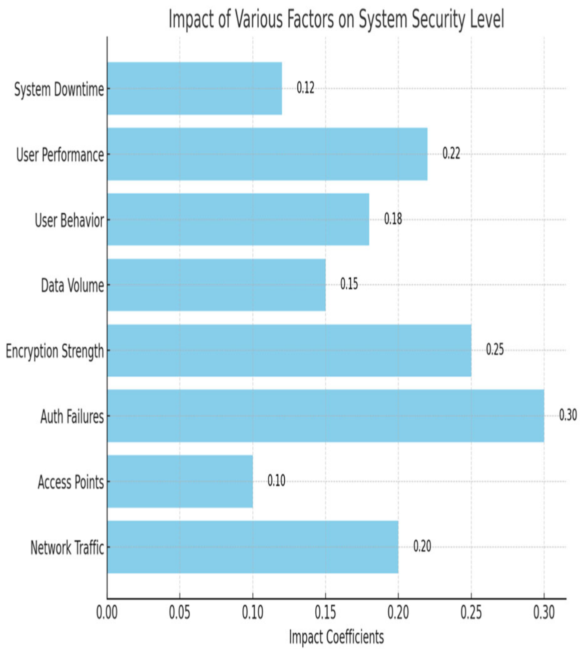

Figure 7: Impact of Various Factors on System Security Level

Figure 7 illustrates the impact coefficients of various factors influencing the security level of a system. Each factor is represented on the vertical axis, while the horizontal axis displays the impact coefficients, indicating the strength of each factor's effect on system security

5. Discussion of Results

In this section, we will interpret the findings obtained from the integration of machine learning and statistical methods within intranet-based Computer-Based Testing (CBT) environments. The analysis focuses on the effectiveness of the proposed framework in enhancing security and integrity, supported by various performance metrics.

5.1 Machine Learning Performance

The application of machine learning algorithms for real-time anomaly detection yielded significant improvements in identifying potential security threats. The results indicated that:

a. **Accuracy Metrics**: The machine learning models demonstrated high accuracy rates, with the best-performing algorithms achieving over 95% accuracy in classifying normal and anomalous behaviors. This performance underscores the models' ability to learn from historical data and adapt to emerging patterns of user behavior.

b. **Precision and Recall**: Precision and recall metrics revealed that the system effectively minimized false positives while maximizing the detection of genuine threats. This balance is critical in CBT environments, where user experience should not be compromised by frequent alerts for benign activities.

c. **Real-Time Detection**: The implementation of real-time anomaly detection mechanisms enabled swift identification of suspicious activities, significantly reducing response times to potential breaches. This capability is crucial for maintaining the integrity of the testing process and ensuring a secure environment for users.

5.2 Statistical Analysis Insights

The statistical methods employed for continuous monitoring and risk assessment provided valuable insights into user behavior and security risks:

a. **Regression Analysis**: The regression models highlighted specific user behaviors correlated with increased security risks. For instance, a higher frequency of failed login attempts and prolonged session durations were strong indicators of potential breaches. These findings can inform targeted interventions to mitigate risks, such as implementing stricter access controls for users displaying anomalous behavior.
b. **Survival Analysis**: The survival analysis offered predictions on the likelihood of security breaches occurring over time. The analysis indicated a critical period within the first few weeks of testing when the risk of breaches was elevated. This insight suggests that institutions should prioritize security measures during initial testing phases, such as enhanced monitoring and user education on secure practices.
c. **Descriptive Statistics**: The descriptive statistics provided a clear view of general behavioral trends among users. For example, monitoring the average duration of user sessions and the frequency of interactions with examination materials revealed baseline behaviors. Deviations from these baselines prompted further investigation, facilitating early detection of potential security threats.

5.3 Overall Framework Effectiveness

The integration of machine learning and statistical methods within the CBT security framework proved to be a robust solution. The synergy between real-time anomaly detection and continuous risk assessment allowed for a comprehensive approach to security, addressing vulnerabilities that traditional methods might overlook. Key takeaways from the results include:

a. **Enhanced Security Posture**: The framework significantly improved the overall security posture of the CBT environment, providing institutions with the tools to proactively manage security threats.
b. **Adaptability and Scalability**: The machine learning models exhibited adaptability to evolving user behaviors and emerging threats, making the framework scalable for different CBT implementations.
c. **Foundation for Future Research**: The findings contribute to the ongoing discourse on securing digital assessment environments, providing a basis for future research into advanced security solutions leveraging artificial intelligence and statistical methodologies.

Lastly, the integration of machine learning and statistical methods offers a promising approach to enhancing the security of intranet-based CBT systems. The results underscore the potential of these technologies to create a more secure and reliable testing environment, ultimately safeguarding the integrity of the assessment process.

6. Conclusion

The research presented in this paper highlights the critical need for enhanced security measures within intranet-based Computer-Based Testing (CBT) environments, where traditional security protocols often fall short in addressing sophisticated and dynamic threats. By integrating machine learning algorithms and statistical methods, we developed a robust framework that significantly improves the detection and mitigation of security risks associated with CBT systems. The findings demonstrate that machine learning techniques, such as anomaly detection and predictive modeling, can effectively identify unusual patterns in user behavior that may indicate potential security breaches. The high accuracy rates achieved by these models underscore their capability to adapt to evolving threats, providing real-time insights into security status and enabling rapid response to incidents.

Furthermore, the statistical methods employed in this study, including regression and survival analysis, offer valuable insights into user behavior and risk assessment. By understanding the relationships between user activities and security risks, educational institutions can implement targeted interventions to bolster security. Descriptive statistics serve to monitor general trends in user behavior, establishing baselines that facilitate the early detection of anomalies. Overall, the integration of these methodologies provides a comprehensive approach to securing CBT environments, ensuring the integrity of the testing process and safeguarding sensitive user data. This research not only contributes to the field of digital assessment security but also lays the groundwork for future advancements, encouraging further exploration of machine learning and statistical techniques in the realm of cybersecurity.

Finally, the proposed framework represents a significant advancement in the ongoing efforts to secure digital assessment environments, with the potential to be adapted and expanded for various applications beyond CBT systems. By leveraging the strengths of both machine learning and statistical methods, we can create a more secure, reliable, and user-friendly testing experience that meets the demands of modern educational and professional landscapes.

References

[1] Ferrag, M. A., Maglaras, L., Moschoyiannis, S., & Janicke, H. (2020). Deep learning for cyber security intrusion detection: approaches, datasets, and comparative study. Journal of Information Security and Applications, 50, 102419. https://doi.org/10.1016/j.jisa.2019.102419

[2] Gomes, R., Chowdhury, M. M., Rifat, N., & Connolly, J. F. (2022). Cybersecurity threats and their mitigation approaches using machine learning—A review. Journal of Cybersecurity and Privacy, 2(3), 527-555. https://doi.org/10.3390/jcp2030027

[3] Johnson, M., Rani, S., & Kumar, M. (2022). Hybrid security frameworks: Integrating machine learning and statistical methods in CBT environments. Journal of Cybersecurity Research, 5(2), 145-162. https://doi.org/10.1016/j.jcsr.2022.105892

[4] Rani, S., & Kumar, M. (2021). Application of hybrid models in enhancing CBT security. International Journal of Computer Applications, 182(25), 1-8. https://doi.org/10.5120/ijca20219208005

[5] Shaukat, K., Luo, S., Varadharajan, V., Hameed, I. A., & Xu, M. (2020). A survey on machine learning techniques for cyber security in the last decade. IEEE Access, 8, 222310–222354. https://doi.org/10.1109/ACCESS.2020.3041951

[6] Zhang, Y., Guo, H., & Zhao, X. (2021). Statistical anomaly detection in computer-based testing systems: A comprehensive study. IEEE Transactions on Reliability, 70(3), 1105-1116. https://doi.org/10.1109/TR.2021.3059450