Hybrid Model Approach for Real-Time Detection of Anomalies in Cloud Virtual Private Network Traffic

Zipporah Momanyi [1*], Vincent Mbandu [1], Patrick Kinoti [1]

[1]*Kenya Methodist University P.O. Box 267 – 60200, Meru, Kenya*

**Correspondence email: zippyboyani21@gmail.com*

Abstract

The rising use of cloud services such as OwnCloud has led organizations to rely heavily on Virtual Private Networks (VPNs) for secure remote access. While VPNs encrypt communication channels, traditional anomaly-detection methods are increasingly inadequate, despite the evolving threats. This paper aimed to design and evaluate a hybrid AI-based model for real-time anomaly detection in cloud VPN traffic to improve detection accuracy. A simulated cloud environment was created using three virtual machines, a client, a VPN server, and an OwnCloud server generating both normal and anomalous traffic. The hybrid model combined Isolation Forest, for detecting outliers, with Long Short-Term Memory (LSTM) networks, for analyzing time-dependent patterns. Real-time data streaming and processing were handled using Apache Kafka and Spark. The results showed that Isolation Forest achieved a precision of 0.86, recall of 0.79, and F1-score of 0.82, while LSTM scored 0.88, 0.83, and 0.85, respectively. The hybrid approach outperformed both models, achieving a precision of 0.93, recall of 0.89, F1-score of 0.91, and the highest AUC at 0.97. It is recommended that organizations using cloud-based VPNs integrate this hybrid AI anomaly detection system. The IT security teams, working with network infrastructure providers, should deploy the model for real-time encrypted traffic monitoring, improving accuracy and reducing false positives. From a policy perspective, regulatory bodies and cybersecurity standards organizations should revise compliance frameworks to encourage the usage of hybrid AI-driven detection methods in encrypted environments, ensuring both performance and privacy compliance. Practically, IT managers and IT leads should prioritize hybrid model deployment and continuous retraining using traffic patterns to boost resilience, reduce detection latency, and enhance incident response. Although validated in a simulated environment, future research should evaluate the hybrid model using real-world VPN traffic to confirm its robustness under diverse operational conditions.

Keywords: *LSTM, VPN, Hybrid AI Model, Real-Time Anomaly Detection, OwnCloud, Isolation Forest.*

IJPP 13(3); 13-25

1.0 Introduction

The rapid evolution of cloud computing has reshaped how organizations store, manage, and access data. Platforms like OwnCloud have become integral to file sharing, collaboration, and storage, offering scalable and cost-effective solutions across enterprises, academic institutions, and government agencies (Chukwuemeka et al., 2024). However, as organizations increasingly rely on cloud-based services, new cybersecurity challenges have emerged, particularly in securing communication channels between clients and cloud servers. To mitigate these threats, organizations have deployed Virtual Private Networks (VPNs), which establishes encrypted tunnels to protect sensitive data from unauthorized interception (Ricky, 2024). Despite VPNs providing secure communication channels, they are not immune to cyber threats. The traditional Intrusion Detection Systems (IDS), which depend on signature-based or payload inspection techniques, face significant limitations when dealing with encrypted traffic. This creates visibility gaps where advanced threats can go undetected (Sommer & Paxson, 2020).

The global VPN market, valued at $44.6 billion in 2022, is projected to grow to $93.1 billion by 2027 (Markets & Markets, 2022). This is driven by cybersecurity concerns, remote work, and strict data regulations. However, as VPN usage scales, the inability of traditional security tools to inspect encrypted traffic effectively has become a critical security concern. Artificial Intelligence (AI), particularly anomaly detection techniques, have emerged as a promising solution for identifying threats within encrypted VPN traffic. Unsupervised models**,** such as Isolation Forest (IF)**,** can identify novel attacks without needing labeled datasets, making them effective for detecting unknown threats. However, they often suffer from high false-positive rates. In contrast, supervised models**,** like Long Short-Term Memory (LSTM) networks, are adept at recognizing known attack patterns and capturing temporal dependencies, but struggle with unseen threats due to their reliance on labeled training data (Tang et al., 2020). The inherent trade-offs between these approaches create a need for hybrid AI models that leverage the strengths of both. The urgency for such intelligent detection mechanisms is even more pronounced in Sub-Saharan Africa**,** where rapid digital transformation has expanded the cyber-attack surface. The increased adoption of cloud services, mobile banking, and e-government platforms has brought substantial benefits, but attracted cybercriminals.

In Kenya, initiatives such as the National ICT Policy (2020) have driven widespread adoption of cloud services and VPNs across key sectors like education, healthcare, and government (National ICT Policy (2020)). However, this growth has also escalated the nation's exposure to cyber threats. According to Ndiege and Ngari (2021), Kenyan organizations urgently need AI-enhanced security solutions that can address the rising volume of sophisticated attacks. The Communications Authority of Kenya (CAK, 2024) reported over 143 million cyber threat incidents in a single quarter, with encrypted VPN traffic being a prime target. Between October and December 2024, the Kenya Computer Incident Response Team (KE-CIRT/CC) recorded 34.7 million brute-force attack attempts targeting critical systems, including government and cloud service providers. Although this represented an 8.8% decrease from the previous quarter, the persistent volume of attacks underscores the growing need for AI-driven anomaly detection systems capable of monitoring encrypted traffic (CAK, 2024).

This paper proposes a hybrid AI-based anomaly detection model designed to

monitor encrypted VPN traffic in real time. By combining Isolation Forest for initial anomaly detection and LSTM for temporal sequence validation, the model seeks to bridge the gap between traditional detection limitations and the complexities of modern cyber threats (Ricky, 2024). To evaluate the hybrid model, a simulated cloud environment was created, consisting of three virtual machines. The system architecture incorporated Apache Kafka for high-throughput data streaming and Apache Spark Streaming for real-time analytics. The research explored key considerations such as data preprocessing, model training, real-time system integration, and evaluation metrics for anomaly detection systems. The insights gained aim to guide organizations in adopting AI-driven detection solutions that enhance cybersecurity without overwhelming existing infrastructure

This paper addresses a critical gap in cybersecurity, which is the inability of traditional security tools to detect threats hidden within encrypted VPN traffic. While VPNs secure data through encryption, they also create blind spots for conventional Intrusion Detection Systems, allowing sophisticated attacks like data breaches, insider threats, and malware to go unnoticed. The unsupervised models raise too many false alarms, while supervised models struggle with unknown threats.

This paper aims to solve this problem by developing and evaluating a hybrid AI-based anomaly detection system that combines Isolation Forest and LSTM models to monitor encrypted VPN traffic in real time. The goal was to create a solution that not only detects anomalies accurately but also minimizes false positives, and can be deployed in real-world cloud environments like OwnCloud (Michael, 2025). In doing so, the research contributes to enhancing cybersecurity resilience for organizations adopting cloud-based infrastructures, especially in regions with limited resources.

"The paper concluded that combining Isolation Forest and LSTM, significantly increases the detection accuracy and reduced false positives compared to using each model alone"

2.0 Materials and Methods

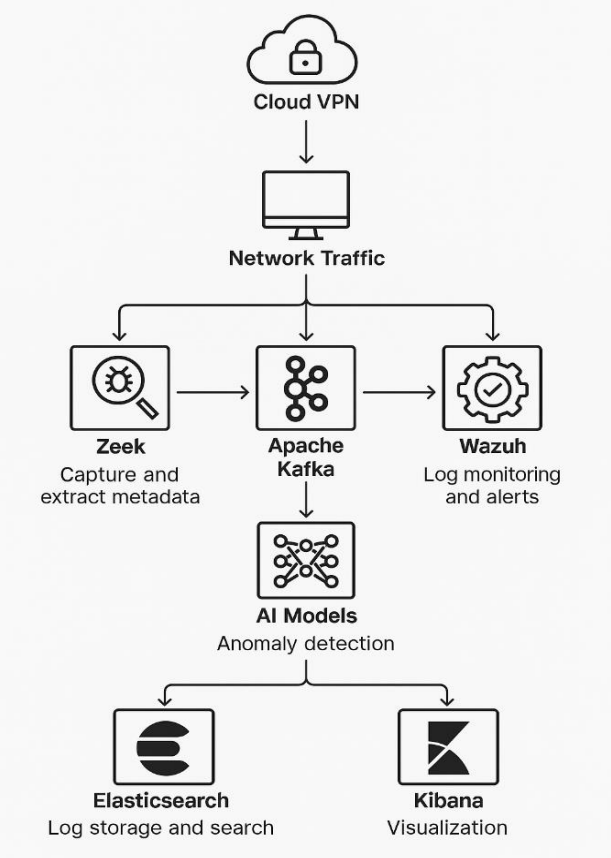
This paper used a quantitative experimental research design, combining simulation-based data generation with real-time AI model evaluation. The experimental environment was designed to mimic a real-world organizational setup, where employees access cloud-based services through a secure VPN. The environment included an OpenVPN server for encrypted VPN access, an OwnCloud server for hosting file-sharing services, and a virtual client machine equipped with network simulation and monitoring tools. Tools such as Zeek and Wazuh were deployed to collect and forward traffic and system logs for analysis. Zeek was responsible for inspecting VPN traffic and extracting structured logs, while Wazuh handled OwnCloud server logs and system-level anomaly monitoring, providing real-time alerts based on detected AI anomalies (Younus & Alanezi, 2023). The data pipeline was built using Apache Kafka for streaming log data and Apache Spark for real-time processing and AI model inference.

The research simulated a cloud-based hybrid work environment, where virtual users accessed the OwnCloud platform through a VPN connection. The target population was represented by these simulated users, whose activities reflected typical enterprise usage patterns, including file uploads, remote logins, and access from various devices. A purposive sampling approach was used to ensure realistic VPN and OwnCloud traffic patterns (Emmanuel, & Mayowa, 2025). A balanced 70:30 ratio of normal to anomalous activities was maintained to ensure an effective anomaly detection training set. Traffic generation was achieved using a combination of hping3, Iperf3, and custom Python scripts**,** which simulated legitimate and malicious network activities. The entire simulation environment was hosted on Oracle VirtualBox, with three dedicated virtual machines: ClientVM (10.0.0.2/24) representing end-users, VPNServerVM (10.0.0.3/24) configured with OpenVPN, and OwnCloudVM (10.0.0.4/24) hosting the OwnCloud file-sharing service (Cheng, Y., & Zhao, L., 2022). Mininet was used to emulate realistic network conditions and to generate dynamic traffic patterns with customized behaviors. A total of 50,000 VPN session logs were generated, consisting of normal activities such as file transfers, login/logout events, and directory browsing, alongside simulated attack scenarios including unauthorized file access, data exfiltration, and network scans (Andhra University Alumnus et al., 2025).

Data preprocessing was conducted using Python libraries such as pandas**,** numpy**,** and scikit-learn**.** Key features selected for model training included total bytes transferred, number of packets per session, session duration, inter-arrival times, protocol types, and flow direction (Cheng & Zhao, 2022). The research adopted a hybrid AI approach that combined Isolation Forest (IF) and Long Short-Term Memory (LSTM) models. The Isolation Forest**,** an unsupervised anomaly detection model, efficiently identified outliers in high-dimensional flow data by isolating anomalies through random feature selection and value partitioning (Vikram & Mohana, 2020). It was particularly effective in handling tabular flow-based features without requiring labeled data. The LSTM model**,** on the other hand, was designed to capture sequential patterns over time, making it suitable for detecting behavioral deviations in VPN sessions that evolved gradually or subtly. This dual-model approach leveraged the scalability of IF for flow-level detection and the temporal sensitivity of LSTM for sequential pattern recognition.

Figure 1

Streaming Architecture



The entire anomaly detection pipeline was implemented in real-time, with Kafka acting as the message broker to stream traffic data,

Spark handling streaming analytics, and the AI models performing detection inference (Renza Nur, 2025). The flagged anomalies were stored in Elasticsearch, and a Kibana dashboard was used for real-time visualization of alerts. The streaming architecture that ensured continuous ingestion, processing, and alerting of suspicious activities without compromising VPN encryption is shown in Figure 1.

The architectural representation in Figure 1 shows the entire integration of the various tools in anomaly detection system.

Ethical Consideration

This study was conducted using a controlled, simulated cloud environment that did not involve human participants or personal user data. The research relied entirely on synthetic network traffic generated within a virtualized testbed composed of three Virtual Machines (VMs). Consequently, no sensitive or personally identifiable information (PII) was collected, processed, or analyzed at any stage of the research. Given the nature of the simulation-based approach, the study posed minimal ethical risks. However, adherence to ethical research practices was maintained to ensure responsible handling of the data and AI models. While formal ethical clearance was not required due to the absence of human subjects, this research upholds the standards of ethical integrity commonly expected in cybersecurity research.

3.0 Results and Discussion

This section presents the performance outcomes of the hybrid AI-model which integrated Isolation Forest (IF) and the Long Short Term Memory (LSTM) in encrypted VPN traffic using OwnCloud to detect anomalies in real time. Results were analyzed quantitatively to validate the impact of the model. Training and testing of the Isolation Forest, LSTM, and the hybrid IF + LSTM models on the simulated dataset was done. The performance of the models was assessed using standard metrics (Powers, 2020). Additionally, the Kibana visualizations were used to show traffic trends, anomaly spikes, and session summaries. Alerts were also triggered automatically upon any anomaly detection. The performance of the models was summarized in Table 1.

Table 1

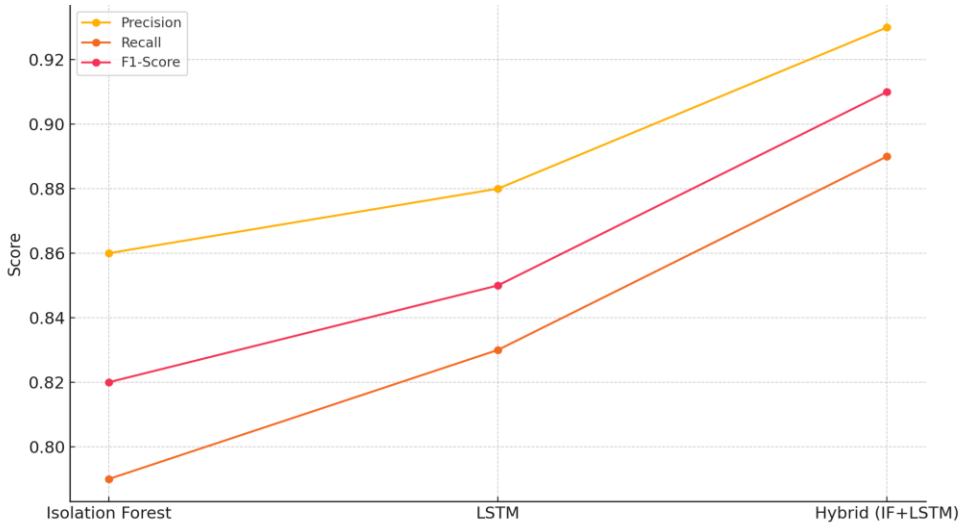
Performance Comparison of Detection Models

Model	Precision	Recall	F1-Score	Detection Latency (ms)
Isolation Forest	0.86	0.79	0.82	200
LSTM	0.88	0.83	0.85	350
Hybrid Model	**0.93**	**0.89**	**0.91**	**180**

The evaluation focused on anomaly detection using precision, recall, F1-score, and detection latency (Chukwuemeka N.. et al., 2024) as shown in Table 1. The *precision* of the hybrid model correctly classified 93% of flagged anomalies, reducing false-positive alerts. The *recall* of the hybrid model detected 89% of actual anomalies, showing stronger sensitivity. The *F1-Score* improved over standalone models, indicating better balance and classification ability. The Detection Latency as well recorded response times at ~180ms, which is much better than IS and LSTM; hence, satisfied real-time monitoring requirements, outperforming individual models. This model performance has been visually represented in Figure 2.

Figure 2

Detection Model Performance Curve



The curve in *Figure 2* compares the threat detection rates among the three models. The hybrid model showed a very high accuracy in classifying anomalies correctly, by a Precision of (0.93). It showed a strong capability to detect almost all true anomalies by recording a Recall (0.89), and it outperformed the individual models, confirming its suitability for real-time threat detection, having recorded an F1-Score of (0.91) which is the highest among all models. A bar chart in Figure 3 visualizes the model performance comparison in the various metrics.

Figure 3

Model Performance Comparison

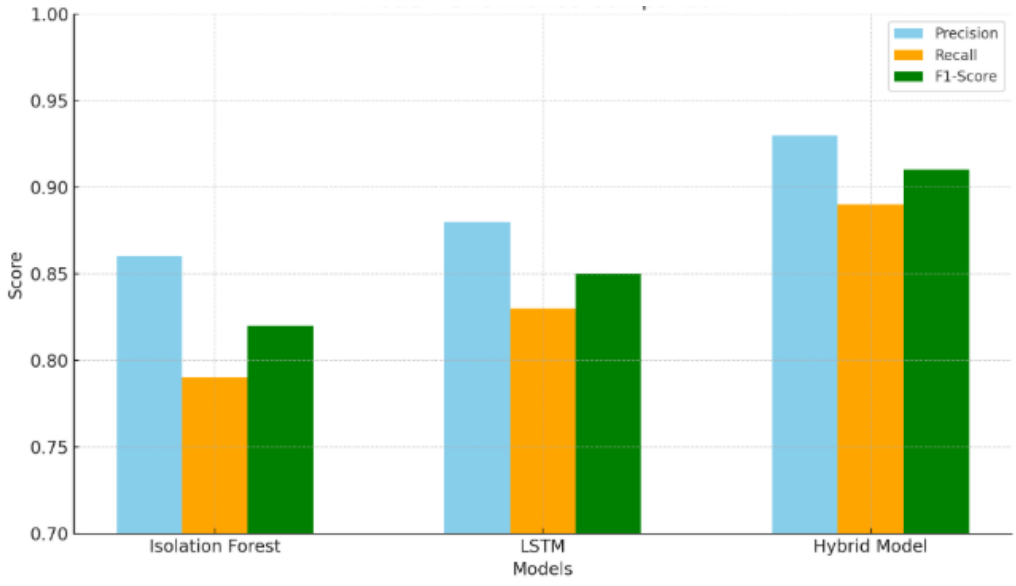
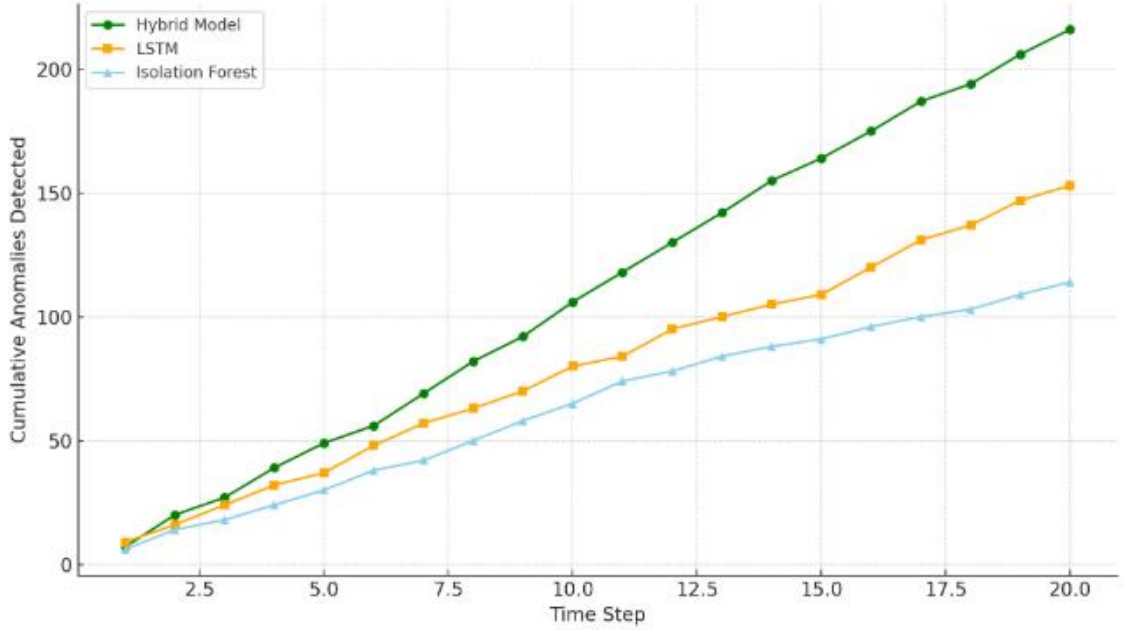


Figure 3 also compares the performance of the three models, (Isolation Forest, LSTM, and the Hybrid) based on the Recall, F1-Score, and Precision. The hybrid model clearly outperformed the individual models across all the three metrics. It recorded 0.93 for precision, 0.89 for recall, and 0.91 for F1-score. This shows a more balanced and accurate detection mechanism, minimizing the false negatives and positives. The Recall, Precision, and F1-Score comparisons in this bar chart highlight consistent improvement in anomaly detection in the hybrid model. The comparison on anomaly detection rates of the three models was also done as shown in Figure 4.

Figure 4

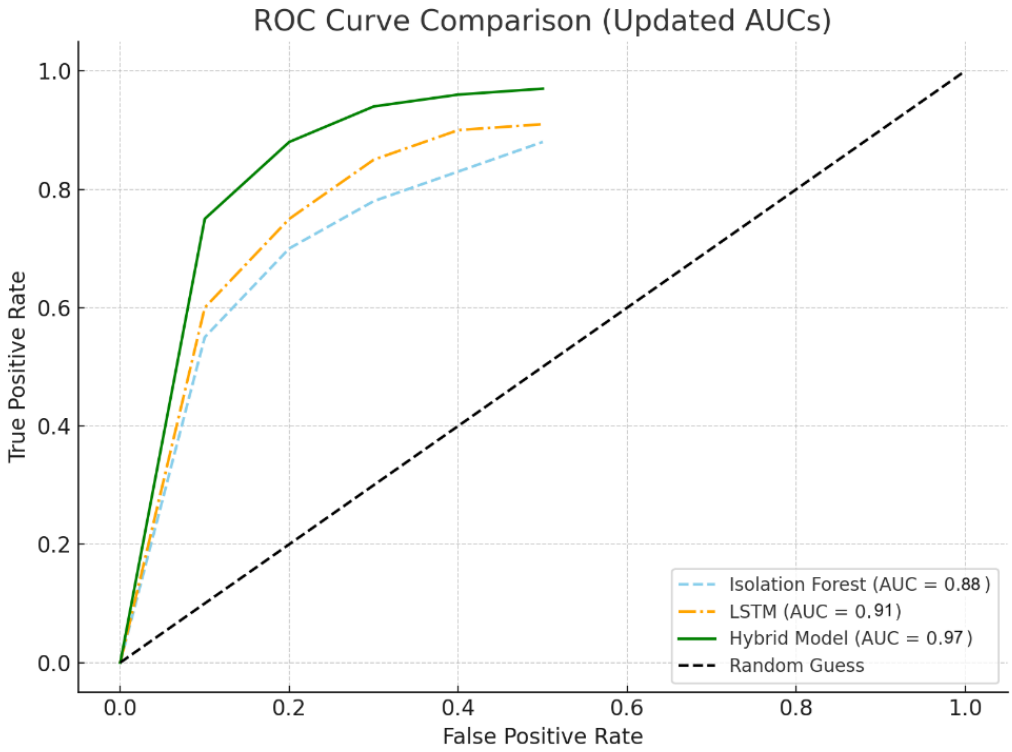
Detection trends over time



The hybrid model detected the highest number of anomalies at every time step; hence, showing a consistently steep upward trajectory as shown in Figure 4. It reached over 220 cumulative detections by the 20th time step, unlike the IS and LSTM that recorded 110 and 150 detections respectively. These results suggest that the hybrid model is more sensitive and robust at identifying both sudden and evolving threats in VPN traffic. This detection trend illustrates the model's responsiveness across ten different time intervals. A ROC curve showing the trade-off between detecting true anomalies and mistakenly flagging normal traffic as anomalies was presented in Figure 5.

Figure 5

ROC Curve



This Receiver Operating Characteristic (ROC) curve in Figure 5, recorded 0.88 for Isolation Forest, 0.91 for LSTM, and 0.97 for the Hybrid Model. This clearly attested to the superior True Positive Rate vs. False Positive Rate trade-offs. The steeper curve of the hybrid model confirms its superior classification power in distinguishing normal from anomalous traffic. The models' ability to distinguish between positive (anomaly) and negative (normal) cases across all possible decision thresholds are presented in Figure 6.

Figure 6

Area Under the Curve (AUC)

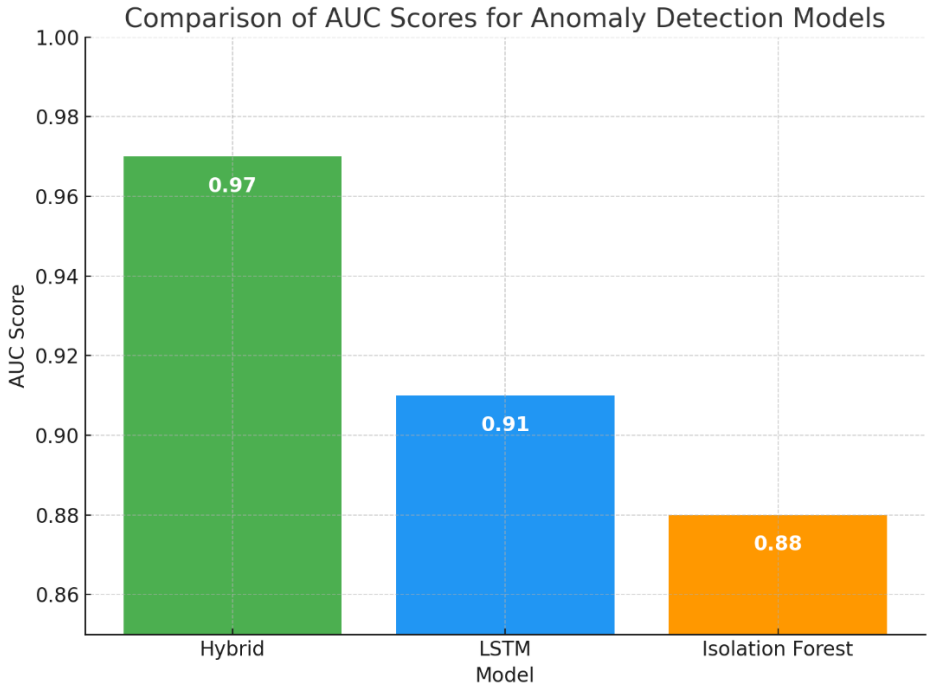
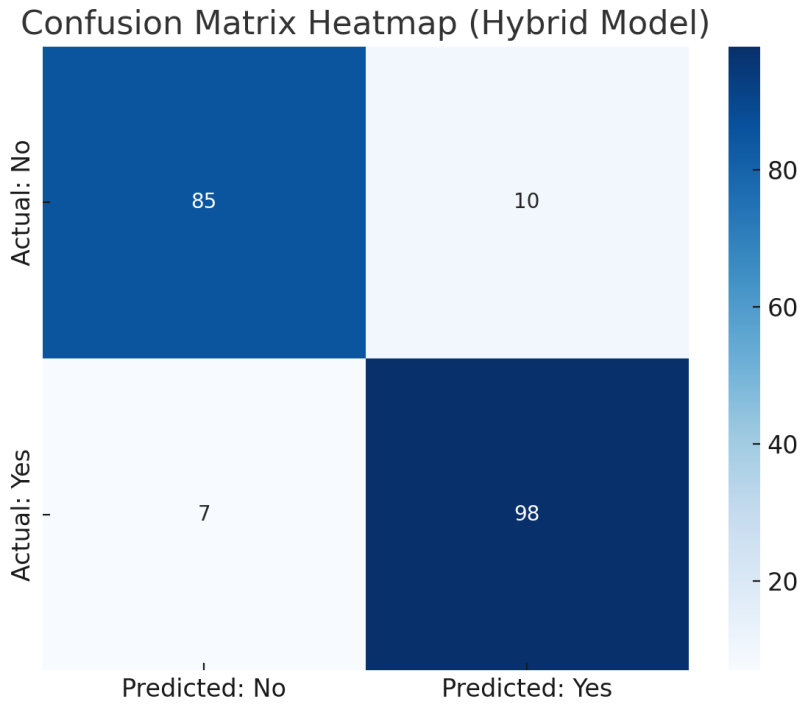


Figure 6 shows Area under the Curve (AUC) at 0.88 for Isolation Forest, 0.91 for LSTM and 0.97 for the Hybrid Model. The steady improvement from individual models to the Hybrid Model (0.97) demonstrated the added value of model integration. The Hybrid approach leveraged the strengths of both models, capturing both statistical irregularities and temporal anomalies. A confusion matrix pointed to a balanced and accurate hybrid model that excels in detecting anomalies without overburdening analysts with false alerts as shown in Figure 7.

Figure 7

Confusion Matrix for Hybrid Model



The confusion matrix for the hybrid model in Figure 7 indicates 85 true negatives, 98 true positives, 10 false positives, and only 7 false negatives. This distribution confirmed that the hybrid model is not only precise but also sensitive to subtle anomaly patterns in VPN traffic. This made it suitable for real-time applications.

Analysis and Insights

By combining IF and LSTM, the model capitalized on both statistical and temporal detection capabilities of IF that handles irregular and high-volume deviations efficiently, and LSTM that addresses subtle and evolving patterns (Paulikas et al., 2022). This combined capability reduces shortcomings of either model when used alone. Benchmark studies noted that Isolation Forests, a fast and lightweight, was ideal for constrained environments, while LSTMs sacrificed some speed for behavioral depth (Paulikas et al., 2022). However, the hybrid model balanced both, maintaining low latency (180ms) and strong accuracy.

Operating within a Kafka – Spark pipeline, the model practically supports scalable real-time deployment. This aligns with hybrid intrusion detection systems studied in similar contexts. The hybrid model achieves responsiveness (<200ms latency) critical for timely alerting without overwhelming security teams. This low latency shows that the pipeline is suitable for real-time applications (Renza Nur, 2025). Unlike deep packet inspection tools, which struggle with encrypted traffic, the hybrid model uses only

flow metadata, preserving privacy and operational efficiency.

Discussion

The hybrid model's higher accuracy and lower latency are in line with research by Cheng, Y., & Zhao, L., 2022 who showed that integrating deep learning and statistical techniques improves anomaly detection. While the LSTM captured subtle temporal patterns, the Isolation Forest component successfully identified abrupt outliers. This dual approach enhances generalization to evolving attacks and minimizes model blind spots. These findings are particularly relevant for environments with encrypted traffic, such as VPN tunnels, where payload inspection is limited. As discussed by Cheng, Y., & Zhao, L., 2022, metadata-driven models, like this hybrid model have become essential. Furthermore, in the African context, where infrastructure is limited, models that require only flow metadata and minimal computational resources are more deployable (Ndiege& Ngari, 2021).

With the tests on OwnCloud using simulated traffic, the hybrid model validates the viability of implementing intelligent, lightweight detection frameworks that protect user privacy without sacrificing security. These findings support the theory that in encrypted cloud VPN environments, hybrid AI approaches offer anomaly detection that is more precise, timely, and resource-efficient than standalone AI models.

4.0 Conclusion

The proposed hybrid AI model approach effectively enhances real-time threat detection in encrypted cloud VPN traffic. By combining Isolation Forest and LSTM, the system leveraged both statistical anomaly detection and temporal behavior learning. Results show a significant increase in detection accuracy and reduced false positives compared to using each model alone. This model demonstrates a viable solution for modern cloud environments where encryption limits traditional detection techniques.

5.0 Recommendations

Based on the findings, it is recommended that organizations using cloud-based VPNs embrace this hybrid AI anomaly detection system. The IT security teams, working with network infrastructure providers, should deploy the hybrid model for real-time encrypted traffic monitoring, improving accuracy and reducing false positives. The cybersecurity regulatory bodies and standards organizations should revise compliance frameworks to encourage the usage of hybrid AI-driven detection methods in encrypted environments, ensuring both performance and privacy compliance. Although validated in a simulated environment, future research should evaluate the hybrid model using real-world VPN traffic to confirm its robustness under diverse operational conditions. It is also crucial to do a Multi-Cloud Integration to extend the framework to support multi-cloud environments like AWS, Azure, and Google Cloud; and different VPN protocols like WireGuard, and IPSec, to ensure broad applicability. This will bring great contribution in the cloud security

References

Andhra University Alumnus, Alang, K. S., Kushwaha, P. (Dr) A. S., & Sharda University. (2025). Stream Processing with Apache Kafka: Real-Time Data Pipelines. *International Journal of Research in Modern Engineering & Emerging Technology*, *13*(3), 216–227. https://doi.org/10.63345/ijrmeet.org.v13.i3.13

CAK. (2024). *The National KE-CIRT/CC* (pp. 7–11). Communications Authority of Kenya. https://ke-cirt.go.ke/wp-content/uploads/2025/01/2024-25-Q2-Cyber-Security-Report.pdf

Cheng, Y., & Zhao, L. (2022). Real-time anomaly detection using hybrid deep learning in encrypted cloud environments. *Future Generation Computer Systems, 128, 12–23.*

Chukwuemeka Nwachukwu, Kehinde Durodola-Tunde, & Chukwuebuka Akwiwu-Uzoma. (2024). AI-driven anomaly detection in cloud computing environments. *International Journal of Science and Research Archive*, *13*(2), 692–710. https://doi.org/10.30574/ijsra.2024.13.2.2184

Communication Authority of Kenya. (2021). *Quarterly sector statistics report Q1 2021/2022*. Communications Authority of Kenya. (2021).

Emmanuel Ok, & Mayowa Emmanuel. (2025). *Real-Time Network Traffic Anomaly Detection Using Hybrid Deep Learning Models*. https://www.researchgate.net/publication/390034032_Real-Time_Network_Traffic_Anomaly_Detection_Using_Hybrid_Deep_Learning_Models

Markets and Markets. (2022). *Virtual private network (VPN) market with COVID-19 impact analysis.* (MarketsandMarkets). https://www.marketsandmarkets.com/

Michael Stephen. (2025). AI-Based Anomaly Detection for Cloud Cost Spikes in SaaS Environments. *Researchgate*.

National ICT Policy (2020). (n.d.). https://www.ca.go.ke/sites/default/files/CA/Statutes%20and%20Regulations/National-ICT-Policy-Guidelines-2020.pdf

Ndiege, J. & Ngari, J. M. (2021). Adoption of cloud computing services by Kenyan SMEs. *African Journal of Business Management*, 35–45.

Paulikas, G., Sandonavičius, D., Stasiukaitis, E., Vilutis, G., & Vaitkunas, M. (2022a). Survey of Cloud Traffic Anomaly Detection Algorithms. In A. Lopata, D. Gudonienė, & R. Butkienė (Eds.), *Information and Software Technologies* (Vol. 1665, pp. 19–32). Springer International Publishing. https://doi.org/10.1007/978-3-031-16302-9_2

Paulikas, G., Sandonavičius, D., Stasiukaitis, E., Vilutis, G., & Vaitkunas, M. (2022b). Survey of Cloud Traffic Anomaly Detection Algorithms. In A. Lopata, D. Gudonienė, & R. Butkienė (Eds.), *Information and Software Technologies* (Vol. 1665, pp. 19–32).

Springer International Publishing. https://doi.org/10.1007/978-3-031-16302-9_2

Powers, D. M. W. (2020). *Evaluation: From precision, recall and F-measure to ROC, informedness, markedness and correlation*. https://doi.org/10.48550/ARXIV.2010.16061

Renza Nur. (2025). Leveraging Apache Kafka and Apache Flink for Optimized Real-Time Data Pipelines: A Performance Comparison in Machine Learning Workflows. *Researchgate*.

Ricky Johnny. (2024). Machine Learning-Driven Anomaly Detection for Proactive Performance Optimization in Cloud Environments. *Researchgate*.

Sommer, R., & Paxson, V. (2020). Outside the Closed World: On Using Machine Learning for Network Intrusion Detection. *2020 IEEE Symposium on Security and Privacy*, 305–316. https://doi.org/10.1109/SP.2010.25

Tang, T. A., Mhamdi, L., McLernon, D., Zaidi, S. A. R., Ghogho, M., & El Moussa, F. (2020). DeepIDS: Deep Learning Approach for Intrusion Detection in Software Defined Networking. *Electronics*, *9*(9), 1533. https://doi.org/10.3390/electronics9091533

Vikram, A. & Mohana. (2020). Anomaly detection in Network Traffic Using Unsupervised Machine learning Approach. *2020 5th International Conference on Communication and Electronics Systems (ICCES)*, 476–479. https://doi.org/10.1109/ICCES48766.2020.9137987

Younus, Z. S., & Alanezi, M. (2023). Detect and Mitigate Cyberattacks Using SIEM. *2023 16th International Conference on Developments in eSystems Engineering (DeSE)*, 510–515. https://doi.org/10.1109/DeSE60595.2023.10469387