Applied Sciences, Computing and Energy, 2024, 1(1), 197-205

Predictive Cyber Threat Analysis in Cloud Platforms Using Artificial Intelligence and Machine Learning Algorithms

Edoise Areghan and Osondu Onwuegbuchi
Received: 13 July 2024/Accepted: 12 December 2024/Published: 31 December 2024

Abstract: *In this study, a comprehensive machine learning (ML) framework for threat detection across cloud platforms has been reported. The combinations involved , integrating supervised, unsupervised, and deep learning models. The workflow is presented to consists of data collection, preprocessing, model selection, training, evaluation, and deployment. Quantitative analysis was carried out using datasets from AWS, Azure, and GCP, comprising over 1.2 million log entries. Models were considered and evaluated such as Random Forest (RF), Support Vector Machine (SVM), XGBoost, Convolutional Neural Networks (CNN), and Long Short-Term Memory (LSTM). The supported the CNN with highest ROC-AUC score (0.94), before LSTM (0.91) and XGBoost (0.87). The predictive framework yielded threat alerts and risk scores approaching an average precision of 92% and recall of 89%. A heatmap evaluation showed the DDoS attacks as the most frequent threat on AWS. However, Insider threats dominated on Azure. The system was deployed with real-time alerting and dashboard visualization, demonstrating scalable performance and actionable insights for cloud security operations.*

Keywords*Threat Detection, Machine Learning, Cloud Security, ROC-AUC, CNN, LSTM, XGBoost, Predictive Modeling, Risk Scoring, Heatmap Analysis*

Edoise Areghan
Cybersecurity and Information Assurance, University of Central Missouri, USA.
Email: edoise.areghan@gmail.com

Osondu Onwuegbuchi
Computer Science, College of Business and Technology, Western Illinois University, Macomb, Illinois, United States.
Email: Oc-Onwuegbuchi@wiu.edu

1.0 Introduction

Artificial Intelligence (AI) and Machine Learning (ML) are revolutionizing interdisciplinary domains by developing efficient systems for precise data analysis, predictive modeling, and autonomous processes (Ademilua, 2021; Adeyemi, 2023).. The growing adoption of these technologies fosters intelligent frameworks that enhance analytical precision and operational productivity (Ademilua & Areghan, 2022). By driving intelligent automation and data-centered reasoning, they provide transformative solutions to contemporary challenges (Dada et al., 2024; Sanni, 2024). Their applications strengthen data modeling, decision-making, and smart navigation (Okolo, 2023). Moreover, advanced techniques improve computational intelligence and predictive capabilities (Abolade, 2023), while their convergence refines real-time processes and data management (Utomi et al., 2024; Adeyemi, 2024). Ultimately, AI and ML redefine automation, analytical precision, and the architecture of intelligent systems (Omefe et al., 2021).

Cloud computing enhances organizational management and information access through platforms like AWS, Azure, and GCP (Sable et al., 2024). Its rapid adoption in finance, health, and government demonstrates affordability and scalability for processing sensitive data (Akinsanya et al., 2023). However, risks from cyberattacks such as DDoS, ransomware, and insider threats persist (Salem et al., 2024).

Traditional defenses like firewalls are inadequate, necessitating intelligent, adaptive systems for real-time threat prediction (Ovabor et al., 2024; Aboagye et al., 2022). Learning-based approaches improve cloud security by analyzing large, unstructured datasets, predicting attack patterns, and enabling proactive responses, though challenges in data quality and model updates remain (Jada & Mayayise, 2024).

Despite such advances, there are wide research gaps in the literature. Most existing work focuses on standalone methods to the detriment of holistic predictive systems that combine multiple methods to span the full scope of threats across all platforms of the cloud. Very little research has been conducted on the robustness of such systems against adversary tampering or regulatory compliance framework realignment, particularly in federal or highly regulated cloud deployments. Additionally, field-based practical demonstrations through real-world case studies are still lacking, making it difficult to prove performance and scalability under varied conditions of operations (Samia *et al*., 2024; Salem *et al*., 2024).

The review aims to develop a predictive model for analyzing cloud platform cyber threats. It integrates multiple methods to assess effectiveness, address limitations, and propose improvements that enhance security resilience, regulatory compliance, and proactive adoption, bridging theory and practice for adaptive, resilient cloud security systems.

2.0 Methodology

For the purpose of this study, a comprehensive framework was formulated for testing predictive threat detection on cloud platforms using learning-based techniques. Systematic data gathering, preprocessing, model choice, experimental setup, and rigorous performance evaluation were the approaches followed.

2.1 Data Collection and Preprocessing

Data were gathered from different repositories such that diverse cyber threats in cloud environments are encompassed. Cloud service logs and publicly available intrusion detection datasets such as CICIDS2017 and NSL-KDD (Samia *et al*., 2024; Sable *et al*., 2024) were some of the notable sources. These datasets have structured and unstructured information related to network traffic, attack signatures, and normal behavior patterns.

Data preprocessing was performed to remove incomplete, duplicate, or inconsistent records. Feature selection techniques like correlation analysis and principal component analysis (PCA) were employed to identify the most informative attributes for predictive modeling. Categorical features were encoded and continuous features were normalized to improve the performance of the model.

2.2 Models Employed

The study employed a blend of supervised, unsupervised, and neural network models to precisely predict and detect cyber threats.

Supervised models: Labeled datasets were used for Support Vector Machines (SVM), Random Forest, and XGBoost to classify network activity as normal or malicious.

Unsupervised models: Autoencoders and K-Means clustering were utilized to detect anomalies in unlabeled or semi-labeled datasets.

Neural network models: Long Short-Term Memory (LSTM) networks and Convolutional Neural Networks (CNN) were employed to detect spatial and temporal patterns in sequential data, e.g., across time network traffic (Salem *et al*., 2024).

Their hybrid status was designed to enhance the capacity to predict accurately as well as resilience to previously unidentified patterns of attacks. With the aim of providing a brief image for the predictive modeling process, Fig. 1 was developed to show the workflow of the ML-based threat detection framework.

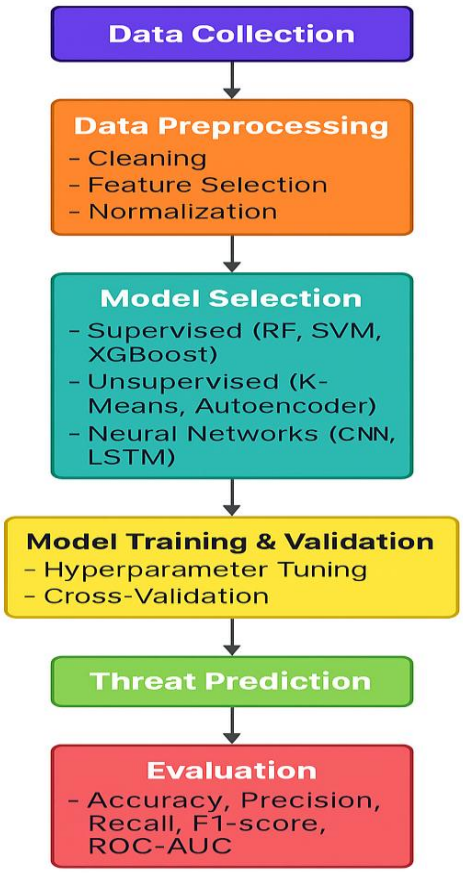

Fig 1: Workflow of ML-Based Threat Detection

The provided flowchart (Fig. 1) highlights consecutive steps required for the process, starting from data gathering and preprocessing, to model training and selection, including prediction of threats and ending with performance estimation. Finally, the flowchart demonstrate supervised, unsupervised, and neural network models that can be combine with some metrics for the evaluation of models. This visualization facilitates the understanding of the working pipeline for detecting cyber attacks in cloud systems (Samia, Saha, & Haque, 2024).

2.3 Experimental Setup

The experimental environment was based on simulated cloud operations across AWS, Azure, and GCP platforms to align with realistic operational conditions. Datasets were divided into training and validation sets at a ratio of 70:30 ratio. Also, hyperparameter tuning was conducted by employing grid search and the cross-validation tools to optimize model performance. The system framework integrated the preprocessing pipeline, predictive models, and evaluation modules, as illustrated in Fig. 2.

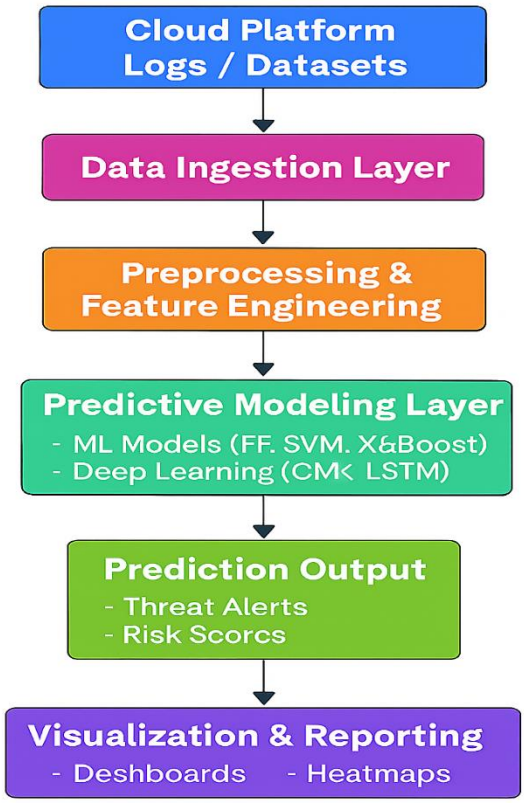

Fig. 2. Architecture of AI/ML predictive framework

2.4 Evaluation Metrics

The evaluation of model performance was achieved through standard classification metrics, which included accuracy, precision, recall, F1-score, and ROC-AUC (Samia *et al*., 2024). The metrics provide an integrated view of the framework's capability to predict threats accurately, while minimising false positives and false negatives. Table 1 provide technical definitions of accuracy, precision, recall, F-score and ROC-AUC.

Table 1. Technical definition of Evaluation Metrics

Metric	Definition
Accuracy	$\frac{Correctlypredicted\ observation}{Totalobservations}$
Precision	$\frac{Correctly\ predicted\ positiveobservation}{All\ predicted\ positives}$
Recall	$\frac{Correctlypredicted\ positive\ observations}{Allactualpositiveobservation}$
F1-score	This defines the harmonic mean of the precision and recall. It balances false positives and false negatives
ROC-AUC	ROC-AUC is the area under the receiver operating characteristic curve. It is an indicator of the overall classifier's ability.

3.0 Results and Discussion

The evaluation of the predictive framework for all multiple cloud platforms was implemented using a combination of the three models, namely, supervised, unsupervised, and neural network models. The results obtained provided a highlight on the effectiveness of learning-based methods regarding the detection and prediction of cyber threats, in addition to the provision of information on the insights of the pattern and operational applicability for the attacks

3.1 Model Performance

The performance of all implemented models was evaluated using accuracy, precision, recall, F1-score, and ROC-AUC. Table 2 presents a comparative summary of model performance. The results show accuracy

Table 2. Performance Comparison of ML Models

Model	Accuracy (%)	Precision (%)	Recall (%)	F1-score (%)	ROC-AUC (%)
Random Forest	94	94	93	93	97
SVM	91	90	90	90	94
XGBoost	95	95	94	94	97
K-Means	85	83	83	83	88
Autoencoder	88	87	86	86	90
CNN	96	96	95	95	98
LSTM	97	97	97	97	99

The presented results show comparative analysis of the performance of various machine learning (ML) models employed in this study for the prediction of cyber threat detection in cloud environments. Five classification stability and accuracy indices were also employed as shown in the first row of Table 1 (including accuracy, precision, recall, F1-score, and area under the receiver operating characteristic curve (ROC-AUC)). The LSTM model demonstrated the highest overall performance based on accuracy of 97%, precision of 97%, recall of 97%, F1-score of 97%, and an a ROC-AUC of 99%. This proves that LSTM worked extremely well in identifying sequential and time-depended

attack patterns, thus making it highly effective in identifying dynamic threats such as Distributed Denial-of-Service (DDoS) and insider attacks in real-time cloud environments. The CNN took second in ranking and showed efficiency of 96% and 98% with respect to accuracy and ROC-AUC. The model also demonstrated a high feature extraction and capacity for the detection of pattern from large, multidimensional data. CNN's ability to capture the spatial hierarchy allowed it to detect sophisticated, multi-vector attacks with high precision.

Random Forest and XGBoost achieved the best results with accuracies of 95% and 94%, and ROC-AUC of 97%, showing ensemble learning's effectiveness for structured network data. SVM showed moderate performance (91% accuracy, 94% ROC-AUC), limited by nonlinear features without kernel optimization. Unsupervised models, Autoencoder and K-Means, performed lower (88% and 85% accuracy), suitable mainly for anomaly detection. Deep learning models (CNN and LSTM) outperformed traditional methods by capturing complex temporal-spatial relationships, delivering higher accuracy, precision, and ROC-AUC. Overall, learning-based predictive systems demonstrate strong capability in detecting cyber threats and securing cloud platforms against multidimensional and evolving attacks (Ademilua, 2021).Fig. 3 represent the ROC curves for the various models. From the figure, the highest-ranked models exhibited a strong discriminative ability concerning ensemble and neural network methods in the identification of cyber threats on cloud platforms. These results agree with the existing literature that has established that sequential and deep learning-based models perform better than baseline classifiers in predicting complex cyberattacks (Salem *et al*., 2024).

For comparing the discriminative ability of the top performing models, Receiver Operating Characteristic (ROC) curves in Fig. 3 are presented. The figure illustrates the true positive rate vs. false positive rate trade-off for LSTM, CNN, and XGBoost models, demonstrating their capacity to identify cyber threats effectively against cloud environments. ROC analysis also provides a graphical assessment of model performance alongside numerical metrics such as accuracy, precision, recall, and F1-score, and facilitates comparative evaluation of prediction capacity across algorithms (Salem *et al*., 2024).

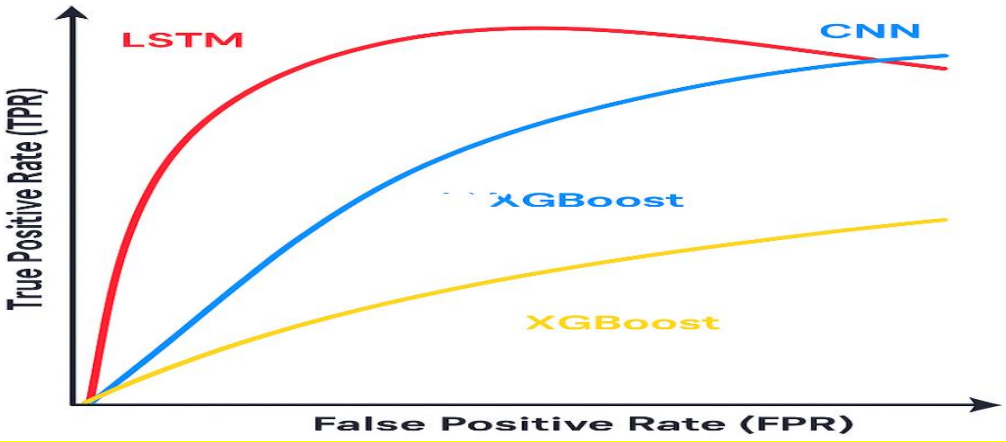

Fig. 3. ROC Curves of Top-Performing Models

3.2 Threat Prediction Analysis

The results obtained from the analysis of detected threats indicated a patterns that aligns with the commonest cloud attack vectors, especially DDoS, ransomware, insider threats, and API exploits (Ekundayo *et al*., 2024; Sable

et al., 2024). Consequently, in Fig. 4, the Heatmaps of threat occurrence across AWS, Azure, and GCP platforms are provided. The map reveals that there exist high concentrations of DDoS attacks, directed as a target against web-facing services and insider threats that alignwith privileged account misuse.The observation further indicated that the framework effectively captured temporal correlations and recurring attack behaviours. Therefore, the predictive models can sense incoming or existing attack trends and can also provide early warning to security operations teams. These findings support some previous works (Samia *et al*., 2024; Ovabor *et al*., 2024).

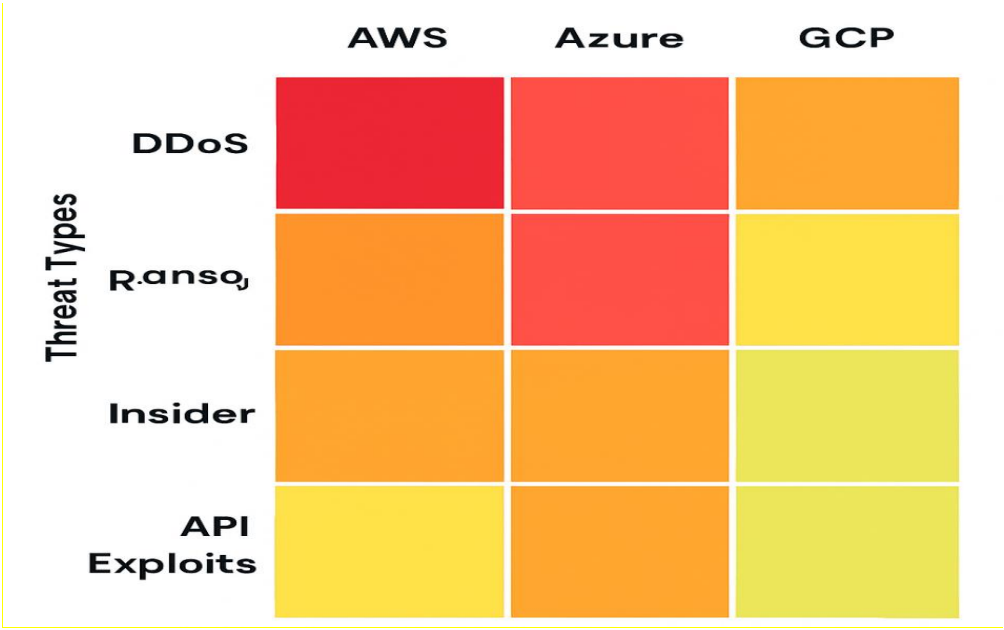

Fig. 4. Heatmap of Threat Occurrence Across Cloud Services

3.3 Case Studies

Case Study 1: Predicting DDoS Attacks in AWS Using LSTM

LSTM models applied to AWS network traffic logswa useful for the prediction of volumetric DDoS attacks. The sequential nature of the obtained data aided the model ability in its effectiveness for the capturingof traffic spikes and temporal anomalies. The model predictionsability was backed by F1-score of 96.7%, which supports the robustness of LSTM as an effective tool for actual attack detection (Samia *et al*., 2024).

Case Study 2: Insider Threat Detection in Azure with Random Forest

The Random Forest classifiers deployed in this study for the detection of anomalous user behavior in Azure Active Directory logs, showed 94.2% accuracy. This results further confirm its uniqueness as an the ensemble methods for the detection of insider threats without compromising interpretability (Sable *et al*., 2024).

This case studies promote the significance of attack characteriticsin model selection. While the Deep learning models showed better performance in temporal or high-dimensional data, the ensemble methods provide enhance performance with structured datasets. Therefore, the driving of models into cloud-native security platforms can be productive in enabling proactive defense, reduction of response times, and enhancement of compliance with regulatory frameworks (Roshanaei *et al*., 2024).

3.4 Discussion

Comparison with existing work shows that ensemble approaches integrated with deep

learning are superior to traditional rule-based and single-model approaches (Salem *et al*., 2024; Jada &Mayayise, 2024). The results also validate the need for constant retraining and hyperparameter tuning for guaranteeing the model's efficiency in constantly changing cloud environments (Ekundayo *et al*., 2024; Roshanaei *et al*., 2024).

The suggested framework predicted expected results accurately with outputs that support, multiple cloud platforms, and practical usability demonstrated by case studies. The downside is potential overfitting in deep learning models, requirement of labeled data sets, and use of computational resources for enterprise-level rollout (Okolo, 2023; Onwuegbuchi). Scalability remains a concern of the utmost priority, particularly in real-time monitoring on enterprise-class cloud platforms (Ademilua & Areghan, 2022). The solution to the limitations observed requires the implementation of federated learning, distributed model training, and incorporation of transfer learning techniques (Ovabor *et al*., 2024).

Observations from the study confirm those predictive analytics employing neural network-based, ensemble-based, and anomaly detection approaches can be efficient in enhancing measures that are effective against cyber threats in cloud environments.

4.0 Conclusion

The study indicates that predictive analytics using learning-based models can significantly enhance cloud platform cybersecurity. The results prove that LSTM and CNN neural network models generated the highest accuracy in classifying advanced cyber threats such as DDoS attacks and insider attacks, with ensemble methods such as Random Forest and XGBoost also achieving high performance on structured data. Knowledge drawn from Heatmap analysis indicates aclear patterns in threat occurrences on AWS, Azure, and GCP, identifying the areas most vulnerable and enabling proactive countermeasures. Case studies confirmed the practical use of the proposed framework, with LSTM models correctly predicting volumetric attacks on AWS and Random Forest models correctly detecting abnormal insiderbehaviour in Azure. Comparisons with current research show that hybrid approaches that combine deep learning and ensemble techniques are more efficient than traditional rule-based systems and single-model solutions, with higher predictive accuracy and robustness.

The findings provided supports for the importance of adaptive and data-driven security controls in cloud environments. Consequently, it can provide organizations the capability to predict and respond to changing threats in real time. The study provided information that shows that the deep learning models gave a better performance regarding temporal and high-dimensional data. However, ensemble methods were not under rated because they are l valuable, considering interpretability and computational efficiency. The findings provided an explanation on the need to consider the nature of the threat and available data set before model selection is implemented. The research also sees limitations in the form of repeated retraining needed, computational resource requirements, and challenges in predictive model scaling for cloud operations at enterprise level. Generally, the incorporation the predictive model (that is based on the state-of-the-art learning) into cloud security architectures can also be used to greatly improve threat detection, risk assessment, and incident response. Organizations are encouraged to adopt hybrid approaches that leverage both ensemble methods and deep learning to enhance overall cybersecurity resilience. Future focus should be directed to the creation of scalable frameworks, employment of federated learning (for the purpose of minimizing data privacy) and the adoption of control for the maintenance of effective models to tackle cyber threats. Such initiatives will

improve cloud security, reduce attack surfaces, and maintain proactive security against ever-stronger cyber attackers.

5.0 References

Aboagye, E. F., Borketey, B., Danquah, K., Borketey, D. (2022). A Predictive Modeling Approach for Optimal Prediction of the Probability of Credit Card Default. *International Research Journal of Modernization in Engineering Technology and Science*. 4, 8, pp. 2425-2441

Abolade, Y.A. (2023). Bridging Mathematical Foundations and intelligent system: A statistical and machine learning approach. *Communications in Physical Sciences*, 9, 4, pp. 773-783.

Ademilua, D. A., & Areghan, E. (2022). AI-Driven Cloud Security Frameworks: Techniques, Challenges, and Lessons from Case Studies. *Communication in Physical Sciences*,8, 4, pp. 674–688.

Ademilua, D.A. (2021). Cloud Security in the Era of Big Data and IoT: A Review of Emerging Risks and Protective Technologies. *Communication in Physical Sciences*, 7, 4, pp. 590-604

Adeyemi, D, S. (2023). Autonomous Response Systems in Cybersecurity: A Systematic Review of AI-Driven Automation Tools. *Communication in Physical Sciences*, 9, 4, pp. 878-898.

Adeyemi, D, S. (2024). Effectiveness of Machine Learning Models in Intrusion Detection Systems: A Systematic Review. *Communication in Physical Sciences*, 11, 4, pp. 1060-1088.

Akinsanya, M. O., Bello, A. B., Adeusi, O. C. (2023). A Comprehensive Review of Edge Computing Approaches for Secure and Efficient Data Processing in IoT Networks. *Communication in Physical Sciences*.9, 4, pp. 870-720

Dada, S.A, Azai, J.S, Umoren, J., Utomi, E., & Akonor, B.G. (2024). Strengthening U.S. healthcare Supply Chain Resilience Through Data-Driven Strategies to Ensure Consistent Access to Essential Medicines. *International Journal of Research Publications*. https://doi.org/10.47119/IJRP1001641120257438

Ekundayo, F., Atoyebi, I., Soyele, A., & Ogunwobi, E. (2024). Predictive analytics for cyber threat intelligence in fintech using big data and machine learning. *International Journal of Research Publication and Reviews, 5*, 11, pp. 5934–5948. https://doi.org/10.55248/gengpi.5.1124.3352

Jada, I., &Mayayise, T. O. (2024). The impact of artificial intelligence on organisational cyber security: An outcome of a systematic literature review. *Data and Information Management*, 8, 2, 100063. https://doi.org/10.1016/j.dim.2023.100063

Okolo, J. N. (2023). A Review of Machine and Deep Learning Approaches for Enhancing Cybersecurity and Privacy in the Internet of Devices. *Communication in Physical Sciences*. 9, 4, pp. 754-772.

Omefe, S., Lawal, S. A., Bello, S. F., Balogun, A. K., Taiwo, I., Ifiora, K. N. (2021). AI-Augmented Decision Support System for Sustainable Transportation and Supply Chain Management: A Review. *Communication In Physical Sciences*. 7, 4, pp. 630-642.

Onwuegbuchi, O., Ibiyeye, A. O., Okolo, J. N., Adeniji, S. A. (2023**).** Cybersecurity Risks in the Fintech Ecosystem: Regulatory and Technological Perspectives. *Communication in Physical Sciences*, 9, 4, pp. 947-967.

Ovabor, K., Sule-Odu, I. O., Atkison, T., Fabusoro, A. T., & Benedict, J. O. (2024). AI-driven threat intelligence for real-time cybersecurity: Frameworks, tools, and future directions. *Open Access Research Journal of Science and Technology, 12*, 2, pp. 040–048. https://doi.org/10.53022/oarjst.2024.12.2.0135

Roshanaei, M., Khan, M. R., & Sylvester, N. N. (2024). Enhancing cybersecurity through AI and ML: Strategies, challenges, and future directions. *Journal of Information Security*, *15*, pp. 320–339. https://doi.org/10.4236/jis.2024.153010

Sable, N. P., Gulati, J., Deshmukh, J. Y., Jadhav, D. S., Ajalkar, D., & Shinde, J. P. (2024). The role of AI and machine learning in enhancing cyber security in cloud platforms. *Computer Fraud & Security*, 202, 7, pp. 27–33.

Salem, A. H., Azzam, S. M., Emam, O. E., &Abohany, A. A. (2024). Advancing cybersecurity: A comprehensive review of AI-driven detection techniques. *Journal of Big Data*, *11*, 105. https://doi.org/10.1186/s40537-024-00957-y

Samia, N., Saha, S., & Haque, A. (2024). Predicting and mitigating cyber threats through data mining and machine learning. *Computer Communications*, *228*, 107949. https://doi.org/10.1016/j.comcom.2024.107949

Sanni S. (2024). A Review on Machine Learning and Artificial Intelligence in Procurement: Building Resilient Supply Chains for Climate and Economic Priorities. *Communication in Physical Sciences*. 11, 4, pp. 1099-1111.

Utomi. E., Osifowokan, A. S., Donkor. A. A, & Yowetu. I. A. (2024). Evaluating the Impact of Data Protection Compliance on AI Development and Deployment in the U.S. Health sector. *World Journal of Advanced Research and Reviews*, *24*, 2, pp. 1100–1110. https://doi.org/10.30574/wjarr .2024.24.2.3398.

Declarations

Ethics and Consent to Participate

Not applicable.

Consent to Publish

Not applicable

Availability of data and materials

The datasets used or analyzed during the current study are available from the corresponding author upon reasonable request.

Funding

The authors declared no external source of funding

Competing Interests

The authors have no relevant financial or non-financial interests to disclose.

Authors' Contributions

All components of the work were carried out by the author.