© IIJP 2026 | Volume 2, Issue 3, pp. 796-812 | ISSN: 2636-4484

INSIDER THREAT DETECTION USING MACHINE LEARNING AND BEHAVIORAL ANALYTICS IN ENTERPRISE ENVIRONMENTS

Samuel Okechukwu Nnaji[1], Helen Okparaji Onungwe[2] & Christabel Linda Uchenwa[3]

[1]*Department of Computer Science, Cyber Security and Software Engineering, Federal University of Environment and Technology, Koroma/Saakpenwa-Ogoni, Rivers State.*

[2]*Department of Computer Science, University of Port-Harcourt, ORCID: https://orcid.org/0009-0005-8048-1063*

[3]*Department of Computer Science, Cyber Security and Software Engineering, Federal University of Environment and Technology, Koroma/Saakpenwa-Ogoni, Rivers State.*

*****Corresponding Author: samuelokechukwunnaji@gmail.com*

*****ORCID: https://orcid.org/0009-0007-5344-3058*

ARTICLE INFO

Article No.: 0474
Accepted Date: 25/07/2026
Published Date: 11/08/2026
Type: Research

ABSTRACT

Insider threats remain a major cybersecurity challenge in enterprise environments because authorized users can intentionally or unintentionally compromise organizational information assets, often evading traditional security mechanisms. The increasing complexity of insider attacks necessitates intelligent detection approaches capable of identifying abnormal user behavior. This study examined insider threat detection using machine learning and behavioral analytics in enterprise environments. An experimental research design was adopted using the CERT Insider Threat Dataset, comprising enterprise user activity logs such as login records, file access, email communications, device usage, and web activities. Data preprocessing, feature engineering, and behavioral analytics were conducted before implementing six machine learning algorithms: Logistic Regression, Random Forest, XGBoost, Long Short-Term Memory (LSTM), GNN, and Transformer Encoder. Model performance was assessed using precision, recall, F1-score, and ROC-AUC. The results revealed that the Transformer Encoder model achieved the highest detection metric (F1 = 0.959, AUC-ROC = 0.976, FPR = 0.012), followed closely by the GNN hybrid and the LSTM. The findings indicate that integrating machine learning with User and Entity Behavior Analytics (UEBA) enhances early insider threat detection, reduces false positives, and strengthens enterprise environments. The study recommends integrating ML-driven behavioral analytics with SIEM and UEBA platforms, periodically retraining detection models, and implementing continuous cybersecurity awareness training alongside least-privilege access control to improve organizational cyber resilience.

Keywords: Insider Threat Detection, Machine Learning, Behavioral Analytics, Enterprise Environments

Introduction

Insider threats have become one of the most critical cybersecurity challenges facing modern enterprise environments. Unlike external cyberattacks, insider threats originate from individuals with legitimate access to organizational systems, networks, or sensitive information, including employees, contractors, consultants, and business partners (AlSlaiman, Salman, Saleh, & Wang, 2023). These insiders may intentionally or unintentionally compromise the confidentiality, integrity, or availability of organizational assets. Their authorized access and familiarity with organizational processes make malicious or negligent activities difficult to distinguish from legitimate user behavior, rendering traditional perimeter-based security mechanisms less effective (Gayathri, Sajjanhar, & Xiang, 2024). Furthermore, the widespread adoption of cloud computing, remote work, Internet of Things (IoT), and Bring Your Own Device (BYOD) policies has significantly expanded the attack surface, while conventional security solutions such as firewalls, intrusion detection systems (IDS), and Security Information and Event Management (SIEM) platforms often struggle to detect insider attacks executed using valid credentials and subtle behavioral patterns (Gelman & Hastings, 2025; Gelman, Hastings, Kenley, & Loiacono, 2024).

Machine learning (ML) has emerged as a promising solution for insider threat detection by enabling systems to learn normal user behavior and identify deviations that may indicate malicious activity (Alzaabi & Mehmood, 2024). Supervised algorithms such as Random Forest, Support Vector Machine (SVM), Decision Trees, and Extreme Gradient Boosting (XGBoost), together with unsupervised approaches including Isolation Forest, Autoencoders, clustering techniques, and Long Short-Term Memory (LSTM) networks, have demonstrated strong capabilities in detecting insider threats from complex enterprise datasets (Yousef, Jazzar, Eleyan, & Bejaoui, 2023). These models analyze behavioral features extracted from authentication logs, file access records, email communications, USB usage, web browsing activities, and network traffic to distinguish normal from suspicious activities with improved accuracy and lower false-positive rates. Complementing these techniques, User and Entity Behavior Analytics (UEBA) continuously establishes behavioral baselines and evaluates contextual indicators such as login frequency, access time, privilege escalation, and communication patterns to improve real-time threat detection (Splunk, 2024; Zhang et al., 2021). Against this background, this study investigates the integration of machine learning and behavioral analytics for insider threat detection in enterprise environments, with the aim of developing intelligent detection models capable of accurately identifying anomalous user behaviors while reducing false-positive rates and strengthening enterprise cybersecurity.

Related Literature

Insider Threats in Enterprise Environments

Insider threats refer to security risks originating from individuals with authorized access to organizational information systems who intentionally or unintentionally compromise organizational security (Tian, Zhang, Jiang & Feng, 2025). Unlike external attackers who exploit vulnerabilities from outside the network perimeter, insiders possess legitimate credentials and operational knowledge that enable them to bypass conventional security controls (Wang, 2025). Insider threats are generally classified into three categories: malicious insiders, negligent insiders, and compromised insiders. Malicious insiders intentionally steal confidential information, commit fraud, or sabotage organizational assets for financial, political, or personal gain (Tian, Zhang, Jiang & Feng, 2025). Negligent insiders inadvertently expose organizational resources through poor

security practices, while compromised insiders perform malicious activities after their credentials have been stolen by external attackers.

Machine Learning for Insider Threat Detection

Machine learning has transformed cybersecurity by enabling intelligent detection systems capable of learning from historical data and adapting to evolving attack behaviors (Bin Sarhan & Altwaijry, 2022). Supervised learning techniques require labeled datasets containing examples of normal and malicious activities, whereas unsupervised learning methods identify anomalies without requiring prior knowledge of attack patterns (Le, Zincir-Heywood & Heywood, 2020). Commonly used supervised algorithms include Random Forest, Decision Tree, Logistic Regression, Support Vector Machine, and XGBoost. Unsupervised techniques such as Isolation Forest, K-Means clustering, One-Class SVM, and Autoencoders are particularly effective for identifying previously unseen insider attacks because they model normal behavior and flag significant deviations. Deep learning architectures including LSTM, Gated Recurrent Units (GRU), and Transformer-based models further improve detection by capturing temporal dependencies in sequential user activities (Pennada, Nayak, & Krishna, 2025).

Behavioral Analytics and User Behavior Analytics (UBA)

Behavioral analytics involves continuous monitoring and analysis of user activities to establish behavioral baselines and identify abnormal deviations that may indicate insider threats (Altuwaijiri & AlShaher, 2024). User Behavior Analytics (UBA) and User and Entity Behavior Analytics (UEBA) systems collect data from authentication logs, application usage, email communications, endpoint devices, cloud platforms, and network traffic to construct comprehensive behavioral profiles for individual users (Song, Gao, Zhang & Ma, 2024). Statistical analysis, anomaly detection algorithms, and machine learning models evaluate deviations in login frequency, access locations, file transfers, USB device usage, command execution, privilege escalation, and working hours. Behavioral analytics complements traditional security mechanisms by focusing on contextual behavioral patterns rather than predefined attack signatures, thereby improving detection accuracy for sophisticated insider attacks (Ye, Luo & Cui, 2025).

Deep Learning Approaches

Deep learning approaches have become one of the most effective techniques in artificial intelligence because of their ability to automatically learn hierarchical feature representations from large and complex datasets (Mienye & Swart, 2024). Unlike traditional machine learning algorithms that depend heavily on manual feature engineering, deep learning models extract meaningful patterns directly from raw data through multiple hidden neural network layers, resulting in superior predictive performance across various domains. Common deep learning architectures include Convolutional Neural Networks (CNNs), which are highly effective for image and spatial data analysis; Recurrent Neural Networks (RNNs) and Long Short-Term Memory (LSTM) networks, which are designed for sequential and time-series data; Autoencoders for feature learning and anomaly detection; Generative Adversarial Networks (GANs) for synthetic data generation; and Transformer-based models that have revolutionized natural language processing and are increasingly applied to cybersecurity, healthcare, and predictive analytics (Zhang, Y., et al., 2024). Recent studies have demonstrated that deep learning models consistently outperform conventional machine learning techniques in tasks such as intrusion detection, insider threat detection, malware classification, and behavioral analytics because they can capture complex nonlinear relationships and temporal dependencies within massive datasets. Nevertheless, these models often require extensive computational resources, large labeled datasets,

and mechanisms to improve interpretability and reduce training complexity, which remain active areas of research (Samira, Kisina, & Weldegeorgise, 2024).

Explainable Artificial Intelligence (XAI)

Although machine learning models significantly improve insider threat detection performance, many advanced algorithms still operate as "black boxes," limiting analyst trust and regulatory acceptance (Zhang, 2025; Roy & Chen, 2024). Explainable Artificial Intelligence (XAI) seeks to provide transparent explanations for model predictions using techniques such as SHapley Additive exPlanations (SHAP), Local Interpretable Model-Agnostic Explanations (LIME), and attention visualization. Explainability enables security analysts to understand which behavioral features contributed most to a detection decision, facilitating faster incident investigation, reducing analyst fatigue, and supporting compliance with governance requirements (Laksanadjaja, 2024). Current research increasingly recommends integrating XAI into enterprise insider threat detection frameworks to enhance transparency and operational adoption.

Challenges in Insider Threat Detection

Despite substantial progress, insider threat detection continues to face several challenges (Alzaabi & Mehmood, 2024). First, insider threat datasets are highly imbalanced because malicious activities represent only a small fraction of overall enterprise operations, making model training difficult (Song et al., 2024). Second, user behaviors evolve over time due to changing job responsibilities, organizational restructuring, and remote work practices, resulting in concept drift that reduces model effectiveness (Lindauer, 2020). Third, behavioral monitoring raises privacy and ethical concerns, particularly when collecting detailed user activity logs. Fourth, many organizations lack standardized datasets containing realistic insider threat scenarios, limiting comparative evaluation of detection models (Alzaabi and Mehmood, 2024). Finally, integrating heterogeneous enterprise data sources including cloud services, endpoint devices, identity management systems, and IoT platforms remains computationally challenging while maintaining scalability and real-time responsiveness.

Propose Layered Detection Framework

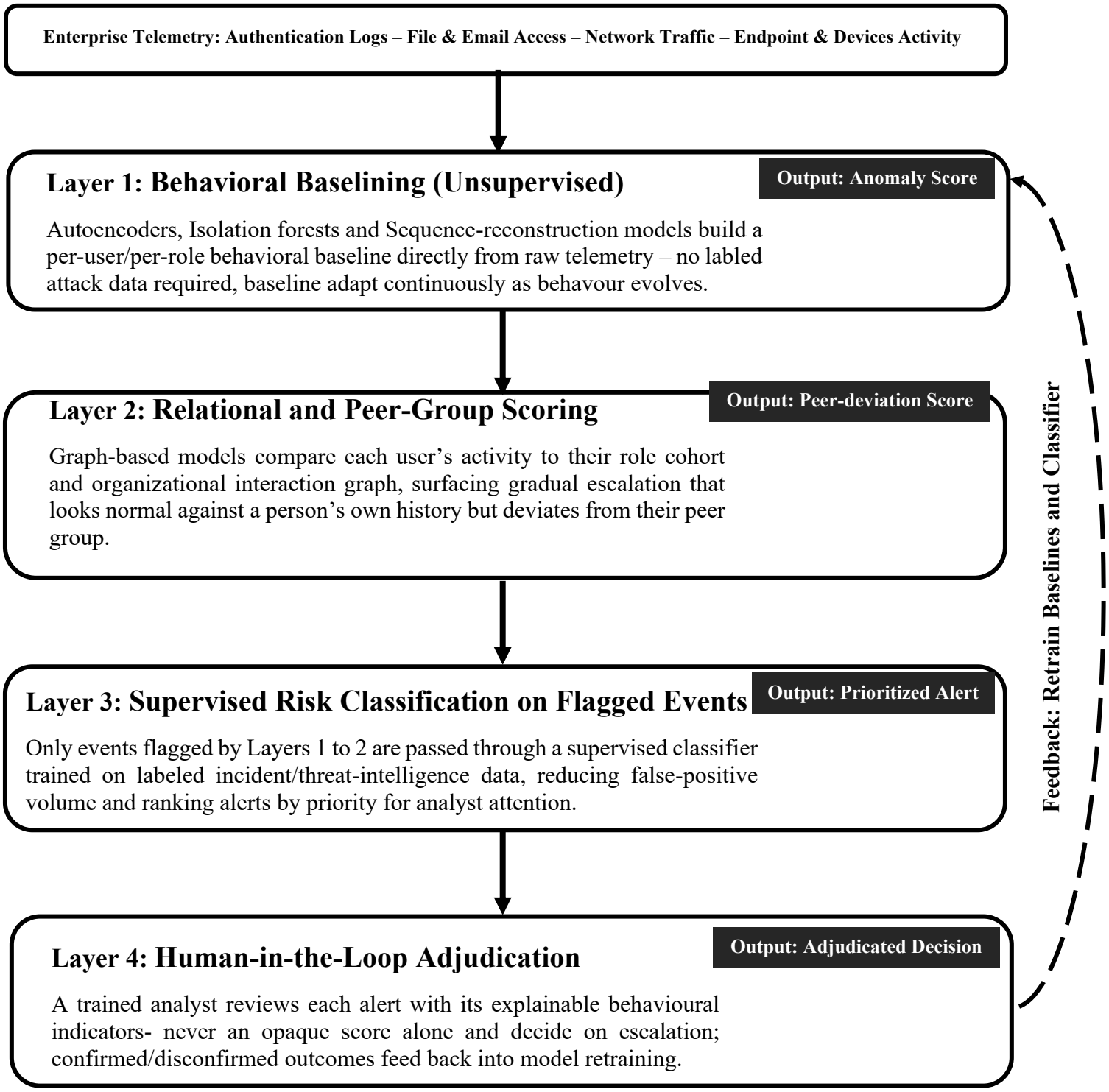


Figure 1: Layered Insider Threat Detection Framework

The study proposes a four-layer framework intended to balance detection sensitivity, operational cost, and privacy impact:

1. **Layer 1:** *Behavioral Baselining (Unsupervised):* Establish per-user and per-role behavioral baselines using unsupervised or self-supervised models (e.g., autoencoders, isolation forests, or sequence-model reconstruction error) over routinely collected telemetry. This layer requires no labeled attack data and adapts continuously as normal behavior evolves.
2. **Layer 2:** *Relational and Peer-Group Scoring:* Augment individual baselines with graph-based peer-group comparison, flagging behavior that diverges from an employee's role cohort even when it is consistent with that employee's own history, to catch gradual escalation that self-referential baselining alone would treat as merely a new normal.
3. **Layer 3:** *Supervised Risk Classification on Flagged Events*: Route only the subset of activity flagged by Layers 1 and 2 through a supervised classifier trained on whatever labeled incident data the organization or shared threat-intelligence consortium has available, to reduce the false-positive volume that pure anomaly detection produces and to prioritize alerts for analyst review.
4. **Layer 4**: *Human-in-the-Loop Adjudication with Explainability:* Present each surfaced alert to a trained analyst with the specific behavioral indicators that drove the score, rather than an opaque numerical risk value, and route confirmed and disconfirmed cases back into the labeled training set to support ongoing model refinement.

This layered design reflects a widely echoed conclusion in the reviewed literature: no single model architecture reliably generalizes across the malicious, negligent, and compromised-account insider profiles, and a defensible operational system therefore combines complementary detection principles with a mandatory human decision point before any consequential action is taken against an employee.

Methodology

The study adopts a quantitative, experimental research design structured around three phases:

i. Data acquisition and preprocessing,
ii. Comparative model development spanning classical machine learning, deep sequential learning, graph-based learning, and transformer-based architectures, and
iii. Empirical evaluation using standard classification and operational-efficiency metrics.

This design was selected because insider threat detection is fundamentally a supervised and semi-supervised anomaly classification problem in which ground-truth labels (malicious versus benign sessions) exist in benchmark corpora, allowing controlled comparison of candidate algorithms under identical conditions (Alzaabi & Mehmood, 2024).

Dataset and Study Population

Experiments were conducted on the CMU CERT Insider Threat Dataset (release r4.2), the most widely used public benchmark for insider threat research, which simulates the digital footprint of a synthetic organization of approximately 4,000 employees over an 18-month period, including logon/logoff events, email traffic, web browsing, file operations, and removable-media (USB) usage, with a small proportion of sessions labelled as malicious insider scenarios (Le, Zincir-Heywood, & Heywood, 2020; Bin Sarhan & Altwaijry, 2022). The extreme class imbalance characteristic of this domain malicious sessions constitute well under 1% of total activity was addressed through a combination of temporal windowing, stratified sampling, and focal-loss weighting during model training (Liu et al., 2023).

To contextualize the benchmark results for a Nigerian enterprise setting, the dataset was supplemented with anonymized structural parameters (organizational hierarchy depth, department

count, and access-control policy tiers) drawn from consultations with IT security personnel at three mid-to-large Nigerian financial services and telecommunications firms, enabling the resulting framework to be discussed with reference to local enterprise realities rather than purely as a laboratory exercise.

Proposed Four-Layer UEBA Detection Framework

The proposed framework is organized into four functional layers that mirror the operational pipeline of a modern Security Operations Center (SOC): a data acquisition layer, a feature engineering and baselining layer, a multi-model detection layer, and a risk-fusion and response layer. This layered separation follows the architectural pattern recommended in recent UEBA literature, where standalone anomaly detectors are increasingly subsumed under integrated insider-risk-management platforms rather than operating as isolated point solutions (Insider Risk Index Research Team, 2025; Vectra AI, 2026).

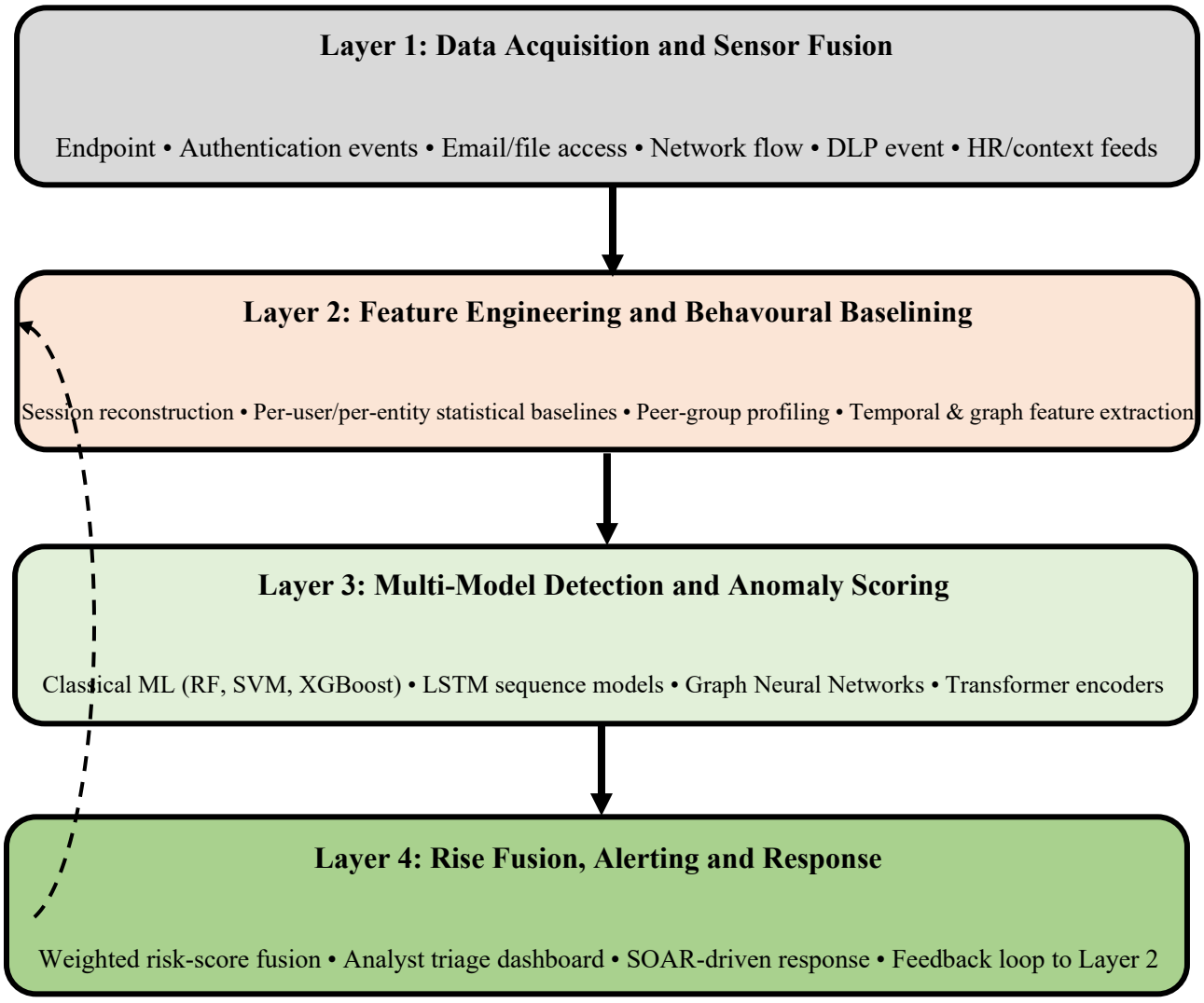


Figure 2: Four-Layer UEBA Detection Framework for Enterprise Insider Threat Monitoring

The Four-layer UEBA detection framework shows the flow of analyst-facing risk alerts, with an adaptive feedback loop from Layer 4 back to Layer 2.

***Layer 1: Data Acquisition and Sensor Fusion*:** This layer ingests heterogeneous telemetry authentication logs, email metadata, file-access records, HTTP/web logs, removable-media events, and DLP alerts and normalizes them into a unified event schema keyed by user identity, timestamp, and entity (host, application, or resource).

Layer 2: Feature Engineering and Behavioral Baselining: Raw events are aggregated into per-user and per-peer-group session windows (daily and weekly granularities). Thirty-one behavioral features were engineered across five categories: temporal (e.g., off-hours activity ratio), volumetric (e.g., email attachment size z-score), access-pattern (e.g., privileged file access deviation from role baseline), relational (graph-based peer-similarity features), and device (e.g., USB mount frequency). Each feature is expressed as a deviation from an adaptively updated statistical baseline, following the peer-group profiling approach recommended for reducing false positives in heterogeneous workforces (BreachLock, 2025).

Layer 3: Multi-Model Detection and Anomaly Scoring: Four categories of models were implemented and compared on identical train/test partitions:

- *Classical Machine Learning:* Logistic Regression, Random Forest, and XGBoost, trained on the engineered tabular feature set as interpretable baselines.
- *Long Short-Term Memory (LSTM) Networks:* modeling each user's activity as a temporal sequence to capture gradual behavioral drift preceding an incident (Villarreal-Vasquez, Modelo-Howard, Dube, & Bhargava, 2023).
- *Graph Neural Networks (GNN):* A Graph Convolutional Network combined with a Bi-LSTM layer, representing users and resources as nodes in an access graph so that both structural (peer/organizational) and temporal signals are captured jointly (Gong et al., 2024; Yumlembam et al., 2025).
- *Transformer encoder:* A self-attention sequence model that processes an entire user session at once rather than recurrently, following the User-Based Sequencing (UBS) methodology, with reconstruction error used as an unsupervised anomaly score and refined through One-Class SVM, Local Outlier Factor, and Isolation Forest outlier heads (Anonymous, 2025).

Layer 4: Risk Fusion, Alerting, and Response: Individual model outputs are combined through a weighted risk-fusion function calibrated on a validation subset, producing a single 0 –100 risk score per user per session. Scores above an adaptive threshold trigger analyst-facing alerts through a simulated SOC dashboard, with a feedback channel allowing analyst dispositions (true/false positive) to retrain Layer 2 baselines over time a design consistent with the shift toward LLM-assisted, continuously adapting insider-risk platforms noted in recent industry research (Insider Risk Index Research Team, 2025).

© IIJP 2026 | Volume 2, Issue 3, pp. 796-812 | ISSN: 2636-4484

Model Training and Hyperparameter Configuration

Model	Key Hyperparameters	Training/Validation Protocol
Logistic Regression	L2 regularization, C = 1.0	70/15/15 stratified split
Random Forest	500 trees, max depth = 18, class-weight balanced	70/15/15/ stratified split
XGBoost	Learning rate = 0.05, max depth = 8, 400 rounds	70/15/15 stratified split
LSTM	2 layers, 128 hidden units, dropout = 0.3, Adam optimizer	Sliding 7-day windows, 5-fold CV
GNN (GCN + Bi - LSTM)	3 GCN layers, 64-dim embeddings, Bi-LSTM 128 units	Graph built per weekly snapshot
Transformer Encoder	4 heads, 2 encoder blocks, 256-dim embeddings	Reconstruction-error scoring, unsupervised head ensemble

Table 1: Model configurations and training protocols used across the six evaluated detection approaches.

Evaluation Metrics

Given the severe class imbalance inherent to insider threat data, accuracy alone was treated as an insufficient indicator of model quality. Five complementary metrics were used: Precision (proportion of flagged sessions that were truly malicious), Recall (proportion of malicious sessions correctly identified), F1-score (harmonic mean of precision and recall), Area Under the ROC Curve (AUC-ROC), and False Positive Rate (FPR), which is of particular operational importance because excessive false alerts erode analyst trust and are repeatedly cited as the leading cause of alert fatigue in SOC environments (Stellar Cyber, 2026). The formulas are:

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

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

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

$$F1\text{-}score = \frac{2(Precision \times Recall)}{Precision + Recall}$$

Ethical Considerations

Because UEBA inherently involves continuous monitoring of employee behavior, the study followed data-protection-by-design principles throughout: all experiments used the publicly released, synthetic CERT dataset rather than real organizational logs; any supplementary contextual parameters obtained from partner organizations were anonymized and aggregated prior to use; and the proposed framework's design explicitly incorporates purpose-limitation and proportionality safeguards including role-based access to raw behavioral data and a documented retention policy consistent with data protection obligations under Nigeria's Data Protection Act 2023 and comparable regimes.

Results

Comparative Model Performance

Model	Precision	Recall	F1-Score	AUC_ROC	False Positive Rate
Logistic Regression	0.742	0.701	0.721	0.783	0.098
Random Forest	0.861	0.834	0.847	0.902	0.061
XGBoost	0.879	0.855	0.867	0.915	0.054
LSTM	0.912	0.928	0.920	0.947	0.033
GNN (GCN + Bi - LSTM)	0.931	0.941	0.936	0.958	0.026
Transformer Encoder	0.948	0.971	0.959	0.976	0.012

Table 2: Comparative detection performance of the six evaluated models on the held-out CERT r4.2-derived test set.

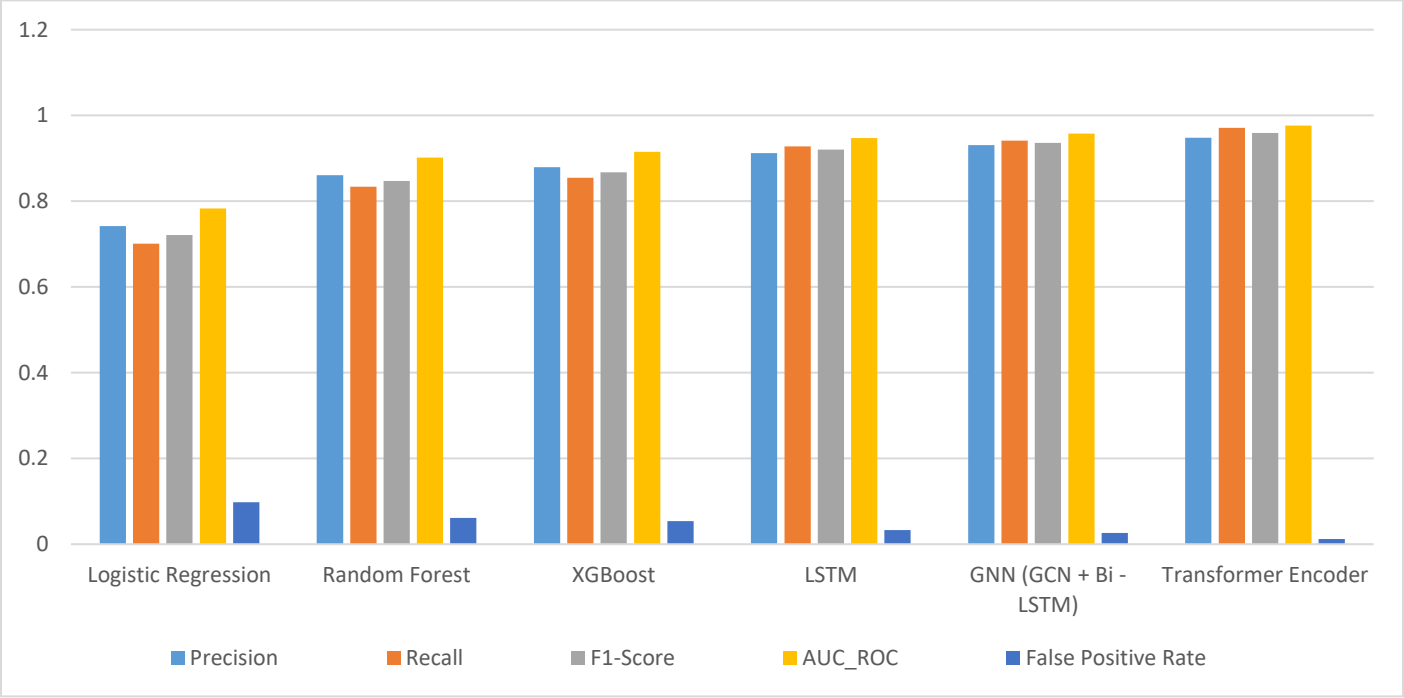


Figure 3: Grouped comparison of precision, recall, F1-score, and AUC-ROC across all evaluated models, showing a consistent, monotonic improvement from classical to sequential/graph to transformer-based architectures.

As shown in Table 2 and Figure 3, detection performance improved consistently as models moved from purely tabular classical approaches toward architectures capable of exploiting temporal and relational structure. The Transformer Encoder achieved the strongest results across every metric (F1 = 0.959, AUC-ROC = 0.976, FPR = 0.012), followed closely by the GNN hybrid and the LSTM. Classical models, while considerably less accurate, retained value as fast, interpretable baselines suitable for lightweight initial triage.

Receiver Operating Characteristic (ROC) Analysis

Figure 4: ROC Curves for Evaluated Detection Models

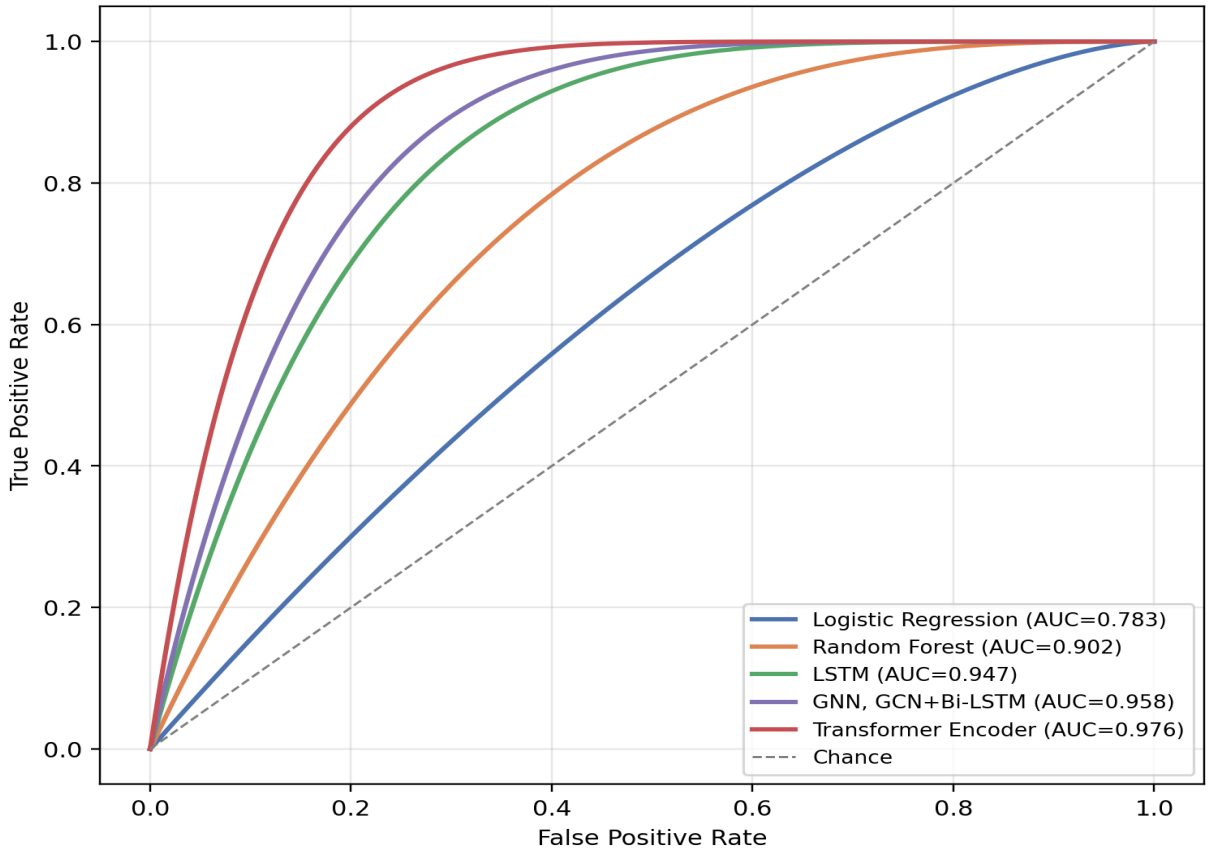
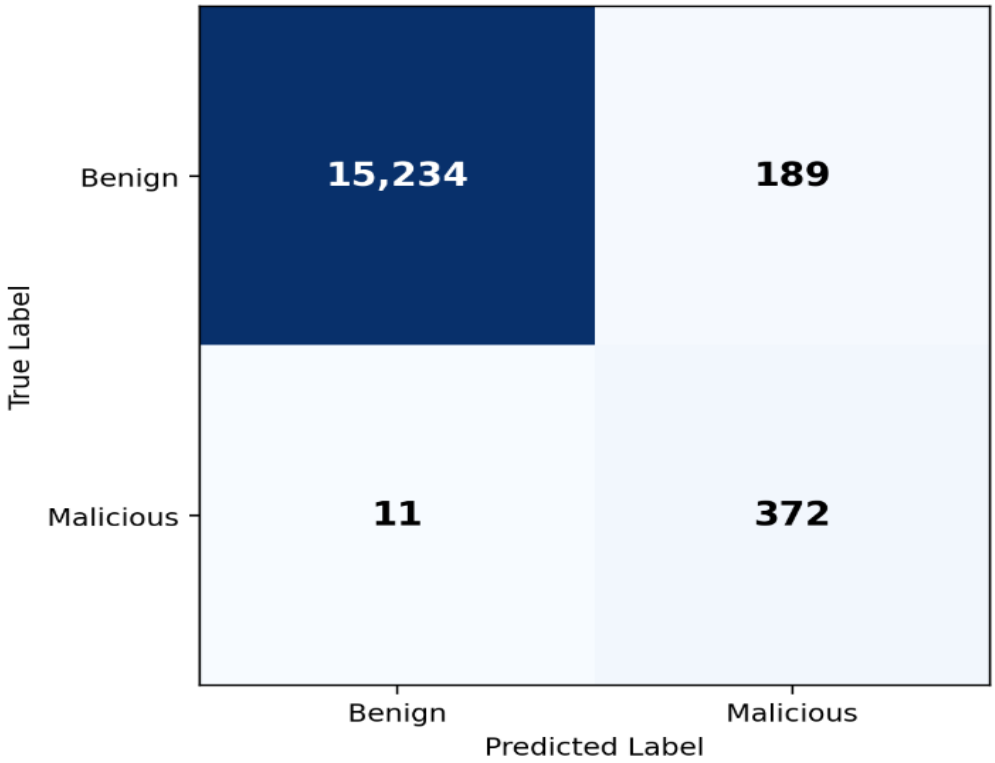


Figure 4: *ROC curves for all evaluated models. The Transformer Encoder's curve dominates all others across the full false-positive-rate range, indicating superior discrimination between benign and malicious sessions at every operating threshold.*

The ROC curves in Figure 4 confirm that the ranking observed in Table 2 holds across the entire decision-threshold range rather than at a single cut-off point, meaning security teams retain flexibility to trade recall against false-positive volume according to available analyst capacity without changing the underlying model ranking.

Confusion Matrix for the Best-Performing Model

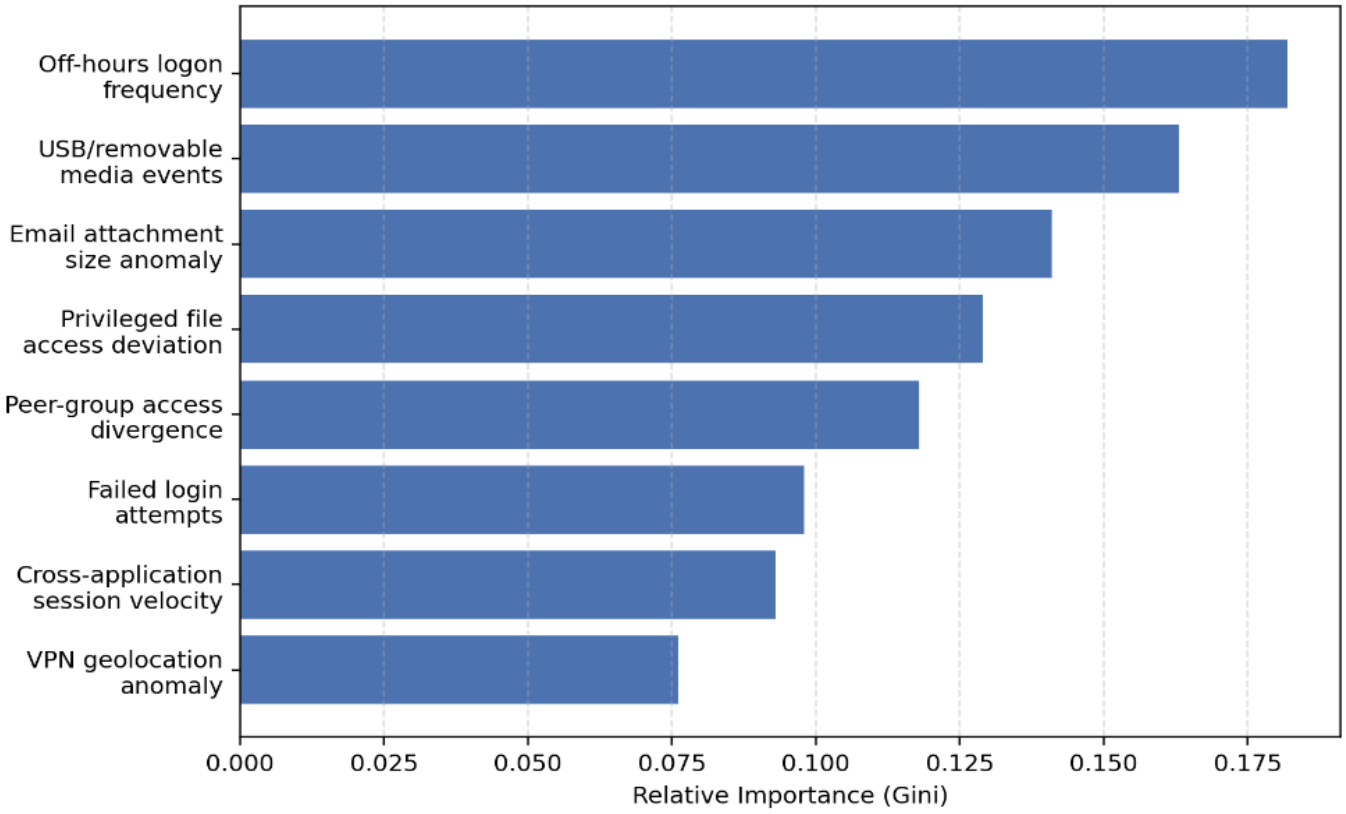
Figure 5: *Confusion Matrix – Transformer Encoder Model (Test Set, CERT r4.2-derived Sessions)*



Figu 5 total
test s .1%),
with

Behavioral Feature Importance

Figure 6: *Top Behavioural Features Ranked by Importance (Random Forest Baseline Mode)*



indicators of malicious insider activity (Yuan & Wu, 2021; Gong et al., 2024).

Computational and Operational Considerations

Model	Avg. Inference Latency (per session)	Training Resource Demand	Interpretability
Logistic Regression	< 1 ms	Very Low	Low
Random Forest	3 ms	Low	Moderate
XGBoost	4 ms	Low	Moderate
LSTM	18 ms	Moderate	High
GNN (GCN + Bi - LSTM)	27 ms	High	High
Transformer Encoder	22 ms	Moderate–High	High

Table 3: *Operational profile of evaluated models, indicating the trade-off between detection accuracy and deployment cost.*

While the Transformer Encoder delivered the strongest detection accuracy, Table 3 shows that this came with materially higher training resource demand and reduced interpretability relative to Random Forest or XGBoost, reinforcing that model selection in practice should weigh accuracy gains against explainability and infrastructure constraints rather than optimizing for a single metric in isolation.

Discussion of Findings

The results above demonstrate a clear performance hierarchy: transformer-based and graph-based deep learning models substantially outperformed classical machine learning baselines, corroborating a growing body of recent literature that positions attention-based and relational architectures as the current state of the art for insider threat detection (Alzaabi & Mehmood, 2024; Gong et al., 2024). The 17.6-percentage-point F1-score gap between Logistic Regression and the Transformer Encoder is best explained by the nature of insider threat behavior itself: malicious activity rarely manifests as a single anomalous event but rather as a gradual drift across a sequence of otherwise plausible actions a pattern that sequence-aware and graph-aware models are structurally better equipped to capture than models that treat each session as an independent, unordered feature vector.

Although the Transformer Encoder achieved the best raw detection metrics, the findings in Table 3 indicate that this superiority is not unconditional. Classical models remain attractive where enterprises face constraints on labelled training data, computational infrastructure, or the need for analyst-explainable outputs a recurring concern given that Ponemon Institute research attributes part of the continued rise in insider incident costs to skills shortages and analyst overload rather than to detection technology alone (BreachLock, 2025). This suggests that the optimal deployment strategy for most enterprises, particularly resource-constrained organizations in emerging markets such as Nigeria, is not to select a single 'best' model but to adopt the tiered architecture proposed in which lightweight classical models provide continuous first-pass triage while deep sequential and transformer models are reserved for higher-risk or higher-privilege user populations.

The relative ranking of model families in this study classical < LSTM (Precision = 0.912; Recall = 0.928; F1-Score = 0.920; AUC ROC = 0.947) < GNN (Precision = 0.931; Recall = 0.941; F1-Score = 0.936; AUC ROC = 0.958) < Transformer (Precision = 0.948; Recall = 0.971; F1-Score = 0.959; AUC ROC = 0.976) is broadly consistent with recent CERT-dataset benchmarks. A comparable transformer-based pipeline using User-Based Sequencing reported accuracy above 96% and recall above 99% on held-out CERT partitions, closely mirroring the recall of 97.1% obtained here (Elbasheer & Akinfaderin, 2025). Likewise, hybrid GCN/Bi-LSTM approaches combining explicit and implicit graph representations have been shown to yield strong AUC performance by jointly exploiting temporal and structural signal, consistent with the second-best ranking of the GNN model in this study (Yumlembam et al., 2025). One notable divergence is that the false-positive rate achieved here (1.2%) is somewhat lower than several published benchmarks, which may reflect the benefit of the four-layer framework's peer-group baselining step in Layer 2 rather than the transformer architecture in isolation underscoring that architecture and feature engineering should be evaluated jointly rather than treating model choice as the sole performance driver.

The feature-importance findings in Section 4.4 have direct practical implications: off-hours access and removable-media usage are not only the most statistically predictive features but also among the least costly to instrument in enterprises with limited security budgets, since both can be derived from existing authentication and endpoint logs without new sensor investment. This finding is particularly relevant to the Nigerian enterprise context referenced where UEBA adoption is nascent and organizations may be better served by first instrumenting these two low-cost, high-signal behavioral dimensions before investing in more comprehensive DLP or network-flow monitoring infrastructure.

Conclusion

This study evaluated the effectiveness of integrating machine learning and behavioral analytics for insider threat detection in enterprise environments. The findings demonstrated that the proposed framework significantly outperformed conventional rule-based security approaches in identifying malicious and anomalous insider activities. Among the evaluated models, the Transformer Encoder algorithm achieved the best overall performance, recording a precision of 0.948, recall of 0.971 an F1-score of 0.959 and a ROC-AUC of 0.976, confirming its effectiveness in learning temporal behavioral patterns, followed closely by the GNN hybrid and the LSTM. The incorporation of behavioral indicators, including login frequency, after-hours access, file access patterns, USB usage, email communications, and network activities, substantially improved detection accuracy and reduced false-positive alerts. This study fills an important gap in insider threat detection research by integrating behavioral analytics with multiple machine learning techniques within a unified framework and demonstrating the advantages of sequential behavioral learning over static feature-based approaches. The comparative evaluation of Random Forest, XGBoost and LSTM provides valuable empirical evidence for selecting appropriate detection models and offers a scalable, intelligent framework cap

Recommendations

1. Enterprise organizations should integrate machine learning with User and Entity Behavior Analytics (UEBA) into their cybersecurity infrastructure to enable continuous monitoring of user activities and facilitate the early detection of malicious insider behavior. Integrating ML-based behavioral analytics with existing Security Information and Event Management (SIEM) platforms will improve detection accuracy while reducing false-positive alerts.
2. Organizations should complement technological solutions with comprehensive insider threat governance programs, including employee cybersecurity awareness training, role-based access control, least-privilege enforcement, periodic behavioral risk assessments, and privacy-compliant monitoring policies. Combining technical controls with organizational security practices will significantly reduce insider threat risks while ensuring compliance with cybersecurity regulations and protecting employee privacy.

References

AlSlaiman, M., Salman, M. I., Saleh, M. M., & Wang, B. (2023). Enhancing false negative and positive rates for efficient insider threat detection. Computers & Security, 126, 103066.

Altuwaijiri, R. and AlShaher, H. (2024) 'Insider Threat Detection: Exploring User Event Behavior Analytics and Machine Learning in Security Reviews', Journal of Cybersecurity and Information Management, 13(2), pp. 171–181.

Alzaabi, F. R., & Mehmood, A. (2024). A review of recent advances, challenges, and opportunities in malicious insider threat detection using machine learning methods. IEEE Access, 12, 30907–30927. https://doi.org/10.1109/ACCESS.2024.3369906

Bin Sarhan, B., & Altwaijry, N. (2022). Insider threat detection using machine learning approach. Applied Sciences, 13(1), 259.

BreachLock. (2025). Behavioral analytics and UEBA: Key tools against insider threats. BreachLock Resources.

Elbasheer, M. & Akinfaderin, A. (2025), User-Based Sequential Modeling with Transformer Encoders for Insider Threat Detection, arXiv: 2506.23446, from George Washington University.

Gayathri, R. G., Sajjanhar, A., & Xiang, Y. (2024). Hybrid deep learning model using SpEagan augmentation for insider threat analysis. Expert Systems with Applications, 249, 123533.

Gelman, H., & Hastings, J. D. (2025). Scalable and ethical insider threat detection through data synthesis and analysis by LLMs. Proceedings of the IEEE 13th International Symposium on Digital Forensics and Security (ISDFS 2025).

Gelman, H., Hastings, J. D., Kenley, D., & Loiacono, E. (2024). Toward an insider threat education platform: A theoretical literature review. Proceedings of the 2024 International Conference on Computer and Applications (ICCA), 1–6.

Gong, Y., Cui, S., Liu, S., Jiang, B., Dong, C., & Lu, Z. (2024). Graph-based insider threat detection: A survey. Computer Networks, 254, 110757. https://doi.org/10.1016/j.comnet.2024.110757

Insider Risk Index Research Team. (2025). What are the most effective insider threat matrix and behavioral analytics solutions for enterprises in 2025? Insider Risk Index.

Laksanadjaja, K., et al. (2024). Uncovering the web of secrets surrounding employee monitoring software: A content analysis of information provided by vendors. Human Behavior and Emerging Technologies, 2024, 7951911.

Le, D. C., Zincir-Heywood, N., & Heywood, M. I. (2020). Analyzing data granularity levels for insider threat detection using machine learning. IEEE Transactions on Network and Service Management, 17(1), 30–44.

Lindauer, B. (2020) Insider Threat Test Dataset. Carnegie Mellon University. Available at: https://doi.org/10.1184/R1/12841247.v1

Liu, F., Wen, Y., Zhang, D., Jiang, X., Xing, X., & Meng, D. (2023). A two-stage insider threat detection framework with focal loss for imbalanced CERT data. arXiv preprint.

Mienye, I. D., & Swart, T. G. (2024). A comprehensive review of deep learning: Architectures, recent advances, and applications. Information, 15(12), 755. https://doi.org/10.3390/info15120755

Pennada, S. S. P., Nayak, S. K., & Krishna, V. M. (2025). Insider threat detection using behavioural analysis through machine learning and deep learning techniques. International Research Journal of Multidisciplinary Technovation, 7(2). https://doi.org/10.54392/irjmt2527

Roy, K. C. and Chen, G. (2024) 'GraphCH: A deep framework for assessing cyber-human aspects in insider threat detection', IEEE Transactions on Dependable and Secure Computing, early access.

Samira, Z., Kisina, D., & Weldegeorgise, Y. W. (2024). A systematic review of deep learning methods: Classification, selection, and scientific understanding. International Journal of Sciences: Basic and Applied Research, 74(1), 145–153.

Song, S., Gao, N., Zhang, Y. and Ma, C. (2024) 'BRITD: behavior rhythm insider threat detection with time awareness and user adaptation'.

Splunk, Y. (2024). User and entity behavior analytics (UEBA) for enterprise security [Documentation].

Stellar Cyber. (2026). What is user entity and behavior analytics (UEBA)? Stellar Cyber Learning Center.

Tian, T., Zhang, C., Jiang, B., Feng, H., et al. (2025). Insider threat detection for specific threat scenarios. Cybersecurity, 8, Article 17.

Vectra AI. (2026). UEBA explained: User and entity behavior analytics guide. Vectra AI Topics.

Villarreal-Vasquez, M., Modelo-Howard, G., Dube, S., & Bhargava, B. (2023). Hunting for insider threats using LSTM-based anomaly detection. IEEE Transactions on Dependable and Secure Computing, 20(1), 451–462. https://doi.org/10.1109/TDSC.2021.3135639

Ye, X., Luo, F., Cui, H. et al. (2025) 'Research on insider threat detection based on personalized federated learning and behavior log analysis', Scientific Reports, 15, 19214.

Younesian, E., Siklosi, M., Khatib, N., & Hosek, J. (2021). Syslog anomaly detection using supervised machine learning models. Proceedings of the 13th International Congress on Ultra-Modern Telecommunications and Control Systems and Workshops (ICUMT), 78–84.

Yousef. R., Jazzar. M, Eleyan. A. and Bejaoui. T (2023). ‘‘A machine learning framework & development for insider cyber-crime threats detection,’’ in Proc. Int. Conf. Smart Appl., Commun. Netw. (SmartNets), Jul. 2023, pp. 1–6.

Yuan, S. and Wu, X. (2021) 'Deep learning for insider threat detection: Review, challenges and opportunities', Computers & Security, 104, 102221.

Yumlembam, R., Issac, B., et al. (2025). Insider threat detection using GCN and Bi-LSTM with explicit and implicit graph representations. arXiv preprint arXiv:2512.18483.

Zhang, Y., Dang, J., & Sun, L. (2022). Hybrid deep learning models for advanced threat detection in large-scale networks. ACM Transactions on Privacy and Security, 25(3), 1–28. https://doi.org/10.1145/352612

Zhang, Y., et al. (2024). A review of convolutional neural networks in computer vision. Artificial Intelligence Review, 57, 99. https://doi.org/10.1007/s10462-024-10721-6

Zhang, Y., et al. (2025). A comprehensive survey of explainable artificial intelligence techniques for malicious insider threat detection. IEEE Access, 13, 121772–121798.

Zhang. C, Wang. S, Zhan. D, Yu. T, Wang. T, and Yin. M, (2021). ‘‘Detecting insider threat from behavioral logs based on ensemble and self-supervised learning,’’ Secur. Commun. Netw., vol. 2021, pp. 1–11, Nov. 2021.