AN AI-DRIVEN APPROACH TO INSIDER THREAT DETECTION THROUGH BEHAVIOURAL ANALYTICS

C S N DURGHA
Research Scholar
Shri JJT University
Rajasthan.

Dr. PRASADU PEDDI
Supervisor
Shri JJT University
Rajasthan.

Dr. VENKATA RAJANI KATURI
Co-Supervisor
Shri JJT University
Rajasthan.

Abstract

Insider threats represent a complex cybersecurity challenge because individuals with legitimate access to organizational systems can potentially misuse their privileges, intentionally or unintentionally, to compromise information assets. Traditional security mechanisms based primarily on authentication, access control, signatures, and predefined rules may have difficulty identifying subtle deviations in legitimate user activity. Artificial intelligence (AI) and behavioural analytics provide an opportunity to address this challenge by continuously analysing user and entity activities and identifying patterns that differ from established behavioural baselines. Recent research highlights the use of supervised, unsupervised, deep-learning, anomaly-detection, and hybrid approaches for insider-threat detection, while also identifying persistent challenges involving class imbalance, false positives, limited labelled data, and interpretability.

This article examines an AI-driven approach in which behavioural data from authentication systems, endpoints, networks, applications, file-access systems, and data-transfer activities are analysed to identify potentially suspicious patterns. The proposed approach combines behavioural profiling, feature extraction, anomaly detection, machine-learning-based risk assessment, and human-led investigation. The study uses a conceptual and secondary research methodology to examine existing literature and develop a framework for AI-assisted insider-threat detection. The findings indicate that behavioural analytics can enhance the identification of subtle and evolving anomalies, particularly when multiple behavioural indicators are analysed together. However, AI-based detection should complement rather than replace human investigation and organizational security controls. Privacy, explainability, data quality, model bias, false positives, and adversarial adaptation remain important considerations. The study proposes a human-centred AI framework for improving proactive insider-threat detection while maintaining appropriate cybersecurity governance.

Keywords: *Artificial Intelligence, Insider Threat, Behavioural Analytics, Machine Learning, User Behaviour Analytics, Anomaly Detection, Cybersecurity, Threat Detection, Risk Assessment, User and Entity Behaviour Analytics.*

1. Introduction

The increasing dependence of organizations on digital infrastructure has expanded both the volume and value of information processed through enterprise networks. Cloud computing, remote work, mobile devices, interconnected applications, and digital collaboration platforms have created highly distributed computing environments. These developments have also increased the importance of protecting organizational systems from threats originating from individuals who possess legitimate access.

An insider threat differs from many conventional external cyberattacks because the individual or compromised account may already possess valid credentials and authorized access. Consequently,

conventional perimeter-based security mechanisms may not identify suspicious activity when it occurs within an authorized session. Research from the Software Engineering Institute has emphasized that insider-threat detection requires the integration and analysis of information from multiple technical and organizational sources rather than relying on isolated security events. (SEI)

Behavioural analytics provides an alternative approach by establishing a baseline of normal user activity and identifying deviations from that baseline. Relevant indicators may include login times, authentication frequency, files accessed, applications used, network connections, data transfers, privilege usage, and endpoint interactions.

AI and machine learning can extend behavioural analytics by processing large volumes of heterogeneous security data. Rather than requiring security analysts to define every possible suspicious behaviour in advance, machine-learning models can identify statistical patterns, clusters, correlations, and anomalies. Recent research reviews have identified machine learning, deep learning, anomaly detection, graph-based approaches, and multimodal analysis as important directions in insider-threat research.

The Software Engineering Institute has also noted that AI and machine learning can examine behavioural patterns that may be difficult to identify through simple threshold-based monitoring, including activity distributed over longer periods.

However, AI-driven insider-threat detection presents important challenges. A deviation from normal behaviour does not necessarily indicate malicious activity. Employees may legitimately change their working patterns, access new systems, work outside conventional hours, or transfer large volumes of information as part of their duties. Therefore, an effective system should prioritize suspicious activity for human investigation rather than automatically treating an anomaly as proof of malicious intent.

2. Objectives

The study has the following objectives:

1. To examine the nature and characteristics of insider threats in modern digital organizations.
2. To analyse the role of behavioural analytics in identifying abnormal user activity.
3. To examine the application of artificial intelligence and machine learning in insider-threat detection.
4. To develop a conceptual AI-driven framework for behavioural insider-threat detection.
5. To identify behavioural indicators that can contribute to insider-risk assessment.
6. To examine the potential benefits of supervised, unsupervised, and hybrid machine-learning techniques.
7. To analyse challenges associated with false positives, data imbalance,

privacy, explainability, and model reliability.

8. To propose a human-centred approach for integrating AI-generated alerts with security-analyst investigation.

3. Research Methodology

3.1 Research Design

The study follows a **descriptive, analytical, and conceptual research design** based primarily on secondary sources. The methodology examines existing research on insider threats, behavioural analytics, artificial intelligence, machine learning, anomaly detection, and user/entity behaviour analytics.

Recent research has specifically reviewed machine-learning approaches for insider-threat detection and identified issues such as data imbalance, feature heterogeneity, limited labelled datasets, and interpretability.

3.2 Data Sources

Secondary information was examined from:

- Peer-reviewed research articles
- IEEE publications
- Cybersecurity research reports
- Government cybersecurity guidance
- Carnegie Mellon University Software Engineering Institute publications
- Research on machine learning and anomaly detection
- Publicly discussed insider-threat datasets and benchmarks

The CERT/SEI research ecosystem is particularly relevant because it has developed extensive research and case-based resources for studying insider threats. (SEI)

3.3 Behavioural Data

The proposed framework considers multiple categories of behavioural information:

Behavioural Dimension	**Illustrative Indicators**
Authentication	Login time, login frequency, failed attempts
File Activity	Files opened, copied, modified, deleted
Network Activity	Connections, bandwidth, unusual destinations
Application Usage	Applications accessed and usage frequency
Privilege Usage	Administrative operations and privilege changes
Data Transfer	Uploads, downloads, removable-media activity
Endpoint Behaviour	Processes, devices and system interactions
Temporal Behaviour	Activity outside established patterns
Access Behaviour	Systems and resources accessed

Behavioural Dimension	Illustrative Indicators
Session Behaviour	Duration, interaction and usage patterns

3.4 AI and Machine-Learning Approaches

The framework considers three major learning approaches.

Supervised Learning

Supervised algorithms can be trained using labelled examples of normal and suspicious activity. Classification models can then estimate whether new observations resemble previously identified threat patterns.

Unsupervised Learning

Unsupervised algorithms can identify unusual patterns without requiring every threat example to be labelled. This is particularly relevant because confirmed insider-threat incidents can be relatively limited compared with normal organizational activity.

Hybrid Learning

A hybrid architecture can combine anomaly detection with classification and behavioural risk scoring. Recent research has identified hybrid and multimodal approaches as promising directions for addressing the limitations of individual detection techniques.

4. Results and Discussion

Because this research is based on secondary literature and a conceptual framework rather than a newly collected experimental dataset, the results represent **analytical findings from the reviewed literature and proposed framework**, rather than newly measured model accuracy.

4.1 Behavioural Analytics Improves Contextual Detection

The analysis indicates that behavioural analytics can provide greater contextual information than isolated security alerts. A single failed login may have little significance, whereas a combination of unusual login activity, access to unfamiliar resources, large-scale data retrieval, and abnormal network communication may warrant further investigation.

SEI research notes that behavioural analytics and user/entity behaviour analytics can incorporate multiple indicators and data sources to provide a broader picture of insider risk. (SEI)

4.2 AI Enables Continuous Behavioural Analysis

AI-based systems can continuously process security events and identify relationships that may be difficult for human analysts to detect manually. This is particularly important in large organizations where security systems can generate extremely large volumes of events.

Research on machine learning and insider threats has highlighted the ability of AI/ML approaches to analyse disparate datasets and identify potentially meaningful behavioural patterns. At the same time, SEI cautions that machine-generated alerts can be noisy and should support, rather than replace, human analysis.

4.3 Behavioural Indicators for Risk Assessment

The proposed approach treats insider risk as a combination of multiple indicators rather than a single suspicious event.

For example:

Normal behaviour:
A user accesses systems and files consistent with their normal role.

Behavioural deviation:
The user begins accessing resources that are unusual for their established activity profile.

Elevated-risk pattern:
The user combines unusual resource access with abnormal data downloads, privilege use, and external data transfers.

The final interpretation should depend on organizational context and human investigation.

4.4 Comparative Analysis of Detection Approaches

Detection Approach	Major Capability	Key Challenge
Rule-Based Detection	Identifies predefined conditions	Limited against unknown behaviours
Signature-Based Detection	Detects known threat patterns	Weak against novel activity
Behavioural Analytics	Identifies deviations from normal behaviour	Can produce false positives
Supervised ML	Learns from labelled examples	Requires reliable labelled data
Unsupervised ML	Detects previously unknown anomalies	Interpretation can be difficult
Deep Learning	Models complex relationships	High computational and data requirements
Hybrid AI	Combines multiple detection methods	Greater architectural complexity

Recent reviews similarly identify false positives, class imbalance, limited labelled data, and interpretability as continuing challenges in machine-learning-based insider-threat detection.

4.5 Proposed AI-Driven Detection Framework

The proposed framework can be represented as:

Data Sources
↓
Data Collection and Integration
↓
Data Pre-processing
↓

User Behaviour Profiling
↓
Feature Extraction
↓
AI/ML Anomaly Detection
↓
Behavioural Risk Scoring
↓
Alert Prioritization
↓
Security Analyst Investigation
↓
Incident Response and Model Feedback

This architecture supports continuous improvement because analyst feedback can be incorporated into future model development.

4.6 Human-in-the-Loop AI

An important finding is that AI should function as an **analytical decision-support mechanism** rather than an autonomous authority.

For example, an AI system may identify that a user's behaviour has deviated substantially from their historical pattern. The system can assign an elevated risk score and generate an alert. A security analyst can then examine additional contextual information before determining an appropriate response.

This approach is consistent with SEI research emphasizing that machine-learning outputs can help prioritize investigative attention while human analysts remain important for interpreting the results.

4.7 Key Challenges

False Positives

Legitimate behavioural changes can be incorrectly classified as suspicious. Reducing false positives is therefore a major requirement for practical deployment.

Data Imbalance

Normal user activity generally constitutes a much larger volume of data than confirmed malicious insider activity. Recent research identifies class imbalance as a significant machine-learning challenge.

Explainability

Security teams need to understand why an AI model generated an alert. Explainable AI can help analysts interpret important features and behavioural deviations.

Privacy

Behavioural monitoring can involve sensitive information concerning employees and organizational activities. Data collection should therefore follow applicable privacy, security, and governance requirements.

Model Adaptation

User behaviour changes over time. A model that does not adapt may incorrectly classify legitimate behavioural changes as anomalies.

Adversarial Adaptation

A sophisticated threat actor may deliberately alter activity patterns to avoid detection. Continuous model evaluation is therefore necessary.

5. Discussion

The findings indicate that AI-driven behavioural analytics can strengthen insider-threat detection by moving organizations

from static monitoring toward **continuous, adaptive, and context-aware analysis**.

Traditional controls may identify individual events, while AI-based behavioural analytics can examine relationships among events over time. For example, an individual data-transfer event may not be suspicious on its own, but a sustained change in access behaviour combined with unusual data movement may provide a stronger basis for investigation.

Recent research also indicates growing interest in multimodal, graph-based, adaptive, and explainable approaches.

The use of AI should nevertheless be accompanied by organizational governance. Security teams should define appropriate data sources, establish behavioural baselines, periodically evaluate model performance, monitor false positives, and ensure that human analysts can investigate alerts.

The approach should also avoid treating statistical anomaly as equivalent to malicious intent. An AI model identifies patterns; organizational investigators must establish context and determine the appropriate response.

6. Conclusion

Insider threats remain an important cybersecurity challenge because legitimate access can be misused without necessarily triggering traditional perimeter-based security mechanisms. Behavioural analytics provides a mechanism for identifying deviations from established user and entity activity, while artificial intelligence and machine learning can automate the analysis of large and complex security datasets.

This study proposed an **AI-driven behavioural analytics framework** consisting of data collection, behavioural profiling, feature extraction, anomaly detection, risk scoring, alert prioritization, and human-led investigation. The reviewed literature indicates that machine-learning and AI techniques can support the detection of subtle and evolving insider-threat patterns, but their effectiveness depends on data quality, appropriate modelling, organizational context, and continuous evaluation. (IEEE Xplore)

The study concludes that the most sustainable approach is not fully automated threat judgement but **human-centred AI-assisted cybersecurity**. AI can process large volumes of behavioural information and prioritize potentially significant anomalies, while security professionals provide contextual interpretation and determine appropriate responses. Future research should focus on explainable AI, multimodal behavioural modelling, privacy-preserving analytics, adaptive learning, graph-based approaches, and rigorous evaluation using representative insider-threat datasets.

References

1. *Alawneh, A., & Al-Dosari, K. (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 (IEEE Xplore)*

2. *Software Engineering Institute. (2017). Machine Learning and Insider Threat.*

Carnegie Mellon University, CERT Division. (SEI)

3. *Software Engineering Institute. (2020). Aggregate Indicator Measurement Method Characterization. Carnegie Mellon University, CERT Insider Threat Center. DOI: 10.1184/R1/13138580. (SEI)*
4. *Software Engineering Institute. (2023). The 13 Key Elements of an Insider Threat Program. Carnegie Mellon University. (SEI)*
5. *A Comprehensive Review of Anomaly Detection Techniques for Insider Threat Identification. (2026). **IEEE Conference Publication**, ICAUC 2026. DOI: 10.1109/ICAUC68182.2026.11440999. (IEEE Xplore)*
6. *Machine Learning-Based Approaches for Malicious Insider Threat Detection: A Survey of Recent Advances and Open Issues. (2025/2026). **IEEE Conference Publication**, ICOIICS. DOI: 10.1109/ICOIICS67115.2025.11390400. (IEEE Xplore)*
7. *Insider Threat Detection Using Machine Learning Models for User Behavior Analysis. (2025). **2025 5th International Conference on Expert Clouds and Applications (ICOECA)**. DOI: 10.1109/ICOECA66273.2025.00143. (IEEE Xplore)*
8. *Deltrux: Insider Threat Detection of End-User Behavior Analysis Using Machine Learning. (2024). **2024 3rd International Conference for Advancement in Technology (ICONAT)**. DOI: 10.1109/ICONAT61936.2024.10775095. (IEEE Xplore)*
9. *AI-Powered Insider Threat Detection Using Behavioral Biometrics. (2026). **2026 IEEE Global Symposium on Emerging and Communication Technologies (GSEACT)**. DOI: 10.1109/GSEACT68539.2026.11619981. (IEEE Xplore)*
10. *Behavioural Analysis in Human-Machine Interaction for Insider Threats. (2025). **2025 IEEE 29th International Conference on Intelligent Engineering Systems (INES)**. DOI: 10.1109/INES67149.2025.11078220. (IEEE Xplore)*
11. *AI Based Methods for Insider Threats Detection for Cloud Risk Mitigation. (2025). **International Conference on IT Innovation and Knowledge Discovery (ITIKD)**. DOI: 10.1109/ITIKD63574.2025.11004994. (IEEE Xplore)*
12. *Towards More Effective Insider Threat Countermeasures: A Survey of Approaches for Addressing Challenges and Limitations. (2024). **2024 IEEE International Systems Conference (SysCon)**. DOI: 10.1109/SysCon61195.2024.10553441. (IEEE Xplore)*