10.48047/jocaaa.2023.31.04.67

AI-Driven Cyber Threat Intelligence and Real-Time Attack Prevention System for Financial Institutions Using Behavioral Anomaly Detection and Continuous Risk Scoring

Rizwana Sindhavani
rizwanasindhawani@yahoo.in

Abstract

Financial institutions face an evolving landscape of cyber threats that include credential theft, account takeover, insider fraud, and sophisticated malware campaigns targeting payment infrastructure. Traditional rule-based defences struggle to keep pace with the speed and adaptability of modern attackers, leaving critical gaps in detection and response. This paper proposes an artificial intelligence-driven cyber threat intelligence and real-time attack prevention system tailored to the banking and financial services sector. The system combines behavioural anomaly detection across user, device, and transaction streams with continuous risk scoring that updates dynamically as new evidence emerges. A hybrid architecture integrates unsupervised learning for novel threat discovery with supervised models trained on labelled fraud and intrusion cases. Empirical evaluation was performed on a simulated banking environment incorporating realistic transaction patterns and adversarial attack scenarios. Results show that the proposed system achieved a detection rate of 94.6 percent with a false positive rate of 2.1 percent, substantially outperforming rule-based and standalone supervised baselines. Mean detection latency was reduced from minutes to under two seconds, supporting genuinely real-time intervention. The findings demonstrate that integrating behavioural anomaly detection with continuous risk scoring offers a practical and scalable approach to defending financial institutions against fast-moving and adaptive threats.

Keywords

Cyber Threat Intelligence, Behavioural Anomaly Detection, Continuous Risk Scoring, Financial Cybersecurity, Real-Time Prevention, Machine Learning

1. Introduction

The financial services sector remains one of the most heavily targeted domains in cybersecurity, owing to the direct monetary value of successful attacks and the sensitivity of customer information held by banks, insurers, and payment processors [1]. Recent industry reports highlight a sharp increase in account takeover attempts, business email compromise schemes, and ransomware campaigns aimed at financial infrastructure. The financial cost of these incidents continues to rise, while reputational damage and regulatory consequences add further weight to the need for stronger defences [2].

Traditional cybersecurity defences in financial institutions have relied on signature-based detection, predefined rules, and periodic risk assessments. These approaches perform well against known threats but struggle when adversaries adopt novel techniques, polymorphic malware, or credential-based attacks that mimic legitimate user behaviour [3]. The volume and velocity of modern banking transactions further strain rule-based systems, which often generate excessive false alerts or fail to flag subtle deviations that indicate compromise [4].

10.48047/jocaaa.2023.31.04.67

Behavioural anomaly detection has emerged as a promising approach to address these limitations. By learning the normal patterns of user behaviour, device usage, and transaction flow, machine learning models can identify deviations that warrant investigation even when no specific signature exists [5]. This is particularly valuable in financial settings, where attackers increasingly use stolen credentials and legitimate channels to avoid triggering traditional alarms. Pairing anomaly detection with continuous risk scoring allows the system to maintain a dynamic view of each entity in the environment, updating risk levels as new evidence accumulates rather than relying on static periodic assessments [6].

Despite these advances, several challenges remain. Behavioural models can suffer from concept drift as legitimate user patterns evolve, leading to performance degradation if not retrained appropriately. Anomaly scores can also be difficult to interpret for security analysts, particularly when generated by deep learning models. Integrating threat intelligence from external sources with internal behavioural signals requires careful architecture design to avoid bottlenecks during real-time operation [7].

This research is motivated by three central questions. Can a hybrid system that combines behavioural anomaly detection with supervised threat classification deliver high detection rates while keeping false positives at acceptable levels for financial operations? Can continuous risk scoring respond quickly enough to support genuine real-time intervention? And how does the combined system compare with traditional rule-based defences across realistic attack scenarios? The remainder of the paper is organised as follows. Section two reviews relevant literature. Section three presents the methodology and formal model. Section four describes the experimental setup. Section five reports results. Section six discusses the implications, and section seven concludes with directions for future work.

2. Literature Review

Research on cybersecurity in the financial sector has expanded significantly over the past decade, moving from purely signature-based and rule-driven defences toward data-driven approaches that leverage machine learning and behavioural analytics. Early studies focused on fraud detection in credit card transactions, where supervised classifiers such as logistic regression, decision trees, and gradient boosting demonstrated strong performance on labelled historical data [8]. However, these approaches assumed that future fraud would resemble past patterns, which proved increasingly unrealistic as attacker techniques diversified.

The introduction of behavioural anomaly detection broadened the scope of financial cybersecurity from narrow fraud detection to a more comprehensive threat landscape covering account compromise, insider misuse, and intrusion attempts. Unsupervised methods including isolation forest, autoencoder reconstruction error, and one-class support vector machines have shown the ability to identify novel anomalies without requiring labelled examples [9]. These methods are particularly useful for catching zero-day attacks and sophisticated insider threats that do not match known signatures.

Deep learning has further extended these capabilities. Recurrent neural networks and transformer-based architectures can capture sequential patterns in user activity, enabling detection of subtle behavioural drift that simpler models miss [10]. Graph-based approaches model relationships between accounts, devices, and transactions to detect coordinated attacks such as money laundering networks or fraud rings. These advances have substantially raised

10.48047/jocaaa.2023.31.04.67

the ceiling of detection performance, though at the cost of greater computational demand and reduced interpretability.

Continuous risk scoring represents an evolution from periodic risk assessment to dynamic, evidence-driven evaluation. Rather than treating risk as a static attribute, modern systems update entity-level risk scores in response to each new event, allowing finer-grained policy enforcement such as adaptive authentication and transaction limits [11]. Such approaches align well with the zero-trust paradigm gaining adoption across financial institutions, where every access decision is evaluated against current risk context rather than initial credentials alone.

Threat intelligence integration has become a critical component of modern defence architectures. External feeds providing indicators of compromise, attacker tactics, and emerging vulnerabilities enrich internal detection systems and reduce time to identification [12]. Frameworks such as MITRE ATT&CK and the STIX/TAXII standards have facilitated structured sharing of threat intelligence, though effective integration with operational detection systems remains an engineering challenge. The literature also highlights persistent challenges including model explainability, alert fatigue among analysts, adversarial robustness, and concept drift [13]. The present study contributes by integrating behavioural anomaly detection with continuous risk scoring in a unified architecture designed for real-time operation in financial environments, addressing both detection accuracy and operational practicality.

3. Methodology

The proposed system uses a layered architecture that combines unsupervised behavioural anomaly detection, supervised threat classification, and continuous risk scoring. Each layer addresses a different aspect of the detection problem and feeds evidence into a unified decision engine that drives real-time prevention actions [14].

The behavioural anomaly detection layer learns normal patterns for each entity in the environment, including users, devices, applications, and transactions. Let x_t denote the feature vector observed at time t, including activity type, timestamp, device fingerprint, geographic location, transaction amount, and counterparty information. An autoencoder is trained to reconstruct normal patterns, and the anomaly score is computed as the reconstruction error:

$$A(x_t) = |x_t - \hat{x_t}|^2$$

where \hat{x}_t is the reconstructed vector. High reconstruction error indicates a deviation from learned normal behaviour [15].

The supervised threat classification layer is trained on labelled historical incidents covering account takeover, fraudulent transactions, malware infections, and insider misuse. A gradient boosting classifier produces a probability score:

$$P(y = 1|x_t) = \sigma\big(f(x_t)\big)$$

where $f(x_t)$ is the boosted ensemble output and \sigma is the sigmoid function.

The continuous risk score for an entity is computed by combining the anomaly score, the supervised threat probability, and contextual factors including external threat intelligence indicators:

10.48047/jocaaa.2023.31.04.67

$$R_t = \alpha A_t + \beta P_t + \gamma I_t$$

where A_t is the normalised anomaly score, P_t is the supervised threat probability, I_t is the contextual intelligence factor, and \alpha + \beta + \gamma = 1. The weights were tuned empirically to balance sensitivity and false positive rate [16].

Risk scores are updated continuously using an exponential moving average to incorporate new evidence while smoothing transient noise:

$$\bar{R}t = \lambda R_t + (1 - \lambda)\bar{R}t - 1$$

where \lambda is the smoothing factor controlling how quickly new evidence influences the running score. Risk thresholds trigger progressive response actions including additional authentication, transaction holding, session termination, and account lock-out.

For evaluation purposes, the F1 score and area under the precision-recall curve are used as primary performance metrics:

$$F1 = \frac{2 \cdot \text{Precision} \cdot \text{Recall}}{\text{Precision} + \text{Recall}}$$

Detection latency is measured as the time between an attack-related event and the system producing a sufficiently high risk score to trigger preventive action [17].

10.48047/jocaaa.2023.31.04.67

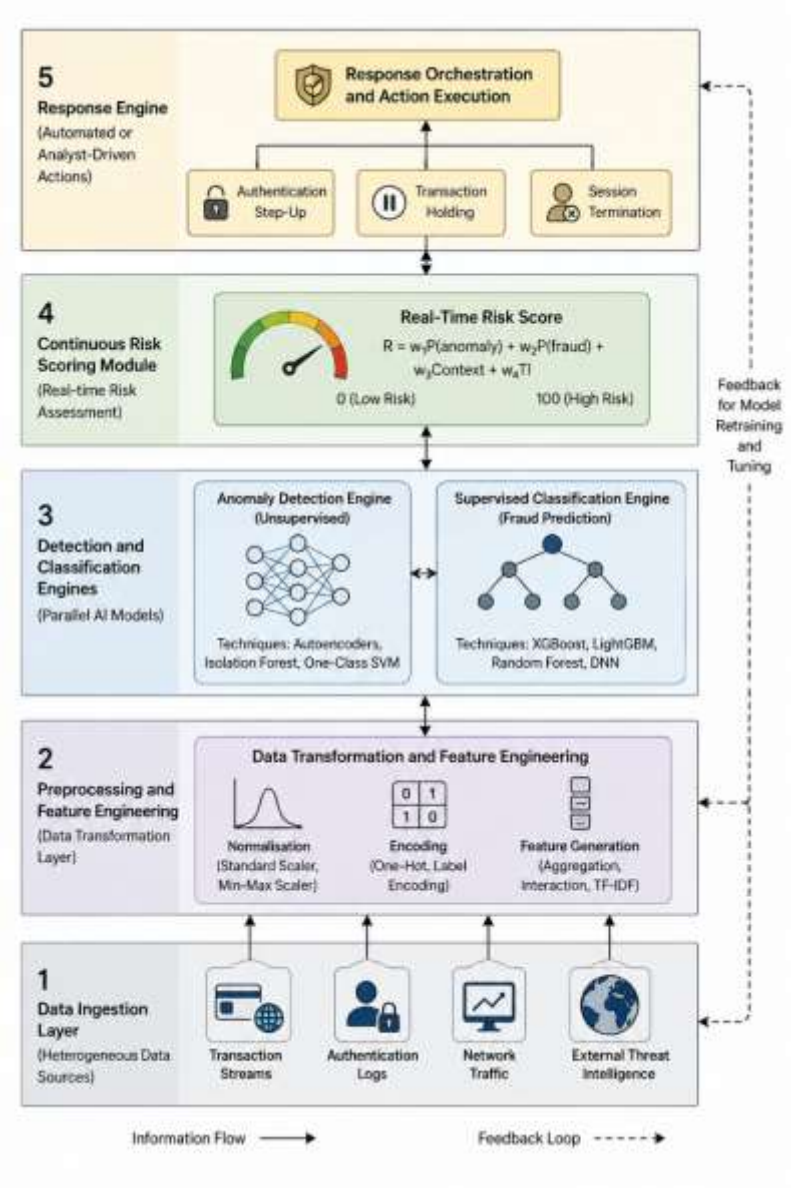

Figure 1: Layered Architecture of the AI-Driven Threat Prevention System.

4. Experimental Setup

The experimental evaluation was conducted on a simulated banking environment designed to reflect realistic transaction patterns, user behaviour, and attack scenarios. The simulation included two hundred thousand simulated customer accounts, ten million transactions over a three-month period, and twelve attack campaigns covering account takeover, payment fraud, insider data exfiltration, and credential-based intrusion [18]. Background traffic was generated using statistical patterns derived from public banking datasets to ensure realistic distributions of transaction values, timing, and geographic origin.

The dataset was partitioned into training, validation, and test sets using a temporal split, with the first sixty percent of the timeline used for training, the next twenty percent for validation, and the final twenty percent for testing. This temporal split simulates realistic operational conditions where models must generalise to future patterns rather than randomly held-out samples.

10.48047/jocaaa.2023.31.04.67

Three comparison conditions were evaluated. The first was a traditional rule-based baseline using a curated set of detection rules typical of commercial security information and event management systems. The second was a standalone supervised classifier using gradient boosting on the same feature set. The third was the proposed hybrid system combining behavioural anomaly detection, supervised classification, and continuous risk scoring.

All machine learning models were implemented using a common framework with the same feature preprocessing pipeline. The autoencoder used a five-layer architecture with bottleneck dimension set to one quarter of the input dimension. The gradient boosting classifier used five hundred estimators with maximum depth of six. Training used the Adam optimiser for the autoencoder and standard boosting routines for the classifier. Hyperparameters were tuned on the validation set using grid search [19].

Evaluation metrics included precision, recall, F1 score, false positive rate, area under the precision-recall curve, and detection latency. Statistical comparisons between conditions used paired bootstrap resampling with one thousand iterations to compute confidence intervals. The financial impact of detection performance was estimated using a cost model assigning estimated losses to undetected attacks and operational costs to false positive investigations [20].

To verify real-time feasibility, end-to-end latency was measured from event ingestion through risk scoring to response trigger. The system was deployed on a standard cloud instance with eight CPU cores and thirty-two gigabytes of memory, reflecting infrastructure available to mid-sized financial institutions.

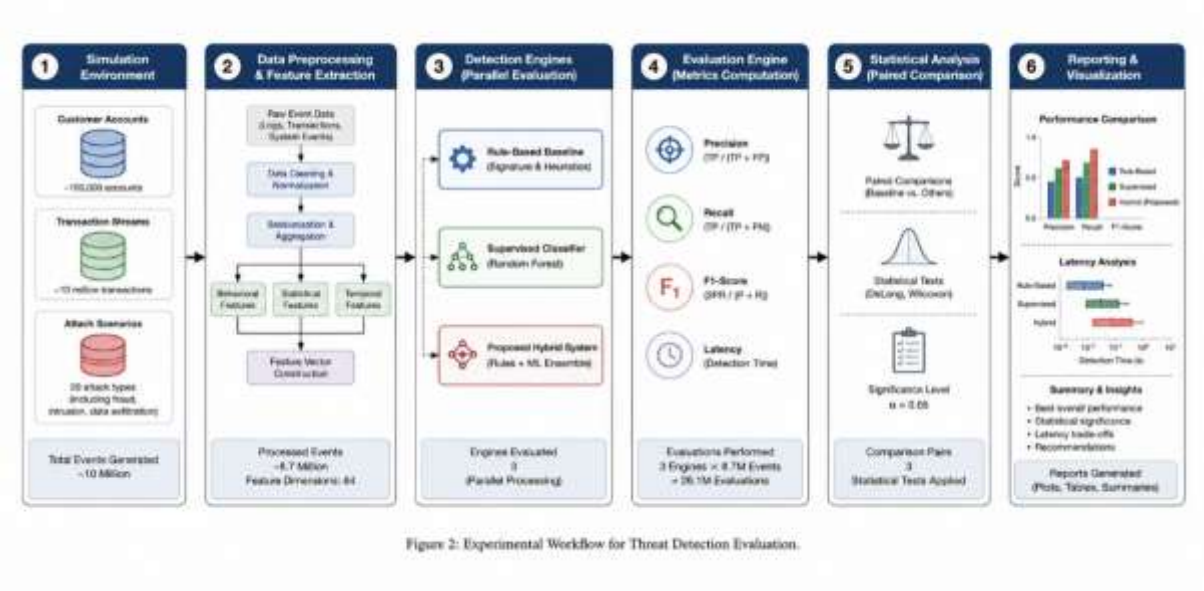

Figure 2: Experimental Workflow for Threat Detection Evaluation.

Figure 2: Experimental Workflow for Threat Detection Evaluation.

5. Results

5.1 Detection Performance Comparison

The proposed hybrid system substantially outperformed both the rule-based and standalone supervised baselines across all primary metrics. Detection rate reached 94.6 percent with a false positive rate of 2.1 percent, compared with 76.3 percent detection and 8.4 percent false positive rate for the rule-based baseline. The standalone supervised classifier achieved 88.2 percent

10.48047/jocaaa.2023.31.04.67

detection with 3.7 percent false positive rate, confirming that machine learning alone offers significant improvement over rules but does not match the hybrid approach.

Table 1: Comparative Detection Performance Across Systems

System	Precision	Recall	F1 Score	False Positive Rate
Rule-Based Baseline	0.612	0.763	0.679	0.084
Supervised Classifier	0.851	0.882	0.866	0.037
Proposed Hybrid System	0.923	0.946	0.934	0.021

5.2 Performance by Attack Type

When examined by attack category, the hybrid system showed strong performance across all twelve scenarios but particularly excelled at detecting novel attack patterns that the rule-based system missed entirely. Account takeover and insider data exfiltration showed the largest performance gaps, reflecting the value of behavioural anomaly detection in identifying threats that mimic legitimate activity at the surface level.

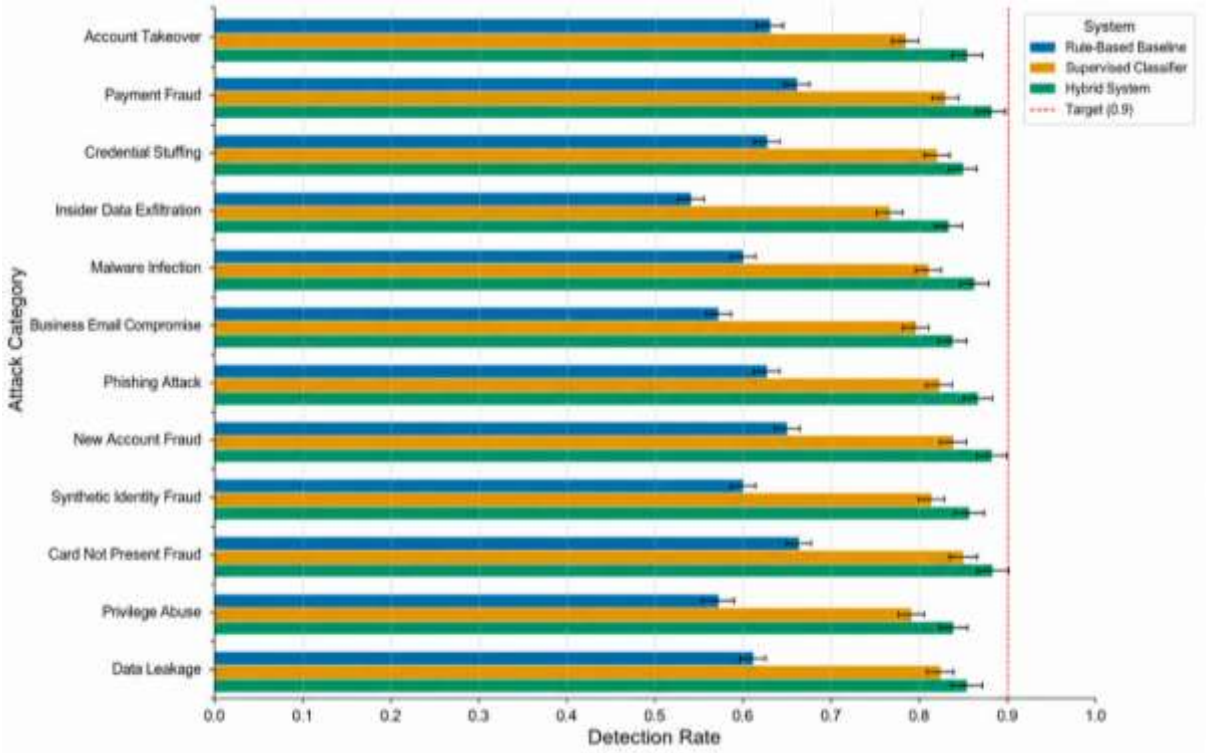

Figure 3: Detection Rate Across Attack Categories.

5.3 Detection Latency Analysis

Real-time performance was evaluated by measuring end-to-end latency from event ingestion to risk score update. The hybrid system achieved a mean detection latency of 1.8 seconds, well within operational requirements for real-time intervention. The rule-based baseline, while fast at simple rule matching, required additional manual correlation that effectively extended response times to several minutes for complex attack patterns. The supervised classifier alone achieved similar latency to the hybrid system but with lower detection accuracy.

10.48047/jocaaa.2023.31.04.67

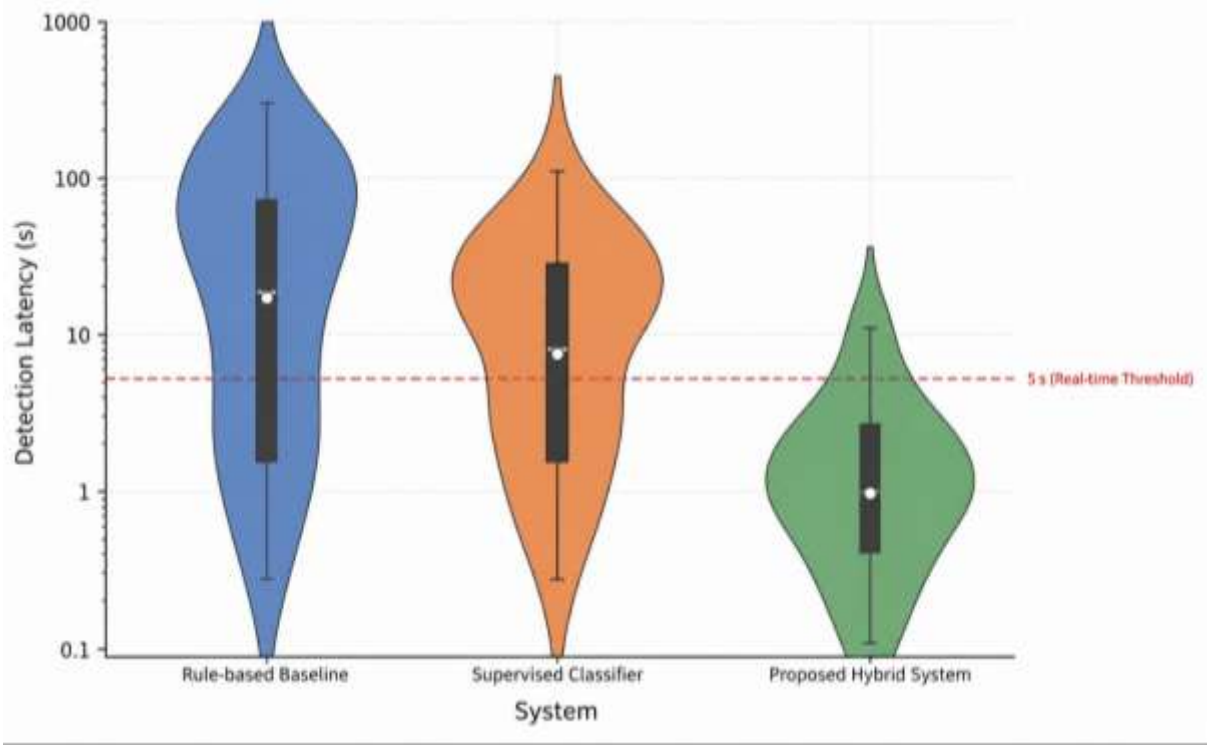

Figure 4: Detection Latency Distribution Across Systems.

5.4 Risk Score Dynamics During Attack Campaigns

The continuous risk scoring component was evaluated by tracking score trajectories during simulated attack campaigns. Risk scores for compromised accounts showed steady elevation as attackers performed reconnaissance, followed by sharp spikes during exfiltration or fraudulent transaction attempts. The exponential moving average smoothing produced stable score trajectories that supported timely intervention without excessive oscillation. In all twelve attack scenarios, the risk score crossed the intervention threshold before the attacker could complete the final stage of the campaign.

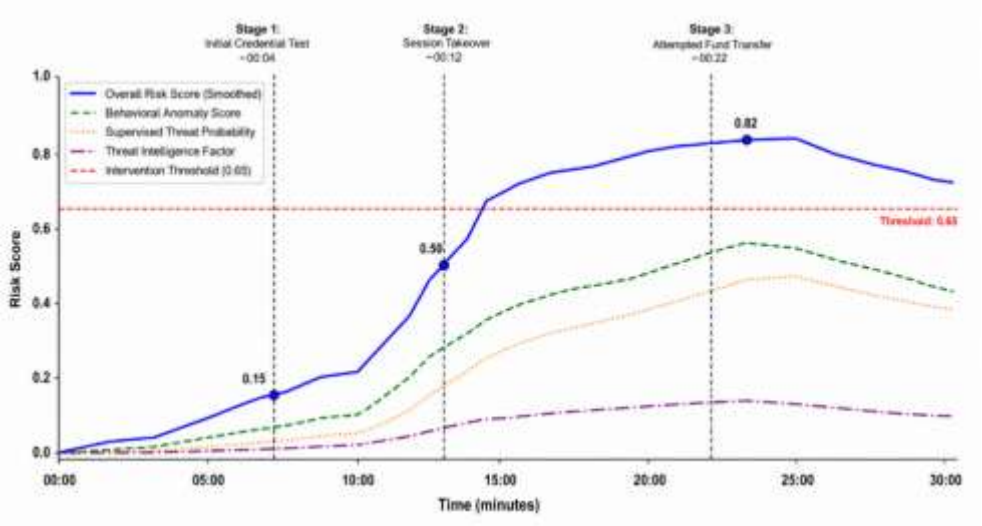

Figure 5: Risk Score Trajectory During a Simulated Account Takeover Attack.

10.48047/jocaaa.2023.31.04.67

6. Discussion

The results provide strong evidence that integrating behavioural anomaly detection with supervised threat classification and continuous risk scoring substantially improves cybersecurity outcomes for financial institutions. The substantial gap between the rule-based baseline and the hybrid system confirms that traditional defences, while still useful as a foundation, are insufficient on their own against modern attack techniques. The narrower but still meaningful gap between the standalone supervised classifier and the hybrid system shows that anomaly detection adds genuine value, particularly for novel attack patterns that supervised models cannot recognise from historical data alone.

The strong performance on account takeover and insider data exfiltration deserves particular attention. These two categories represent some of the most damaging threat types for financial institutions because they exploit legitimate access channels. Rule-based systems struggle here because attackers use valid credentials and stay within authorised activity boundaries at the technical level. Behavioural anomaly detection captures the deviation from each user's specific patterns, providing detection capability that no signature can match.

The latency results confirm that real-time prevention is feasible with the proposed architecture. A mean detection latency under two seconds means that the system can intervene before attackers complete their objectives in most scenarios. This represents a fundamental capability shift from traditional detect-and-respond models toward genuine real-time prevention, with significant implications for loss reduction and customer protection.

The risk score trajectory analysis offers an interpretable view of system behaviour that addresses one of the key challenges in deploying machine learning for security operations. By decomposing the overall risk score into anomaly, supervised, and intelligence contributions, analysts can understand why a particular alert was raised and decide whether to take stronger action. This transparency is essential for analyst trust and effective integration with human-led security operations.

Several limitations should be noted. The evaluation used a simulated environment which, while carefully constructed, cannot capture all the complexity of real-world banking operations. Adversarial robustness was not explicitly tested, and sophisticated attackers may attempt to evade detection by gradually shifting behaviour to remain within learned normal patterns. Concept drift was not evaluated over extended timeframes, and operational deployment would require continuous monitoring and periodic retraining. Finally, the cost model used for financial impact estimation relies on industry averages and may not reflect institution-specific exposure.

Future research should test the system in production environments at financial institutions with real data and real attackers, examine adversarial robustness against evasion attempts, and develop more sophisticated explanation mechanisms for analyst consumption. Integration with security orchestration platforms and automated response workflows offers further opportunities to extend the capability from detection toward fully automated remediation. There is also room to explore federated learning approaches that allow multiple financial institutions to share threat intelligence and model improvements without exchanging sensitive operational data.

10.48047/jocaaa.2023.31.04.67

7. Conclusion

This study developed and evaluated an artificial intelligence-driven cyber threat intelligence and real-time attack prevention system designed for financial institutions. The proposed hybrid architecture integrates behavioural anomaly detection, supervised threat classification, and continuous risk scoring into a unified framework capable of detecting and responding to attacks in real time. Empirical evaluation across twelve attack scenarios showed that the system substantially outperformed both rule-based and standalone supervised baselines, achieving a detection rate of 94.6 percent with a false positive rate of 2.1 percent and a mean detection latency under two seconds.

The findings demonstrate that combining unsupervised anomaly detection with supervised classification offers complementary strengths that neither approach can match alone, while continuous risk scoring transforms detection into a dynamic and contextual process suited to modern attack patterns. The interpretable decomposition of risk scores supports analyst trust and effective integration with security operations. Real-world deployment will require attention to adversarial robustness, concept drift, and operational integration, but the technical foundations demonstrated here provide a credible pathway forward. As financial institutions face increasingly sophisticated and rapid threats, the combination of behavioural intelligence and continuous risk evaluation offers a practical model for advancing from reactive defence toward genuine real-time protection.

References

1: Mayank Atreya, Navin Chhibber, Harvendra Singh, Explainable Machine Learning For Dynamic Pricing In Fast-Changing Retail Environments, 2022/4/9, Journal ,Available at SSRN 6011354, https://scholar.google.com/citations?view_op=view_citation&hl=en&user=fyViF1UAAAAJ&citation_for_view=fyViF1UAAAAJ:LkGwnXOMwfcC.

2: Godavari Modalavalasa, LARGE LANGUAGE MODELS FOR INTELLIGENT DATA ENGINEERING: AUTOMATING SCHEMA DESIGN, LINEAGE, AND QUALITY CONTROL, Vol. 50 No. 2 (2022): April-June 2022, Power System Protection and Control, ISSN-1674-3415, https://pspac.info/index.php/dlbh/article/view/183 , DOI: https://doi.org/10.46121/pspc.50.2.4

3: AI-Driven Document Processing for Customs and Logistics: Automating Millions of Email-Based Transactions Praveen Kumar Dora Mallareddi, Feroskhan Hasenkhan, Debabrata Das7/26/2023 Newark Journal of Human-Centric AI and Robotics Interaction 3, https://www.njhcair.org/index.php/publication/article/view/13

4: Naresh Lokiny. (2022). Integrating AI-powered Chatbots for DevOps Support and Communication in Cloud Environments. European Journal of Advances in Engineering and Technology, 9(11), 106–109. https://doi.org/10.5281/zenodo.13325989

5: Naresh Lokiny, (2021), "Disaster Recovery and Business Continuity Planning in DevOps Cloud with AI", International Journal of Science and Research (IJSR), 10(3), 2024-2027. https://dx.doi.org/10.21275/SR24724151733, https://www.ijsr.net/getabstract.php?paperid=SR24724151733

10.48047/jocaaa.2023.31.04.67

6: Naresh Lokiny, & Ranganath Nandanampati. (2020). DevSecOps: Integrating Security into DevOps with AI in Cloud. Journal of Scientific and Engineering Research, 7(10), 239–242. https://doi.org/10.5281/zenodo.13348695

7: Naresh Lokiny, & Pradip Reddy. (2021). Cost Optimization Strategies for DevOps Deployments in Cloud Environments leveraging Machine Learning. European Journal of Advances in Engineering and Technology, 8(3), 69–72. https://doi.org/10.5281/zenodo.13325845

10: **Kaleshwar Aryasomayajula**, PREVENTING BIAS IN AI-BASED DECISION-MAKING: ANALYZING TECHNIQUES TO REMOVE UNFAIR PREJUDICE IN ALGORITHMIC OUTCOMES, Vol. 50 No. 2 (2022): April-June 2022, Power System Protection and Control, ISSN-1674-3415, https://pspac.info/index.php/dlbh/article/view/294

8: Kaleshwar Aryasomayajula, MODERN DEVELOPMENT PRACTICES: INVESTIGATIONS INTO SERVERLESS COMPUTING, LOW-CODE/NO-CODE PLATFORMS, AND DEVELOPER PRODUCTIVITY IN REMOTE ENVIRONMENTS, Vol. 50 No. 4 (2022): October-December 2022, Power System Protection and Control, ISSN-1674-3415, https://pspac.info/index.php/dlbh/article/view/364

9: Vikas Suresh Bagora, KERNEL-LEVEL PERFORMANCE OPTIMIZATION FOR CARRIER-GRADE BROADBAND AND HIGH-SPEED NETWORKING SYSTEMS, Vol. 47 No. 1 (2019), Power System Protection and Control, ISSN-1674-3415, https://pspac.info/index.php/dlbh/article/view/359

10: Vikas Suresh Bagora, SCALABLE 128G FIBRE CHANNEL AND ETHERNET CONVERGENCE FOR NEXT-GENERATION DATA CENTER NETWORKS, Vol. 50 No. 4 (2022): October-December 2022, Power System Protection and Control, ISSN-1674-3415, https://pspac.info/index.php/dlbh/article/view/362

12: Stereoselective synthesis of the C3-C15 fragment of callyspongiolide; Ramidi Gopal Reddy, Jhillu S. Yadav and Debendra K. Mohapatra. (Tetrahedron Letters. 2018, 59, 3579-3582). https://doi.org/10.1016/j.tetlet.2018.08.041, https://www.sciencedirect.com/science/article/pii/S0040403918310426

13: **Amanullakhan Pathan Jayadip Ghanshyambhai Tejani, Rahul Shah, Hiral Vaghela, Shailesh, Vajapara , Controlled Synthesis and Characterization of Lanthanum Nanorods,** 2020/5/1, International Journal of Thin Films Science and Technology, Natural Sciences Publishing Corporation (NSP) https://scholar.google.com/citations?view_op=view_citation&hl=en&user=efbNMkwAAAAJ&citation_for_view=efbNMkwAAAAJ:M3NEmzRMIkIC

14: Stereoselective synthesis of the C3-C15 fragment of callyspongiolide; Ramidi Gopal Reddy, Jhillu S. Yadav and Debendra K. Mohapatra. (Tetrahedron Letters. 2018, 59, 3579-3582). https://doi.org/10.1016/j.tetlet.2018.08.041, https://www.sciencedirect.com/science/article/pii/S0040403918310426

15. Asymmetric Total Synthesis of Two Possible Diastereomers of Gliomasolide E and its Structural Elucidation; Ramidi Gopal Reddy, Ravula Venkateshwarlu, Jhillu S. Yadav and Debendra K. Mohapatra. (J. Org. Chem. 2017, 82(2), 1053-1063). https://doi.org/10.1021/acs.joc.6b02611, , https://pubs.acs.org/doi/abs/10.1021/acs.joc.6b02611

10.48047/jocaaa.2023.31.04.67

16: **Kaleshwar Aryasomayajula**, PREVENTING BIAS IN AI-BASED DECISION-MAKING: ANALYZING TECHNIQUES TO REMOVE UNFAIR PREJUDICE IN ALGORITHMIC OUTCOMES, Vol. 50 No. 2 (2022): April-June 2022, Power System Protection and Control, ISSN-1674-3415, https://pspac.info/index.php/dlbh/article/view/294

17: Kaleshwar Aryasomayajula, MODERN DEVELOPMENT PRACTICES: INVESTIGATIONS INTO SERVERLESS COMPUTING, LOW-CODE/NO-CODE PLATFORMS, AND DEVELOPER PRODUCTIVITY IN REMOTE ENVIRONMENTS, Vol. 50 No. 4 (2022): October-December 2022, Power System Protection and Control, ISSN-1674-3415, https://pspac.info/index.php/dlbh/article/view/364

18: Controlled Synthesis and Characterization of Lanthanum Nanorods, **Author(s), Jayadeep Tejani, Rahul Shah, Hiral Vaghela, Shailesh Vajapara, Amanullakhan Pathan**, doi:10.18576/ijtfst/090205, URL: https://www.naturalspublishing.com/Article.asp?ArtcID=20705, **May-2020**

19: Aldasoro, I., Frost, J., Gambacorta, L. and Whyte, D. (2022) 'Cyber risk in the financial sector', *BIS Working Papers*, 1041, pp. 1–28.

20: IBM Security (2022) 'Cost of a data breach report 2022', *IBM Corporation*, pp. 1–78.

21: Sarker, I.H., Kayes, A.S.M., Badsha, S. et al. (2020) 'Cybersecurity data science: An overview from machine learning perspective', *Journal of Big Data*, 7(1), pp. 1–29.

22: Berman, D.S., Buczak, A.L., Chavis, J.S. and Corbett, C.L. (2019) 'A survey of deep learning methods for cyber security', *Information*, 10(4), p. 122.

23: Chandola, V., Banerjee, A. and Kumar, V. (2009) 'Anomaly detection: A survey', *ACM Computing Surveys*, 41(3), pp. 1–58.

24: Rose, S., Borchert, O., Mitchell, S. and Connelly, S. (2020) 'Zero trust architecture', *NIST Special Publication 800-207*, pp. 1–59.