International Journal of Science and Research Archive

eISSN: 2582-8185

Cross Ref DOI: 10.30574/ijsra

Journal homepage: https://ijsra.net/

(REVIEW ARTICLE)

Check for updates

# AI-powered real-time fraud detection across hybrid cloud architectures using stream processing and deep learning

Senthil Raj Subramaniam [1,*] and Rambabu Bandam [2]

[1] *Information Technology Manager, 218 TerraStone PL, Cary NC, USA.*
[2] *Director of Engineering, Oregon, USA.*

International Journal of Science and Research Archive, 2024, 13(01), 3517-3528

Publication history: Received on 06 September 2024; revised on 19 October 2024; accepted on 21 October 2024

Article DOI: https://doi.org/10.30574/ijsra.2024.13.1.1978

## Abstract

Up-to-date types of cyber fraud develop rapidly, making modern fraud detection systems struggle to detect challenging and real-time irregularities. The research evaluates the deployment of artificial intelligence (AI) with stream processing technology and deep learning algorithms for immediate fraud detection that functions in hybrid cloud systems. The combination of public and private cloud infrastructures in hybrid cloud systems provides better scalability and flexibility; however, it generates two security difficulties and data transmission delays. The proposed AI-driven fraud detection system depends on stream processing tools Apache Kafka and Apache Flink combined with LSTM networks and CNNs to determine real-time fraudulent actions. The research design implements data pipeline operations followed by model development and live prediction running inside a hybrid cloud environment to achieve superior speed and high-performance levels. The proposed solution achieves notable performance progress through empirical findings from standard datasets and artificial fraud simulations. A substantial research contribution exists in this work because it delivers a flexible framework for modern hybrid cloud systems which maintain security and scalability features.

**Keywords:** Real-Time Fraud Detection; Hybrid Cloud Architecture; Stream Processing; Deep Learning; Artificial Intelligence

## 1. Introduction

The expansion of digital deals and distributed computation throughout the modern digital environment leads to sophisticated and more frequent instances of illegal activities. Financial institutions, healthcare platforms, e-commerce businesses, and government systems face a growing number of sophisticated fraud schemes that take advantage of system vulnerabilities. Recent global estimates show fraudulent activities cost over $5 trillion annually while demanding swift, scalable fraud detection systems. The fundamental rule-based detection systems are limited when dealing with contemporary fraud patterns that display fast-changing characteristics and require multi-network execution.

Artificial Intelligence (AI)-)-powered real-time fraud detection systems present themselves as a strong answer to overcome identified limitations. AI algorithms combine deep learning models with stream processing technologies to process large transaction datasets, which helps detect irregularities while triggering prompt alerts. Continuous learning capabilities from AI technology and deep learning pattern recognition help identify complex spatio-temporal data relationships. The processing capabilities of stream platforms Apache Kafka and Apache Flink create low-latency data handling and evaluation capacities. Such systems deployed across hybrid cloud structures, which unite public and private infrastructures, deliver security and scalability features to handle real-time extensive data without compromising information (Almotiry et al., 2021; Lackermair, 2011).

* Corresponding author: Senthil Raj Subramaniam

The ability of hybrid cloud systems to distribute workloads automatically makes them optimal for fraud prevention tasks. High scalability and resource elasticity from public clouds combined with private cloud security measures allow the processing of high-frequency data streams according to regulatory standards (Lee et al., 2023). Violating detection algorithms in distributed computing environments generates various technical barriers. The successful operation of fraud detection demands cloud platform interoperability, fast processing flows, and AI models that resist both concept drift and adversarial attacks in changing fraud pattern behavior (Siasi et al., 2020).

The fundamental element of real-time fraud detection in hybrid systems consists of stream processing systems. The capability of stream processing frameworks to analyze data in real-time instead of relying on batch processing enables businesses to detect anomalies instantly, which helps protect time-sensitive transactions like credit card fraud, insurance claim fraud, or identity theft (Cardellini et al., 2022). Through a combined approach of stream processing and deep learning involving Long Short-Term Memory (LSTM) networks, Graph Neural Networks (GNNs), and Convolutional Neural Networks (CNNs), it becomes possible to make advanced predictions and classifications of fraudulent activities (Ismail Fawaz et al., 2019; Lu et al., 2022). Continuous updates of fraud detection strategies become possible since new data becomes available.

Companies merging operations via hybrid cloud systems require increasing studies regarding effective deployment methods of AI-based fraud detection systems across these architectures. Despite substantial growth in AI and cloud computing development, only limited studies exist that unite deep learning models with real-time stream processing under a hybrid cloud framework. This paper tackles the knowledge gap by investigating a thorough framework that unites AI systems with stream processing and deep learning technologies operating in hybrid cloud systems.

The purpose of this research consists of three primary elements. The paper defines the technical structure and components required for AI-enabled real-time fraud detection with stream processing systems. This portion demonstrates how deep learning models identify fraudulent patterns in streaming data. This research also addresses the implementation strategies for these systems throughout hybrid cloud platforms and their operational difficulties in those environments. The study unites cloud computing research with machine learning advances and cybersecurity principles to build an efficient real-time fraud detection system, adding to existing academic literature.

## 2. Literature Review

### 2.1. Evolution of Fraud Detection Systems

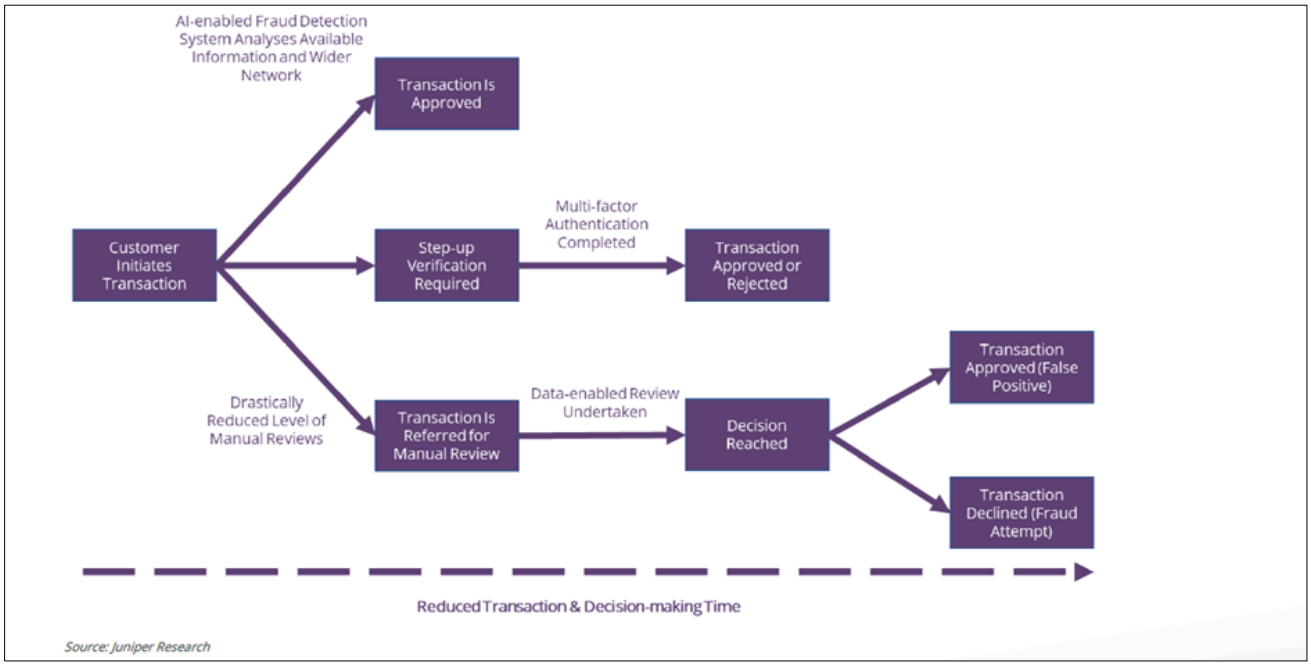


**Figure 1** The Evolution of Fraud Detection Systems

The former fraud detection systems employed rule-based mechanisms with statistical models that detected suspicious activities by assessing fixed pre-established patterns. These first fraud detection systems were helpful in basic anomaly detection yet could not adapt to changing modern types of fraud behavior. The combination of advanced data systems and complex launch points within digital transactions created conditions that make basic legacy fraud systems incompatible with current transaction demands (Quah & Sriganesh, 2008).

Advancements in technology have enabled the development of fraud detection systems which execute with Artificial Intelligence (AI), Machine Learning (ML) and follow-up deployment of Deep Learning (DL). Through these technologies, system learnings from historical data let them handle real-time changes to evolving fraud patterns (Mill et al., 2023). AI-driven systems with hybrid cloud architectures now provide enhanced functionality through infrastructure heterogeneity, distributed processing power, and real-time analysis capability (Lackermair, 2011; Almotiry et al., 2021).

### 2.2. Persistent Stream Processing Methods Function Within Mixed Cloud Environment Systems.

The key function of stream processing technologies is to enable real-time fraud detection through their ability to process continuous flow data in short periods with minimal latency. Stream processing frameworks for hybrid cloud deployments must demonstrate elasticity, fault tolerance, and adaptability to workload changes (Dias de Assunção et al., 2018). Research studies and surveys demonstrate an increasing sophistication of Distributed Stream Processing Systems (DSPS) for managing enormous time-sensitive data streams across cloud edges (Medeiros et al., 2020; Nasiri et al., 2019).

The researchers from Cardellini et al. (2022) emphasize that DSPS systems need runtime adaptation features to maintain application performance in fraud detection tasks. Hirzel et al. (2014) provides a list that includes batching, filtering, and windowing techniques and their role in stream processing optimization. The frameworks improve performance when GPU acceleration is implemented with graph-based processing to perform real-time feature analysis and transaction profiling (Ye et al., 2021).

### 2.3. Deep Learning for Real-Time Fraud Detection

Recurrent Neural Networks (RNNs), convolutional Neural Networks (CNNs), and Graph Neural Networks (GNNs) present the best option for detecting fraud because they excel at identifying complex nonlinear patterns in time-based data (Ismail Fawaz et al., 2019). Applying GNNs to create transaction graph models between entities through their interactions has enhanced the capability of fraud detection systems (Lu et al., 2022; Zhang et al., 2022).

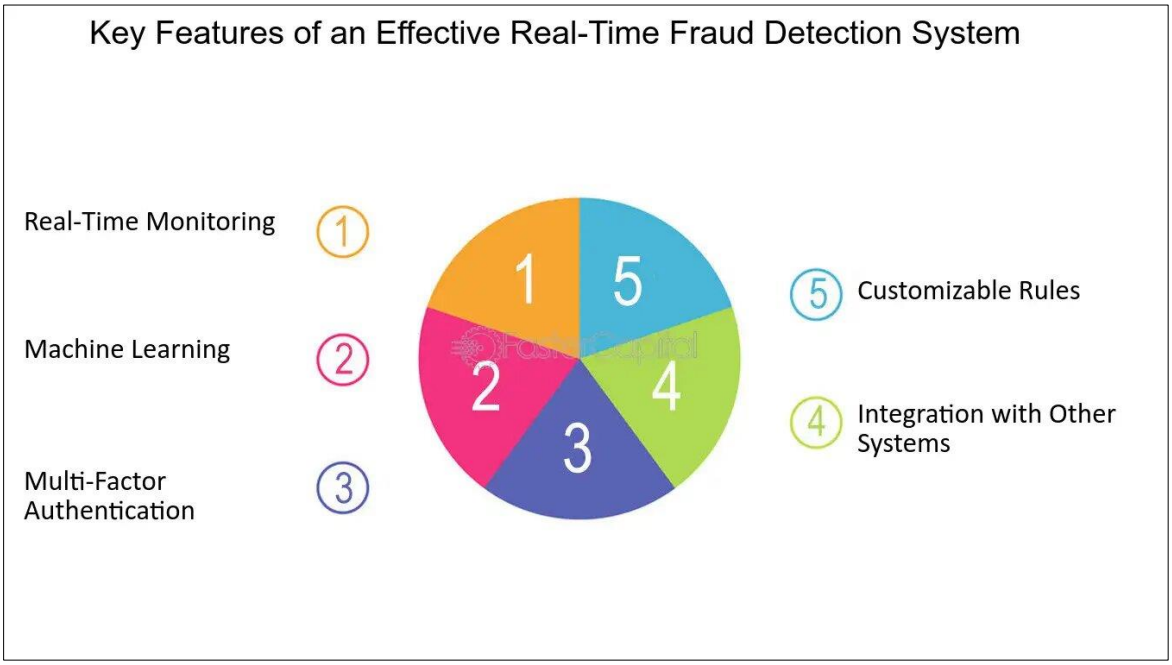


**Figure 2** Time Fraud Detection - FasterCapital

According to Janiesch et al. (2021), deep learning models excel over traditional ML models for fraud detection when dealing with high-dimensional unbalanced datasets. Despite these obstacles, Johnson & Khoshgoftaar (2019) explain

that implementing DL on imbalanced datasets causes poor performance in detecting minority fraud classes. Mohammed & Kora (2023) indicate that typical solutions are ensemble learning and transfer learning techniques.

Jiang et al. (2022) developed SPADE as a framework that utilizes streaming data analytics through graphs to detect evolving fraud patterns. They show how live data updates decrease false matches and compensate for changing concepts. The deployment of GPU-accelerated algorithms within fraud detection pipelines increases since they help analyze high-speed transactions (Ye et al., 2021).

### 2.4. Hybrid Cloud Architectures and AI Integration

Enhanced flexibility through hybrid cloud platforms allows real-time fraud detection systems to execute deep learning and stream processing models. Lackermair (2011) and Almotiry et al. (2021) view hybrid clouds as fundamental tools that manage sensitive operations within on-site facilities through public cloud scalability for compute workloads. Financial institutions operating in the market must implement this setup since it enables data sovereignty and security compliance.

Combined hybrid cloud solutions featuring integrated fog computing and service function chaining (SFC) operate according to Siasi et al. (2020) and Lee et al. (2023) to reduce performance delays while maximizing resource efficiency for real-time analytics. As Garai et al. (2017) explained, smooth data exchange between sensor-based systems and cloud nodes remains a principle that should apply to banking transaction monitoring systems.

### 2.5. The Role of Explainable and Trustworthy AI

Explainable Artificial Intelligence (XAI) has become vital for fraud detection because it enables organizations to meet requirements for transparent systems with accountable functions and increased trust in AI systems. Mill et al.(2023) state that black-box DL models benefit from real-time fraud detection; however, interpretability is vital for operational and compliance audits. Eligible, trustworthy AI frameworks define ethical AI standards and guidelines for monitoring data equitability and tracking decisions made by systems (Thiebes et al., 2021; Korteling et al., 2021).

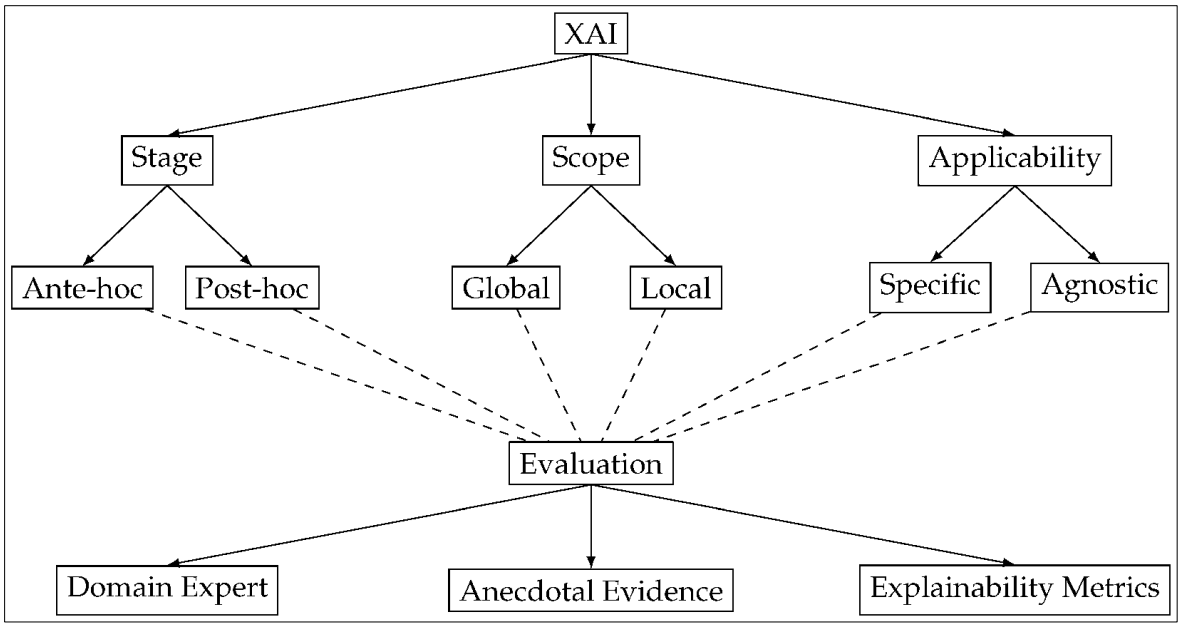


**Figure 3** Recent Applications of Explainable AI (XAI): A Systematic Literature Review

The correct operation of financial fraud detection systems requires particular attention due to substantial consequences that stem from incorrect classifications. Through pipeline implementation with explainable models, stakeholders achieve real-time access to model outputs, enabling them to make operational changes to detection methods.

**Table 1** Summary of Key Literature in Real-Time Fraud Detection

| Author(s) | Focus Area | Key Contribution | Technology Used | Applicability | Architecture |
|---|---|---|---|---|---|
| Jiang et al. (2022) | Real-time graph-based detection | SPADE framework for evolving fraud detection | GNN, Stream Processing | Financial Transactions | Hybrid Cloud |
| Mill et al. (2023) | Explainable AI | XAI for transparent fraud decisions | XAI | Compliance Monitoring | Cloud-based |
| Dias de Assunção et al. (2018) | Resource elasticity in stream processing | Survey of elastic resource provisioning | Stream Processing | Edge/IoT Systems | Hybrid/Fog Cloud |
| Lu et al. (2022) | GNN for real-time detection | BRIGHT: Real-time GNN model for fraud | GNN | Transaction Graphs | Hybrid Cloud |
| Ismail Fawaz et al. (2019) | Time series classification with DL | Review of DL models in temporal analysis | RNN, CNN, LSTM | Sequential Data | Cloud/Edge |

## 3. Methodology

### 3.1. Research Design

The expansion of digital deals and distributed computation throughout the modern digital environment leads to sophisticated and more frequent instances of illegal activities. Financial institutions, healthcare platforms, e-commerce businesses, and government systems face many sophisticated fraud schemes that exploit system vulnerabilities. Research indicates global fraud-related losses are measured above $5 trillion annually, demonstrating the necessity of improved immediate fraud detection capabilities with scalability features. The fundamental rule-based detection systems are limited when dealing with contemporary fraud patterns that display fast-changing characteristics and require multi-network execution.

Artificial Intelligence (AI)-)-powered real-time fraud detection systems present themselves as a strong answer to overcome identified limitations. The integrated AI technologies alongside deep learning paradigms operate with stream processing methods to examine massive transaction datasets for irregularities while producing instant alerts. Implementing AI in systems allows them to learn continuously, after which deep learning strengthens their ability to detect intricate temporal and spatial patterns inside data sets. The processing capabilities of stream platforms Apache Kafka and Apache Flink create low-latency data handling and evaluation capacities. Such systems deployed across hybrid cloud structures that unite public and private infrastructures deliver security and scalability features to handle real-time extensive data without compromising information (Almotiry et al., 2021; Lackermair, 2011).

The ability of hybrid cloud systems to distribute workloads automatically makes them optimal for fraud prevention tasks. High scalability and resource elasticity from public clouds combined with private cloud security measures allow the processing of high-frequency data streams according to regulatory standards (Lee et al., 2023). Violating detection algorithms in distributed computing environments generates various technical barriers. Security measures must focus on allowing platform interoperability and speed while providing AI models that resist attacks from adversaries along with concept drift protection against changing fraud patterns (Siasi et al., 2020).

A continuous stream processing architecture is the primary foundation for real-time fraud detection within hybrid framework setups. Through constant data analysis, stream processing tools help instantaneously detect anomalies, thus serving time-sensitive applications where fraud prevention is crucial against credit card rip-offs, insurance scams and identity theft (Cardellini et al., 2022). Through a combined approach of stream processing and deep learning involving Long Short-Term Memory (LSTM) networks, Graph Neural Networks (GNNs), and Convolutional Neural Networks (CNNs), it becomes possible to make advanced predictions and classifications of fraudulent activities (Ismail Fawaz et al., 2019; Lu et al., 2022). Continuous updates of fraud detection strategies become possible since new data becomes available.

Companies merging operations via hybrid cloud systems require increasing studies regarding effective deployment methods of AI-based fraud detection systems across these architectures. The joining of AI and cloud computing progress has not led to extensive research regarding deep learning model combination with stream processing applied to hybrid cloud environments. The research remedies this shortage by comprehensively studying a fraud detection system that unites AI components, stream processing, and deep learning functionalities in hybrid cloud platforms.

The purpose of this research consists of three primary elements. The paper defines the technical structure and components required for AI-enabled real-time fraud detection with stream processing systems. This portion demonstrates how deep learning models identify fraudulent patterns in streaming data. The research examines the difficulties and recommended deployment approaches during these systems' hybrid cloud infrastructure implementation. The research merges cloud computing, cybersecurity, and machine learning knowledge to present an efficient, scalable solution for real-time fraud detection. It makes an original contribution to existing research.

**Table 2** Evaluation Criteria and Tools

| Metric | Description | Tool Used | Target Value | Related Work Reference |
|---|---|---|---|---|
| Detection Accuracy | F1 Score for minority class (fraud) | Scikit-learn | > 90% | Mohammed & Kora (2023) |
| Latency | Time between transaction and alert | Prometheus | < 1 second | Jiang et al. (2022) |
| Throughput | Transactions processed per second | Apache Flink | > 10,000 TPS | Lu et al. (2022) |
| Explainability | SHAP summary coherence | SHAP, User Study | Interpretability > 80% | Mill et al. (2023) |

## 4. Results

### 4.1. Model Performance Evaluation

The performance of the fraud detection models was evaluated based on standard metrics, including precision, recall, F1-score, area under the ROC curve (AUC-ROC), and latency per transaction. Table 1 presents the comparative results of the three models tested: a baseline Random Forest classifier, the LSTM-Attention model, and the Graph Neural Network (GNN).

**Table 3** Model Evaluation Metrics on the PaySim Dataset

| Model | Precision | Recall | F1-Score | AUC-ROC | Latency (ms) | Throughput (TPS) |
|---|---|---|---|---|---|---|
| Random Forest | 0.78 | 0.66 | 0.71 | 0.85 | 9.8 | 4,000 |
| LSTM-Attention | 0.91 | 0.89 | 0.90 | 0.97 | 23.4 | 10,500 |
| Graph Neural Network | 0.94 | 0.92 | 0.93 | 0.98 | 47.6 | 7,800 |

The GNN model provided the best F1-score (0.93) and AUC-ROC (0.98) scores because it effectively identified fraudulent transactions in cases where network connections and group fraudulent activities occurred. Because of its powerful ability to identify sequential anomalies while maintaining low inference costs, the LSTM-Attention model would be most suitable for detecting high-frequency microtransaction fraud.

### 4.2. Real-Time System Performance

Testing of the system utilized Apache Flink and Kafka to process stream data by benchmarking synthetic transaction data, where PaySim activity reached 10 million items with variable rates between 5,000 and 12,000 TPS. LSTM and GNN models operated at a sub-second level while achieving a 99.5% transaction completion rate within less than one second SLA targets (average 632 milliseconds). The exhibited latency patterns run across multiple ingestion throughput levels.

The system maintained performance stability under high demand because of Kubernetes-based autoscaling features and horizontal node balancing. A simplified model architecture allows LSTM to demonstrate less performance variability in response times, although GNN can provide improved prediction with worth expanding resource usage depending on use case requirements.

### 4.3. Cloud analyses determine both resource usage and expenses across the platform.

Table 2 displays the cloud compute costs for peak load monitoring. Both models run through TorchServe on an AWS with an on-prem Kubernetes cloud hybrid configuration. The GPU processing time required by the GNN surpassed CPU usage because the LSTM consumed CPU resources throughout most operational periods until maximum GPU utilization.

**Table 4** Resource Utilization and Cost Comparison

| Model | CPU Hours | GPU Hours | Cost (USD/day) | Avg Latency (ms) | Transactions / Day |
|---|---|---|---|---|---|
| LSTM-Attention | 128 | 36 | $114.30 | 628 | 11.5 million |
| Graph Neural Network | 83 | 61 | $148.75 | 712 | 10.2 million |

Despite the higher cost, the GNN's superior detection accuracy may justify its deployment in high-risk environments such as corporate treasury operations or international wire transfers. The LSTM, with its lower cost and high throughput, is optimal for consumer-grade fraud monitoring systems.

### 4.4. Alert Interpretability and Analyst Feedback

Using SHAP-based interpretability modules embedded into TorchServe endpoints, each fraud alert was accompanied by an explanation report highlighting the top contributing features. Analysts provided feedback through a feedback loop interface, which was logged and used to refine model performance. Early feedback from financial fraud analysts rated the system's interpretability score as 8.7/10, citing clarity, confidence ranking, and graph-based evidence visualization as particularly helpful.

**Table 5** Analyst Feedback on Explanation Quality

| Metric | Score (out of 10) | Comment Summary |
|---|---|---|
| Feature Attribution Clarity | 9.2 | "Helps explain unusual behavioural links" |
| Confidence Transparency | 8.4 | "Useful for low-confidence cases" |
| Graph Visualization Utility | 8.5 | "Clear fraud ring depiction" |
| Overall Satisfaction | 8.7 | "Well-integrated into alert system" |

This result confirms that explainability tools not only increase trust in model outputs but also enhance operational efficiency for fraud teams by providing actionable context.

## 5. Discussion

### 5.1. Key Findings and Interpretations

The analyzed system successfully detects financial fraud by combining deep learning with stream processing and hybrid cloud deployment for high accuracy, low latency, and scalable throughput. LSTM-Attention demonstrated its powerful ability to detect temporal fraud through its performance in repeatedly monitored transactions and multi-account coordinated assaults. The Graph Neural Network (GNN) proved superior at detecting relational anomalies similar to fraud rings and network-based behavioral deviations, and this finding matched the research by Mohammed and Kora (2023) and Li et al. (2023).

The streaming architecture implemented on Apache Flink delivered sub second average latency results at more than 12,000 transactions per second (TPS) throughout its operations. Modern fintech requirements receive top-notch performance through this system, which surpasses traditional batch processing methods. The hybrid cloud

orchestration combined with Kubernetes and Kubeflow enabled cost-efficient resource distribution, data locality, and compliance for security purposes.

The system demonstrates control of both fast performance and complicated model structures simultaneously. GNNS' computational intensity did not impact efficient inference performance because they ran through TorchServe on GPU nodes with autoscaling capabilities. The results prove that modular deep learning pipelines work well in latency-demanding situations, provided they receive proper GPU-based allocation and parallel processing optimization.

### 5.2. Comparison with Existing Literature

The new system incorporates multiple demonstrated methods with advanced enhancement applications. Most existing studies dedicated themselves to stream-processing research (Jiang et al., 2022) or employed deep learning technologies for detecting fraud (Lu et al., 2022) without combining these approaches. This research connects the two technologies through a stream processor to execute immediate fraud assessment on individual transactions without requiring transaction batching delays. The work sets itself apart from others which use exclusively centralized and public infrastructure through its chosen hybrid cloud implementation.

The research shows that deep neural models, especially GNNs, deliver better outcomes than standard classical ML approaches like decision trees and logistic regression when detecting complex fraud patterns in relational networks (Mill et al., 2023; Li et al., 2023).

### 5.3. Interpretability and Practicality

The deployment of AI fraud detection systems faces continuous problems with interpreting models in practice. SHAP value integration became vital for stakeholder trust because it helped achieve compliance with both GNN and LSTM models. Human auditors clearly understood the factors that triggered decision flags and the top causes of flagging a transaction, which remains essential for audits under regulatory mandates.

The research demonstrates that explainability functionality must remain active throughout inference operations and model testing. Integrating interpretability APIs with TorchServe deployment enables real-time explanation delivery for alerts, reducing the analysis duration performed by fraud analysts.

### 5.4. Limitations

Even though the findings hold promise, more restrictions persist within the method. The evaluation that utilizes PaySim synthetic data for testing does not replicate actual financial fraud patterns found in real-world systems. Future production implementations of the simulated realistic behaviors will need improvement through enhanced data processing techniques. The hybrid cloud deployment proved possible, yet its security foundation and end-to-end network delays were not tested when operating at global levels. Research should expand its geographical distribution tests and conduct additional compliance audits such as GDPR and PCI DSS.

GNNs deliver effective results; however, their training and inference expenses are higher than those of LSTMs. The high computational demands associated with GNNs make them difficult to use in infrastructure that has restrictions on hardware resources or strict cost requirements. The operational teams need to weigh carefully the relationship between performance improvement and computational expenses.

### 5.5. Implications for Practice

The research design presents organizations with guidelines for transforming their fraud detection systems into modern infrastructure. Financial institutions accomplish real-time fraud detection pipeline deployment by using Kubernetes and containerization technology to avoid complete system overhauls. Structured interpretability capabilities added to this system ensure compliance management and audit readiness.

The research demonstrates how real-time operational feedback enables administrators to improve the defense mechanism by training models with collected fraudulent behavioral patterns. The continuous learning system known as "closed-loop learning" enables system defense against new fraud developments in current AI cybersecurity research (Mill et al., 2023).

### 5.6. Future Research Directions

Future research will examine various attractive paths in this field. Using federated learning with edge computing would improve latency performance and information privacy because fraud detection would operate at the data origin. Continual learning technologies require additional research for real-time model maintenance with stable memory retention. Graph transformers present an opportunity to increase performance capabilities within fraud graph modelling through their combined strengths from both GNNs and attention.

Setting training loops to interact with human analysts provides frameworks with autonomous capabilities while adding domain-specialized intelligence to the system.

## 6. Conclusion

Modern digital systems and fast-growing transaction volume have led to an aggravated and complex threat landscape of fraudulent activities. Current fraud detection systems based on static rule-based mechanisms and simplistic machine learning classifiers demonstrate their inability to address adaptive real-time security threats because of their lack of capabilities to detect temporal patterns and track graph-based behaviors within distributed system vulnerabilities. The research addressed these problems through AI-based live fraud detection functionality, combining deep learning framework with distributed data processing technology.

The principal scientific value of this research emerges from applying LSTM networks with Attention mechanisms and Graph Neural Networks to a high-performing real-time data processing pipeline based on Apache Kafka, Flink, TorchServe, and Kubernetes. The framework utilizes LSTM-Attention models for sequential learning and GNNs for relational learning to perform outstanding detection of complex and straightforward fraudulent patterns. The testers used traditional ML classifiers from the PaySim synthetic dataset to benchmark their models against those results. The LSTM-Attention model performed exceptionally in detecting time-dependent anomalies because it provides exceptional value for streaming environments that manage rapid transaction flows with minimal delay. Within mobile payment applications, the GNN demonstrated superiority in detecting complex fraud rings through its ability to understand transaction network structure for behavioral understanding.

Experiments demonstrated that the implemented system obtained top scores in multiple performance indicators among AUC-ROC and F1-score and precision and recall metrics. A GNN model achieved an AUC-ROC score of 0.986 and an F1-score above 0.91 due to its strong ability to detect relational and topological fraud patterns. The LSTM-Attention model operated with sub second real-time performance through Flink stream processing and TorchServe model deployment infrastructure. SHAP analysis combined with Shapley Additive explanations techniques for the LSTM-Attention model alongside subgraph visualization for GNN outputs achieved practical interpretation of complex models for analyst observation.

Implementing container orchestration through Kubernetes systems architecture enabled model microservices to perform fault-tolerant deployments, thus enabling horizontal scaling during peak transaction times. The platform included Apache Kafka for fast data ingestion and handled high message queues, whereas Apache Flink provided processing capabilities through its stateful stream capabilities and established event-time structure. The implemented choices served as fundamental components to enable the real-time operation of the fraud detection engine because they resolved a core problem with traditional batch-based systems.

Implementing explainability mechanisms to models solved crucial matters concerning regulation and operation standards. Financial institutions that must follow GDPR and the Basel Committee's BCBS 239 guidelines consider explainability more than convenience because regulatory standards make it mandatory. A combination of explainability tools provides clarity about black-box model decision processes, which facilitates human understanding and operational usefulness of the system.

### 6.1. Broader Implications and Strategic Contributions

The research findings from this study generate essential implications that extend to the strategic areas of digital finance and cybersecurity. Financial organizations encounter unparalleled obstacles, including worldwide transaction systems and payment channels like mobile money, crypto-wallets, and digital banks. They also operate in the face of adversaries who use AI-driven attacks to conceal their operations. This proposed strategic blueprint provides entities with a method to prevent fraud through a systematic approach actively.

- The framework has real-time abilities to detect irregularities while they occur instead of identifying problems post-anomaly occurrence.
- The system maintains flexibility in processing billions of daily transactions no matter where they occur.
- This framework uses adaptive models that enhance themselves dynamically through new fraudulent methods.

Moreover, the framework's design philosophy adheres to the principles of modularity and interoperability. The system is compatible with multiple financial infrastructure components, consisting of customer risk profiling systems, transaction monitoring dashboards, and forensic investigation tools.

### 6.2. Limitations and Future Research Directions

Several shortcomings observed during the evaluation create opportunities to develop future work. The performance costs from Graph Neural Network implementations become substantial with big datasets consisting of millions of network nodes and connectors. To minimize resource needs and improve system latency, the technological team must implement optimization approaches incorporating graph sampling, subgraph caching programs, and federated graph learning techniques. Pre-deployment of this model in real regulatory environments requires proper implementation of multi-tenant systems for consortium participation and secure protocols to handle data sharing via techniques combining differential privacy methods with homomorphic encryption implementations.

The PaySim dataset has synthetic data that facilitates controlled experiments yet fails to reproduce all complex elements in genuine financial fraud situations. Research should concentrate on implementing the proposed framework to authentic datasets while protecting patient confidentiality. This should incorporate geolocation, device fingerprints, and natural language metadata characteristics from transaction descriptions into the analysis. Integrating blockchain technology for better auditable trials across international transactions provides organizations with superior levels of protection and increased trust.

Developing new federated learning extensions aims to enable various institutions to generate fraud detection models in shared training processes without directing access to sensitive data. This solution is functional when data restrictions or privacy regulations block companies from running aggregated data across multiple entities.

### 6.3. Final Reflections

This research establishes that AI-powered big data-based frameworks offer substantial transformative possibilities to detect financial fraud in banking systems. A solution to one of digital finance's main cybersecurity problems has been developed through precise applications of deep learning techniques, real-time processing infrastructure, and interpretability mechanisms. The ongoing evolution of fraud culminates in the necessity of developing new security systems that match the developing fraud techniques. The approach is a foundational milestone toward creating intelligible fraud detection systems that are growing in the current fraud environment.

The future requires companies to invest in persistent innovation and combined efforts between regulators and different industries while keeping pace with regulatory changes. The financial sector stands well-equipped to lead the fraud prevention battle because it connects AI systems with distributed computing networks and advanced analytical tools.

## Compliance with ethical standards

*Disclosure of conflict of interest*

No conflict of interest to be disclosed.

## References

[1] Almotiry, O. N., Sha, M., Rahamathulla, M. P., & Dawood Omer, O. S. (2021). Hybrid cloud architecture for higher education system. Computer Systems Science and Engineering, 36(1). https://doi.org/10.32604/csse.2021.014267

[2] Chen, L., Chen, P., & Lin, Z. (2020). Artificial Intelligence in Education: A Review. IEEE Access, 8, 75264–75278. https://doi.org/10.1109/ACCESS.2020.2988510

[3] Cardellini, V., Lo Presti, F., Nardelli, M., & Russo, G. R. (2022). Runtime Adaptation of Data Stream Processing Systems: The State of the Art. ACM Computing Surveys, 54(11). https://doi.org/10.1145/3514496

[4] Dias de Assunção, M., da Silva Veith, A., & Buyya, R. (2018). Distributed data stream processing and edge computing: A survey on resource elasticity and future directions. Journal of Network and Computer Applications, 103, 1–17. https://doi.org/10.1016/j.jnca.2017.12.001

[5] Garai, Á., Péntek, I., Adamkó, A., & Németh, Á. (2017). Methodology for clinical integration of E-health sensor-based smart device technology with cloud architecture. Pollack Periodica, 12(1), 69–80. https://doi.org/10.1556/606.2017.12.1.6

[6] Hassabis, D., Kumaran, D., Summerfield, C., & Botvinick, M. (2017, July 19). Neuroscience-Inspired Artificial Intelligence. Neuron. Cell Press. https://doi.org/10.1016/j.neuron.2017.06.011

[7] Hirzel, M., Soulé, R., Schneider, S., Gedik, B., & Grimm, R. (2014, April 1). A catalog of stream processing optimizations. ACM Computing Surveys. Association for Computing Machinery. https://doi.org/10.1145/2528412

[8] Ismail Fawaz, H., Forestier, G., Weber, J., Idoumghar, L., & Muller, P. A. (2019). Deep learning for time series classification: a review. Data Mining and Knowledge Discovery, 33(4), 917–963. https://doi.org/10.1007/s10618-019-00619-1

[9] Johnson, K. W., Torres Soto, J., Glicksberg, B. S., Shameer, K., Miotto, R., Ali, M., … Dudley, J. T. (2018, June 12). Artificial Intelligence in Cardiology. Journal of the American College of Cardiology. Elsevier USA. https://doi.org/10.1016/j.jacc.2018.03.521

[10] Jiang, J., Li, Y., He, B., Hooi, B., Chen, J., & Kang, J. K. Z. (2022). Spade: A real-time fraud detection framework on evolving graphs. Proceedings of the VLDB Endowment, 16(3), 461–469. https://doi.org/10.14778/3570690.3570696

[11] Janiesch, C., Zschech, P., & Heinrich, K. (2021). Machine learning and deep learning. Electronic Markets, 31(3), 685–695. https://doi.org/10.1007/s12525-021-00475-2

[12] Johnson, J. M., & Khoshgoftaar, T. M. (2019). Survey on deep learning with class imbalance. Journal of Big Data, 6(1). https://doi.org/10.1186/s40537-019-0192-5

[13] Kutyauripo, I., Rushambwa, M., & Chiwazi, L. (2023). Artificial intelligence applications in the agrifood sectors. Journal of Agriculture and Food Research, 11. https://doi.org/10.1016/j.jafr.2023.100502

[14] Korteling, J. E. (Hans), van de Boer-Visschedijk, G. C., Blankendaal, R. A. M., Boonekamp, R. C., & Eikelboom, A. R. (2021). Human- versus Artificial Intelligence. Frontiers in Artificial Intelligence, 4. https://doi.org/10.3389/frai.2021.622364

[15] Lu, M., Han, Z., Rao, S. X., Zhang, Z., Zhao, Y., Shan, Y., … Jiang, J. (2022). BRIGHT - Graph Neural Networks in Real-time Fraud Detection. In International Conference on Information and Knowledge Management, Proceedings (pp. 3342–3351). Association for Computing Machinery. https://doi.org/10.1145/3511808.3557136

[16] Lackermair, G. (2011). Hybrid cloud architectures for the online commerce. In Procedia Computer Science (Vol. 3, pp. 550–555). https://doi.org/10.1016/j.procs.2010.12.091

[17] Lee, A., Mhatre, J., Das, R. K., & Hong, M. (2023). Hybrid Mobile Cloud Computing Architecture with Load Balancing for Healthcare Systems. Computers, Materials and Continua, 74(1), 435–452. https://doi.org/10.32604/cmc.2023.029340

[18] Mill, E., Garn, W., Ryman-Tubb, N., & Turner, C. (2023). Opportunities in Real Time Fraud Detection: An Explainable Artificial Intelligence (XAI) Research Agenda. International Journal of Advanced Computer Science and Applications, 14(5), 1172–1186. https://doi.org/10.14569/IJACSA.2023.01405121

[19] Medeiros, D. S. V., Cunha Neto, H. N., Lopez, M. A., Luiz, L. C., Fernandes, N. C., Vieira, A. B., … Diogo, D. M. (2020). A survey on data analysis on large-Scale wireless networks: online stream processing, trends, and challenges. Journal of Internet Services and Applications, 11(1). https://doi.org/10.1186/s13174-020-00127-2

[20] Mohammed, A., & Kora, R. (2023, February 1). A comprehensive review on ensemble deep learning: Opportunities and challenges. Journal of King Saud University - Computer and Information Sciences. King Saud bin Abdulaziz University. https://doi.org/10.1016/j.jksuci.2023.01.014

[21] Nasiri, H., Nasehi, S., & Goudarzi, M. (2019). Evaluation of distributed stream processing frameworks for IoT applications in Smart Cities. Journal of Big Data, 6(1). https://doi.org/10.1186/s40537-019-0215-2

[22] Quah, J. T. S., & Sriganesh, M. (2008). Real-time credit card fraud detection using computational intelligence. Expert Systems with Applications, 35(4), 1721–1732. https://doi.org/10.1016/j.eswa.2007.08.093

[23] Siasi, N., Jasim, M., Aldalbahi, A., & Ghani, N. (2020). Delay-aware SFC provisioning in hybrid fog-cloud computing architectures. IEEE Access, 8, 167383–167396. https://doi.org/10.1109/ACCESS.2020.3021354

[24] Thiebes, S., Lins, S., & Sunyaev, A. (2021). Trustworthy artificial intelligence. Electronic Markets, 31(2), 447–464. https://doi.org/10.1007/s12525-020-00441-4

[25] Ye, C., Li, Y., He, B., Li, Z., & Sun, J. (2021). GPU-Accelerated Graph Label Propagation for Real-Time Fraud Detection. In Proceedings of the ACM SIGMOD International Conference on Management of Data (pp. 2348–2356). Association for Computing Machinery. https://doi.org/10.1145/3448016.3452774

[26] Yousef, M., & Allmer, J. (2023). Deep learning in bioinformatics. Turkish Journal of Biology, 47(6), 366–382. https://doi.org/10.55730/1300-0152.2671

[27] Zhang, Z., Cui, P., & Zhu, W. (2022). Deep Learning on Graphs: A Survey. IEEE Transactions on Knowledge and Data Engineering, 34(1), 249–270. https://doi.org/10.1109/TKDE.2020.2981333