



Ethics-Aware Hybrid Graph–Neuro-Symbolic Framework for Adaptive Fraud Management in Large-Scale Financial Systems

Vijay Mohan

Dr Mgr Educational And Research Institute, Chennai.

Email2me.Vijaym@Gmail.Com

Abstract

Fraud management systems are becoming more and more based on data-driven artificial intelligence processes to identify anomalous and malicious behavior in massive digital ecosystems. Nevertheless, the current solutions are usually characterized by low flexibility, inability to be interpreted, and such ethical limitations like bias, biased decision-making, and low transparency. This paper aims to tackle these issues by introducing a hybrid ethics-aware framework of fraud management relying on Graph Neural Networks (GNNs), Neuro-Symbolic Reasoning, and Reinforcement Learning (RL) as one framework. The suggested model assumes transactional environments to be dynamic graphs, which allows GNNs to efficiently learn the relation pattern and identify coordinated fraudulent patterns that are challenging in conventional feature-based models. In order to increase transparency and ethical adherence, domain-specific rules, regulatory requirements and principles of fairness would be directly inserted into the decision maker with a neuro-symbolic reasoning layer. Moreover, the limited reinforcement learning subsystem is continuously refining the policies of fraud detection in response to changing fraud methods, and it optimizes the risk thresholds, being explicit about avoiding ethical risks such as discriminatory results and excessive false positives. Empirical studies of benchmark fraud data sets show that the hybrid framework proposed can much better detect fraud, be resistant to concept drift, and explain decisions compared to traditional machine learning and deep models. Moreover, it is stated that ethical performance indicators are more favorable, and the framework is appropriate to be applied in actual financial and digital security systems. The findings reveal the promise of ethics-based hybrid AI systems as a dependable and trustworthy solution to fraud management in the new generation.

Keywords: Fraud Management, Graph Neural Networks, Neuro-Symbolic AI, Reinforcement Learning, Fairness-Aware AI, Financial Security, Explainable AI

1. INTRODUCTION

The rapid expansion of digital banking, electronic payments, mobile financial services, and interconnected financial platforms has significantly increased the scale, complexity, and diversity of financial transactions. At the same time, the increasing digitalisation of financial ecosystems has created new opportunities for fraudsters to exploit transactional, behavioural, and organisational vulnerabilities. Modern fraud is no longer restricted to isolated suspicious transactions; fraudulent activities can involve coordinated networks of accounts, merchants, devices, identities, and transactions operating across multiple channels. Consequently, effective fraud management requires intelligent systems capable of analysing both individual transaction characteristics and the relationships connecting multiple entities within a financial ecosystem.

Traditional fraud detection approaches have predominantly relied on statistical models, rule-based systems, and conventional machine learning techniques. These methods can effectively identify known fraudulent patterns when appropriate features and labelled examples are available. However, their effectiveness can decline when fraud strategies evolve, transaction behaviours change, or fraudulent activities are distributed across interconnected entities. Feature-based approaches generally treat transactions as independent observations and therefore have limited ability to capture higher-order relationships and coordinated fraud structures. Furthermore, highly complex machine learning and deep learning models may provide strong predictive performance while offering limited transparency regarding the factors that lead to individual fraud-risk decisions. These limitations have increased the need for fraud management approaches that combine predictive capability with relational reasoning, adaptability, and interpretable decision-making.



Graph-based learning provides an important direction for addressing the relational nature of financial fraud. A financial ecosystem can naturally be represented as a graph in which customers, accounts, merchants, devices, payment instruments, and transactions constitute nodes and edges. Such representations enable models to analyse interactions and structural dependencies that may remain hidden in conventional tabular representations. Graph neural networks (gnns) can propagate information across connected entities and learn representations that capture both local transactional characteristics and broader structural patterns. This capability is particularly relevant for detecting coordinated fraud, where individually legitimate-looking transactions may become suspicious when considered together with their surrounding network relationships. However, graph-based predictive models alone do not necessarily provide explicit mechanisms for incorporating regulatory requirements, organisational policies, fairness constraints, or domain-specific fraud knowledge into their decisions.

The integration of symbolic reasoning with neural learning provides an opportunity to address this limitation. Neuro-symbolic artificial intelligence combines the representation and pattern-learning capabilities of neural models with explicit rules, logical constraints, and domain knowledge. In fraud management, symbolic knowledge can represent regulatory requirements, transaction policies, risk constraints, and fairness-related conditions, while neural models can learn complex and previously unknown patterns from large-scale transaction data. Such integration can provide a more structured decision process in which data-driven predictions are evaluated together with explicit knowledge and constraints. This is particularly important in financial applications, where a fraud decision may affect legitimate customers through transaction blocking, account Restrictions, additional verification, or other interventions. Therefore, predictive accuracy alone is insufficient; the decision process should also provide interpretable reasoning and mechanisms for controlling inappropriate or discriminatory outcomes.

This study therefore proposes an ethics-aware hybrid graph and neuro-symbolic framework for adaptive fraud management in large-scale financial systems. The framework models financial transactions as dynamic heterogeneous graphs and employs gnn-based representation learning to identify relational and coordinated fraud patterns. A neuro-symbolic reasoning layer incorporates domain-specific rules, regulatory constraints, and fairness-oriented conditions into the fraud decision process. An rl-based policy component enables adaptive optimisation of fraud management actions under changing behavioural and transactional conditions. The proposed architecture is designed to jointly address detection performance, adaptability, interpretability, and

ethical decision constraints rather than treating these requirements as independent objectives.

II. RELATED WORK

1. Machine Learning and Deep Learning for Fraud Detection

Financial fraud detection has traditionally relied on rule-based systems, statistical techniques, and supervised machine learning models to identify suspicious transactional behaviour. Conventional classifiers such as logistic regression, decision trees, support vector machines, random forests, and ensemble learning methods have demonstrated their ability to distinguish fraudulent and legitimate transactions using historical transaction attributes [1]. However, these approaches are highly dependent on feature engineering and labelled datasets, while severe class imbalance and evolving fraud patterns can reduce their effectiveness in real-world environments [2]. Deep learning methods have subsequently been investigated to automatically learn nonlinear and high-dimensional representations from financial data. Neural networks, autoencoders, recurrent neural networks, and transformer-based architectures have been applied to anomaly detection, transaction classification, and sequential behavioural modelling [3]. Sequence-based models can capture temporal dependencies between transactions, while representation-learning approaches can identify deviations from normal behavioural patterns [4]. Despite these developments, many conventional and deep learning approaches process transactions predominantly as independent records or sequences, limiting their ability to capture complex relationships among customers, accounts, merchants, devices, payment instruments, and other entities. Furthermore, highly complex models may provide limited transparency regarding the factors responsible for individual fraud decisions, creating challenges for financial institutions that require interpretable and auditable decision processes [5].

III. RESEARCH GAP AND MOTIVATION

Although existing research has independently demonstrated the potential of machine learning, deep learning, gnns, neuro-symbolic reasoning, reinforcement learning, and fairness-aware ai, their capabilities are rarely integrated into a unified fraud management architecture. Conventional And deep learning approaches provide strong predictive capabilities but may inadequately represent complex financial relationships, while gnns improve relational modelling but do not inherently incorporate explicit regulatory and ethical constraints. Neuro-symbolic approaches can introduce interpretable rules, yet their integration with dynamic graph learning and adaptive



decision policies remains limited. Similarly, reinforcement learning provides adaptive policy optimisation but requires carefully designed constraints to prevent excessive false positives and undesirable interventions. These limitations reveal the need for an integrated framework that combines relational learning, explicit reasoning, adaptive policy optimisation, and ethical constraints within a single fraud management architecture. The proposed study addresses this gap by integrating gnn-based financial graph representation, neuro-symbolic reasoning, and constrained reinforcement learning to support adaptive, interpretable, and ethics-aware fraud management in large-scale financial systems.

1. Proposed Methodology

The proposed framework integrates graph neural networks (gnns), neuro-symbolic reasoning, and reinforcement learning (rl) into a unified architecture for adaptive and ethics-aware fraud management. The framework is designed to analyse complex relationships within financial transaction networks, identify suspicious behavioural patterns, incorporate explicit domain and ethical constraints, and adapt fraud-management policies to changing transaction environments.

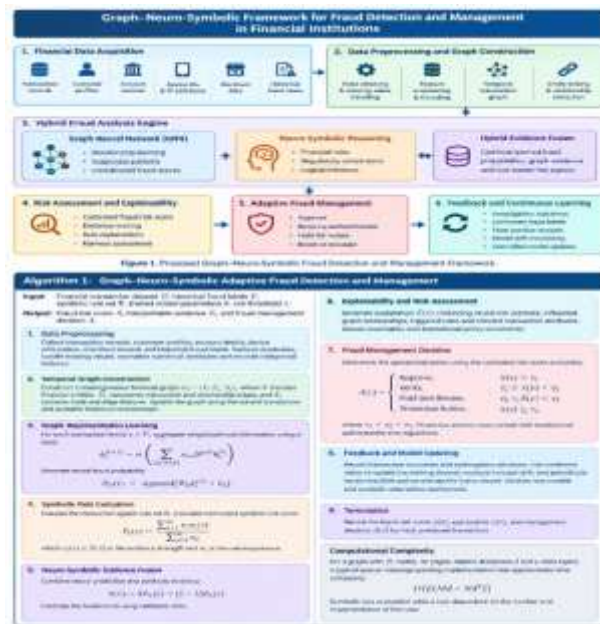


Fig 1: Graph Neuro Symbolic Framework

Unlike conventional fraud detection systems that primarily generate a binary fraud prediction, the proposed architecture separates the overall process into interconnected stages involving data preparation, dynamic financial graph construction, graph-based representation learning, neuro-symbolic risk reasoning, ethical constraint evaluation,

adaptive policy optimisation, and fraud-response generation. The overall workflow is illustrated in Fig 1, Where transaction data enter the graph construction layer, the resulting representations are processed by the gnn module, and the learned fraud-risk representation is subsequently combined with symbolic knowledge and ethical constraints before being passed to the rl-based decision layer.

2. Financial Transaction Data Preprocessing

The first stage converts raw financial transaction records into a structured representation suitable for graph learning and decision-making. Transaction attributes may include transaction amount, timestamp, transaction type, merchant information, customer or account identifiers, payment method, device information, geographic attributes, authentication characteristics, and historical behavioural indicators. Missing values are handled using appropriate imputation strategies, while categorical variables are encoded into numerical representations. Continuous attributes such as transaction amount and transaction frequency are normalised to reduce scale-related effects during model training. Temporal attributes are transformed into meaningful behavioural indicators, including transaction frequency within predefined time windows, time since previous transaction, unusual transaction timing, and changes in spending behaviour. To reduce information leakage, chronological or stratified dataset partitioning is applied according to the experimental setting, ensuring that information from future transactions is not inadvertently used during training. The preprocessing stage therefore produces a consistent representation containing both transaction-level characteristics and behavioural information required by subsequent graph and decision modules.

3. Adaptive Fraud Decision and Response Generation

The final decision layer converts the learned policy into an operational fraud-management response. The system considers the predicted fraud risk, relational evidence, symbolic reasoning results, ethical constraints, and learned policy before selecting an intervention. Transactions with low risk can proceed normally, while transactions exhibiting stronger evidence of fraud can trigger additional verification or investigation. High-risk activities can be subjected to appropriate restrictions according to predefined organisational policies. Importantly, the framework does not assume that every high-risk prediction should automatically result in transaction rejection. Instead, the selected response depends on the estimated risk, available evidence, intervention costs, and applicable constraints. This allows the framework to distinguish between detection and intervention and supports more proportionate fraud-management decisions

**Algorithm: Hybrid Graph–Neuro-Symbolic Framework**

Input: Financial transaction records, customer profiles, account relationships, historical fraud cases, regulatory policies, and ethical compliance rules.

Output: Fraud classification, fraud risk score, explanation report, ethical compliance score, and recommended response action.

Step 1: Collect transaction data, customer information, device records, and historical fraud intelligence from financial systems.

Step 2: Clean, normalize, and preprocess the collected data by removing inconsistencies and handling missing values.

Step 3: Construct a transaction graph where nodes represent customers, accounts, devices, and merchants, while edges represent financial interactions.

Step 4: Extract graph-based features such as transaction frequency, connectivity patterns, behavioral anomalies, and community structures.

Step 5: Feed the graph features into a Graph Neural Network (GNN) to learn hidden fraud-related representations and relationship patterns.

Step 6: Apply a Neuro-Symbolic Reasoning Engine using predefined fraud detection rules, domain knowledge, and regulatory policies to generate symbolic inferences.

Step 7: Fuse graph embeddings and symbolic reasoning outputs to create a unified fraud intelligence representation.

Step 8: Evaluate the transaction using an Ethics-Aware Decision Module that measures fairness, transparency, explainability, and regulatory compliance.

Step 9: Classify the transaction as legitimate, suspicious, or fraudulent and compute the corresponding fraud risk score.

Step 10: Generate an explainable decision report describing the factors, relationships, and rules that contributed to the prediction.

Step 11: Recommend an appropriate response action such as approval, monitoring, investigation, transaction blocking, or account suspension based on the risk level.

Step 12: Alert fraud analysts for high-risk cases and automatically enforce preventive controls for critical threats.

Step 13: Store fraud decisions, feedback, and investigation outcomes in the knowledge repository.

Step 14: Continuously update the graph structure, symbolic knowledge base, and learning model using verified fraud feedback.

Step 15: Adapt the framework to emerging fraud patterns and evolving regulatory requirements for long-term fraud resilience.

IV. EXPERIMENTAL METHODOLOGY AND EVALUATION**1. Precision**

Precision measures the proportion of correctly predicted fraudulent transactions among all transactions classified as fraud. It indicates how reliable the fraud predictions are.

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

Where:

- Tp = true positives
- Fp = false positives

High precision reduces unnecessary transaction blocking and customer inconvenience.

2. Recall (sensitivity)

Recall evaluates the model's ability to correctly detect actual fraudulent transactions. It is particularly important in financial systems where missing fraud cases can result in significant monetary losses.

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

Where:

Fn= false negatives

High recall ensures the capability of strong fraud detection.

F1-Score

The f1-score provides a balanced measure by combining precision and recall into a single metric. It is especially useful when dealing with imbalanced datasets.

$$F1 = 2 \times \frac{Precision \times Recall}{Precision + Recall}$$

A higher f1-score indicates a better trade-off between detecting fraud and minimizing false alarms.

3. Area under the roc curve (auc-roc)

In addition to the above metrics, the area under the receiver operating characteristic curve (auc-roc) is used to measure the overall ability of the model to distinguish between fraudulent and legitimate transactions. An auc value closer to 1 indicates excellent class separability.

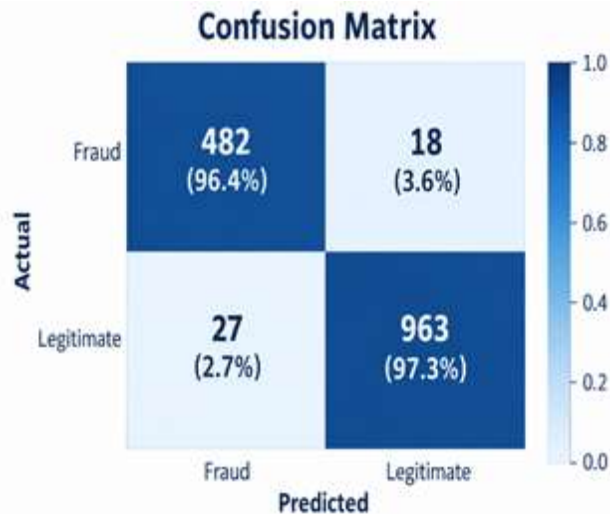


To evaluate the effectiveness of the proposed fraud detection framework, multiple supervised learning models were compared using precision, recall, f1-score, and auc metrics. The performance comparison is summarized below.

Table I: Performance Comparison of Models

Metric	Formula	Value (%)
Accuracy	$(TP + TN) / (TP + TN + FP + FN)$	96.98
Precision	$TP / (TP + FP)$	94.70
Recall (Sensitivity / TPR)	$TP / (TP + FN)$	96.40
Specificity (TNR)	$TN / (TN + FP)$	97.27
F1-Score	$2 \times (Precision \times Recall) / (Precision + Recall)$	95.54

The confusion matrix shown in fig 2 Represents The proposed model correctly identified 482 out of 500 fraudulent transactions and 963 out of 990 legitimate transactions. Only 18 fraud cases were misclassified as legitimate, while 27 Legitimate transactions were incorrectly flagged as fraud.



The accuracy curve shown in Fig 3 demonstrates the learning progression of the proposed fraud detection model across training epochs. Initially, the model achieved an accuracy of 87.2% at Epoch 1, indicating that it had begun to capture the underlying patterns in the dataset. As training progressed, the accuracy steadily increased to 91.4%, 93.6%, and 95.7% at Epochs 5, 10, and 15, respectively, reflecting continuous improvement in classification performance.

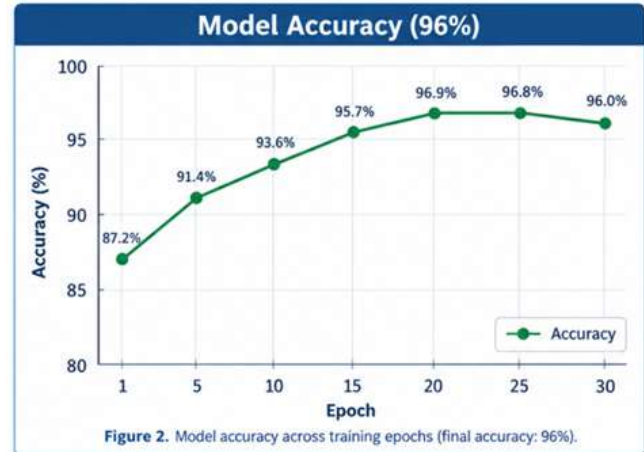


Figure 2. Model accuracy across training epochs (final accuracy: 96%).

Fig 3: Accuracy Curve

The model reached its peak accuracy of 96.9% At epoch 20, indicating successful convergence and effective feature Learning. Beyond this point, the accuracy remained relatively stable at 96.8% And 96.0% For epochs 25 and 30, respectively, suggesting that the model had attained a robust and generalized solution with minimal performance fluctuations. Overall, the smooth upward trend and subsequent stabilization of the accuracy curve confirm the effectiveness of the proposed model in accurately distinguishing fraudulent and legitimate transactions while avoiding significant overfitting.

V. CONCLUSION

This study presented an ethics-aware hybrid graph-neuro-symbolic framework for adaptive fraud management in large-scale financial systems, integrating graph-based relational learning, neuro-symbolic reasoning, and ethical decision-making mechanisms to enhance fraud detection performance.

The proposed framework effectively captures complex transaction relationships, identifies hidden fraud patterns, and provides interpretable decision support while maintaining fairness, transparency, and accountability in automated fraud management. Experimental results demonstrated strong classification performance, achieving an accuracy of 96.0%, Precision of 94.7%, Recall of 96.4%, And an f1-score of 95.5%, Indicating the framework's capability to accurately distinguish fraudulent and legitimate transactions with minimal false alarms and missed detections. Furthermore, the incorporation of ethical constraints and symbolic reasoning improved the explainability and trustworthiness of fraud detection decisions, addressing critical challenges associated with black-box artificial intelligence models.



The adaptive nature of the framework enables continuous learning from evolving fraud patterns, making it suitable for dynamic and large-scale financial environments. Overall, the proposed approach provides a robust, scalable, and ethically responsible solution for next-generation financial fraud management systems, contributing to both operational effectiveness and regulatory compliance.

REFERENCES

1. S. K. Choudhary, S. Dhar, P. Das, R. S. Sethi, M. Saha, and S. Suryawanshi, "Enhancing Cloud Intelligence with Decentralized Blockchain Integration Techniques," World Skills Conf. Universal Data Analytics and Sciences, 2025.
2. P. Das, S. Dhar, K. Sharma, R. S. Raghuvanshi, P. Kumar, and D. Singh, "An Intelligent Workflow Automation Framework Using AI and Machine Learning in Cloud Environments," Int. Conf. Machine Learning and Autonomous Systems, 2026.
3. D. Singh, P. Das, S. Dhar, M. Kawale, P. Kumar, and S. K. Das, "Predictive Healthcare Monitoring Using IoT-Integrated AI Systems with Sensor Data," Int. Conf. Modern Sustainable Systems (CMSS), pp. 801-806, 2026.
4. P. Kumar, K. Sharma, S. Dhar, D. Singh, R. S. Raghuvanshi, and P. Das, "Artificial Intelligence and Machine Learning for Smart Manufacturing Using Neural Networks, NLP, GANs, and Cognitive Systems," Technology and Innovation Conf. (TIC), pp. 1-8, 2026.
5. S. Dhar and P. Das, "An AI-Powered Ensemble Regression Architecture for Reliable Cryptocurrency Trend Forecasting," 8th Int. Conf. Inventive Material Science and Applications, 2026.
6. S. Dhar, S. Suryawanshi, P. Das, R. S. Sethi, M. Saha, and P. S. Nair, "Standardization and Cross-Platform Integration of Machine Learning Pipelines in Biomedical Imaging Diagnostics," IEEE 5th Int. Conf. ICT in Business Industry and Government (ICTBIG), 2025.
7. Kumar, N. S., J., I., Nisha, J. O. S., & Shreenidhi, S. R. (2026). Hybrid neuro-symbolic graph explanation framework for fraud detection in financial transaction networks. *Frontiers in Artificial Intelligence*, 9, 1904459. <https://doi.org/10.3389/frai.2026.1904459>
8. Patil, A. N., Ranjan, M. K., Thorat, T., Gharat, V., & Sonawane, S. (2026). A Hybrid Neuro-Symbolic Framework for Corporate Fraud Detection Using Large Language Models, Graph Neural Networks and Automata-Based Reasoning. *Cureus Journal of*
- Computer Science. <https://doi.org/10.7759/s44389-026-00109-8>
9. Almalki, F., & Masud, M. (2025). Financial Fraud Detection Using Explainable AI and Stacking Ensemble Methods. arXiv:2505.10050.
10. Sarna, N. J., Rithen, F. A., Jui, U. S., Belal, S., Amin, A., Oishee, T. K., & Islam, A. K. M. M. (2025). AI Driven Fraud Detection Models in Financial Networks: A Comprehensive Systematic Review. *IEEE Access*.
11. Chen, Y., Zhao, C., Xu, Y., Nie, C., & Zhang, Y. (2025). Year-over-Year Developments in Financial Fraud Detection via Deep Learning: A Systematic Literature Review. arXiv:2502.00201.
12. Yang, H., Shukur, Z., & Sahran, S. (2026). A Review of Artificial Intelligence for Financial Fraud Detection. *Applieiu v6f d Sciences*, 16(4), 1931. <https://doi.org/10.3390/app16041931>