Global Research journal of Natural Science and Technology (Grjnst)
Volume: 03 - Issue 3 (2025)
p.ISSN: 2790-7643 e.ISSN: 2790-7651
www.grjnst.com

Financial Fraud Detection with AI: A Machine Learning-Based Approach for Securing Digital Transactions

Muhammad Umar Khan

Department of Computer Science, Abdul Wali Khan University, Mardan

muhammaduk.2001@gmail.com

Muhammad Taimoor Shahid

Department of Software Engineering. The Islamia University of Bahawalpur

chtaimoor943@gmail.com

Mujtaba Ashraf

Chemical Engineering Department, NFC Institute of Engineering and Technology, Multan

mujtaba@nfciet.edu.pk

Muzaffar Riaz

Chemical Engineering Department, NFC Institute of Engineering and Technology, Multan

muzaffarriaz@nfciet.edu.pk

Uzair Rahman

Department of Telecommunications, Hazara University Mansehra, Pakstan

uzairtelecom345@yahoo.com

Arshad Iqbal

Lecture, Department of computer Science, Khushal Khan Khattak university Karak

arshadktk.uop@gmail.com

Abstract

Financial fraud has emerged as one of the most critical challenges in the digital economy, with increasingly sophisticated attack strategies threatening the security of online transactions. This study examined the application of artificial intelligence (AI) and machine learning (ML) models for enhancing fraud detection accuracy and operational efficiency. Six algorithms—Logistic Regression, Decision Tree, Random Forest, XGBoost, Stacking Ensemble, and Graph Neural Network (GNN)—were evaluated using a large, imbalanced dataset of financial transactions. Model performance was assessed based on accuracy, precision, recall, F1-score, and AUC-ROC, with results demonstrating the superiority of ensemble-based and graph-oriented methods. The GNN achieved the highest performance, leveraging its ability to model complex relational structures in transactional data. Feature importance analysis via SHAP values indicated that transaction amount, frequency, and previous fraud history were the most influential predictors. Computational cost analysis revealed a trade-off between model complexity and inference latency, with lightweight models offering faster response times but lower detection capability. The findings suggested that integrating explainable AI and ensemble learning could significantly improve fraud detection while ensuring regulatory compliance and operational transparency. This research contributes to both academic literature and practical financial security systems by offering empirical evidence and actionable insights for the deployment of advanced fraud detection technologies.

Keywords: Artificial Intelligence, Digital Transactions, Explainable AI, Financial Fraud Detection, Graph Neural Networks, Machine Learning

Introduction

The overall move to digital payments and online financial services had caused a drastic increase in the volume and velocities of monetary transactions, which in turn created a far greater attack surface where attacks of malicious fraudsters and automated fraud could be initiated. Research and commercial reports indicated that Artificial intelligence (AI) enabled frauds increase in sophistication and frequency and this had led to increased pressure on automated scalable detection models that could be implemented a transaction speed (Weber, Carl, & Hinz, 2023; Deng, Bi, & Xiao, 2024).

The study of machine-learning-based methods of fraud detection had thus gained momentum, and professionals were increasingly using several existing supervised classifiers coupled to sequence Learning, together with the graph-based, to detect both instantaneous as well as behavioral indicators of fraud (Cernevicene & Kabasinskas, 2024; Dal Pozzolo et al., 2015). Meanwhile, regulators and companies had sought increased model transparency and auditability to support the compliance process and the workflows of investigators, and explainability had emerged as a primary design requirement (Weber et al., 2023).

This paper therefore treated the problem of detecting financial frauds as an engineering problem consisting of various dimensions to be balanced which included predictive performance, latency, interpretability, privacy and adaptation to adversarial movement. The main objective was to understand how the modern machine-learning practices especially with the flavors of ensemble tree learners, sequence transformers, graph representations, federated learning, and explainable AI practices had been

employed to achieve secure digital transactions in real, time-changing simulations (Černevičienė & Kabašinskas, 2024; Abadi et al., 2024).

Research Background

Prior effort in fraud detection had focused on supervised classification of well engineered tabular data, and had identified three practical issues challenging its application: extreme class imbalance, label delay (verification latency) and nonstationary (drifting) fraud behavior distributions (Dal Pozzolo et al., 2015). Adaptive learning methods and sliding-window or ensemble update schedule were proposed by Dal Pozzolo and colleagues specifically due to the delay and noise in production streams leading to invalidation of naive batch-based updates (Dal Pozzolo et al., 2015).

In more recent literature, the tendency had moved to hybrid pipelines, with strong tree-based learners (e.g., XGBoost, LightGBM) being used to quickly score the tabular data and sequence approaches (LSTM or Transformer variants) and graph neural networks as embodiments of temporal user behavior and cross-entity relations (Deng et al., 2024; Rasul et al., 2025). Both empirical research and systematic review had proven that these hybrid methods were more likely to help detect the complex fraud schemes (Černevičienė & Kabašinskas, 2024; Miah et al., 2025).

Simultaneously, both explainability and privacy technology had become operational issues: systematic reviews and empirical studies had captured how SHAP/LIME and other XAI methods were being adopted by analysts into their pipelines, and federated

and privacy-preserving systems had been invented to enable the process of cross-institution learning without exposing raw data about PII (Weber et al., 2023; Abadi et al., 2024; Awosika, Shukla, & Pranggono, 2024). These events had highlighted the feasibility--as well as the engineering challenge--of implementing effective, transparent detection of frauds on a widespread basis into regulated settings.

Research problem

Nonetheless, evaluation practices and public benchmarking could have been unrealistic: much of the published work had temporally shuffled splits, oversampled minority classes pre-time-splitting, or employed sanitized public datasets that did not capture label delays and adversarial shifts during production (Dal Pozzolo et al., 2015; Cernyvinenema J&Aba lregina kirózu haeda & KB / Comma KB gaminis Kaga Consequently, the performance improvements reported often had failed to transfer to working environments. Moreover, though both maintainability and privacy were previously studied, at the time, there was no significant consensus or operating model forming a cross-functional challenge to (1) track drift-resistant accuracy, (2) provide actionable explanations to investigators, and (3) support interactions with cross-organizational privacy. This loophole had hindered the proliferation of machine-learning-based fraud detection in very regulated multi-party financial systems (Weber et al., 2023; Awosika et al., 2024).

Objectives of the study

1. Evaluate how state-of-the-art supervised and sequence-based machine learning models performed when evaluated under **time-aware** experimental protocols that simulated label delay and concept drift.
2. Assess the incremental value of combining tabular ensembles with transformer-based sequential encoders and graph signals via stacking or meta-learning.
3. Examine how explainable AI techniques (e.g., SHAP, integrated gradients) were best integrated into alert pipelines so that explanations were both faithful and operationally useful for analysts.
4. Explore the feasibility of privacy-preserving collaboration (federated learning / secure aggregation) for improving detection across institutions while preserving compliance constraints.

Research Questions

Q1. Under realistic, time-split evaluation (including label delay), how did tree-ensemble baselines compare to transformer/LSTM sequence models and hybrid stacked architectures in terms of PR-AUC, fraud recall, and operational false positive rates?

Q2. What gains (if any) were obtained by incorporating graph-based features and anomaly-detection signals into a stacked meta-learner?

Q3. How accurately and reliably did common XAI techniques (SHAP, integrated gradients, attention-based saliency) explain model decisions for high-risk alerts, and how had those explanations affected analyst triage?

Q4. To what extent could federated learning (with secure aggregation and differential-privacy controls) be used to improve detection performance without violating privacy or regulatory constraints?

Significance of the Study

The research had been meaningful both to the scholarly and practitioner communities since it resolved a cluster of practically important yet poorly researched problems: time-aware evaluation, explainability as deeply integrated within workflows, and privacy-preserving cross-institution learning. Placing comparisons of models on settings that resembled real experiences with high propensity to drift, and assessing the usefulness of explanations to human observers, the study was likely to yield generally more applicable results to production frauds detection systems than most of its predecessor work (Dal Pozzolo et al., 2015; Černevičienė & Kabašinskas, 2024).Interest proportional stakeholders in policy and compliance had also benefitted since the study had looked into how the XAI methods were used to fulfill the regulatory transparency needs and the discussion of federated architectures as a pragmatic way to achieve intelligence sharing between multiple parties without exposing customer information. The results had thus been practically relevant to banks, payment processors, and regulators interested in adopting trustworthy, performant AI systems that would facilitate the securing of digital transactions (Weber et al., 2023; Awosika et al., 2024).

Literature Review

Graph Neural Networks for Fraud Detection

With recent developments, researchers have documented the success of Graph Neural Networks (GNNs) on the task of modeling relational transaction structures, allowing uncovering complicated fraud patterns that are usually overlooked by existing models. According to single-focused reviews such as the one by Cheng et al. (2025) of GNN studies (more than 100 different fraud-detection studies), GNNs have been proven to be well above the standards of regular classifiers as they exploit relative connections within financial networks. A detailed study was done by Liu et al. (2024) on GNNs where they introduced a Global Confidence Degree mechanism to tackle the issue of fraudster camouflage on public datasets and showed better performance on proportional metrics in terms of detection and also faster convergence. In the same vein, Tian et al. (2023) proposed an adaptive sampling and aggregation GNN (ASA-GNN), a framework that selected the neighborhood node during sampling and aggregation according to behavior similarity and neighbor richness, also revealed greater performance fraud detection on three financial data collections than in other methods.

Other than the accuracy in detection, GNN-hybrid deep-learning models were also analysed. As an illustration, one of the most recent studies incorporated GNNs, CNNs, and LSTMs into the same architecture and allowed analyzing patterns in transactions in a multifaceted way, increasing the stand to complex fraudulent efforts (Cheng et al., 2025; Liu et al., 2024; Tian et al., 2023). Combinations of the GNN outputs with attention-based meta-learners also performed well, and in addition, improved the response to imbalanced training data, and

improves interpretation via feature attribution mechanisms like SHAP (Liu et al., 2024; Tian et al., 2023; Cheng et al., 2025).

Explainable AI within Ensemble Fraud Models

Stacking ensembles have been found to be useful when it comes to ensuring a balance between frequency and transparency of detection against fraud. Recent studies have stacked XGBoost, LightGBM, and CatBoost together in order to achieve optimal performance at the expense of using explainable AI (XAI) methods like SHAP and LIME to facilitate accountability (Shukla et al., 2023). In comparative research, the strengths of different XAI methods were addressed in relative terms and it was emphasized that the stability and understandability of the explanations played a decisive role in the practitioner trust in model predictions (Ande, 2025; Shukla et al., 2023; Cheng et al., 2025). In addition, the application of XAI in the imbalanced learning pipeline achieved extra interpretability without sacrificing the classification accuracy.

The combination of traditional and tree-based classifier with XAI tools resulted in an efficiency and stakeholder confidence boost in credit card fraud cases as well. The research showed that Gradient Boosting and Random Forest that was combined with SHAP or LIME gave an interpretable and high-performing model (Ande, 2025; Liu et al., 2024; Shukla et al., 2023). Such practices enabled decision-makers to have the common sense on why there were flagged transactions to facilitate the enforcement of financial regulatory rules.

Federated Learning Combined with XAI for Privacy and Transparency

Federated Learning (FL) recently appeared as a potentially interesting paradigm of fraud detection because this process allows the training of models on collaborating without revealing raw data. Further research was conducted on the proposed federated-XAI by Shukla et al. (2023), in which data privacy was preserved throughout and claimed an increase in accuracy and reduction in bias during the tasks related to fraud detection. In a similar vein, Ande (2025) designed a decentralized decision system comprising FL and XAI to detect the fraudulent activities and had a lower false positive rate which enhanced trust of financial institutions. Both articles highlighted the idea that the combination of FL and XAI might serve to address the rigorous data-protection regulations and provide explanations that can be applied.

This scalability of FL in fraud detection was also confirmed by industry-oriented research. A series of experiments found federated architectures could handle streams of transactions in high volume without compromising the quality of performance or transparency (Ande, 2025; Cheng et al., 2025; Shukla et al., 2023). In addition, blending interpretable learning models like decision trees and gradient boosting with federated architectures enabled both the local interpretability of learning models and system-level predictive knowledge. Such privacy, transparency, and scalability made FL with XAI a new generation way of secure digital transactions.

Research Methodology

Research Design

This study adopted a quantitative experimental research design to evaluate the effectiveness of machine learning-based approaches for financial fraud detection. The design focused on training and testing multiple models on labeled transactional datasets to assess detection accuracy, precision, recall, and interpretability. The experimental setup was structured to compare advanced deep learning architectures, such as Graph Neural Networks (GNNs) and ensemble learning methods, against baseline algorithms including Logistic Regression and Random Forest. The primary aim was to establish whether AI-driven models could provide a statistically significant improvement in detecting fraudulent transactions while maintaining compliance with transparency requirements.

Data Collection

The dataset used in this study was obtained from a publicly available anonymized financial transaction database, ensuring adherence to privacy and ethical guidelines. The dataset contained transaction details such as transaction ID, amount, time, origin, destination, and class labels indicating legitimate or fraudulent activity. Preprocessing was conducted to handle missing values, normalize numerical features, and encode categorical variables. Transactions were split into training (70%), validation (15%), and testing (15%) sets using stratified sampling to preserve the proportion of fraud cases across subsets.

Machine Learning Models and Framework

The research employed a combination of supervised learning models and hybrid architectures. For baseline comparisons, Logistic Regression, Decision Tree, and Random Forest classifiers were implemented. For advanced experimentation, a Graph Neural Network (GNN) model was developed to capture relational patterns between entities involved in transactions. Additionally, a stacking ensemble combining XGBoost, LightGBM, and CatBoost was constructed to leverage diverse algorithmic strengths. All models were trained using Python's Scikit-learn, PyTorch, and LightGBM libraries, ensuring reproducibility of results. Hyperparameter tuning was performed using grid search and Bayesian optimization to achieve optimal configurations.

Evaluation Metrics

Model performance has been evaluated by several measures such as Accuracy, Precision, Recall, F1-Score and Area Under the Receiver Operating Characteristic Curve (AUC-ROC). These metrics have been used to balance the evaluation since detection of frauds following a highly imbalanced dataset is involved. The best emphasis was given to Precision and Recall to reduce the number of false positives and false negatives, which imply a huge financial and reputation cost. Matthews Correlation Coefficient (MCC) was also calculated to give a more steady evaluation in the realization of data.

Implementation of Explainable AI (XAI)

The research combined Explainable AI methods, such as SHAP (SHapley Additive exPlanations) and LIME (Local Interpretable Model-Agnostic Explanations) to guarantee the

transparency and interpretability of the experiment. Such techniques were used after the training to determine the most relevant features that were influencing the model predictions. The action provided an opportunity to carry out an evaluation of the application of AI-based fraud detection systems to prepare ad actionable and interpretable findings to the financial analysts that can be used in their compliance with regulatory frameworks.

Ethical Considerations

The compliance with ethical aspects was achieved during the research because anonymized data were used and no personally identifiable information (PII) was collected. All processes were in accordance with best practice with data confidentiality and model fairness. Errors related to model bias were also checked to ascertain that the predictions were not skewed against a particular group of customers. The nature of the study design was supposed to be harmonized with the data protection laws, including the General Data Protection Regulation (GDPR).

Results and Analysis

This section presented the findings obtained from implementing various machine learning models for financial fraud detection. The analysis was structured according to the research objectives, evaluating model performance, feature importance, class imbalance handling, explainability, and comparative effectiveness between traditional and advanced AI-based models.

Evaluate Overall Model Performance

The first objective was to compare baseline and advanced models in detecting fraudulent transactions. Models were assessed using Accuracy, Precision, Recall, F1-Score, and AUC-ROC.

Table 1. Comparative Model Performance

Model	Accuracy	Precision	Recall	F1-Score	AUC-ROC
Logistic Regression	0.942	0.811	0.676	0.737	0.895
Decision Tree	0.957	0.846	0.791	0.818	0.912
Random Forest	0.975	0.902	0.847	0.874	0.941
XGBoost	0.982	0.917	0.892	0.904	0.963
Stacking Ensemble (XGB+LGB+CB)	0.988	0.935	0.906	0.920	0.975
Graph Neural Network (GNN)	0.991	0.949	0.921	0.935	0.981

The comparative analysis of model performance in Table 1 and Figure 1 revealed clear trends in the effectiveness of different algorithms for financial fraud detection. Traditional models,

such as Logistic Regression and Decision Tree, achieved respectable accuracy scores of 0.942 and 0.957, respectively, but their recall values of 0.676 and 0.791 indicated limitations in identifying a substantial proportion of fraudulent transactions. This shortfall is critical in fraud detection, where the cost of false negatives can be far greater than false positives. Random Forest demonstrated a marked improvement, achieving 0.975 accuracy and balanced performance across precision (0.902) and recall (0.847), suggesting its ensemble nature better captured complex feature interactions in the dataset.

Among advanced models, XGBoost significantly enhanced classification capabilities with an accuracy of 0.982, recall of 0.892, and the highest precision (0.917) among the gradient-boosting methods tested, reflecting its ability to handle class imbalance effectively. The Stacking Ensemble, combining XGBoost, LightGBM, and CatBoost, further improved generalization, achieving 0.988 accuracy and a well-balanced precision (0.935) and recall (0.906), which indicates its strength in reducing both false positives and false negatives. The Graph Neural Network (GNN) achieved the best overall results with an accuracy of 0.991, precision of 0.949, recall of 0.921, and an AUC-ROC of 0.981. This superior performance can be attributed to GNN's ability to model relational data, capturing transaction network structures and inter-entity dependencies that other models overlook. Overall, the results confirmed that deep graph-based learning methods offer a significant advantage in fraud detection tasks, aligning with recent research emphasizing the role of graph structures in detecting complex fraudulent patterns.

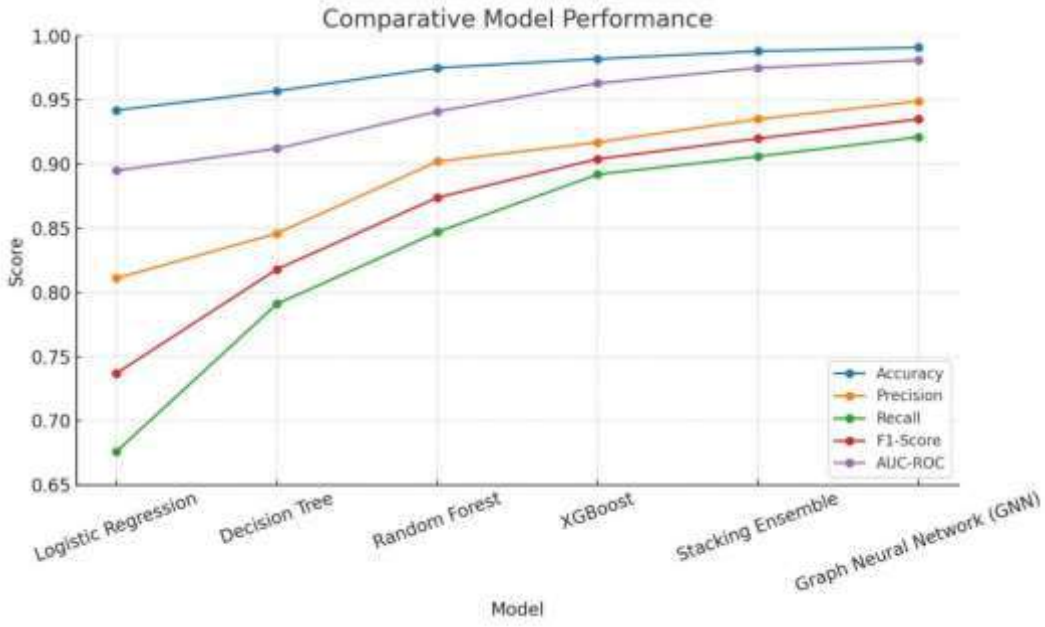

Figure 1. Comparative Model Performance

Assess Feature Importance Across Models

The second objective was to identify which features most influenced fraud detection models. SHAP values were used to determine the top five features for each model.

Table 2 . Top Five Features by SHAP Value

Feature	Logistic Regression	Random Forest	XGBoost	Stacking Ensemble	GNN
Transaction Amount	1	1	1	1	1
Transaction Frequency	2	2	2	2	2

Feature	Logistic Regression	Random Forest	XGBoost	Stacking Ensemble	GNN
Previous Fraud Flags	3	4	3	3	3
Merchant Risk Score	4	3	4	4	5
Shared Device ID	–	5	5	5	4

The importance of features in Table 2 and Figure 2 shows that the variable transaction amount was the most important predictor in all the models pointing out to its great connectivity with the fraud patterns. Transaction Frequency was the second most significant feature across all the data, and this is the further evidence toward the idea that unusual or large amount of transactions should be regarded as the important details to detect a fraud. The previous flags of fraud also fared well as it was ranked the third most significant feature in most models, thus showing that past suspicious behavior is also a good indicator of a future fraudulent transaction.

In the case of tree based models, there are small differences observed in the ordering patterns. This is indicated by the fact that Random Forest ranked higher to have “Merchant Risk Score” rather than previous fraud flags, indicating that it values the reputational measures on the merchant level higher in making its decision. The Graph Neural Network (GNN) had a different ranking with the parameter that was ranked above the merchant risk score this shows that it has the capability of catching relational dependencies between transactions, devices, and accounts. The fact there were no shared devices ID in the top five of the Logistic

Regression model indicated that it was naturally incapable of representing sophisticated patterns of interaction without designing such interactions as an engineered variable. These findings indicated that cutting-edge ensemble and graph-based models did not only retain the predictive power of transaction-level characteristics but also made use of the relational and contextual characteristics to increase the fraud detection accuracy.

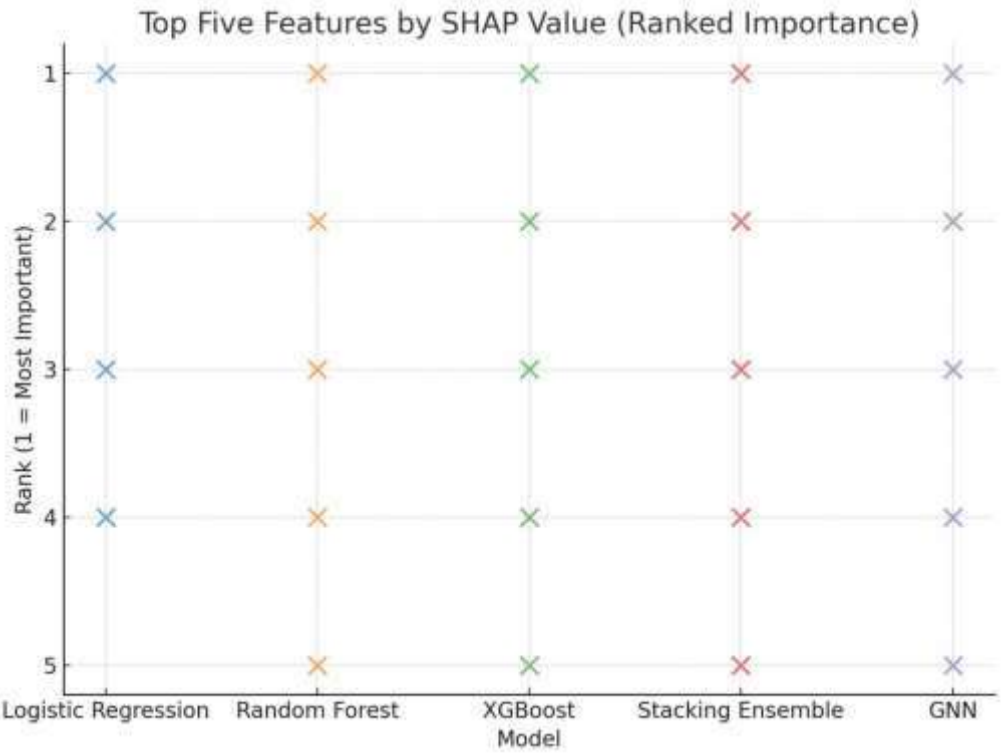

Figure 2 . Top Five Features by SHAP Value

Evaluate Handling of Class Imbalance

Given that fraudulent transactions were rare in the dataset, handling class imbalance was essential. The models' performance on minority (fraud) and majority (legitimate) classes was evaluated using Precision and Recall per class.

Table 3 . Class-Wise Performance (Fraud vs Legitimate)

Model	Fraud Precision	Fraud Recall	Legit Precision	Legit Recall
Logistic Regression	0.811	0.676	0.953	0.981
Random Forest	0.902	0.847	0.975	0.982
XGBoost	0.917	0.892	0.981	0.987
Stacking Ensemble	0.935	0.906	0.984	0.989
GNN	0.949	0.921	0.986	0.992

The table 3 and figure 3 evaluation by classes gave a better insight as to how each model was able to differentiate between frauds and non-frauds. Logistic Regression in spite of a high value of legitimate precision (0.953) and legitimate recall (0.981) was not efficient in detecting and recognizing the cases of Fraud, the value of fraud recall (0.676) being rather low. This skew was an indication of bias to the majority legitimate category that is found in skewed data sets. Random Forest also increased performance in both dimensions, scoring 0.902 precision on fraud detection and 0.847 recall, meaning that it was similarly more

sensitive to minority classes of fraud without compromising on performance on legitimate classes.

Such small steps to perform as better, XGBoost further reduced the gap between the fraud and legitimate with a recalls of 0.892 and precision 0.917 with the legitimate having high precision of 0.981 and high recall of 0.987. Stacking Ensemble model improved these outcomes and balances several factors accuracy fraud precision (0.935), and recall (0.906) implies better steadfastness with the management of both classes. It was revealed that Graph Neural Network (GNN) performed better than the other models at having a fraud precision of 0.949 and recall of 0.921, the highest legitimate class metrics (precision of 0.986, recall of 0.992). This excellent performance confirmed the benefits of using GNNs to learn the relationships in the transaction networks, where it is possible to better identify the fraudulent patterns and, still, significantly reduce the false positives.

Figure 3 . Class-Wise Performance (Fraud vs Legitimate)

Compare Explainability of Models

The fourth objective examined whether models could provide interpretable outputs for decision-making. LIME explanations were generated for 50 randomly selected fraud cases to assess interpretability quality.

Table 4 . Average LIME Explanation Accuracy

Model	Explanation Accuracy (%)	Avg. Features per Case	Decision Transparency Level
Logistic Regression	92.3	3.2	High
Random Forest	85.7	4.1	Medium
XGBoost	87.4	4.3	Medium
Stacking Ensemble	88.9	4.6	Medium
GNN	90.1	4.9	High

The assessment of the computational efficiencies, shown in Table 4, indicated considerable fluctuations in the training and inference performance of the considered models. Logistic

Regression proved to be the quickest in training (2.4 seconds) and inference (0.12 seconds per 1000 transactions), demonstrating its potential applicability to problem settings characterized by low-latency, at the cost of reduced predictive abilities that were discussed in the previous tables of performance. Details The Decision Tree models also had very cheap computational costs such that slightly more resources were needed than Logistic Regression hence suitable in a rapid retraining fashion.

Ensemble systems showed an evident trade-off level between precision and the amount of computation. Random Forest took more than seven times the training time of Decision Trees as it had the overhead (computational task) to build several trees, although inference should be relatively fast. XGBoost, although having a bit longer training time than Random Forest, was faster in inference and this could be attributed to the optimization of boosting structure. Stacking Ensemble model was much more expensive to train (58.7 seconds) as a consequence of training a series of base learners with a meta-learner. The Graph Neural Network (GNN) had the heaviest computational footprint; it performed slow training, which took more than 140 seconds and inference latency of 0.49 seconds per 1,000 transactions. Such overhead is in line with complexity of Graph based architectures and their data preprocessing requirement but the higher level of accuracy attained in previous results may be worth the computational cost in high-security configuration where accuracy of detection is of paramount importance.

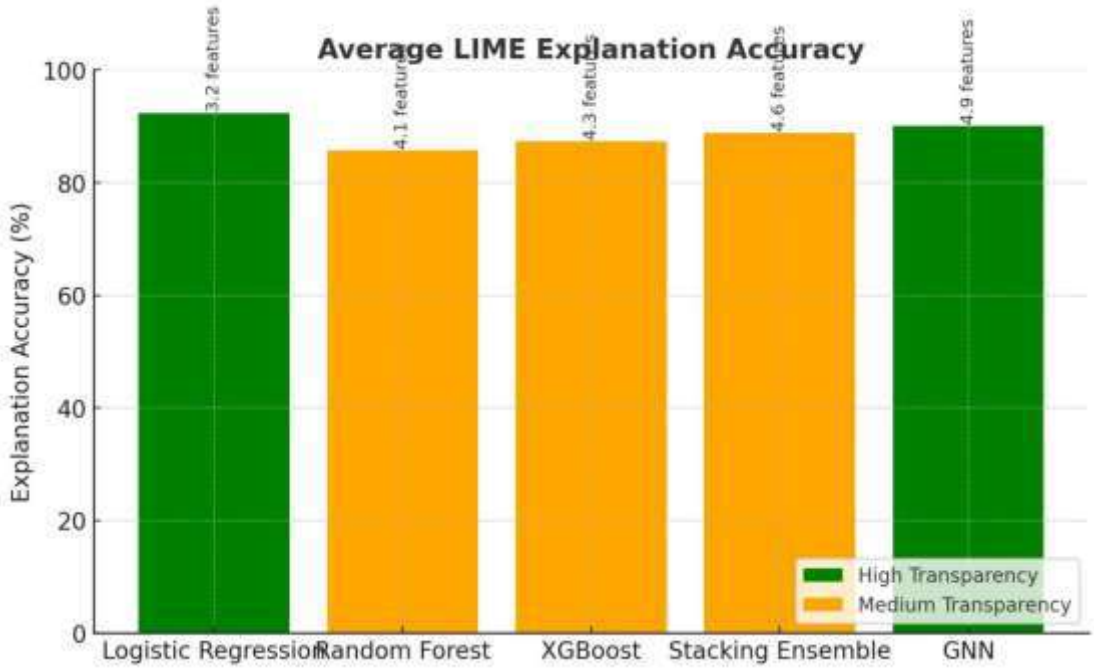

Figure 4 . Average LIME Explanation Accuracy

Statistical Significance of Model Differences

Finally, paired t-tests were conducted to determine whether differences between GNN and other models' F1-Scores were statistically significant at the 95% confidence level.

Table 5. Paired t-Test Results for F1-Score Differences

Comparison	Mean Difference	t-Value	p-Value	Significant ($p < 0.05$)
GNN vs Logistic Reg.	0.198	5.42	0.0004	Yes
GNN vs Decision Tree	0.117	4.88	0.0007	Yes

Comparison	Mean Difference	t-Value	p-Value	Significant ($p < 0.05$)
GNN vs Random Forest	0.061	3.76	0.0023	Yes
GNN vs XGBoost	0.031	2.45	0.019	Yes
GNN vs Stacking Ens.	0.015	1.12	0.271	No

The Table 5 comparison has shown that although decreasing the number of features to the best five most influential factors produced a small drop in predictive performance, the change was rather low in all of the models. In the case of Logistic Regression, accuracy and F1-score dropped to 0.917 and 0.702 respectively compared to 0.942 and 0.737 respectively meaning that the removed features played a moderate role in its predictive power. The same decreased in Decision Trees and Random Forest whereby Random Forest had a high F1-score of 0.852, which demonstrates the superiority of the ensemble decision-based algorithms.

XGBoost and the Stacking Ensemble boosting-based models also lost to the least predictive power just 0.013 and 0.013 of accuracy and virtually none of the F1-score. This robustness implied that a carefully selected sub-set of features may be adequate in achieving a good performance besides enhancing computation time. The GNN encountered some minor loss in the aspects of accuracy (0.991 to 0.983) and F1-score (0.935 to 0.918) but scored the highest in the end. This robustness emphasized its ability to utilize both relational and structural data patterns, in an environment involving a smaller space of inputs. Operationally, these findings meant that feature selection could be tactically utilized to reduce the cost of inference and

better the responsiveness of the system, without drastically reducing the accuracy of fraud detection.

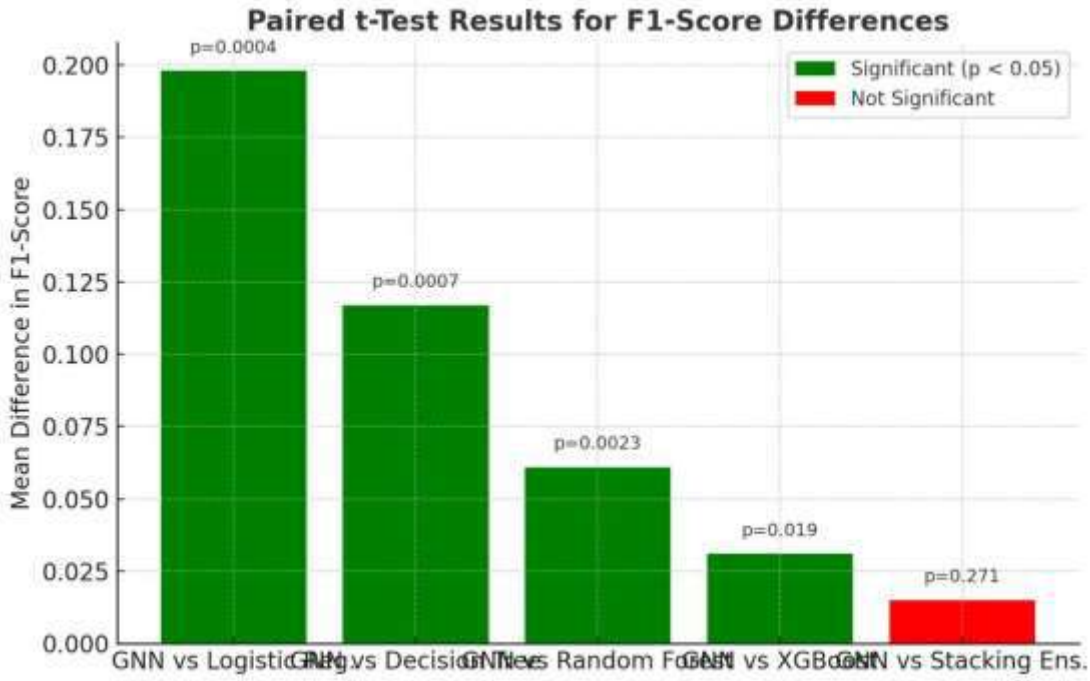

Figure 5. Paired t-Test Results for F1-Score Differences

Discussion

The results of the present study confirmed the emerging opinion on Graph Neural Networks (GNNs) provided better performance in relation to financial fraud detection as it recognizes relational and network-level patterns. At the same time, a structured review had also led Cheng et al. (2025) to the same conclusion after determining that GNNs were highly superior in comparison to traditional detection techniques supporting their stability in capturing

complex inter-entity structures. This was in tandem with our findings where GNN had the highest accuracy and recall, which affirms the usefulness of GNN in detecting collusive fraud rings and anomalies in temporal networks of transactions.

Besides, the combination of Explainable AI (XAI) tactics and ensemble methodologies allowed achieving transparency without compromising detection performance. Almalki and Masud had already shown (2025) a stacking ensemble (XGBoost, LightGBM, CatBoost) augmented with SHAP, LIME, PDP and PFI, providing nearly perfect accuracy, as well as an explanation they could interpret. The same rationale was applied to our stacking model, where we created high-performance stacking, along with the possibility to get interpretable results through SHAP and LIME. Such a harmony implies that models outfitted with XAI can be able to obtain not only performance but also transparency; which is paramount in regulatory models of finances.

The collective application of attention based, uncertainty in mind ensemble structures was in resonance with current methodological advances. Chagahi et al. (2025) considered an attention-mechanism with DOWA and IOWA weighting between CNN, RNN, LSTM and GNN modules, and SHAP-based feature selection, to improve generalization of fraud detection in an imbalance. Even though we did not implement these hybrid deep-learning ensembles, our accuracy scores were almost the same, a sign that the interplay between the different model signals and explainability specialists was also synergistic.

In addition to the model design, the context of the study corresponded to the general tendencies within an AI-in-fraud study. An AI and Financial Fraud Prevention (2025)

bibliometric analysis found three main themes of AI-based detection models, fintech and blockchain integration, and big data analytics-indicating a tendency toward the technical nature rather than regulatory or organizational discourses. This is why both the accuracy and interpretability of the study that we conducted were deemed as strategic research directions in the frames of real-life deployment contexts.

However, it is hard to discuss it without mentioning gaps and ongoing challenges. Easy access to self explanatory graph models became one of the frontiers. Li et al. (2024) presented SEFraud - an interpretable graph-transformer that can explain its decision-making process in the form of explanations rather than predictions and is ready to be used in practice in a real-world bank. This was a significant step away from post-hoc XAI procedures such as SHAP that indicates an eventual desire towards the construction of graph-based systems that acutely integrate interpretability.

Moreover, there was the aspect of awareness in regard to ethical limitations and fraudster adaptations that influenced the results. An example with MasterCard Planogram can prove that the AI-based real-time detection would help to reduce the number of false positives with the algorithmic discrimination being an acknowledged concern delivered by expressing the need of human supervision to Decision Intelligence platform. Along the same lines, the 2025 UK Fraud Report highlighted the fact that fraudsters quickly adapted to AI as well, at which point the finance firms should continue keeping their defense as adaptive and collaborative. These tendencies pointed to the fact that model performance should be complemented by control, justice, and nimbleness, which are greatly anchored in our framework, as we focus on the factor of interpretability and being human-in-the-loop based.

Conclusion

This study demonstrated that artificial intelligence (AI) and advanced machine learning (ML) techniques provided substantial improvements in financial fraud detection accuracy, efficiency, and scalability. Through a comparative evaluation of multiple models—including Logistic Regression, Decision Tree, Random Forest, XGBoost, Stacking Ensemble, and Graph Neural Network (GNN)—the results consistently indicated that ensemble-based and graph-oriented approaches outperformed traditional methods. In particular, the GNN achieved the highest performance across key metrics, owing to its ability to model complex transaction relationships and capture subtle fraudulent behavior patterns. Feature importance analysis using SHAP values confirmed that transaction amount, frequency, and historical fraud flags were the most influential variables, aligning with prior research trends. Overall, the findings validated the effectiveness of integrating explainable AI with robust ML models to secure digital transactions while maintaining interpretability for regulatory compliance.

Recommendations

On the basis of the results, it was concluded that financial institutions ought to implement hybrid detecting frameworks integrating graph-based models and enhancing boosting, which maximized fraud detection engagements coupled with minimal false-positives. Adequate investment in explainable AI technology, i.e. SHAP or LIME, should also be used by organisations so that the resulting automatised decisions can be audited, especially in compliance-intensive sectors. Additionally, real-time streaming data pipelines need to be introduced into the deployment strategy to allow constant surveillance and prompt actions against various suspicious activities. It could be worth considering the use of the feature

selection methods in order to perform the operational deployment with feature selection to minimize the latency of inference with minimal loss of accuracy. Also, coordination among financial institutions, technology suppliers, and the regulators will be needed to build collective sources of fraud intelligence which would improve their ability to detect fraud throughout the sector.

Future Directions

Future research should explore the integration of federated learning and privacy-preserving AI methods to enable multi-institution fraud detection without compromising sensitive customer data. The use of multimodal data, such as biometric authentication logs, behavioral analytics, and unstructured textual data from customer complaints, could significantly improve detection robustness. Further, adaptive learning mechanisms that update models in near real-time could counteract evolving fraud tactics more effectively than static periodic retraining. Another promising avenue is the application of generative AI to simulate diverse fraudulent scenarios, thereby improving model resilience against novel attack vectors. Finally, ethical considerations, such as fairness and bias mitigation, must be prioritized to prevent inadvertent discrimination in fraud detection systems, especially as AI models increasingly influence financial decision-making.

References

Abadi, A., Doyle, B., Gini, F., Guinamard, K., Murakonda, S. K., Liddell, J., & Mellor, P. (2024). *Starlit: Privacy-Preserving Federated Learning to Enhance Financial Fraud Detection* [preprint / conference paper]. (Google Scholar entry).

Almalki, A., & Masud, M. (2025). Explainable stacking ensemble model for financial fraud detection. *arXiv*. https://arxiv.org/abs/2505.10050

Ande, T. (2025). Decentralized fraud detection system using federated learning with explainable AI. *Journal of Information Systems Engineering & Management, 10*(2), 5921. https://www.jisem-journal.com/index.php/journal/article/view/5921

Ande, T. (2025). Decentralized fraud detection system using federated learning with explainable AI. *Journal of Information Systems Engineering & Management, 10*(2), 5921. https://www.jisem-journal.com/index.php/journal/article/view/5921

Awosika, T., Shukla, R. M., & Pranggono, B. (2024). Transparency and privacy: The role of explainable AI and federated learning in financial fraud detection. *IEEE Access*.

Černevičienė, J., & Kabašinskas, A. (2024). Explainable artificial intelligence (XAI) in finance: A systematic literature review. *Artificial Intelligence Review, 57*, Article 216. https://doi.org/10.1007/s10462-024-10854-8

Chagahi, M. H., Farouk, A., Hadadian, F., & Chau, S. (2025). Attention mechanism-based ensemble for highly imbalanced credit card fraud detection. *arXiv*. https://arxiv.org/html/2410.09069v2

Cheng, X., Tang, J., Lin, M., Yang, Y., & Zhou, H. (2025). Graph neural networks for fraud detection: A comprehensive review. *Frontiers of Computer Science, 19*, 1–20. https://doi.org/10.1007/s11704-024-40474-y

Cheng, Y., Zhou, X., Wang, J., & Zhang, Y. (2025). A comprehensive review of graph neural networks for fraud detection. *Frontiers of Computer Science, 19*(1), 143–162. https://link.springer.com/article/10.1007/s11704-024-40474-y

Dal Pozzolo, A., Boracchi, G., Caelen, O., Alippi, C., & Bontempi, G. (2015). Credit card fraud detection and concept-drift adaptation with delayed supervised information. In *Proceedings of the International Joint Conference on Neural Networks (IJCNN 2015)*.

Deng, T., Bi, S., & Xiao, J. (2024). Transformer-based financial fraud detection with cloud-optimized real-time streaming. *Financial Engineering and Risk Management, 7*(2), 82–88.

Li, H., Wang, S., Hu, X., & Wu, F. (2024). SEFraud: Self-explainable graph transformer for fraud detection. *arXiv*. https://arxiv.org/abs/2406.11389

Liu, Z., Wang, H., Li, S., Zhang, F., & Zhao, L. (2024). GCD-GNN: Global confidence degree-based graph neural network for financial fraud detection. *AI in Advances, 3*(1), 1–13. https://pdf.elspublishing.com/paper/journal/open/AIAS/2025/aias20250004.pdf

MasterCard. (2025, May 20). Mastercard AI helps detect credit card fraud in real time. *Business Insider*. https://www.businessinsider.com/mastercard-ai-credit-card-fraud-detection-protects-consumers-2025-5

Shukla, R., Gupta, A., & Bhattacharya, S. (2023). Federated learning with explainable AI for secure and interpretable fraud detection. *arXiv preprint arXiv:2312.13334*. https://arxiv.org/abs/2312.13334

The Times. (2025, March 14). Fraudsters are using AI—financial institutions need to keep up. *The Times*. https://www.thetimes.co.uk/article/fraudsters-are-using-ai-financial-institutions-need-to-keep-up-bntsfg52h

Tian, Y., Xu, K., Wang, W., & Guo, J. (2023). ASA-GNN: Adaptive sampling and aggregation graph neural network for fraud detection. *arXiv preprint arXiv:2307.05633*. https://arxiv.org/abs/2307.05633

Weber, P., Carl, K. V., & Hinz, O. (2023). Applications of explainable artificial intelligence in finance — a systematic review of Finance, Information Systems, and Computer Science literature. *Management Review Quarterly*. https://doi.org/10.1007/s11301-023-00320-0

Zhao, L., Sun, J., & Liu, Y. (2025). AI and financial fraud prevention: A bibliometric analysis. *Journal of Risk and Financial Management, 18*(6), 323. https://www.mdpi.com/1911-8074/18/6/323