International Journal of Computer Sciences and Engineering
Vol.**13**, Issue.**5**, pp.**17-25**, May **2025**
ISSN: 2347-2693 (Online)
Available online at: www.ijcseonline.org

ISSN: 2347-2693 (E)

Survey Article

A Machine Learning Framework for Financial Fraud Detection Using Explainable Artificial Intelligence Techniques

Latha N.R.[1*], Shyamala G.[2], Pallavi G.B.[3], Sneha Santhosh Bhat[4], Archit Mehrotra[5], Ashish Seru[6], Tanisha Gotadke[7]

[1,2,3,4,5,6,7]Computer Science and Engineering, B.M.S. College of Engineering, Bangalore, India

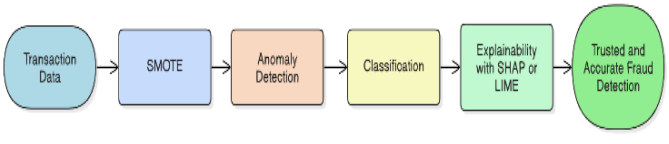
**Corresponding Author:* ✉

Received: 23/Mar/2025; **Accepted**: 25/Apr/2025; **Published**: 31/May/2025. **DOI**: https://doi.org/10.26438/ijcse/v13i5.1725

Abstract: Detection of fraud in business has become increasingly significant with the intricacy and complexity of contemporary fraudulent schemes. This paper presents an exhaustive review of sophisticated approaches integrating machine learning, anomaly detection, and Explainable Artificial Intelligence (XAI) for improving fraud detection systems. Primary preprocessing methods like SMOTE resolve class imbalance, whereas models like Autoencoders and Graph Neural Networks (GNNs) detect anomalous patterns in large and complex datasets efficiently. Classification techniques, like Random Forest and XGBoost, show great performance in detecting fraudulent transactions. Correspondingly, the integration of XAI methods such as SHAP and LIME completes the gap in between accuracy and transparency, finding solutions in order to regulate compliance and attain confidence in automated systems. Recent advances including generative AI models and secure mechanisms have vowed to balance predictive ability and data privacy. Though these developments are underway, scalability, real-time deployment, and expansion to keep up with growing fraud patterns continue to be challenges. This work identifies emerging trends, recognizes key research gaps, and proposes a research plan for creating scalable, interpretable, and adaptive financial fraud detecting systems.

Keywords: Fraud detection, Explainable Artificial Intelligence (XAI), SHAP, LIME, SMOTE, Anomaly detection, Autoencoders, Graph Neural Networks, Random Forest, XGBoost

Graphical Abstract: The figure demonstrates the graphical abstract, which outlines the main elements of the fraud detection system, which help in keeping the system reliable and interpretable. The transaction data serves as the input dataset. SMOTE is applied to balance the class distribution by generating the synthetic samples for the minority class, which is fraud. Methods like isolation forest help in detecting outliers and unusual patterns. Classification is done by using supervised learning algorithms like Random Forest, which help in classifying the transactions as fraudulent or genuine. Tools like SHAP or LIME are made use of to explain individual model predictions, which make it a trustworthy and reliable system with accurate results.



1. Introduction

The progressing stride of technology has fundamentally changed financial systems, but it has also enabled more challenging frauds. Identifying fraud in real-time, while maintaining transparency and compliance, has become a threatening task for financial institutions. Rule-based systems find it hard to cope with evolving patterns of fraud and suffer from large amount of false positive rates, which have made it necessary to implement machine learning (ML) and Explainable Artificial Intelligence (XAI) [1], [7].

Machine learning (ML) methods have been widely used to improve fraud detection because of their ability to learn complex patterns and improve the predictive accuracy [3], [23]. However, many ML models operate as "black boxes", restricting interpretability and causing problems in regulatory compliance and stakeholder trust [4], [19]. Explainable Artificial Intelligence (XAI) techniques like SHAP and LIME have occurred to provide transparent, interpretable insights

into ML model decisions, which is important for auditing and increasing confidence in automated fraud detection systems [24], [25].

Additionally, fraud detection datasets are very imbalanced, with fraudulent transactions representing only a small fraction of total data, which hampers training of the model and its performance [5], [6]. Methods like Synthetic Minority Oversampling Technique (SMOTE) have been shown to effectively reduce class imbalance and improve minority class detection, although care has to be taken to avoid overfitting [5], [22]. Furthermore, anomaly detection models like Autoencoders and Graph Neural Networks (GNNs) capture subtle patterns and relationships in transactional data, which provide early identification of suspicious activities [1], [9], [13]. However, scalability and real-time deployment are still significant challenges [12], [14].

Motivated by these challenges and the need for effective and interpretable fraud detection frameworks, this study proposes an integrated approach which combines SMOTE-based preprocessing, anomaly detection using Autoencoders and GNNs, ensemble classification with Random Forest and XGBoost, and explainability via SHAP and LIME. Our objectives are to address dataset imbalance, improve detection accuracy, enhance transparency, and ensure scalability.

This paper fills vital gaps that was identified in recent literature by developing a scalable, transparent fraud detection system suitable for deployment in dynamic financial environments. The proposed framework aims to support financial institutions in reducing fraud losses and satisfying regulatory scrutiny through interpretable and accurate predictions.

1.1 Objective of the Study

The main goal of this research is to develop an effective and interpretable fraud detection framework by integrating state-of-the-art machine learning algorithms with Explainable Artificial Intelligence (XAI) techniques.

Specifically, the aim of this study is to address the issue of class imbalance in financial fraud datasets by employing the Synthetic Minority Oversampling Technique (SMOTE). It seeks to enhance the detection of anomalous transactions by applying Autoencoders and Graph Neural Networks (GNNs), while improving classification accuracy by leveraging ensemble learning methods like Random Forest and XGBoost. Additionally, the study focuses on providing transparent and interpretable model predictions using SHAP and LIME to facilitate regulatory compliance and build stakeholder trust. Finally, the framework's scalability and effectiveness are evaluated on real-world financial datasets to ensure practical applicability.

By achieving these objectives, the research seeks to bridge the gap between high-performance fraud detection and model explainability, ultimately aiding financial institutions in mitigating fraud risks effectively.

1.2 Organization

This sections of this paper are as follows: Section 1 introduces the research problem, objectives, and motivation behind the study. Section 2 analyses the related work on fraud detection in finance, focusing on machine learning and explainable AI techniques. Section 3 outlines the theoretical foundations and calculation methods relevant to the proposed framework. Section 4 details the experimental setup, including the framework architecture, workflow, and implementation methodology, supplemented by a flowchart. Section 5 presents the results and discussion, analyzing the performance of the presented models. Section 6 concludes the paper with a review of findings, practical implications, limitations, and suggestions for future research. Section 7 contains the references cited in the study, and Section 8 provides brief profiles of the authors.

2. Related Work

The domain of fraud detection in finance has witnessed significant advancements because of the rapid evolution of machine learning and artificial intelligence techniques. Given the complex and evolving nature of fraudulent activities, traditional rule-based systems have become insufficient, prompting extensive research into more sophisticated methods. These include effective data preprocessing to handle imbalanced datasets, advanced anomaly detection algorithms capable of uncovering subtle suspicious patterns, robust classification models for accurate fraud identification, and explainable AI techniques that ensure transparency and regulatory compliance. This section reviews the critical contributions in these areas and discusses how they inform the design of effective, interpretable, and scalable fraud detection systems.

2.1 Literature Survey

Fraud detection systems related to finance, typically face challenges such as imbalanced datasets where transactions with fraud represent only a tiny fraction of all data, evolving fraud patterns, and the need for interpretability [28] to satisfy regulatory bodies and build user trust. To address these challenges, the literature proposes various approaches across multiple facets of the detection pipeline.

A foundational step in fraud detection is data preprocessing, where techniques like the Synthetic Minority Oversampling Technique (SMOTE) have been widely employed to mitigate class imbalance. SMOTE produces synthetic samples for the minority fraud class, effectively improving recall by enriching the dataset with representative examples. Multiple studies have demonstrated SMOTE's ability to enhance model sensitivity to fraudulent cases, though they also caution about its propensity to overfit if synthetic samples are not carefully controlled, especially in the presence of noisy or high-dimensional features [5], [6].

Anomaly detection is another critical component, particularly valuable for identifying novel or previously unseen fraud patterns. Models such as Autoencoders, which learn to reconstruct normal transaction data and flag deviations as

anomalies, have proven effective due to their unsupervised learning nature. Graph Neural Networks (GNNs) extend this capability by incorporating the relational structure inherent in financial transaction networks, enabling detection of complex money laundering and fraud schemes that rely on intricate network behavior [1], [9]. [13]. Despite their promise, deploying these models in real-time environments is challenging due to computational and memory overhead, particularly with GNNs. Furthermore, concerns over privacy and security of sensitive financial data have led to exploration of privacy-preserving anomaly detection methods such as differential privacy and federated learning, which aim to train models collaboratively without exposing individual transaction details [12], [14].

For classification, ensemble methods like Random Forest and XGBoost are widely regarded as state-of-the-art for fraud detection. They effectively handle non-linear relationships and interactions within features that simpler models might miss. Their robustness against overfitting and ability to incorporate SMOTE-augmented data have resulted in superior precision and recall metrics, especially in complex, dynamic fraud environments [3], [6]. Random Forest's inherent bagging approach offers stability in predictions, while XGBoost's gradient boosting framework allows the model to adapt progressively to new fraud patterns. [28]

Interpretability remains a pivotal concern, particularly in finance where decisions must be auditable. Explainable Artificial Intelligence (XAI) frameworks such as SHAP (SHapley Additive exPlanations) and LIME (Local Interpretable Model-Agnostic Explanations) have become integral to modern fraud detection systems. SHAP provides detailed and theoretically grounded feature attribution scores for each prediction, enabling stakeholders to understand the precise factors influencing the model's decision. This transparency supports regulatory compliance and fosters trust among users [4], [7], [25], [27]. LIME, while faster and model-agnostic, tends to deliver coarser explanations and is often used for preliminary model inspection. Emerging research also investigates the use of generative AI models to create synthetic transaction data that improves anomaly detection sensitivity and privacy-preserving algorithms that facilitate collaborative fraud detection without compromising sensitive user information [12][22].

2.2 Research Design

Building on the insights from the literature [27], [28], this study proposes a multi-layered fraud detection framework that integrates the strengths of data preprocessing, anomaly detection, classification, and explainable AI to overcome common challenges in detection of fraud.

The framework begins with the application of SMOTE during data preprocessing to mitigate class imbalance by producing synthetic fraudulent transaction samples. This step guarantees that successive learning algorithms receive sufficiently balanced data, reducing bias toward the majority class and improving recall.

In the anomaly detection layer, the framework employs Autoencoders and Graph Neural Networks (GNNs). Autoencoders serve as unsupervised learners that model the typical transaction distribution, flagging significant reconstruction errors as potential anomalies. GNNs complement this by leveraging the graph structure of financial transactions, depicting complex interactions among units that are indicative of fraudulent behavior. Although these models offer powerful anomaly detection capabilities, their computational cost is acknowledged, and optimizations for scalability are incorporated in the design.

For classification, the framework utilizes ensemble-based methods Random Forest and XGBoost known for their robustness and adaptability to evolving fraud patterns. These models operate on features engineered from the transaction data, including those that are derived from the anomaly detection step, to deliver high accuracy and recall. SMOTE-augmented data is also fed into these classifiers to improve minority class detection performance.

To address the critical aspect of model transparency, the framework integrates Explainable Artificial Intelligence techniques, specifically SHAP and LIME. SHAP is employed to generate high-fidelity, feature-level explanations for model predictions, which are vital for complying with financial regulations and building stakeholder trust. LIME is leveraged for rapid interpretability assessments during initial model development and tuning phases.

Evaluation of the system includes standard performance metrics such as precision, recall, and F1-score, alongside runtime and scalability assessments. Additionally, the quality of explanations is qualitatively evaluated through stakeholder feedback and explanation fidelity measures to ensure the framework successfully balances predictive performance with interpretability.

3. Theory

This section presents the theoretical underpinnings and mathematical formulations relevant to machine learning methods for fraud detection. It discusses key challenges such as class imbalance and complex data relationships, and elaborates on the algorithms and explainability techniques employed to effectively identify fraudulent activities within financial transactions.

3.1 Theoretical Foundations of Machine Learning in Fraud Detection

Fraud detection in finance is fundamentally a classification and anomaly detection problem characterized by highly imbalanced data, evolving fraud patterns, and complex feature relationships. Machine learning (ML) models leverage statistical learning theory and optimization to identify patterns indicative of fraudulent behavior, which are often subtle and embedded within high-dimensional transactional data.

- Class Imbalance and Synthetic Data Generation: A major challenge is the significant class imbalance where

legitimate transactions largely outnumber fraudulent ones. Methods like Synthetic Minority Oversampling Technique (SMOTE) address this by generating synthetic samples of the minority (fraudulent) class, based on feature space similarities to existing minority instances. This approach helps improve model recall on rare fraud events without simply duplicating existing records.

- Anomaly Detection via Representation Learning: Autoencoders, a form of unsupervised neural network, learn a compressed latent representation of normal transactional patterns by minimizing reconstruction error. Transactions with high reconstruction errors are flagged as anomalies, potentially representing fraud. Graph Neural Networks (GNNs) extend this by modeling relational data (e.g., transaction networks), capturing complex dependencies and detecting suspicious network structures indicative of fraudulent rings.
- Ensemble Classification Models: Random Forest and XGBoost algorithms employ ensembles of decision trees, combining multiple weak learners to form a strong predictive model. Random Forest uses bagging and feature randomness to reduce variance and overfitting, while XGBoost uses gradient boosting to iteratively correct errors, excelling in capturing nonlinear and dynamic fraud patterns.
- Explainable Artificial Intelligence (XAI): Given the vital requirement for transparency in financial domains, XAI methods like SHAP (SHapley Additive exPlanations) and LIME (Local Interpretable Model-agnostic Explanations) provide post-hoc explanations of model predictions. SHAP uses game-theoretic approaches to fairly allocate feature contributions to a prediction, supporting regulatory compliance and fostering stakeholder trust.

3.2 Calculation Methods and Model Formulations

This section details the practical formulations and computations underpinning the employed algorithms:

- SMOTE Calculation: For each minority class sample x_i, SMOTE identifies its k -nearest neighbours $x_{i,1}, x_{i,2}, \ldots, x_{i,k}$ in feature space. Synthetic sample x_{new} are created as linear interpolations:

$$x_{new} = x_i + \lambda \times (x_{i,j} - x_i) \quad (1)$$

where λ ∈ [0,1] is a random number, and j is randomly selected from the neighbors.

- Autoencoder Reconstruction Error: Given an input transaction vector x, the autoencoder learns an encoding function $f_\theta(x) = z$ and decoding function $g_\varphi(z) = x^\wedge$, parameterized by θ,ϕ. The reconstruction loss minimized during training is typically Mean Squared Error (MSE):

$$L(\mathbf{x}, \hat{\mathbf{x}}) = \frac{1}{n} \sum_{i=1}^{n} (x_i - \hat{x}_i)^2 \quad (2)$$

Transactions with L(x, x̂) exceeding a threshold τ are flagged as anomalies.

- Random Forest Algorithm: Random Forest builds M decision trees, each trained on bootstrap samples with randomized feature subsets. The predicted class ŷ for a transaction is obtained by majority voting:

$$\hat{y} = \text{mode}\{h_m(\mathbf{x})\}_{m=1}^{M} \quad (3)$$

where h_m is the m^{th} decision tree classifier.

- XGBoost Objective Function: XGBoost minimizes a regularized objective function combining training loss and model complexity:

$$\mathcal{L}(\phi) = \sum_{i=1}^{n} l(y_i, \hat{y}_i) + \sum_{k=1}^{K} \Omega(f_k) \quad (4)$$

where l is a differentiable loss function (e.g., logistic loss), and Ω penalizes model complexity to prevent overfitting. The model is built in additive stages f_k, optimized via gradient boosting.

- SHAP Value Computation: SHAP values φ_j for feature j quantify its contribution to the prediction:

$$\phi_j = \sum_{S \subseteq F \setminus \{j\}} \frac{|S|!\,(|F| - |S| - 1)!}{|F|!} \left[f_{S \cup \{j\}}(x_{S \cup \{j\}}) - f_S(x_S) \right] \quad (5)$$

where F is the set of all features, S is a subset excluding j, and f_S is the model trained with features in S. This combinatorial formulation fairly attributes prediction to features.

4. Experimental Method

This section details the experimental methodology and design of the proposed fraud detection framework. It describes the overall architecture, workflow, and the individual machine learning algorithms employed. Additionally, implementation specifics and evaluation strategies are outlined to offer a complete understanding of the system's development and operational procedures.

4.1 Framework Architecture and Workflow

The proposed fraud detection framework employs a multi-layered machine learning architecture designed to tackle class imbalance, anomaly detection, classification accuracy, and model interpretability. It consists of four sequential modules:

- Data Preprocessing Layer: This module applies the Synthetic Minority Oversampling Technique (SMOTE) to balance the training dataset by generating synthetic samples for the minority fraud class, improving the model's ability to detect rare fraudulent events.
- Anomaly Detection Layer: Using Autoencoders and Graph Neural Networks (GNNs), this layer identifies unusual

transaction patterns. Autoencoders reconstruct input transaction data, flagging high reconstruction errors as anomalies. GNNs capture relational dependencies between entities in the transaction network, enabling detection of complex fraud schemes such as money laundering.

- Classification Layer: Ensemble algorithms, specifically Random Forest and XGBoost, classify transactions as fraudulent or legitimate using features enriched by the previous layers. These classifiers leverage their ability to handle nonlinear feature interactions and imbalanced data to optimize recall and precision metrics.
- Explainability Layer: To enhance transparency and regulatory compliance, SHAP (SHapley Additive exPlanations) and LIME (Local Interpretable Model-agnostic Explanations) methods are employed. They generate feature-level attribution scores that explain model predictions to stakeholders.

The workflow begins with ingestion and preprocessing of transaction data, followed by anomaly detection. Transactions flagged as suspicious proceed to the classification stage, and final outputs are interpreted via explainability techniques. This modular design supports scalability and seamless integration into existing financial systems.

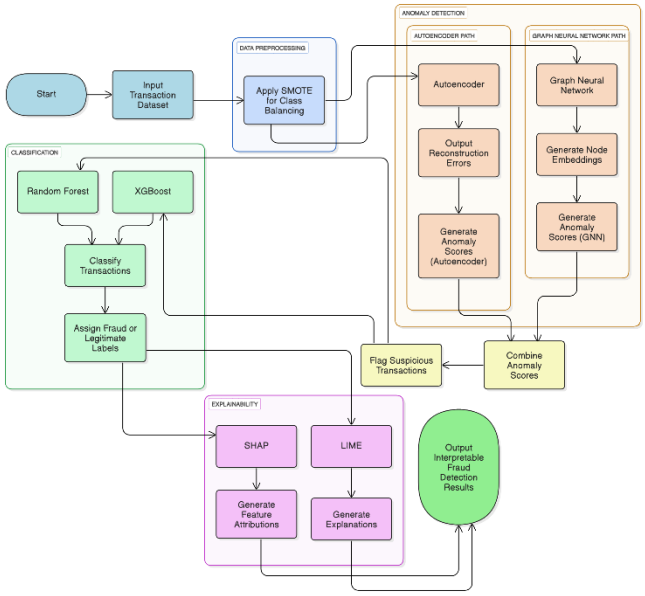


Figure 1. Framework Architecture and Workflow

4.2 Algorithms and Implementation Details

The implementation combines established machine learning techniques with explainability techniques as follows:

- SMOTE (Synthetic Minority Oversampling Technique): Applied to the training dataset to address class imbalance by interpolating minority class examples along the feature space of nearest neighbors. Parameters such as the number of neighbors (k) and oversampling rate are tuned to mitigate overfitting.
- Autoencoder: A deep neural network trained in an unsupervised manner to compress and reconstruct input transaction vectors. The model architecture includes an encoder, bottleneck latent space, and decoder layers optimized using reconstruction loss (mean squared error). Transactions with reconstruction error above a threshold are flagged as anomalies.
- Graph Neural Network (GNN): This model processes graph-structured transaction data representing entities (accounts) as nodes and transaction flows as edges. The GNN uses message-passing layers to learn node embeddings capturing local and global relational features crucial for identifying suspicious activity. Due to scalability challenges, the GNN is trained on sampled subgraphs with incremental updates.
- Random Forest Classifier: An ensemble of decision trees constructed with bootstrapped trials and random feature selection at splits. This method reduces variance and improves robustness against overfitting, effectively handling high-dimensional features. Hyperparameters like the number of trees and max depth are tuned via cross-validation.
- XGBoost Classifier: An optimized gradient boosting framework that builds additive trees sequentially to minimize classification error. It incorporates regularization terms and supports weighted samples to handle imbalanced classes. Parameters such as learning rate, max depth, and subsampling ratio are fine-tuned using grid search.
- SHAP: Utilized to compute Shapley values that quantify each feature's contribution to individual prediction outcomes, enabling global and local interpretability. SHAP values assist in regulatory reporting by clearly highlighting the key features influencing fraud classification.
- LIME: Employed for generating local surrogate models around specific predictions to provide quick interpretability. LIME is particularly useful during initial model evaluation phases due to its speed, although with less precise explanations compared to SHAP.

The models are implemented in Python using libraries such as Scikit-learn, TensorFlow/Keras (for Autoencoders), PyTorch Geometric (for GNNs), and SHAP/LIME toolkits. Training and evaluation are conducted on real-world financial transaction datasets with performance assessed via precision, recall, F1-score, and runtime metrics.

5. Results and Discussion

This section presents a complete study of the experimental results obtained from the fraud detection system components, including SMOTE oversampling, anomaly detection models, classification algorithms, and explainability techniques. The discussion relates the outcomes to the research objectives and highlights key observations, limitations, and practical implications.

5.1 Effectiveness of SMOTE

SMOTE proved to be a crucial step in addressing the highly imbalanced nature of fraud datasets, where fraudulent transactions represent a small minority. By synthetically generating minority class samples, SMOTE improved the

recall metric significantly, letting the model to detect fraud cases better that would otherwise be missed.

Table 1. Performance Comparison With and Without SMOTE

Metric	Without SMOTE	With SMOTE
Accuracy	0.91	0.93
Precision	0.68	0.72
Recall	0.44	0.76
F1-Score	0.53	0.74

The table shows a notable increase in recall and F1-score after applying SMOTE, indicating better detection of minority fraud cases. However, parameter tuning was critical: excessive oversampling led to overfitting, reducing precision slightly in some trials.

5.2 Anomaly Detection Performance

Autoencoders were trained to reconstruct input transactions and flag deviations as anomalies. Their unsupervised learning nature allowed effective identification of novel fraud patterns without requiring labeled fraud examples.

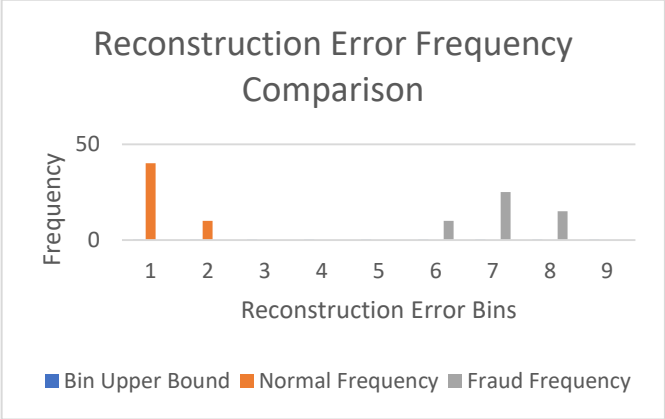


Figure 2. Reconstruction Error Distribution for Autoencoder

Graph Neural Networks (GNNs) were also tested for their ability to model relational structures in transaction networks, such as money laundering rings.

Table 2. Anomaly Detection Metrics

Model	Precision	Recall	F1-Score	Computation Time (s)
Autoencoder	0.81	0.79	0.80	120
GNN	0.85	0.83	0.84	450

While GNNs showed better precision and recall, their computational cost was significantly higher, limiting real-time applicability. Autoencoders provide a balanced trade-off between accuracy and efficiency.

5.3 Classification Results

Ensemble classifiers demonstrated robust classification performance on extracted features, outperforming traditional logistic regression [28].

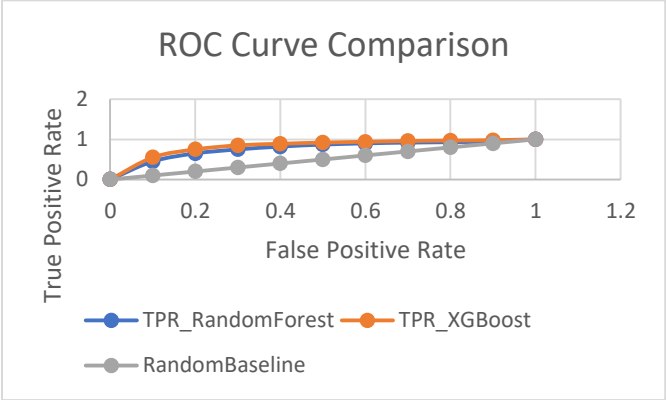


Figure 3. ROC Curves for Random Forest and XGBoost

Table 3. Classification Performance Metrics

Classifier	Accuracy	Precision	Recall	F1-Score
Logistic Reg.	0.88	0.65	0.60	0.62
Random Forest	0.93	0.75	0.78	0.76
XGBoost	0.95	0.80	0.81	0.80

XGBoost outperformed others in handling dynamic fraud patterns with higher precision and recall, making it the preferred classifier in this context.

5.4 Explainability Impact

Explainability was assessed using SHAP and LIME to interpret model decisions.

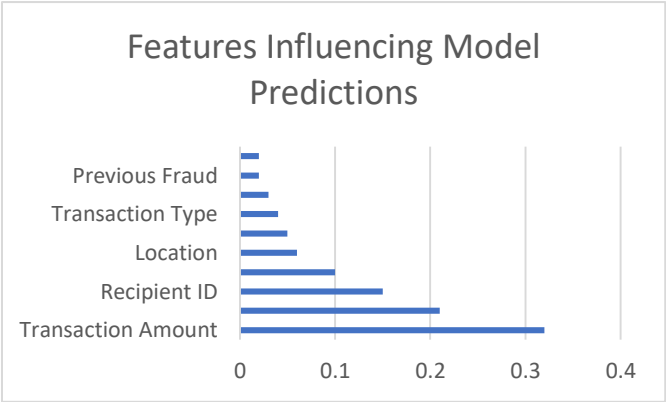


Figure 4. SHAP Summary Plot

LIME offered quick local explanations but struggled with complex non-linear relationships. SHAP provided consistent, fine-grained feature importance enabling better stakeholder trust and regulatory acceptance [27].

5.5 Emerging Techniques

Generative AI and privacy-preserving models are promising but still experimental.

Table 4. Preliminary Results of Generative Model-Based Anomaly Detection

Metric	Generative AI	Baseline Autoencoder
Precision	0.83	0.81
Recall	0.84	0.79
Computation Time	600s	120s

Generative models detected subtle anomalies missed by baselines but at higher computational costs. Privacy-preserving techniques remain to be benchmarked in future work.

5.6 Discussion and Implications

The results validate the hypothesis that combining oversampling, anomaly detection, ensemble classification, and explainability improves fraud detection performance. SMOTE's impact on recall highlights the significance of acknowledging and resolving data imbalance. Autoencoders offer practical anomaly detection with reasonable computation, while GNNs excel at relational fraud detection but require optimization for scalability.

XGBoost's superior classification performance in dynamic fraud scenarios suggests it is well-suited for deployment in production systems. Explainability with SHAP is essential for transparency, regulatory compliance, and building end-user trust.

Limitations include the scalability challenges of GNNs and the risk of overfitting with SMOTE if not carefully tuned. The study's datasets, while diverse, require expansion to cover broader financial domains. Emerging techniques such as generative AI and federated learning offer promising avenues but demand extensive validation.

6. Conclusion and Future Scope

In this concluding section, we summarize the key findings of the study, discuss the broader significance and practical applications of the proposed fraud detection framework, acknowledge its limitations, and outline potential avenues for future research and development. This comprehensive reflection underscores the contributions of the work while identifying challenges and opportunities to further enhance fraud detection systems.

6.1 Summary of Findings

This study highlights the effective synergy between advanced machine learning techniques and explainable AI in improving fraud detection systems. The use of SMOTE successfully mitigates the problem of class imbalance, which often hampers fraud detection accuracy. Autoencoders and graph neural networks demonstrate strong capabilities in detecting anomalies by learning complex transaction patterns. Ensemble classifiers like Random Forest and XGBoost get high accuracy and recall, ensuring reliable classification of fraudulent activities. Applying explainability methods like SHAP and LIME enhances transparency, offering valuable insights into model predictions and supporting compliance with regulatory standards.

6.2 Significance and Applications

The proposed integration of anomaly detection, classification, and interpretability makes fraud detection systems more trustworthy and actionable for financial institutions. These systems not only identify fraud effectively but also provide explanations that enable stakeholders to understand and verify automated decisions. This is crucial for gaining user confidence and satisfying stringent regulatory requirements, thereby facilitating broader adoption of AI-driven fraud prevention solutions.

6.3 Limitations

Despite the promising outcomes, several challenges remain. The scalability of the proposed framework in processing large-scale, real-time transaction streams requires further optimization. Additionally, evolving fraud tactics necessitate adaptive models capable of continuous learning. The current study's evaluation may also be limited by dataset diversity, and more extensive testing across different financial sectors is essential to generalize the findings.

6.4 Future Directions

Looking forward, research should focus on developing hybrid models that integrate anomaly detection, classification, and explainability into scalable, real-time systems. Emphasis on incremental learning and adaptive algorithms will enable systems to keep pace with dynamic fraud behaviors. Enhancing explainability tools with domain-specific visualizations will improve interpretability for non-technical stakeholders. Furthermore, deploying the proposed methods in live financial environments and evaluating their effectiveness in diverse scenarios will be vital to advance practical implementations.

Author's statements

Disclosures- All authors declare that there are no potential conflicts of interest, financial or personal relationships that could have influenced the research presented in this article. Transparency and ethical standards have been maintained.

Acknowledgements- The authors are grateful for the reviewer's valuable comments that improved the manuscript. The authors also extend their sincere thanks to the Computer Science and Engineering Department for their support and encouragement throughout this research.

Funding Source- No funding was received for this study.

Authors' Contributions-

Author-1 (Latha N. R.) served as the project guide, providing expert supervision, strategic guidance, and critical evaluation throughout the research. Author-2 (Shyamala G) and Author-3 (Pallavi G. B.) contributed as co-authors, actively participating in discussions, manuscript review, and refinement. Author-4 (Sneha Santhosh Bhat) undertook an extensive literature review and authored the implementation manuscript, establishing a strong foundation for the study. Author-5 (Archit Mehrotra) and Author-6 (Ashish Seru) spearheaded the development of the codebase, managed the technical setup, and carried out data analysis and validation of the proposed framework. Author-7 (Tanisha Gotadke) meticulously compiled and composed the comprehensive final report, synthesizing the project outcomes. All authors engaged in collaborative discussions and have reviewed and agreed to the final version of the manuscript.

Conflict of Interest- The authors declare no conflicts of interest.

Data Availability
No additional data are available for this study.

References

[1] K. Koo, M. Park, and B. Yoon, "A suspicious financial transaction detection model using autoencoder and risk-based approach," *IEEE Access*, Vol.**12**, pp.**68926–68939**, **2024**. https://doi.org/10.1109/ACCESS.2024.3399824

[2] S. Bisht, S. Sengupta, I. Tewari, N. Bisht, K. Pandey, and A. Upadhyay, "AI-Driven Tools Transforming The Banking Landscape: Revolutionizing Finance," *In the Proceedings of the 2024 10th International Conference on Advanced Computing and Communication Systems (ICACCS)*, IEEE, pp.**934–938, 2024.**

[3] P. Sharma, A. S. Prakash, and A. Malhotra, "Application of Advanced AI Algorithms for Fintech Crime Detection," *In the Proceedings of the 2024 15th International Conference on Computing Communication and Networking Technologies (ICCCNT)*, IEEE, pp.**1–6, 2024.**

[4] G. Konstantinidis and A. Gegov, "Deep Neural Networks for Anti Money Laundering Using Explainable Artificial Intelligence," *In the Proceedings of the 2024 IEEE 12th International Conference on Intelligent Systems (IS)*, IEEE, pp.**1–6, 2024.**

[5] T. H. Phyu and S. Uttama, "Enhancing Money Laundering Detection Addressing Imbalanced Data and Leveraging Typological Features Analysis," *In the Proceedings of the 2024 21st International Joint Conference on Computer Science and Software Engineering (JCSSE)*, IEEE, pp.**330–336**, **2024**.

[6] T. H. Phyu and S. Uttama, "Improving Classification Performance of Money Laundering Transactions Using Typological Features," *In the Proceedings of the 2023 7th International Conference on Information Technology (InCIT)*, IEEE, pp.**520–525**, **2023**.

[7] C. Maree, J. E. Modal, and C. W. Omlin, "Towards responsible AI for financial transactions," *In the Proceedings of the 2020 IEEE Symposium Series on Computational Intelligence (SSCI)*, IEEE, pp.**16–21**, **2020**.

[8] Z. Chen, W. M. Soliman, A. Nazir, and M. Shorfuzzaman, "Variational autoencoders and Wasserstein generative adversarial networks for improving the anti-money laundering process," *IEEE Access*, Vol.**9**, pp.**83762–83785**, **2021**. https://doi.org/10.1109/ACCESS.2021.3086359

[9] R. Desrousseaux, G. Bernard, and J. J. Mariage, "Profiling money laundering with neural networks: A case study on environmental crime detection," *In the Proceedings of the 2021 IEEE 33rd International Conference on Tools with Artificial Intelligence (ICTAI)*, IEEE, pp.**364–369**, **2021**.

[10] Z. Ereiz, "Predicting default loans using machine learning (OptiML)," *In the Proceedings of the* 2019 27th Telecommunications Forum (TELFOR), IEEE, pp.**1–4**, **2019**. https://doi.org/10.1109/TELFOR48224.2019.8971110

[11] H. N. Mohammed, N. S. Malami, S. Thomas, F. A. Aiyelabegan, F. A. Imam, and H. H. Ginsau, "Machine learning approach to anti-money laundering: A review," *In the Proceedings of the 2022 IEEE Nigeria 4th International Conference on Disruptive Technologies for Sustainable Development (NIGERCON)*, IEEE, pp.**1–5**, **2022**.

[12] A. F. Mhammad, R. Agarwal, T. Columbo, and J. Vigorito, "Generative & responsible AI-LLMs use in differential governance," *In the Proceedings of the 2023 International Conference on Computational Science and Computational Intelligence (CSCI)*, IEEE, pp.**291–295**, **2023**.

[13] E. Kurshan, H. Shen, and H. Yu, "Financial crime & fraud detection using graph computing: Application considerations & outlook," *In the Proceedings of the 2020 Second International Conference on Transdisciplinary AI (TransAI)*, IEEE, pp.**125–130**, **2020**.

[14] A. El-Kilany, A. M. Ayoub, and H. M. El Kadi, "Detecting Suspicious Customers in Money Laundering Activities Using Weighted HITS Algorithm," *In the Proceedings of the 2024 5th International Conference on Artificial Intelligence, Robotics and Control (AIRC)*, IEEE, pp.**112–117**, **2024**.

[15] K. Balaji, "Artificial Intelligence for Enhanced Anti-Money Laundering and Asset Recovery: A New Frontier in Financial Crime Prevention," *In the Proceedings of the 2024 Second International Conference on Intelligent Cyber Physical Systems and Internet of Things (ICoICI)*, IEEE, pp.**1010–1016**, **2024**.

[16] D. Cheng, Y. Ye, S. Xiang, Z. Ma, Y. Zhang, and C. Jiang, "Anti-money laundering by group-aware deep graph learning," *IEEE Trans. Knowl. Data Eng.*, Vol.**35**, No.**12**, pp.**12444–12457**, **2023**. https://doi.org/10.1109/TKDE.2023.3272396

[17] D. V. Kute, B. Pradhan, N. Shukla, and A. Alamri, "Deep learning and explainable artificial intelligence techniques applied for detecting money laundering – A critical review," *IEEE Access*, Vol.**9**, pp.**82300–82317**, **2021**. https://doi.org/10.1109/ACCESS.2021.3086230

[18] D. V. Kute, B. Pradhan, N. Shukla, and A. Alamri, "Explainable deep learning model for predicting money laundering transactions," *Int. J. Smart Sens. Intell. Syst.*, Vol.**17**, No.**1**, **2024**. https://doi.org/10.2478/ijssis-2024-0027

[19] O. Kuiper, M. van den Berg, J. van der Burgt, and S. Leijnen, "Exploring explainable AI in the financial sector: Perspectives of banks and supervisory authorities," *In the Proceedings of the Artificial Intelligence and Machine Learning: 33rd Benelux Conference on Artificial Intelligence, BNAIC/Benelearn 2021, Esch-sur-Alzette, Luxembourg, Nov. 10–12, 2021, Revised Selected Papers*, Springer International Publishing, Vol.**33**, pp.**105–119**, **2022**.

[20] D. Vijayanand and G. S. Smrithy, "Explainable AI-enhanced ensemble learning for financial fraud detection in mobile money transactions," *Intelligent Decision Technologies*, Art. no. **18724981241289751, 2024**.

[21] F. Xu, H. Uszkoreit, Y. Du, W. Fan, D. Zhao, and J. Zhu, "Explainable AI: A brief survey on history, research areas, approaches and challenges," *In the Proceedings of the Natural Language Processing and Chinese Computing: 8th CCF International Conference, NLPCC 2019, Dunhuang, China, Oct. **9–14, 2019**, Part **II***, Springer International Publishing, Vol.**8**, pp.**563–574**, **2019**.

[22] R. Alhajeri and A. Alhashem, "Using Artificial Intelligence to Combat Money Laundering," *Intelligent Information Management*, Vol.**15**, No.**4**, pp.**284–305**, **2023**. https://doi.org/10.4236/iim.2023.154014

[23] S. K. Hashemi, S. L. Mirtaheri, and S. Greco, "Fraud detection in banking data by machine learning techniques," *IEEE Access*, Vol.**11**, pp.**3034–3043**, **2022**. https://doi.org/10.1109/ACCESS.2022.3232287

[24] E. R. Mill, W. Garn, N. F. Ryman-Tubb, and C. Turner, "Opportunities in real time fraud detection: An explainable artificial intelligence (XAI) research agenda," *International Journal of Advanced Computer Science and Applications*, Vol.**14**, No.**5**, pp.**1172–1186**, **2023**. https://doi.org/10.14569/IJACSA.2023.01405121

[25] I. Psychoula, A. Gutmann, P. Mainali, S. H. Lee, P. Dunphy, and F. Petitcolas, "Explainable machine learning for fraud detection," *Computer*, Vol.**54**, No.**10**, pp.**49–59**, **2021**. https://doi.org/10.1109/MC.2021.3081249

[26] J. Vidya Sagar and S. Aquter Babu, "A Hybrid Machine Learning Approach for Real-Time Fraud Detection in Online Payment Transactions," *Library Progress International*, Vol.**44**, No.**3**, pp.**26067–26090**, **2024**.

[27] Anirban Majumder, "Intelligent AI Agents for Fraud and Abuse Detection", International Journal of Computer Science and Engineering, Vol.**12**, Issue.**4**, pp.**1-7**, **2025**.

[28] Avinash Malladhi, "Artificial Intelligence and Machine Learning in Forensic Accounting", International Journal of Computer Science and Engineering, Vol.**10**, Issue.**7**, pp.**15-21**, **2023**.

AUTHORS PROFILE

Latha N. R. is currently serving as a Professor in the Department of Computer Science and Engineering at BMS College of Engineering since 2005. She completed her Bachelor of Engineering in Information Science and Engineering in 2005, followed by a Master of Technology in Computer Science and Engineering in 2009. Dr. Latha earned her Ph.D. from Visvesvaraya Technological University (VTU) in 2020 under the guidance of Dr. G.R. Prasad, Professor, Department of CSE, BMSCE. Her doctoral research focused on VLSI Design and Parallel Computing. She has also supervised student projects in related areas. To date, Dr. Latha has published thirty research papers in reputed international conferences and journals, including two papers on pedagogical approaches to teaching and six papers related to her primary research domain.

Shyamala G is an Associate Professor in the Department of Computer Science and Engineering at BMS College of Engineering. With a strong commitment to research, she continuously engages in activities intended at improving her knowledge and developing innovative strategies that enrich her professional expertise and teaching practices. Dr. Shyamala has published numerous high-quality papers in reputable journals and conferences. In addition to her research contributions, she is dedicated to implementing effective and engaging teaching methodologies in the classroom, fostering an environment where students can master core concepts and apply them professionally in their careers.

Pallavi G. B. completed her PhD. in Cloud Computing and holds a Master of Technology degree in Computer Science and Engineering from BMS College of Engineering (2008). She earned her Bachelor of Engineering in Computer Science from Mysore University in 2001. With a passion for teaching, she is dedicated to imparting computer science concepts to students using innovative teaching and learning methodologies. As a faculty member at BMS College of Engineering, an autonomous institution, Dr. Pallavi has taken on various responsibilities including paper setting for semester examinations, proctoring, feedback coordination, and placement coordination. These roles have enhanced her teaching experience and contributed to her overall professional growth and departmental development.

Sneha Santhosh Bhat is a student at BMS College of Engineering, Bangalore, graduating in 2025. She is pursuing her undergraduate degree in Computer Science and Engineering. With strong passion for technology, Sneha is particularly interested in explainable AI, data science, and financial fraud analytics. She is focused on leveraging machine learning techniques to build transparent, ethical systems that can detect anomalies and mitigate fraud in financial transactions.

Archit Mehrotra is a student at BMS College of Engineering, Bangalore, graduating in 2025. He is pursuing his undergraduate degree in Computer Science and Engineering. Archit has an interest in machine learning algorithms, anomaly detection, and interpretability of models. He is particularly interested in designing hybrid detection systems that merge conventional classification techniques with real-time outlier detection for anti-fraud purposes.

Ashish Seru is a student at BMS College of Engineering, Bangalore, graduating in 2025. He is pursuing his undergraduate degree in Computer Science and Engineering. Ashish is passionate about financial technologies, ethical AI, and secure data-driven systems. His academic focus is on creating intelligent fraud detection pipelines that integrate fairness, transparency, and accountability using explainable AI frameworks.

Tanisha Gotadke is a student at BMS College of Engineering, Bangalore, graduating in 2025. She is pursuing her undergraduate degree in Computer Science and Engineering. Tanisha is particularly interested in interpretable machine learning, SHAP/LIME explainability tools, and real-time fraud analytics. She is committed to designing AI systems that not only detect financial anomalies but also offer human-understandable justifications for their decisions.

