INTERNATIONAL JOURNAL OF NOVEL RESEARCH AND DEVELOPMENT (IJNRD)
© 2026 IJNRD | Volume 11, Issue 8, August 2026 | ISSN: 2456-4184 | IJNRD.ORG

An Explainable Hybrid AI Framework for suspicious UPI Transaction detection

Jallampally Vasavi
PG Scholar, Department of Computer Science and Engineering,Jawaharlal Nehru Technological University Hyderabad University College of Engineering Jagtial(Autonomous),Nachupally(Kondagattu), Telangana,India -505501. Email: jallampallyvasavi@gmail.com

Dr. B. Sateesh Kumar
Professor, Department of Computer Science and Engineering,Jawaharlal Nehru Technological University Hyderabad University College of Engineering Jagtial(Autonomous), Nachupally(kondagattu),Telangana, India - 505501. Email: sateeshbkumar@jntuh.ac.in

Abstract*— The rapid adoption of Unified Payments Interface (UPI) has significantly transformed digital payment systems by enabling fast, convenient, and cashless financial transactions. However, the increasing volume of online transactions has also led to a rise in fraudulent activities, creating serious challenges for financial institutions in detecting both known and emerging fraud patterns. Conventional fraud detection approaches primarily rely on individual supervised or unsupervised learning models, which often struggle to effectively identify evolving fraud patterns while maintaining high detection accuracy. To address this challenge, this project proposes an enhanced Hybrid Learning Framework for detecting fraud in UPI transactions by integrating supervised and unsupervised learning models through a weighted risk score fusion strategy. The framework performs data cleaning, behavioral and temporal feature engineering, feature preprocessing using StandardScaler and One-Hot Encoding, class balancing using SMOTE, parallel prediction using a Stacking Classifier and a Deep Autoencoder, weighted risk score fusion, and decision-based fraud classification. The developed system is further enhanced with Explainable Artificial Intelligence (XAI) techniques using LIME, SHAP, Partial Dependence Plot (PDP), and Individual Conditional Expectation (ICE) to improve prediction transparency and model interpretability. A FastAPI-based web application is developed to provide an interactive user interface that performs fraud prediction, displays fraud and legitimate probabilities, risk level, explanation visualizations, and personalized fraud prevention recommendations. Experimental evaluation demonstrates that the proposed Hybrid Stacking–Autoencoder framework achieves an accuracy of 96.8%, precision of 98.2%, recall of 95.3%, F1-score of 96.8%, and a ROC-AUC score of 0.996, outperforming the individual baseline and extended models. The developed framework provides an accurate, reliable, and interpretable solution for intelligent UPI fraud detection and effectively supports secure digital payment systems by combining high predictive performance with transparent decision-making and personalized security guidance.

"Keywords— Unified Payments Interface (UPI), Fraud Detection, Hybrid Learning, Stacking Classifier, Autoencoder, Explainable Artificial Intelligence (XAI)."

I. INTRODUCTION

The Unified Payments Interface (UPI) has emerged as one of the most influential digital payment infrastructures, enabling secure, instantaneous, and interoperable financial transactions through mobile platforms. Its widespread adoption has accelerated the transition toward a cashless economy by simplifying person-to-person and person-to-merchant payments across sectors such as retail, healthcare, education, transportation, and public services. The growing acceptance of digital payment technologies has improved financial inclusion and increased the efficiency of electronic transactions, allowing millions of users to perform banking operations conveniently and securely. As digital payment ecosystems continue to expand, ensuring the confidentiality, integrity, and reliability of financial transactions has become a critical requirement for financial institutions, payment service providers, and regulatory authorities. Consequently, intelligent transaction monitoring has gained significant attention as an essential component for maintaining the security and trustworthiness of modern digital payment systems [1] [2] [3].

Despite the remarkable growth of UPI-based transactions, the increasing transaction volume has simultaneously expanded the opportunities for fraudulent activities. Fraudsters continuously adopt sophisticated techniques, including identity impersonation, account takeover, phishing attacks, unauthorized fund transfers, and abnormal transaction behaviors, to exploit weaknesses in digital financial systems. Conventional fraud detection approaches, which often depend on predefined rules or historical fraud patterns, face considerable limitations in adapting to rapidly evolving fraudulent strategies. These approaches may generate excessive false alarms, fail to recognize previously unseen fraud patterns, and reduce the overall efficiency of fraud prevention mechanisms. In addition, many intelligent fraud detection systems provide limited interpretability, making it difficult for financial analysts and end users to understand the reasoning behind prediction outcomes and reducing confidence in automated decision-support systems [4] [5] [6].

To address these challenges, this work aims to develop an intelligent framework capable of improving the reliability of fraud detection in UPI transactions while supporting transparent and trustworthy decision-making. The proposed framework focuses on accurately distinguishing legitimate and fraudulent transactions through comprehensive transaction analysis and effective risk assessment. It further emphasizes generating meaningful prediction explanations that assist users and financial institutions in understanding the factors influencing classification outcomes. In addition to improving prediction reliability, the framework is designed to provide a practical and user-friendly environment for transaction analysis and fraud monitoring, thereby supporting

efficient decision-making in modern digital payment ecosystems [7] [8].

The significance of this work lies in its potential to strengthen transaction security, minimize financial losses, and enhance public confidence in digital payment platforms. By providing a reliable analytical framework for identifying suspicious transaction behavior and supporting interpretable fraud predictions, the proposed approach contributes to the advancement of secure financial technologies and intelligent payment monitoring systems. The flexible design further enables future enhancements through the integration of larger datasets, emerging analytical techniques, and evolving security requirements, ensuring long-term adaptability in rapidly changing financial environments. Consequently, the framework offers valuable support for financial institutions, payment service providers, cybersecurity practitioners, and academic communities engaged in developing trustworthy and resilient digital payment infrastructures [9] [10].

II. Related Work

Recent advancements in machine learning have significantly improved the capability of intelligent fraud detection systems by enabling the automatic identification of complex transaction patterns. Kingma and Welling introduced the concept of variational autoencoders, demonstrating the effectiveness of unsupervised representation learning for identifying abnormal data distributions without requiring extensive labeled samples, thereby providing a strong foundation for anomaly detection in financial applications [11]. Dal Pozzolo et al. proposed adaptive machine learning techniques for credit card fraud detection that continuously adjust to changing transaction behaviors, improving detection performance in dynamic financial environments. However, the effectiveness of the approach remains dependent on the availability of representative historical fraud data and periodic model adaptation [12]. Van Vlasselaer et al. introduced a network-based fraud detection framework that incorporated relationships among transactions to improve fraud identification. While network information enhanced detection accuracy, the computational complexity associated with graph construction limited its applicability for large-scale real-time deployment [13].

With the rapid growth of digital payment systems, scalable fraud detection methods have become increasingly important. Brown et al. presented the SCARFF framework, which utilized distributed computing for real-time streaming fraud detection and demonstrated the feasibility of large-scale transaction monitoring. Although the framework improved scalability, maintaining consistent detection performance under evolving fraud patterns remained a challenge [14]. Bahnsen et al. proposed cost-sensitive decision tree models that considered the financial impact of classification errors, reducing monetary losses caused by fraudulent transactions. Despite improving decision-making from a business perspective, the approach primarily focused on supervised learning and exhibited limited capability in recognizing previously unseen fraudulent behavior [15]. These findings indicate that scalability and cost-awareness are important considerations, but they alone are insufficient to address the continuously evolving nature of financial fraud.

The increasing demand for trustworthy artificial intelligence has encouraged the integration of explainability into fraud detection systems. Lundberg and Lee introduced SHAP, a unified explanation framework that quantifies the contribution of individual features toward model predictions, significantly improving model interpretability across various machine learning applications [16]. Almalki and Masud demonstrated that combining explainable artificial intelligence with stacking ensemble models enhanced both fraud detection performance and user confidence by providing interpretable prediction outcomes. Nevertheless, the proposed framework primarily emphasized supervised learning and relied heavily on labeled transaction data [17]. Aljunaid et al. developed an explainable federated learning framework for financial fraud detection that preserved data privacy while supporting collaborative model training across distributed environments. Although privacy preservation was effectively addressed, the complexity of distributed model coordination and communication introduced additional implementation challenges [18].

Recent developments have further explored advanced learning paradigms for adaptive fraud detection. Cui et al. proposed a graph neural network integrated with reinforcement learning to capture evolving transaction relationships and dynamically adapt to changing fraud behaviors. While the approach achieved promising detection capability, its computational requirements and model complexity may restrict practical deployment in resource-constrained environments [19]. Rajan and Rajest presented a comprehensive overview of machine learning and data science techniques for financial fraud detection, highlighting the importance of intelligent analytics, feature engineering, and predictive modeling for improving transaction security. However, the review emphasized that many existing solutions continue to face challenges related to adaptability, interpretability, and the detection of emerging fraud patterns [20]. Collectively, the existing literature demonstrates substantial progress in fraud detection but reveals persistent gaps in balancing detection accuracy, anomaly recognition, prediction transparency, and practical deployment. The present work addresses these limitations by developing an integrated and interpretable fraud detection framework capable of improving detection reliability while supporting informed decision-making in secure UPI payment environments.

III. Materials And Methods

The proposed system presents a parallel hybrid learning framework for intelligent UPI fraud detection using a publicly available UPI transaction dataset containing both legitimate and fraudulent payment records. The framework processes transaction data through a unified learning pipeline involving generic preprocessing and dataset preparation, followed by behavioral and temporal feature engineering to capture transaction-specific characteristics. Feature normalization is performed using StandardScaler, while categorical attributes are transformed through One-Hot Encoding, and SMOTE is employed to mitigate class imbalance and improve model generalization. The core methodology integrates a supervised Stacking Classifier with an unsupervised Deep Autoencoder operating in parallel, where their outputs are combined using a weighted risk score fusion strategy to accurately identify both known and previously unseen fraud patterns. To enhance transparency and user trust, Explainable Artificial Intelligence techniques, including LIME, SHAP, Partial Dependence Plot (PDP), and Individual Conditional Expectation (ICE), provide comprehensive model interpretability. The framework is deployed as a FastAPI-based web application that delivers fraud predictions, risk assessment, explanation visualizations, and personalized security recommendations, thereby

achieving accurate, robust, and interpretable fraud detection for secure digital payment systems.

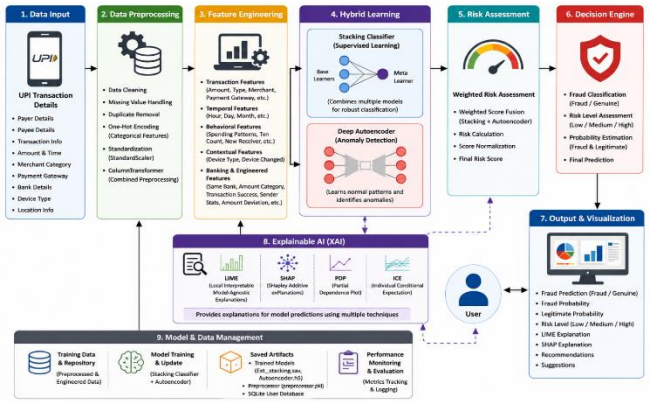

Fig. 1. System Architecture

The proposed system architecture presents an end-to-end hybrid framework for UPI fraud detection. It begins with transaction data acquisition, followed by preprocessing and behavioral feature engineering to prepare high-quality inputs. A parallel Hybrid Learning module combines a supervised Stacking Classifier and an unsupervised Deep Autoencoder using weighted risk score fusion for robust fraud prediction. Explainable AI techniques, including LIME, SHAP, PDP, and ICE, provide transparent decision explanations. The framework is deployed through a FastAPI-based web application that delivers fraud classification, risk assessment, visual analytics, and personalized security recommendations for reliable real-time fraud detection.

A) Dataset Collection

The UPI Transactions dataset was collected from the Kaggle repository and consists of transaction records representing both legitimate and fraudulent UPI payment activities. The dataset contains transaction-level attributes, including financial, behavioral, temporal, and contextual features, along with a binary fraud label used for classification. It exhibits realistic transaction diversity and class imbalance, making it representative of practical fraud detection scenarios. Its comprehensive feature set and labeled transaction records make it well suited for developing, evaluating, and validating intelligent machine learning–based fraud detection frameworks under real-world digital payment conditions.

B) PreProcessing

The preprocessing stage prepares raw UPI transaction data for reliable fraud detection by improving data quality, generating meaningful features, transforming attributes into machine-readable representations, balancing class distributions, and creating representative datasets for robust model training and evaluation.

Data Cleaning: Data cleaning was performed to improve the quality and consistency of the transaction dataset before model development. Missing records, duplicate transactions, and inconsistent entries were identified and eliminated to prevent biased learning and inaccurate predictions. The dataset was subsequently reorganized to maintain a continuous and structured format. This preprocessing stage minimizes data redundancy, improves dataset integrity, and provides a reliable foundation for subsequent feature extraction and machine learning processes, thereby enhancing overall prediction performance.

Feature Engineering: Feature engineering was carried out to derive informative attributes that better represent transaction behavior and financial activity. Additional temporal, transactional, and contextual characteristics were generated from the original records to capture meaningful behavioral patterns associated with legitimate and fraudulent transactions. These engineered features enrich the available information beyond the raw dataset, enabling learning models to identify complex relationships more effectively. Consequently, feature engineering improves the discriminative capability of the framework and supports more accurate fraud detection.

Behavioral Feature Extraction: Behavioral feature extraction was performed to characterize user transaction habits and identify deviations from normal financial behavior. Historical spending patterns, transaction frequency, receiver characteristics, transaction intervals, cumulative spending, and device-related information were incorporated to capture long-term behavioral trends. These behavioral indicators provide valuable contextual information that cannot be obtained from individual transactions alone. By modeling user behavior over time, the framework becomes more effective in recognizing suspicious activities and detecting evolving fraudulent transaction patterns.

Feature Selection: Feature selection was conducted to retain the most informative transaction attributes while eliminating unnecessary or redundant information that could negatively affect model performance. Numerical and categorical variables were identified based on their analytical characteristics, and irrelevant attributes were excluded from further processing. Reducing feature redundancy decreases computational complexity, minimizes overfitting, and improves the efficiency of machine learning models. This step enables the framework to focus on the most relevant information for reliable fraud classification.

Data Pre-processing: Data preprocessing was performed to transform heterogeneous transaction attributes into a standardized representation suitable for machine learning analysis. Numerical variables were normalized to ensure consistent value ranges, while categorical attributes were converted into structured numerical representations for effective model interpretation. Both transformed feature types were integrated into a unified analytical dataset that preserves all relevant transaction information. This preprocessing stage improves data compatibility, enhances learning stability, and contributes to more accurate and robust fraud prediction.

C) Training and Testing

The processed dataset was partitioned into separate training, validation, and testing subsets to ensure reliable model development and unbiased performance evaluation. The training subset was utilized for learning transaction patterns, the validation subset supported model selection and parameter tuning, and the testing subset assessed final predictive performance. Class distributions were preserved across all subsets to maintain representative fraud proportions. This strategy improves model generalization and enables fair evaluation under realistic transaction scenarios.

D) Algorithms

Random Forest Classifier: Random Forest Classifier is employed as a supervised learning model to classify transactions by combining multiple decision trees. Its ensemble learning strategy improves prediction accuracy, reduces overfitting, and enhances robustness by aggregating decisions from diverse trees for reliable fraud identification.

Isolation Forest: Isolation Forest is utilized as an unsupervised anomaly detection algorithm to identify unusual transaction patterns. It isolates anomalous observations based on their uniqueness, enabling effective detection of previously unseen fraudulent activities while complementing supervised classification methods.

Hybrid Model (Random Forest + Isolation Forest): The Hybrid Model combines supervised classification with unsupervised anomaly detection to improve fraud identification. By integrating predictions from both learning paradigms, the model enhances detection reliability, reduces false classifications, and effectively recognizes both known and emerging fraudulent transaction patterns.

SMOTE (Synthetic Minority Over-sampling Technique): SMOTE is applied to address class imbalance by generating representative synthetic samples for the minority class. This balancing strategy improves model learning, minimizes prediction bias toward majority classes, and enhances the detection capability for rare fraudulent transactions.

Stacking Classifier: The Stacking Classifier integrates predictions from multiple complementary learning models through a meta-learning strategy. By combining the strengths of diverse classifiers, it improves prediction accuracy, robustness, and generalization while reducing the limitations associated with individual learning algorithms.

Gradient Boosting: Gradient Boosting is employed as a powerful ensemble learning technique that sequentially refines prediction performance by correcting previous model errors. This iterative learning strategy captures complex transaction patterns, contributing to improved classification accuracy and overall predictive capability.

XGBoost: XGBoost is an optimized gradient boosting algorithm designed for efficient and accurate predictive modeling. Its regularization mechanisms, computational efficiency, and strong learning capability improve model generalization while maintaining high performance in complex classification tasks.

IV. Experimental Results

Accuracy: The accuracy of a test is its ability to differentiate the patient and healthy cases correctly. To estimate the accuracy of a test, we should calculate the proportion of true positive and true negative in all evaluated cases. Mathematically, this can be stated as:

$$Accuracy = \frac{\mathrm{TP} + \mathrm{TN}}{\mathrm{TP} + \mathrm{FP} + \mathrm{TN} + \mathrm{FN}} \quad (1)$$

Precision: Precision evaluates the fraction of correctly classified instances or samples among the ones classified as positives. Thus, the formula to calculate the precision is given by:

$$Precision = \frac{\text{True Positive}}{\text{True Positive} + \text{False Positive}} \quad (2)$$

Recall: Recall is a metric in machine learning that measures the ability of a model to identify all relevant instances of a particular class. It is the ratio of correctly predicted positive observations to the total actual positives, providing insights into a model's completeness in capturing instances of a given class.

$$Recall = \frac{\mathrm{TP}}{\mathrm{TP} + \mathrm{FN}} \quad (3)$$

F1-Score: F1 score is a machine learning evaluation metric that measures a model's accuracy. It combines the precision and recall scores of a model. The accuracy metric computes how many times a model made a correct prediction across the entire dataset.

$$F1\ Score = 2 * \frac{Recall\ \mathrm{X}\ Precision}{Recall + Precision} * 100 \quad (4)$$

AUC-ROC Curve: The AUC-ROC Curve is a performance measurement for classification problems at various threshold settings. ROC plots the True Positive Rate against the False Positive Rate. AUC quantifies the overall ability of the model to distinguish between classes, where a higher AUC indicates better model performance.

$$AUC = \sum_{i=1}^{n-1} (FPR_{i+1} - FPR_i) \cdot \frac{TPR_{i+1} + TPR_i}{2} \quad (5)$$

Table. 1. Performance Evaluation

Model	**Accuracy**	**Precision**	**Recall**	**F1-Score**	**AUC**
Random Forest	0.934	0.544	0.392	0.456	0.884
Isolation Forest	0.897	0.273	0.286	0.279	0.751
Hybrid (RF + ISO)	0.934	0.540	0.373	0.441	0.880
Stacking Classifier	0.967	0.972	0.961	0.966	0.996
Auto Encoder	0.734	0.846	0.572	0.683	0.771
Extension Hybrid	**0.968**	**0.982**	**0.953**	**0.968**	**0.996**

Table 1 compares the performance of baseline, ensemble, and deep learning models for UPI fraud detection. The proposed Extension Hybrid model achieves the highest performance with 96.8% accuracy and 0.996 AUC, demonstrating superior fraud detection capability and reliable classification over the compared approaches.

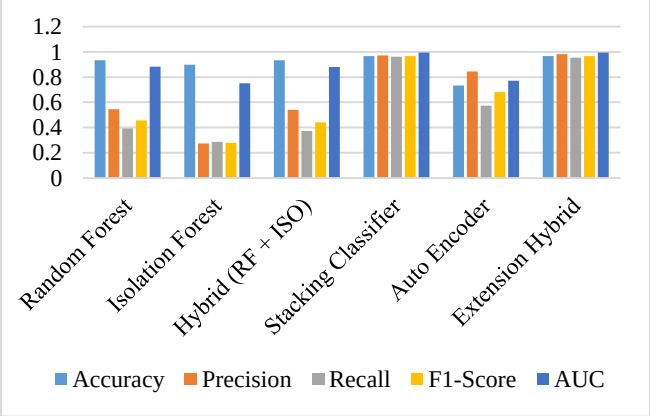

Fig. 2. Comparison Graph

Graph 1 illustrates the comparative performance of all evaluated models using accuracy, precision, recall, F1-score,

and AUC. The proposed Extension Hybrid framework consistently achieves the best overall performance, demonstrating superior accuracy, robustness, and fraud detection capability compared with the baseline and individual models.

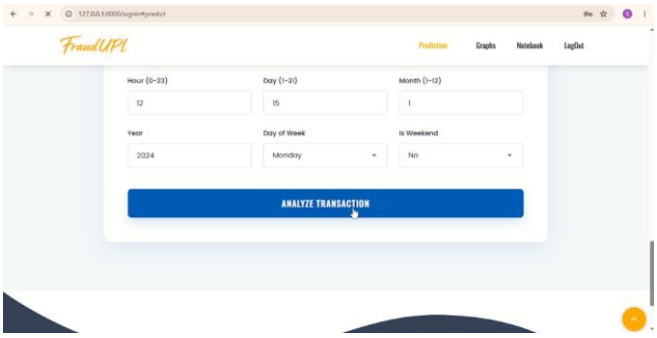

Fig.3 Enter a input data

Fig. 3 illustrates the transaction input interface of the UPI fraud detection application, where users enter temporal and transaction-related details before analysis. After submitting the information, the system processes the input and performs automated fraud prediction through the integrated detection framework.

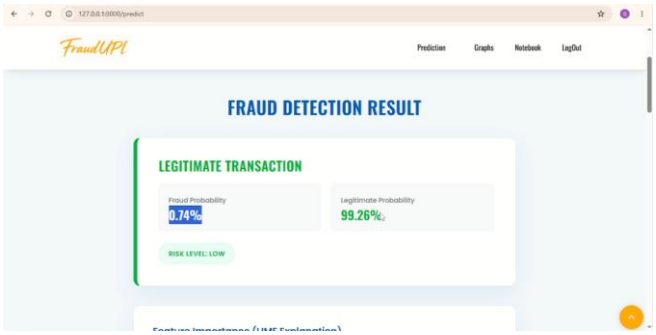

Fig.4 Fraud Detection Result

Fig. 4 illustrates the fraud detection result for a legitimate transaction. The system predicts a legitimate outcome with 99.26% legitimate probability and 0.74% fraud probability, assigning a Low Risk level, demonstrating accurate classification and reliable transaction assessment.

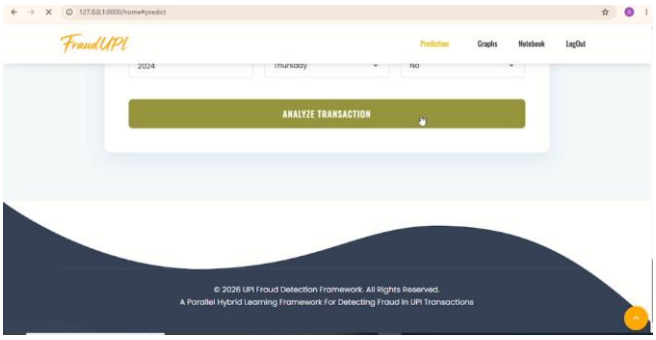

Fig.5 Enter a input data

Fig. 5 illustrates the transaction input interface after entering the required UPI transaction details. Users initiate the fraud analysis by selecting Analyze Transaction, allowing the framework to process the input and generate an intelligent fraud prediction result.

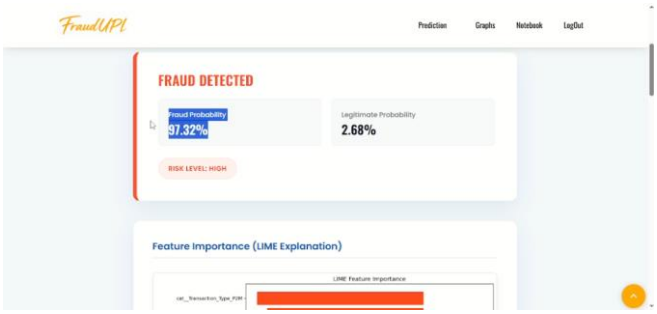

Fig.6 Fraud Detection

Fig. 6 illustrates the fraud detection result for a fraudulent transaction. The system predicts fraud with 97.32% fraud probability and 2.68% legitimate probability, assigning a High Risk level, demonstrating effective identification of suspicious UPI transactions.

V. Conclusion

In conclusion, the proposed work successfully addresses the growing challenge of detecting fraudulent UPI transactions by developing an accurate, reliable, and interpretable hybrid learning framework for secure digital payment systems. The framework utilizes a UPI transaction dataset and combines supervised and unsupervised learning techniques to effectively identify both known and emerging fraud patterns. A parallel architecture integrating a Stacking Classifier and a Deep Autoencoder with a weighted risk score fusion strategy enhances detection capability by leveraging the strengths of both predictive and anomaly detection models. Furthermore, advanced Explainable Artificial Intelligence (XAI) techniques, including LIME, SHAP, Partial Dependence Plot (PDP), and Individual Conditional Expectation (ICE), improve prediction transparency by providing meaningful explanations for model decisions, while the FastAPI-based web application enables practical deployment through an interactive fraud prediction interface. Experimental evaluation demonstrates the effectiveness of the proposed framework by achieving an overall accuracy of 96.8%, confirming its capability to accurately distinguish fraudulent and legitimate transactions under highly imbalanced conditions. The integration of hybrid learning, explainable decision-making, and real-time deployment provides a dependable and user-centric fraud detection solution that strengthens transaction security, supports informed financial decision-making, and contributes to the development of trustworthy and intelligent digital payment ecosystems.

The developed framework can be enhanced by incorporating larger and more diverse real-world UPI transaction datasets to improve generalization across different fraud scenarios. Advanced deep learning and graph-based techniques can be integrated to capture complex transaction relationships and evolving fraud patterns. The system may also be extended with real-time streaming analytics, federated learning for privacy-preserving model training, and adaptive risk assessment mechanisms. Furthermore, integration with banking infrastructures and continuous model updating can improve scalability, robustness, and operational effectiveness in practical financial environments.

References

[1] L. Breiman, "Random Forests," Machine Learning, vol. 45, no. 1, pp. 5-32, 2001.

[2] F. T. Liu, K. M. Ting, and Z.-H. Zhou, "Isolation Forest," in Proceedings of the IEEE International Conference on Data Mining (ICDM), 2008, pp. 413-422.

[3] B. Sateesh Kumar and Md. Sirajuddin, "Hybrid Intrusion Detection Method Based on Improved AdaBoost and Enhanced SVM for Anomaly Detection in Wireless Sensor Networks," International Journal of Advanced Research in Computer Science (IJARCS), vol. 13, no. 5, pp. 38–42, Oct. 2022.

[4] G. Ke et al., "LightGBM: A Highly Efficient Gradient Boosting Decision Tree," in Advances in Neural Information Processing Systems (NeurIPS), vol. 30, 2017.

[5] R. J. Bolton and D. J. Hand, "Statistical Fraud Detection: A Review," Statistical Science, vol. 17, no. 3, pp. 235-249, 2002.

[6] C. Phua, V. Lee, K. Smith, and R. Gayler, "A Comprehensive Survey of Data Mining-Based Fraud Detection Research," arXiv preprint arXiv:1009.6119, 2010.

[7] S. Bhattacharyya, S. Jha, K. Tharakunnel, and J. C. Westland, "Data Mining for Credit Card Fraud: A Comparative Study," Decision Support Systems, vol. 50, no. 3, pp. 602-613, 2011.

[8] F. Carcillo, A. Dal Pozzolo, M. Le Borgne, O. Caelen, Y.-A. Leclercq, and G. Bontempi, "Combining Supervised and Unsupervised Learning in Fraud Detection," Information Sciences, vol. 557, pp. 317-331, 2019.

[9] B. Sateesh Kumar and Md. Sirajuddin, "Intelligent Secure and Malicious-Free Route Management Strategy for IoT-based Wireless Sensor Networks," International Journal of Intelligent Systems and Applications in Engineering (IJISAE), vol. 11, no. 7, pp. 369–380, July 2023.

[10] R. Chalapathy and S. Chawla, "Deep Learning for Anomaly Detection: A Survey," ACM Computing Surveys, vol. 52, no. 1, 2019.

[11] D. P. Kingma and M. Welling, "Auto-Encoding Variational Bayes," arXiv preprint arXiv:1312.6114, 2014.

[12] A. Dal Pozzolo, G. Bontempi, M. Snoeck, and D. Van den Poel, "Adaptive Machine Learning for Credit Card Fraud Detection," IEEE Intelligent Systems, vol. 30, no. 4, pp. 12-19, 2015.

[13] V. Van Vlasselaer et al., "APATE: A Novel Approach for Automated Credit Card Transaction Fraud Detection Using Network-Based Extensions," Decision Support Systems, vol. 75, pp. 38-48, 2015.

[14] G. Brown et al., "SCARFF: Streaming Credit Card Fraud Detection with Spark," in Proceedings of the IEEE International Conference on Big Data, 2018.

[15] A. C. Bahnsen, D. Aouada, B. Ottersten, and A. Stojanovic, "Cost-Sensitive Decision Trees for Fraud Detection," Decision Support Systems, vol. 59, pp. 87-97, 2014.

[16] S. M. Lundberg and S.-I. Lee, "A Unified Approach to Interpreting Model Predictions," in Advances in Neural Information Processing Systems (NeurIPS), vol. 30, 2017.

[17] F. Almalki and M. Masud, "Financial Fraud Detection Using Explainable AI and Stacking Ensemble Methods," arXiv preprint arXiv:2505.10050, 2025.

[18] S. K. Aljunaid et al., "Explainable AI-Driven Federated Learning Model for Financial Fraud Detection," Journal of Risk and Financial Management, vol. 18, no. 4, 2025.

[19] Y. Cui, X. Han, J. Chen, X. Zhang, J. Yang, and X. Zhang, "FraudGNN-RL: A Graph Neural Network with Reinforcement Learning for Adaptive Financial Fraud Detection," IEEE Open Journal of the Computer Society, 2025.

[20] R. Rajan and S. Rajest, "Revolutionizing Credit Card Fraud Detection: Harnessing Machine Learning and Data Science for Enhanced Security," 2024.