ISSN 2319-5991 www.ijerst.com

Peer Reviewed, Referred & Indexed Journal

**Research Paper**

# A Systematic Review of Explainable Hybrid AI Approaches 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 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 techniques using LIME, SHAP, Partial Dependence Plot, and Individual Conditional Expectation 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.

***Index Terms:*** *Unified Payments Interface , Fraud Detection, Machine Learning, Hybrid Learning, Anomaly Detection, Explainable Artificial Intelligence, Digital Payments.*

## INTRODUCTION

The rapid advancement of digital payment technologies has significantly transformed the financial ecosystem by enabling secure, fast, and convenient electronic transactions **[1]**. Among various digital payment platforms, the Unified Payments Interface (UPI) has emerged as one of the most successful real-time payment systems, facilitating seamless fund transfers between individuals, merchants, businesses, and financial institutions **[2]**. Its simplicity, interoperability, and widespread adoption have accelerated the growth of cashless transactions, making UPI an essential component of modern digital commerce. However, the exponential increase in transaction volume has also created new opportunities for financial fraud, posing serious challenges to payment security and consumer trust.

Fraudulent activities associated with digital payment systems continue to evolve in both frequency and sophistication **[3]**. Traditional fraud detection mechanisms, which primarily depend on manually defined rules and fixed thresholds, often struggle to identify newly emerging fraud patterns and adaptive attack strategies **[4]**. As fraudsters continuously modify their techniques to bypass conventional security controls, there is an increasing demand for intelligent fraud detection systems capable of learning complex transaction patterns and identifying suspicious behavior with high accuracy. Consequently, machine learning has become a prominent research area for developing adaptive fraud detection solutions that can automatically distinguish legitimate transactions from fraudulent ones.

A wide range of machine learning techniques has been investigated for financial fraud detection, including supervised learning, unsupervised anomaly detection, ensemble learning, and deep learning approaches **[5]**. Supervised learning algorithms such as Random Forest,

Decision Trees, Support Vector Machines, and Gradient Boosting have demonstrated strong classification performance when sufficient labeled transaction data are available **[6]**. In contrast, unsupervised anomaly detection techniques, including Isolation Forest and Autoencoder-based models, are capable of identifying unusual transaction patterns without relying entirely on labeled fraud instances **[7]**. More recently, hybrid learning frameworks that combine multiple learning paradigms have attracted significant research attention because they integrate the strengths of different algorithms while addressing their individual limitations **[8]**.

Although substantial progress has been achieved, the existing literature remains diverse and fragmented. Many studies evaluate fraud detection techniques using different datasets, feature engineering strategies, evaluation metrics, and experimental settings, making direct comparison challenging **[9]**. Furthermore, issues such as highly imbalanced transaction data, high false positive rates, limited model interpretability, real-time processing requirements, and evolving fraud strategies continue to present significant research challenges. These limitations highlight the necessity for a comprehensive survey that systematically reviews existing approaches, compares their characteristics, and identifies emerging research directions.

This survey presents a comprehensive review of machine learning and hybrid learning techniques for fraud detection in UPI and digital payment transactions **[10]**. Existing studies are categorized based on their learning paradigms, fraud detection strategies, datasets, and evaluation methodologies. The survey further analyzes the strengths and limitations of current approaches, discusses common research challenges, and highlights promising directions such as explainable artificial intelligence, adaptive learning, privacy-preserving techniques, and hybrid fraud detection frameworks. By consolidating recent developments into a structured review, this survey provides a valuable reference for researchers and practitioners working toward the development of accurate, reliable, and interpretable fraud detection systems for secure digital payment environments.

## SURVEY METHODOLOGY

This survey adopts a systematic literature review methodology to provide a comprehensive and objective analysis of existing research on fraud detection in UPI and digital payment systems [11]. The review focuses on machine learning, anomaly detection, ensemble learning, hybrid learning, and explainable artificial intelligence techniques that have been proposed for detecting fraudulent financial transactions. The adopted methodology is designed to ensure

broad coverage of relevant studies while minimizing selection bias and maintaining consistency in the literature analysis [12].

**Data Sources**

The literature reviewed in this survey was collected from well-established academic databases, including **IEEE Xplore, ACM Digital Library, SpringerLink, ScienceDirect, Scopus, and Google Scholar**. These digital libraries were selected because they provide extensive collections of peer-reviewed journals, conference proceedings, and review articles related to machine learning, financial fraud detection, artificial intelligence, and digital payment security.

**Search Strategy**

A keyword-based search strategy was employed to identify relevant publications. Various combinations of keywords were used, including *UPI fraud detection*, *financial fraud detection*, *machine learning*, *hybrid learning*, *ensemble learning*, *anomaly detection*, *digital payment security*, *transaction fraud detection*, *Explainable Artificial Intelligence*, and *fraud analytics [13]*. Boolean operators such as **AND**, **OR**, and **NOT** were utilized to refine search queries, eliminate irrelevant studies, and improve the quality of the retrieved literature [14].

**Inclusion and Exclusion Criteria**

The reviewed studies were selected based on predefined inclusion and exclusion criteria to ensure the relevance and quality of the survey.

**The inclusion criteria include:**

- Peer-reviewed journal articles and conference papers.
- Publications written in English.
- Studies published between **2016 and 2025**.
- Research focusing on fraud detection techniques, machine learning algorithms, anomaly detection, hybrid learning, or digital payment security.

**The exclusion criteria include:**

- Non-peer-reviewed articles, editorials, tutorials, and opinion papers.

- Studies unrelated to financial transaction fraud detection.
- Duplicate publications and incomplete research articles.
- Publications lacking sufficient technical or experimental details.

**Study Selection and Analysis**

The collected publications were initially screened based on their titles and abstracts to determine their relevance to the survey objectives. Subsequently, full-text reviews were conducted to examine the methodologies, datasets, learning techniques, evaluation metrics, and reported findings of the shortlisted studies. The selected literature was then systematically categorized according to the underlying fraud detection approach, including supervised learning, unsupervised anomaly detection, ensemble learning, deep learning, and hybrid learning frameworks. A qualitative comparative analysis was finally performed to identify prevailing research trends, commonly adopted methodologies, existing limitations, and potential future research directions in intelligent fraud detection for digital payment systems.

## TAXONOMY AND CLASSIFICATION OF EXISTING APPROACHES

To systematically analyze the existing literature on fraud detection in UPI and digital payment systems, the surveyed studies are classified based on their primary learning paradigm and fraud detection strategy [15]. This taxonomy provides a structured understanding of the evolution of fraud detection techniques while facilitating comparative analysis across different machine learning approaches. The reviewed literature is broadly categorized into supervised learning, unsupervised anomaly detection, ensemble and hybrid learning, and explainable artificial intelligence-based fraud detection methods [16][17].

### A. Supervised Learning–Based Fraud Detection

Supervised learning approaches represent the most widely adopted category for financial fraud detection due to their ability to learn discriminative patterns from labeled transaction data. Studies in this category employ classification algorithms such as Decision Trees, Random Forest, Support Vector Machines, Logistic Regression, Gradient Boosting, XGBoost, and LightGBM to distinguish fraudulent transactions from legitimate ones. These methods generally achieve high prediction accuracy when sufficient labeled data are available and are capable of learning complex relationships among transaction attributes. However, their

effectiveness often depends on balanced training datasets and the availability of representative fraud examples, making them less adaptable to previously unseen fraud patterns.

### B. Unsupervised Anomaly Detection Approaches

Unsupervised learning techniques focus on identifying abnormal transaction behavior without relying entirely on labeled fraud data. Since fraudulent transactions typically represent only a small proportion of financial datasets, anomaly detection methods have gained considerable attention for detecting rare and emerging fraud patterns. Commonly adopted approaches include Isolation Forest, Local Outlier Factor, clustering algorithms, and Autoencoder-based anomaly detection models. These techniques identify transactions that significantly deviate from normal behavioral patterns, making them effective for discovering novel fraud scenarios. Nevertheless, they may generate higher false positive rates because unusual legitimate transactions can also be classified as anomalies.

### C. Ensemble and Hybrid Learning Frameworks

Recent research increasingly emphasizes ensemble and hybrid learning frameworks that combine multiple learning techniques to improve fraud detection performance. Ensemble methods integrate predictions from several classifiers to enhance robustness, reduce model variance, and improve generalization. Hybrid learning approaches further combine supervised classification with unsupervised anomaly detection, allowing the detection system to identify both previously known fraud patterns and emerging suspicious behaviors simultaneously. These integrated frameworks have demonstrated improved prediction accuracy, reduced false positive rates, and greater adaptability compared with individual machine learning models, making them a promising direction for intelligent financial fraud detection.

### D. Explainable Artificial Intelligence–Based Approaches

As machine learning models become increasingly complex, the interpretability of fraud prediction has emerged as an important research focus. Explainable Artificial Intelligence (XAI) techniques aim to provide transparent explanations for automated fraud detection decisions, enabling analysts and financial institutions to understand the factors influencing model predictions. Commonly used explanation methods include SHAP (Shapley Additive

Explanations), LIME (Local Interpretable Model-Agnostic Explanations), and feature importance analysis. These approaches improve user trust, facilitate regulatory compliance, and support informed decision-making without significantly affecting prediction performance. The integration of explainability with advanced fraud detection models represents an important step toward trustworthy and transparent intelligent financial systems.

## COMPARATIVE ANALYSIS OF SURVEYED STUDIES

This section presents a comparative analysis of the surveyed studies based on their underlying fraud detection approaches, learning strategies, application suitability, and reported limitations [18]. The objective is to highlight the strengths and weaknesses of different machine learning techniques employed for fraud detection in UPI and digital payment systems while identifying the research trends observed across the existing literature [19].

### Comparison of Supervised Learning Approaches

Supervised learning techniques have been extensively employed for fraud detection due to their ability to learn discriminative patterns from labeled transaction data. Algorithms such as Random Forest, Support Vector Machines, Decision Trees, XGBoost, and LightGBM generally achieve high classification accuracy and strong predictive performance when sufficient labeled fraud samples are available. However, their effectiveness depends heavily on the quality and distribution of training data. Highly imbalanced datasets and continuously evolving fraud patterns often reduce their capability to identify previously unseen fraudulent transactions without periodic model retraining.

### Comparison of Unsupervised Anomaly Detection Approaches

Unsupervised anomaly detection techniques identify suspicious transactions by detecting deviations from normal transaction behavior without relying entirely on labeled datasets. Methods such as Isolation Forest, Local Outlier Factor, clustering algorithms, and Autoencoders are particularly useful for detecting emerging fraud patterns that are not represented in historical data. Although these approaches improve adaptability to unknown fraud scenarios, they generally produce higher false positive rates because unusual but legitimate transactions may also be classified as anomalies.

### Comparison of Ensemble and Hybrid Learning Approaches

Ensemble and hybrid learning frameworks combine multiple machine learning models to improve prediction reliability and overcome the limitations of individual algorithms. Ensemble techniques utilize multiple classifiers to enhance generalization performance, whereas hybrid frameworks integrate supervised classification with anomaly detection to simultaneously detect known and unknown fraud patterns. The surveyed literature indicates that these integrated approaches generally achieve higher accuracy, improved recall, lower false positive rates, and greater robustness than standalone machine learning models. Nevertheless, their implementation complexity and computational requirements are comparatively higher.

**Comparative Summary**

Table I summarizes the major categories of fraud detection approaches discussed in the surveyed literature, highlighting their primary techniques, application strengths, and commonly reported limitations.

**Table I. Comparative Summary of Surveyed Studies**

| Approach Category | Representative Techniques | Major Strengths | Reported Limitations |
|---|---|---|---|
| Supervised Learning | Random Forest, SVM, XGBoost, LightGBM | High classification accuracy, effective for labeled fraud detection | Requires labeled data, affected by class imbalance |
| Unsupervised Anomaly Detection | Isolation Forest, Autoencoder, Local Outlier Factor | Detects unknown and emerging fraud patterns | Higher false positive rate, lower interpretability |
| Ensemble Learning | Voting, Bagging, Boosting, Stacking | Improved prediction stability and generalization | Increased computational complexity |
| Hybrid Learning Frameworks | Supervised + Unsupervised Integration | Detects both known and unknown frauds | Higher implementation complexity and |

| | | | |
|---|---|---|---|
| | | with improved reliability | resource requirements |
| Explainable AI-Based Approaches | SHAP, LIME, Feature Importance Analysis | Improves prediction transparency and user trust | Additional computational overhead for explanation generation |

The comparative analysis indicates that no single fraud detection technique is capable of addressing all challenges associated with modern digital payment fraud. Supervised learning models provide excellent predictive performance for known fraud patterns, whereas anomaly detection methods improve the identification of previously unseen attacks. Ensemble and hybrid learning frameworks further enhance overall detection capability by combining the strengths of multiple learning paradigms, while Explainable Artificial Intelligence improves transparency and interpretability. These observations demonstrate an increasing research trend toward integrated, intelligent, and explainable fraud detection frameworks for secure digital payment systems.

## DATASETS AND EVALUATION METRICS

The evaluation of fraud detection techniques largely depends on the availability of representative transaction datasets and standardized performance metrics. Various public and synthetic datasets have been widely utilized by researchers to develop, validate, and benchmark fraud detection models. Likewise, a common set of evaluation metrics is employed to compare the effectiveness of different machine learning approaches under highly imbalanced transaction environments. This section summarizes the datasets and evaluation criteria most frequently reported in the surveyed literature [20].

### Commonly Used Datasets

#### Financial Fraud Detection Datasets

- **IEEE-CIS Fraud Detection Dataset:** A large-scale transaction dataset containing anonymized financial transactions and identity-related attributes. It is widely used

for benchmarking machine learning and ensemble learning models in fraud detection research.

- **Credit Card Fraud Detection Dataset:** A publicly available dataset consisting of European credit card transactions with highly imbalanced fraud instances. It has become one of the most extensively used benchmark datasets for evaluating fraud detection algorithms.
- **PaySim Mobile Money Dataset:** A synthetic mobile payment dataset generated from real financial transaction patterns. It simulates mobile money transfers and fraudulent activities, making it suitable for evaluating fraud detection techniques in digital payment environments.

**UPI Transaction Datasets**

- **UPI Transactions Dataset:** A synthetic dataset designed to simulate realistic Unified Payments Interface (UPI) transactions by incorporating transaction, temporal, demographic, behavioral, and contextual attributes. The dataset supports research on fraud detection, consumer behavior analysis, and intelligent digital payment systems while preserving user privacy through synthetic data generation.

**Transaction Simulation Datasets**

- **Synthetic Financial Transaction Datasets:** Several studies employ programmatically generated transaction datasets to simulate various fraud scenarios while ensuring privacy compliance. These datasets allow researchers to evaluate fraud detection algorithms under controlled environments and diverse transaction characteristics.

**Evaluation Metrics**

The performance of fraud detection models is commonly assessed using standard classification metrics that measure both prediction capability and model reliability. Since fraud detection datasets are typically highly imbalanced, researchers often consider multiple evaluation criteria rather than relying solely on classification accuracy.

**Commonly adopted evaluation metrics include:**

**Classification Performance**

- Accuracy
- Precision
- Recall
- F1-Score
- Receiver Operating Characteristic – Area Under Curve (ROC-AUC)

**Error Analysis**

- False Positive Rate (FPR)
- False Negative Rate (FNR)
- Confusion Matrix

**Computational Performance**

- Prediction Time
- Training Time
- Computational Efficiency

These datasets and evaluation metrics provide a common benchmarking foundation for comparing fraud detection techniques reported in the literature. Their widespread adoption enables researchers to evaluate model performance objectively while facilitating reproducibility, comparative analysis, and the development of more accurate and reliable fraud detection frameworks for digital payment systems.

## CHALLENGES AND OPEN RESEARCH ISSUES

Despite significant advancements in machine learning and artificial intelligence for financial fraud detection, the surveyed literature reveals several persistent challenges that limit the effectiveness of existing fraud detection systems. These challenges arise from the dynamic nature of fraudulent activities, the characteristics of transaction data, and the practical requirements of real-world digital payment environments. Addressing these issues is essential for developing more robust, adaptive, and trustworthy fraud detection frameworks.

### Class Imbalance in Transaction Data

Financial transaction datasets are typically characterized by a highly imbalanced distribution, where fraudulent transactions constitute only a very small percentage of the overall data. This imbalance often biases machine learning models toward the majority class, resulting in reduced fraud detection capability and lower recall for fraudulent transactions. Developing effective imbalance handling techniques remains an important research challenge.

**Detection of Emerging Fraud Patterns**

Fraudsters continuously modify their strategies to evade existing security mechanisms, making fraud detection a constantly evolving problem. Many supervised learning models perform well on previously observed fraud patterns but struggle to identify newly emerging attack strategies that are not represented in the training data. Designing adaptive detection mechanisms capable of recognizing evolving fraud behaviors remains an active area of research.

**False Positive Reduction**

Although many fraud detection models achieve high detection accuracy, they frequently generate false alarms by incorrectly classifying legitimate transactions as fraudulent. High false positive rates may inconvenience customers, increase operational costs, and reduce user confidence in digital payment systems. Achieving an effective balance between fraud detection capability and false positive reduction continues to be a significant research issue.

**Explainability and User Trust**

Advanced machine learning and deep learning models often operate as black-box systems, providing limited insight into the reasoning behind their predictions. Financial institutions increasingly require transparent and interpretable decision-making to support regulatory compliance, facilitate fraud investigation, and improve user confidence. Consequently, integrating explainable artificial intelligence into fraud detection systems has become an important research direction.

**Privacy and Data Security**

Fraud detection systems require access to large volumes of sensitive financial and personal information. Ensuring user privacy while maintaining effective fraud detection remains a major challenge. Researchers continue to investigate privacy-preserving learning techniques that

enable collaborative model development without exposing confidential transaction data or compromising regulatory compliance.

### Lack of Standardized Benchmarking

The surveyed studies employ diverse datasets, feature engineering strategies, evaluation metrics, and experimental settings, making objective comparison of fraud detection techniques difficult. The absence of standardized benchmarking datasets and evaluation protocols limits reproducibility and hinders fair performance comparison across different research works. Establishing common evaluation standards remains an important requirement for future fraud detection research.

## FUTURE RESEARCH DIRECTIONS

The challenges identified in the surveyed literature indicate several promising research directions for developing more intelligent, adaptive, and secure fraud detection systems for digital payment platforms. As financial fraud techniques continue to evolve, future research must focus on improving prediction accuracy, model adaptability, interpretability, and privacy while maintaining efficient real-time transaction processing. Some important research directions are discussed below.

### Adaptive and Online Learning Models

Most existing fraud detection systems are trained using historical transaction data and require periodic retraining to recognize new fraud patterns. Future research should investigate online and incremental learning techniques that continuously update prediction models as new transaction data become available, enabling faster adaptation to evolving fraud strategies.

### Explainable and Trustworthy Artificial Intelligence

Although advanced machine learning models provide high prediction performance, their limited interpretability remains a major concern for financial institutions and regulatory authorities. Future work should focus on developing explainable fraud detection frameworks that provide transparent decision-making while maintaining high predictive accuracy. The

integration of Explainable Artificial Intelligence techniques can improve user trust and facilitate effective fraud investigation.

**Graph-Based Fraud Detection**

Financial transactions naturally form interconnected networks involving customers, merchants, devices, and banking institutions. Future research may explore graph-based learning techniques capable of modeling these complex relationships to identify organized fraud networks, coordinated fraudulent activities, and hidden transaction dependencies that are difficult to detect using conventional machine learning approaches.

**Privacy-Preserving Fraud Detection**

Protecting sensitive financial information remains an important requirement for modern fraud detection systems. Future studies should investigate privacy-preserving machine learning techniques, including federated learning, secure multi-party computation, and privacy-aware model training, to enable collaborative fraud detection without exposing confidential transaction data.

**Real-Time Intelligent Fraud Detection**

The growing volume of digital payment transactions requires fraud detection systems capable of making rapid and accurate decisions with minimal computational delay. Future research should focus on developing lightweight and scalable fraud detection frameworks that support real-time transaction analysis while maintaining high detection accuracy and low false positive rates.

**Standardized Evaluation Frameworks**

The lack of common datasets and benchmarking methodologies makes objective comparison of fraud detection techniques difficult. Future research should emphasize the development of standardized datasets, evaluation protocols, and benchmarking frameworks that enable consistent performance assessment and improve the reproducibility of research findings across different fraud detection approaches.

These research directions highlight the continuing evolution of intelligent fraud detection systems and emphasize the need for adaptive, explainable, privacy-aware, and scalable solutions. Continued research in these areas will contribute to the development of more reliable and trustworthy fraud detection frameworks capable of addressing the increasingly sophisticated challenges of modern digital payment ecosystems.

## DISCUSSION AND CONCLUSION

The surveyed literature demonstrates the growing adoption of machine learning and artificial intelligence for detecting fraudulent activities in digital payment systems. Researchers have explored a wide range of supervised learning, anomaly detection, ensemble learning, and hybrid learning techniques to improve fraud detection performance while addressing the challenges posed by increasingly sophisticated fraud strategies. Although considerable progress has been achieved, existing studies indicate that no single learning approach can effectively address every aspect of financial fraud detection, emphasizing the importance of integrated and adaptive analytical frameworks.

### Trends and Key Insights

The literature reveals a clear transition from traditional rule-based fraud detection methods toward intelligent data-driven approaches capable of learning complex transaction patterns. Supervised learning techniques continue to demonstrate strong performance in detecting known fraudulent activities, while anomaly detection methods enhance the identification of previously unseen fraud patterns. Recent studies increasingly emphasize ensemble and hybrid learning frameworks that combine multiple analytical techniques to improve prediction reliability and reduce false positive rates. Furthermore, Explainable Artificial Intelligence has emerged as an important research area for improving prediction transparency and increasing user confidence in automated fraud detection systems.

### Implications and Future Directions

The findings of this survey indicate that future fraud detection systems should integrate multiple learning paradigms rather than relying on individual machine learning models. Researchers are encouraged to investigate adaptive learning strategies, graph-based transaction analysis, privacy-preserving machine learning, and real-time fraud detection frameworks capable of responding to evolving fraud behaviors. In addition, greater emphasis should be

placed on explainable and trustworthy artificial intelligence to support transparent decision-making, regulatory compliance, and effective financial risk management. The development of standardized datasets and evaluation methodologies will further facilitate objective comparison and reproducibility across future research.

## Summary

This survey systematically reviews existing machine learning and hybrid learning approaches for fraud detection in UPI and digital payment systems. The surveyed studies are categorized according to their learning paradigms, followed by a comparative analysis of their strengths, limitations, commonly used datasets, evaluation metrics, and open research challenges. The review highlights the increasing importance of integrated, adaptive, and explainable fraud detection techniques for securing modern digital payment ecosystems. By consolidating recent research developments and identifying promising future directions, this survey provides a structured reference for researchers and practitioners working toward the development of accurate, reliable, and trustworthy intelligent fraud detection frameworks.

## 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] T. Chen and C. Guestrin, "XGBoost: A Scalable Tree Boosting System," in *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 2016, pp. 785-794.

[4] 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..

[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.