Published by :
http://www.ijert.org

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 14 Issue 11, November 2025

A Comprehensive Survey on Machine Learning Techniques for UPI and Telecommunication Fraud Detection

Ms. Anshika Negi
Assistant Professor, Department of Information Technology,
Dr. Akhilesh Das Gupta Institute of Professional Studies
(ADGIPS), New Delhi, India

Deepak, Manvi Taneja, Neetima Tyagi,
Yash Aggarwal
Student, Department of Information Technology,
Dr. Akhilesh Das Gupta Institute of Professional Studies
(ADGIPS), New Delhi, India

***Abstract* - In today's world, the widespread integration of digital payment systems and mobile communication networks has transformed everyday transactions worldwide. People use online transactions, especially UPI payments, for everything from small to large purchases, and depend on mobile networks for communication. However, with this increasing use, there has also been a rise in fraudulent activities such as fake UPI transactions, phishing, and scam calls. These frauds cause serious financial losses and affect user trust in digital systems.**

This survey paper gives an overview of different research works that use machine learning (ML) and artificial intelligence (AI) techniques to detect such frauds. It mainly focuses on two areas — UPI transaction fraud detection and telecommunication call fraud detection — and studies how different algorithms, like supervised, unsupervised, and hybrid models, have been applied to these problems. The paper also discusses important steps such as data preprocessing, feature extraction, handling imbalanced data, and model evaluation, which are necessary for building accurate detection systems.

Along with reviewing existing research, this paper also points out the common challenges faced in this field, such as the lack of real-world datasets, difficulty in adapting to new types of fraud, and problems in real-time deployment. The goal of this survey is to give readers a clear understanding of the current state of fraud detection techniques and to provide insights for future research on developing more reliable, scalable, and explainable fraud detection systems.

Keywords - UPI Fraud Detection, Call Detail Record (CDR), Telecommunication Fraud, Machine Learning, Ensemble Learning, Neural Networks, Data Preprocessing, Comparative Analysis, Research Challenges

I. INTRODUCTION

The rapid expansion of digital payment infrastructures and telecommunication services has profoundly transformed everyday transactions and communication in India. Among these, the Unified Payments Interface (UPI) has emerged as a cornerstone of the nation's digital economy, enabling seamless peer-to-peer and merchant payments through mobile devices. According to the National Payments Corporation of India (NPCI), UPI handled nearly 19.47 billion transactions worth ₹25.08 lakh crore in July 2025 alone, reflecting a substantial increase from 131.1 billion to 185.8 billion transactions between FY 2024 and FY 2025 [18], [19].
While this demonstrates the growing public confidence in UPI as a reliable platform, the surge in digital activity has simultaneously expanded the attack surface for financial fraud [20].

In parallel, the telecommunication ecosystem continues to face a wide range of fraudulent practices such as vishing (voice phishing), spam and robocalls, SIM swapping, Wangiri fraud (short missed calls luring users to return premium-rate calls), and toll bypassing. These schemes not only lead to financial and reputational losses for consumers and operators but also introduce regulatory and operational challenges. Reports indicate that telecom-related frauds are increasing in frequency and sophistication, driven by social engineering and automated exploitation techniques [21].

Although payment-based and telecommunication-based frauds often appear as separate domains, emerging evidence shows that both share overlapping attacker behaviors and techniques, such as impersonation, identity theft, and behavioral manipulation [22]. For example, phishing or spoofed calls frequently serve as the initial entry point for UPI-related financial scams. This convergence highlights the need for a unified analytical framework that can jointly address fraud across digital payment and telecommunication networks.

This survey aims to fill that gap by systematically reviewing machine learning (ML) and artificial intelligence (AI) approaches used in both UPI transaction fraud detection and telecommunication call fraud detection. It examines supervised, unsupervised, and hybrid modeling paradigms, exploring how they are applied to detect anomalies, classify risky behavior, and adapt to evolving fraud patterns. The paper also evaluates various data preprocessing, feature engineering, and model evaluation strategies that are crucial for practical fraud-detection systems.

The key contributions of this work are summarized as follows:

- It provides a taxonomy and comparative framework for fraud detection methods across payment and telecommunication domains.
- It presents a comprehensive review of state-of-the-art ML and DL models, including their datasets, evaluation metrics, and observed limitations.
- It discusses open research challenges, such as real-time adaptability, lack of labeled data, and model interpretability.
- It outlines future directions for building unified, explainable, and scalable fraud detection architectures that bridge both financial and telecom ecosystems.

The remainder of this paper is structured as follows: Section II introduces key background concepts and evaluation metrics used in fraud detection. Section III describes the taxonomy and classification of detection methodologies. Section IV provides a literature-based review of UPI and telecom fraud studies. Section V offers a comparative analysis of these studies, followed by Section VI, which discusses open challenges and research gaps. Finally, Section VII concludes the paper with insights for future research.

II. BACKGROUND

A. Definition of Fraud Detection

Fraud detection involves recognizing unusual or deceptive actions within a system that aim to misuse financial or operational resources. In digital ecosystems, such activities are often designed to manipulate financial or operational processes for illicit gain. Modern fraud detection systems increasingly depend on Machine Learning (ML) and Artificial Intelligence (AI) techniques, which enable automatic recognition of subtle and complex behavioral patterns in large datasets. These systems can identify anomalies and generate alerts in near real time, greatly reducing the need for manual intervention and improving accuracy compared to traditional rule-based systems.

B. UPI (Unified Payments Interface) Fraud

The Unified Payments Interface (UPI), developed by the National Payments Corporation of India (NPCI), is a real-time payment mechanism that facilitates seamless peer-to-peer and business-to-customer transactions through mobile applications. While UPI has revolutionized India's digital payment landscape, its popularity has also made it a target for several fraudulent schemes. Common UPI-related frauds include:

- Phishing and impersonation: Deceptive communication that tricks users into revealing login credentials or one-time passwords.
- Transaction redirection or manipulation: Altering payment requests or misdirecting funds to unauthorized beneficiaries.
- Application cloning and fake screenshots: Using counterfeit apps or images to simulate completed payments.
- Social engineering and OTP compromise: Psychological manipulation aimed at bypassing user verification layers.

To counter these threats, ML-based UPI fraud detection systems examine transaction-level features, such as transaction frequency, beneficiary behavior, and device identifiers. Algorithms like Random Forest, XGBoost, and ensemble classifiers have been shown to achieve high detection accuracy by identifying abnormal transaction sequences and irregular payment patterns.

C. Telecommunication and Call Fraud

Telecommunication fraud refers to a range of illicit activities that exploit communication networks for monetary or personal advantage. Such frauds compromise user privacy, cause financial losses, and can disrupt network operations. Common instances include:

- Spam and Robocalls: Automated or bulk calls used to spread misleading information, advertisements, or scams.
- Vishing (Voice Phishing) and Impersonation: Fraudulent phone calls designed to trick individuals into sharing confidential data such as account credentials or personal identifiers.
- Wangiri Fraud: A scheme where attackers give a missed call from premium-rate numbers, luring victims into calling back and incurring high charges [21].
- SIM-Box and Toll-Bypass Fraud: Illegal rerouting of international or long-distance calls through unauthorized gateways to evade interconnection fees.

The identification of such fraudulent behavior primarily relies on Call Detail Records (CDRs)—structured datasets that contain attributes like caller and receiver IDs, call duration, time stamps, geographical information, and network details. Analytical and machine learning models examine these records to uncover anomalies in call frequency, duration patterns, or cost fluctuations. Both supervised and unsupervised techniques—such as clustering, neural networks, and autoencoders—are employed to detect and classify unusual communication behaviors indicative of potential fraud.

D. Data Preprocessing and Feature Engineering

Reliable fraud detection depends heavily on the quality of the underlying data. Because both UPI and telecom datasets can contain noise, redundancy, and missing values, a structured data preprocessing pipeline is essential. The process usually involves:

- Data Cleaning: Eliminating duplicate, inconsistent, or incomplete entries.
- Normalization and Encoding: Standardizing numerical scales and converting categorical data into machine-readable formats.
- Feature Selection and Extraction: Deriving meaningful attributes such as transaction frequency, geolocation, device identifiers, or call intensity metrics.
- Handling Class Imbalance: Since fraudulent activities constitute only a small fraction of all transactions or calls, techniques such as SMOTE (Synthetic Minority Oversampling Technique), undersampling, and cost-sensitive learning are commonly used to balance the data.

Well-engineered features not only improve model performance but also enhance interpretability, enabling domain experts to understand the behavioral indicators driving fraud predictions.

E. Evaluation Metrics

Since fraud detection typically involves highly imbalanced datasets, evaluating model performance based only on accuracy can be misleading. Therefore, multiple performance metrics are used to provide a more holistic view of a model's predictive strength:

- Accuracy: Measures the overall proportion of correctly classified instances, covering both fraudulent and legitimate cases.
- Precision: Represents the share of correctly detected fraudulent cases out of all instances the model labeled as fraud.
- Recall (Sensitivity): Denotes the proportion of actual frauds that were successfully identified by the model.
- F1-Score: The harmonic mean of precision and recall, providing a balanced evaluation when considering both false positives and false negatives.
- AUC–ROC (Area Under the Receiver Operating Characteristic Curve): Reflects how effectively the model can differentiate between fraudulent and non-fraudulent activities across various decision thresholds.

Collectively, these evaluation measures highlight the trade-offs between false alarms and missed detections—an important consideration when developing real-time fraud detection systems in financial or telecommunication environments.

III. TAXONOMY / CLASSIFICATION

The surveyed research works on UPI and telecommunication fraud detection employ a wide spectrum of analytical techniques.
These can be classified according to learning paradigm, model family, application domain, and detection objective. This taxonomy helps unify the reviewed literature and reveal major methodological trends.

A. Classification by Learning Paradigm

1. Supervised Learning
   Most UPI-based studies adopt supervised algorithms trained on labeled datasets containing both fraudulent and legitimate transactions.
   Examples include:
   - ○ *S. Jagadeesan et al.* ("UPI Fraud Detection Using Machine Learning") — Random Forest outperformed SVM and Logistic Regression [4].
   - ○ *Miss Sayalee Bodade and P. P. Pawade* ("Implementation Paper on UPI Fraud Detection") — real-time supervised model with evaluated precision, recall, and ROC [6].
   - ○ *Viha Dave and Dhaval Chudasama* ("Fraud Detection in UPI Transactions Using Ensemble Learning") — XGBoost-based ensemble with highest accuracy [7].
   - ○ *Batoul Abo Yehya and Nazih Salhab* ("Telecommunications Fraud Machine Learning-Based Detection") — Random Forest and Logistic Regression on telecom data.
     Supervised models dominate because labeled transaction or call data are available for controlled experimentation [12].

2. Unsupervised and Semi-Supervised Learning
   Applied where labeled fraud cases are scarce.
   - *Ma'shum Abdul Jabbar and Suharjito* used K-Means and DBSCAN clustering on Call Detail Records (CDRs) to isolate anomalous user behavior [9].
   - *Waleed Hilal et al.* reviewed autoencoders and isolation forests as generic anomaly-detection tools. These methods detect deviations without predefined labels, useful for emerging or unseen fraud patterns [1].
3. Hybrid and Ensemble Learning
   Several works integrate multiple models to leverage complementary strengths.
   - *Vitthal Kamble et al.* proposed a stacked-generalization ensemble combining base learners through a meta-model [5].
   - *Viha Dave et al.* demonstrated boosting-based ensembles (XGBoost) outperforming single classifiers [7].
   - *Kavya K. R. and Usha Sree R.* suggested augmenting CNNs with ensemble strategies. Hybrid designs address imbalance, bias-variance trade-off, and adaptive learning challenges [8].
4. Deep Learning and Neural Approaches
   Deep architectures capture sequential or nonlinear patterns in transactional and call data.
   - *Kavya K. R. and Usha Sree R.* employed Convolutional Neural Networks (CNNs) for UPI fraud detection [8].
   - *John C. Daka* used Artificial Neural Networks (ANNs) to analyze telecom traffic patterns [10].
   - *B. Durga Bhavani et al.* applied LSTM networks and NLP to identify fake or scam calls. Such models are data-intensive but excel in pattern discovery and temporal-sequence analysis [11].

B. Classification by Model Family

Table **I**

Model Family	Representative Techniques / Papers	Typical Domain
Statistical & Rule-Based	Logistic Regression (*Yehya et al., Hilal et al.*)	Early fraud screening
Machine Learning	Random Forest, SVM, Decision Tree (*Jagadeesan et al., Bodade et al., Yehya et al.*)	UPI / Telecom
Ensemble & Hybrid	Stacked Generalization, XGBoost, Boosting (*Kamble et al., Dave et al.*)	UPI
Deep Learning	CNN, ANN, LSTM (*Kavya et al., Daka, Bhavani et al.*)	UPI / Call
Clustering / Anomaly	K-Means, DBSCAN *(Jabbar et al.)*	Telecom CDR

C. Classification by Application Domain

1. UPI / Financial Transaction Fraud
   Focuses on identifying abnormal digital-payment behavior using transaction amount, frequency, beneficiary ID, and device fingerprinting.
   Representative works: *Jagadeesan et al.*, *Kamble et al.*, *Bodade et al.*, *Dave et al.*, *Kavya et al.*
2. Telecommunication / Call Fraud
   Utilizes Call Detail Records (CDRs) and voice data to detect anomalies such as SIM-box, Wangiri, spam, or impersonation calls.
   Representative works: *Jabbar et al.*, *Daka*, *Bhavani et al.*, *Yehya et al.*
3. Cross-Domain / General Fraud Detection
   Broader frameworks or reviews spanning finance, telecom, and intrusion detection.
   Representative works: *Hilal et al.*, *Flegel et al.*, *Kou et al.*

D. Classification by Detection Objective

Table **II**

Objective	Description	Example Papers
Binary Classification	Label transactions or calls as "Fraud" / "Legitimate."	*Jagadeesan et al., Dave et al., Yehya et al.*
Anomaly Detection	Identify unusual patterns without explicit labels.	*Jabbar et al., Hilal et al.*
Risk Scoring / Ranking	Compute probability or severity of fraud.	*Kamble et al., Flegel et al.*
Pattern Discovery	Detect coordinated or network-level attacks.	*Kou et al., Daka*

E. Observations

- Supervised ML and ensemble methods dominate UPI-related research due to access to labeled data and structured transaction attributes.
- Unsupervised clustering and neural networks are prevalent in telecom-fraud studies where labeling is scarce and behavior patterns are temporal.
- Hybrid and deep models mark a shift toward adaptive, cross-domain systems capable of detecting evolving fraud strategies.

- Future directions emphasize explainability, real-time adaptability, and integrated multi-channel risk scoring across payment and telecommunication ecosystems.

IV. LITERATURE SURVEY

A. Foundational and Cross-Domain Surveys

Waleed Hilal, S. Andrew Gadsden, and John Yawney — "Financial Fraud: A Review of Anomaly Detection Techniques and Recent Advances" — explores anomaly-detection methods across financial domains and synthesizes recent advances. The paper categorizes anomaly types, explains the typical anomaly-detection pipeline (data preprocessing → feature extraction → model training → scoring → evaluation), and emphasizes domain differences (for example, a sudden physiological change is critical in healthcare while volatility is normal in finance). It surveys fraud applications such as credit-card fraud, insurance fraud, and money-laundering, highlights semi-supervised and unsupervised techniques (autoencoders, isolation forests, clustering), and discusses practical challenges including concept drift, evaluation on proxy datasets, and real-time latency constraints [1].

Ulrich Flegel, Julien Vayssière, and Günter Bitz — "A State of the Art Survey of Fraud Detection Technology" — provides a comprehensive overview of the evolving landscape of fraud detection methods across multiple domains. The authors categorize existing approaches into three main groups: (1) statistical and data-centric techniques, which rely on pattern recognition and rule-based analysis; (2) machine learning and neural network approaches, which learn complex patterns from data; and (3) hybrid and emerging methods that integrate multiple paradigms for improved adaptability. The paper emphasizes the central role of data preprocessing, including data cleaning, feature extraction, and balancing of imbalanced datasets, as foundational to any fraud detection system. It also illustrates how fraud manifests differently across sectors — for instance, in credit-card fraud, sudden deviations in spending behavior are key indicators, while in telecommunication fraud, subscription or account-setup anomalies signal potential misuse. The survey concludes that the dynamic and evolving nature of fraud demands adaptive models capable of continuous learning and cross-domain applicability [2].

Yufeng Kou, Chang-Tien Lu, Sirirat Sinvongwattana, and Yo-Ping Huang — "Survey of Fraud Detection Techniques" — reviews major research contributions in credit card fraud, telecommunication fraud, and computer intrusion detection. The authors analyze statistical, knowledge-based, and machine-learning methods used to identify suspicious behavior in these domains. The paper highlights rule-based systems, data mining algorithms, neural networks, and outlier detection as core methodologies, while also comparing their scalability and real-time detection capabilities. Key research issues identified include limited availability of labeled fraud data, challenges in detecting new fraud patterns (concept drift), and the need for cost-sensitive learning to reduce false alarms. The survey provides one of the earliest structured taxonomies that link diverse fraud domains under a unified analytical framework, making it a foundational reference for modern cross-domain fraud detection studies [3].

B. UPI Fraud Detection Studies

S. Jagadeesan, K. S. Arjun, G. Dhanika, G. Karthikeyan, and K. Deepika — "UPI Fraud Detection Using Machine Learning" — focuses explores of careful preprocessing and feature engineering (e.g., amount transforms, balance-consistency checks, temporal frequency features), then evaluates classical classifiers with emphasis on Random Forest. It reports that Random Forest outperforms baseline models such as Logistic Regression, SVM, and Decision Trees on the dataset used (after addressing class imbalance via weighting/oversampling), and recommends probability calibration and interpretability tools to reduce false positives in deployment [4].
Source: ResearchGate.

Vitthal B. Kamble, Krushna Pisal, Pranav Vaidya, and Sahil Gaikwad — "Enhancing UPI Fraud Detection: A Machine Learning Approach Using Stacked Generalization" — proposes a stacked-generalization (stacking) ensemble for UPI fraud detection and demonstrates improved detection performance over single models. The paper explains stacking mechanics (out-of-fold base model predictions fed to a meta-learner), highlights that class imbalance must be aggressively treated to avoid degraded meta-model performance, and shows that a layered ensemble can yield better precision-recall tradeoffs. It also discusses practical concerns (risk of leakage in stacking pipelines and inference latency) and argues for moving beyond simple single-model baselines to hybrid/ensemble solutions [5].

Miss Sayalee S. Bodade and Prof. P. P. Pawade — "Implementation Paper on UPI Fraud Detection Using Machine Learning" — explores developing a real-time fraud-detection system with a practical user interface and system design. The model analyses each transaction using the user's historical transaction records and individual transaction receipts to identify irregular behavior. The paper evaluates four key performance metrics—precision, recall, false-negative rate, and accuracy—and also presents the ROC curve to assess model robustness. Its main contribution lies in demonstrating an end-to-end prototype that connects data preprocessing, prediction, and visualization for operational UPI fraud detection [6].

Ms. Viha Dave and Mr. Dhaval Chudasama — "Fraud Detection in UPI Transactions Using Ensemble Learning" — emphasizes the superiority of ensemble learning over conventional machine-learning models for UPI fraud detection. The authors compare several algorithms across multiple studies and datasets, showing that gradient-boosting approaches, particularly XGBoost, significantly outperform other ensemble and standalone models. Their analysis considers metrics such as AUC-ROC, F1-score, accuracy, precision, and recall, and highlights the critical influence of features like transaction amount, frequency, time-based patterns, geolocation/device mismatch, and beneficiary history. The study concludes that XGBoost offers strong performance and can be readily integrated into existing UPI systems for real-time detection [7].

Kavya K. R. and Usha Sree R. — "UPI Fraud Detection Using Machine Learning" — investigates the effectiveness of machine-learning and deep-learning models, particularly Convolutional Neural Networks (CNNs), in detecting UPI-based frauds compared with traditional rule-based systems. The paper outlines the proposed CNN architecture, explaining how its adaptive learning enables it to recognize evolving fraud patterns and automatically extract relevant transaction features. Results show that CNNs outperform conventional algorithms due to their ability to learn complex nonlinear relationships. The authors also recommend exploring ensemble learning to further improve generalization and model stability [8].

C. Telecommunication and Call Fraud Detection Studies

Ma'shum Abdul Jabbar and Suharjito — "Fraud Detection Call Detail Record Using Machine Learning in Telecommunications Company" — explores detecting telecommunication fraud and its significant impact on company revenue. The paper applies unsupervised learning techniques to identify anomalies within Call Detail Records (CDRs), using K-Means clustering and DBSCAN for grouping suspicious call patterns. Comparative analysis with labeled data shows that K-Means outperforms DBSCAN, demonstrating clustering as a viable method for early fraud detection. The study also highlights the importance of feature selection in capturing fraudulent behavior and acknowledges limitations, such as potential misclassifications. It concludes that unsupervised clustering can serve as an additional fraud-filtering layer for telecom operators [9].

John C. Daka — "Smart Mobile Telecommunication Network Fraud Detection System Using Call Traffic Pattern Analysis and Artificial Neural Network" — addresses telecommunication frauds such as SIM-box and toll-bypass frauds by developing a neural-network-based detection system. The model analyzes features derived from CDR patterns, emphasizing stages of data preprocessing, feature engineering, and ANN-based classification. The reported results show near-perfect accuracy (100%), which the authors attribute to a small, controlled dataset rather than practical real-world performance. Despite this, the work effectively demonstrates the potential of Artificial Neural Networks for identifying complex fraud behaviors and suggests that telecom operators can integrate such models to mitigate revenue loss and improve network security [10].

B. Durga Bhavani, Uppala Nikitha, Patlolla Nandini, and Nethrika Reddy Gogu — "Artificial Intelligence Based Fake or Fraud Phone Calls Detection" — addresses the problem of fraudulent and impersonation-based phone calls that often lead to financial losses and data breaches. The paper critiques traditional rule-based and human-driven systems for their inability to adapt to rapidly evolving scam patterns. It proposes an AI-driven solution using Natural Language Processing (NLP) and Machine Learning, where a Multinomial Naive Bayes model serves as the baseline and a Long Short-Term Memory (LSTM) network is used to capture sequential dependencies in call or text data. Experimental results show that LSTM significantly outperforms Naive Bayes, providing higher detection accuracy. The authors also develop a Flask-based web application that integrates the trained model for real-time fraud detection, offering instant predictions and a user-friendly interface. The study suggests that hybrid NLP–ML frameworks can evolve dynamically with new fraud strategies, reducing reliance on manual intervention [11].

Batoul Abo Yehya and Nazih Salhab — "Telecommunications Fraud Machine Learning-Based Detection" — presents a comprehensive study of telecommunication frauds and the machine-learning approaches used to detect them. The authors describe various fraud types—such as subscription fraud, SIM-box fraud, and call masking—and examine detection methods including rule-based systems, real-time monitoring, and pattern recognition. Multiple ML algorithms are compared on real datasets, evaluated through metrics such as accuracy, F1-score, and confusion matrix analysis. The findings reveal that Random Forest models achieve superior performance compared to Logistic Regression and other classifiers, owing to their robustness and feature-handling capabilities. The paper also notes that while ML approaches outperform traditional methods, challenges remain regarding scalability, interpretability, and handling highly imbalanced datasets [12].

D. Hybrid and Emerging Methods

With the advancement of machine learning and data privacy techniques, many researchers have started focusing on hybrid and federated approaches that combine multiple learning methods or integrate privacy-preserving technologies such as blockchain and federated learning. These approaches aim to improve scalability, security, and adaptability of fraud detection systems across both UPI and telecommunication domains.

Manisha Julme and Dr. Pankaj Agarkar — "A Study on UPI Fraud Detection Using Blockchain" — proposes a blockchain-augmented UPI monitoring framework that enhances traceability and data integrity by recording transactions in a secure distributed ledger. The study also integrates machine learning screening for identifying suspicious transactions, making it a hybrid model that merges blockchain transparency with intelligent fraud detection. This method demonstrates how blockchain technology can help reduce dependency on centralized verification systems while improving trust and data auditability [13].

Kumrul Hasan, Md Nazmul Hosen, and Kinjol Saha — "Federated Learning for Telecom Fraud Detection: A Privacy-Preserving Approach to Overcoming Data Fragmentation and Enhancing Security" — presents a federated learning-based solution where multiple telecom operators collaboratively train a shared fraud detection model without sharing raw customer data. This approach helps address the problem of data isolation (often referred to as the *data island problem*) and supports compliance with privacy regulations. The model also improves adaptability across regions while maintaining strong data confidentiality, making it suitable for large-scale telecom fraud detection [14].

L. Dhana Lakshmi, B. Sravanthi, A. Sugun Pandu Raju, and B. Gayathri Sadvika — "Hybrid Deep Learning Model for UPI Fraud Detection Using CNN and Random Forest" — describes a hybrid pipeline that combines deep learning and traditional machine learning models. In this work, CNN or LSTM networks are used to automatically learn complex transaction patterns, while Random Forest performs the final classification. The combined model achieves improved recall and reduced false positives compared to single-model baselines. The study highlights how hybrid deep-learning architectures can better capture temporal and behavioral fraud patterns in UPI transactions [15].

Bhraman Sethi, Sarvednya Mhatre, Sachin Yadav, Siuli Das, and Vaishali Jadhav — "Machine Learning-Based UPI Fraud Detection: A Comprehensive Approach Using Random Forest" — focuses on a real-time, hybrid fraud detection system that combines both supervised and unsupervised learning techniques. The proposed model emphasizes continuous monitoring and includes an alert-based user interface for instant notifications of fraudulent activities. This study demonstrates the importance of combining multiple learning paradigms to enhance detection accuracy and system responsiveness in real-world applications [16].

Amogh Deshmukh, Peplluis Esteva de la Rosa, Raul Villamarin Rodriguez, and Sandeep Dasari — "Enhancing Privacy in IoT-Enabled Digital Infrastructure: Evaluating Federated Learning for Intrusion and Fraud Detection" — evaluates the use of federated learning frameworks for detecting fraud and intrusion in distributed IoT systems. The paper discusses how federated models can collaboratively learn across devices and organizations while keeping data localized, thus improving privacy and scalability. This work provides valuable insights for designing next-generation fraud detection frameworks that are both secure and adaptable to evolving digital infrastructures [17].

V. COMPARATIVE ANALYSIS

This section compares the surveyed studies in terms of data domain, learning approach, algorithms used, evaluation metrics, and key observations. The goal is to highlight performance trends, methodological similarities, and gaps across UPI and telecommunication fraud detection research.

A. Comparative Overview of Reviewed Studies

Domain / Paper	Algorithm (s) Used	Learning Type	Dataset Type	Metrics Evaluated	Key Findings / Contributions
Hilal et al. (2023) [1]	Autoencoders, Isolation Forest, Clustering	Unsupervised / Semi-Supervised	Generic financial datasets	Accuracy, Recall, Latency	Provided anomaly-detection taxonomy across domains; emphasized unsupervised methods for emerging frauds.
Flegel et al. (2022) [2]	Statistical, ML, Hybrid	Mixed	Multi-domain (finance + telecom)	Comparative conceptual analysis	Grouped fraud detection into three categories—statistical, ML, hybrid—and discussed adaptability.
Kou et al. (2004) [3]	Rule-based, Neural Net, Outlier detection	Mixed	Credit-card, telecom, intrusion	Accuracy, Detection rate	Early unified taxonomy; introduced cost-sensitive learning and cross-domain perspective.
Jagadeesan et al. (2024) [4]	Random Forest, SVM, LR	Supervised	UPI transaction dataset	Accuracy, Precision, Recall	Random Forest gave best accuracy; stressed preproces

					sing and feature extraction.
Kamble et al. (2024) [5]	Stacked Generalization (Ensemble)	Hybrid Supervised	UPI data (synthetic + real)	F1-score, AUC-ROC	Stacked ensemble improved precision-recall; discussed class-imbalance mitigation.
Bodade & Pawade (2024) [6]	Decision Tree / RF pipeline	Supervised	UPI transaction logs	Precision, Recall, ROC	Built a real-time UI system linking detection with visualization.
Dave & Chudasama (2024) [7]	XGBoost, LightGBM, RF	Supervised / Ensemble	Multi-institution UPI data	Accuracy, F1, ROC-AUC	Gradient boosting outperformed others; emphasized geolocation and device features.
Kavya & Usha Sree (2024) [8]	CNN, Ensemble suggestion	Deep Learning	UPI dataset	Accuracy, Recall	CNNs superior to rule-based models; proposed ensemble-CNN hybrid.
Jabbar & Suharjito (2024) [9]	K-Means, DBSCAN	Unsupervised	CDR dataset (Telecom)	Silhouette, Detection rate	K-Means outperformed DBSCAN; clustering viable for fraud filtering.
Daka (2023) [10]	ANN	Supervised / Deep	CDR synthetic data	Accuracy	ANN reached 100% accuracy (small dataset); shows NN potential in telecom.
Bhavani et al. (2024) [11]	LSTM, Multinomial NB	Deep + NLP	Text / Call transcripts	Accuracy, F1, Precision	LSTM outperformed Naive Bayes; Flask app for real-time fake-call detection.
Yehya & Salhab (2024) [12]	Random Forest, Logistic Regression	Supervised	Telecom operator data	Accuracy, F1, Confusion Matrix	Random Forest best performer; noted scalability and imbalance issues.

B. Analytical Discussion

1. Learning Paradigms
   Most UPI-related papers rely on supervised classification, leveraging structured and labeled transaction data. In contrast, telecom-fraud research often employs unsupervised or deep learning due to limited labeled CDRs.

2. Model Performance Trends

   - Ensemble methods (XGBoost, Stacked Generalization) consistently outperform single classifiers by reducing overfitting and improving recall.

   - Deep networks (CNN, LSTM) achieve high accuracy when sufficient sequential or text data are available but require larger datasets and computational power.

   - Unsupervised clustering provides interpretable anomaly indicators for unlabeled telecom data but suffers from instability in dynamic environments.

3. Data and Feature Challenges
   UPI datasets emphasize transactional variables such as amount, frequency, and beneficiary history, whereas telecom datasets focus on temporal and network-usage statistics. Across both domains, class imbalance and data privacy remain persistent obstacles.

4. Evaluation and Metrics
   Accuracy alone is insufficient in imbalanced scenarios; hence, most effective studies prioritize Precision, Recall, F1-Score, and AUC-ROC to measure discriminatory performance. Few works incorporate cost-based metrics, leaving room for improvement in risk-aware evaluation.

5. Implementation Readiness

- Papers like *Bodade & Pawade* and *Bhavani et al.* advance beyond theoretical modeling by developing deployable systems (Flask or UI-based real-time detection).
- Telecom studies highlight operational scalability issues, emphasizing the need for low-latency detection pipelines capable of handling continuous CDR streams.

6. Cross-Domain Insights
While both UPI and telecom domains face distinct data structures, their fraud patterns exhibit similar behavioral anomalies—frequency bursts, location shifts, or inconsistent device identities—suggesting that hybrid multi-modal models could generalize across domains.

C. Summary of Comparative Findings

- Dominant Approaches: Supervised learning for UPI; unsupervised and neural for telecom.
- Best-Performing Algorithms: XGBoost > Random Forest > SVM > DBSCAN.
- Emerging Techniques: CNN/LSTM and ensemble stacking show strong potential.
- Common Weaknesses: Lack of real-world labeled datasets, imbalance handling, and explainability.
- Research Need: Unified frameworks integrating financial and telecommunication fraud signals for holistic risk scoring.

VI. OPEN ISSUES AND RESEARCH CHALLENGES

Even though many research papers have tried to solve UPI and telecommunication fraud detection using different machine learning and deep learning techniques, there are still many problems and open areas that need improvement. Some of the main challenges are discussed below.

A. Lack of Real Datasets

One of the biggest issues is that real-world data are not available publicly.
Most studies use synthetic or limited private datasets, since real UPI transactions and telecom call data (CDRs) are confidential and cannot be shared due to privacy reasons. Because of this, it becomes hard to compare models fairly or test how they would work in real-life systems.
Future research should focus on creating shared or anonymized datasets and on privacy-preserving learning techniques.

B. Imbalanced and Changing Data

In both UPI and telecom data, fraud cases are very few compared to genuine ones.
This data imbalance causes models to predict “non-fraud” most of the time.
Researchers use sampling methods like SMOTE or weight adjustments to fix this, but these methods don’t always work in dynamic situations.
Also, fraud patterns keep changing (concept drift) — what works today may fail tomorrow — so models need to be updated frequently.

C. Feature Engineering and Model Explainability

Selecting the right features is very important.
For example, in UPI fraud detection, features like transaction frequency, device ID, and beneficiary behavior matter, while in telecom fraud, things like call duration, location, or cost are more useful.
Many deep learning models work well but are hard to explain.
In real organizations, decision-makers want to know *why* a call or transaction was marked as fraud.
Hence, research should focus more on explainable AI models that are both accurate and transparent.

D. Real-Time Implementation

Most of the reviewed papers tested models on offline datasets, but in reality, fraud detection should happen instantly.
Deploying models for real-time detection is difficult because it needs fast computation and low latency.
There is also a need for lightweight and optimized algorithms that can work in mobile or cloud systems for immediate alerts.

E. Combining UPI and Call Data

Another gap is that UPI and call frauds are studied separately.
However, many real scams are connected — for example, a spam or phishing call may lead to a fake UPI transaction.
Future research should try to combine data from both domains to build a more complete fraud-detection system.
Techniques like hybrid models or graph-based learning could help connect behaviour patterns from both areas.

F. Evaluation and Standardization

There is no common standard for evaluating models.
Different studies use different datasets and metrics (accuracy, F1, AUC, etc.), making results hard to compare.
The research community should try to build benchmark datasets and fixed evaluation criteria so future models can be compared more fairly.

G. Privacy and Ethical Concerns

Fraud-detection systems deal with sensitive personal and financial data.
If not handled carefully, it can cause data leaks or bias.
It’s important to include privacy protection methods (like data encryption or anonymization) and also make sure

models don't unfairly target certain users. Future work should consider both accuracy and fairness.

H. Industry Adoption

Many academic models show high accuracy but are not used in real companies. This is because integration with existing systems, regulatory approvals, and maintenance are difficult. There is a need for collaboration between researchers, companies, and regulators to make these solutions more practical and deployable.

VII. CONCLUSION

In this survey paper, we studied and compared different research works related to UPI transaction fraud and telecommunication (call) fraud detection. The main goal was to understand how various machine learning and deep learning techniques are being used to identify fraudulent behavior in these two important domains.

The reviewed papers showed that supervised learning methods, such as Random Forest and XGBoost, are widely used for UPI fraud detection because of the availability of labeled transaction data. On the other hand, unsupervised and deep-learning approaches like K-Means, ANN, CNN, and LSTM are more common in call fraud detection where labeled data are limited. Ensemble and hybrid models also showed better accuracy and generalization compared to single algorithms.

From the analysis, it is clear that while these models perform well on small or controlled datasets, there are still many challenges such as lack of real datasets, data imbalance, model explainability, and real-time implementation. Another key observation is that most studies treat UPI and telecom fraud as separate areas, even though they are often connected in real-world scenarios.

To move forward, future research should focus on creating combined frameworks that can analyze both UPI and call data together, use explainable AI models, and ensure privacy-preserving and fair detection. By solving these challenges, the next generation of fraud-detection systems can become more accurate, adaptive, and practical for real-time use in financial and telecommunication industries.

REFERENCES

[1] W. Hilal, S. A. Gadsden, and J. Yawney, "Financial Fraud: A Review of Anomaly Detection Techniques and Recent Advances," *Expert Systems with Applications*, Elsevier, 2023. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0957417421017164

[2] U. Flegel, J. Vayssière, and G. Bitz, "A State of the Art Survey of Fraud Detection Technology," *ResearchGate*, 2022.

[3] Y. Kou, C.-T. Lu, S. Sinvongwattana, and Y.-P. Huang, "Survey of Fraud Detection Techniques," *Department of Computer Science, Virginia Polytechnic Institute and State University; Tatung University, Taipei*, 2004.

[4] S. Jagadeesan, K. S. Arjun, G. Dhanika, G. Karthikeyan, and K. Deepika, "UPI Fraud Detection Using Machine Learning," *ResearchGate Preprint*, 2024. [Online]. Available: https://www.researchgate.net/publication/385968094_UPI_fraud_detection_using_machine_learning

[5] V. B. Kamble, K. Pisal, P. Vaidya, and S. Gaikwad, "Enhancing UPI Fraud Detection: A Machine Learning Approach Using Stacked Generalization," *IJMSM*, 2024.

[6] S. S. Bodade and P. P. Pawade, "Implementation Paper on UPI Fraud Detection Using Machine Learning," *Journal of Emerging Technologies and Innovative Research (JETIR)*, vol. 11, no. 4, pp. 1450–1455, Apr. 2024. [Online]. Available: https://www.jetir.org/papers/JETIR2404A10.pdf

[7] V. Dave and D. Chudasama, "Fraud Detection in UPI Transactions Using Ensemble Learning," *ResearchGate*, 2024. [Online]. Available: https://www.researchgate.net/publication/391399357_Fraud_Detection_In_UPI_Transactions_Using_Ensemble_Learning

[8] K. K. R. and U. Sree, "UPI Fraud Detection Using Machine Learning," *International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)*, vol. 4, no. 6, pp. 97–104, Nov. 2024. [Online]. Available: https://www.ijarsct.co.in/Paper22521.pdf

[9] M. A. Jabbar and Suharjito, "Fraud Detection Call Detail Record Using Machine Learning in Telecommunications Company," *Advances in Science, Technology and Engineering Systems Journal (ASTESJ)*, vol. 5, no. 4, pp. 123–130, 2024. [Online]. Available: https://www.astesj.com/publications/ASTESJ_050409.pdf

[10] J. C. Daka, "Smart Mobile Telecommunication Network Fraud Detection System Using Call Traffic Pattern Analysis and Artificial Neural Network," *ResearchGate*, 2023. [Online]. Available: https://www.researchgate.net/publication/370245266

[11] B. D. Bhavani, U. Nikitha, P. Nandini, and N. R. Gogu, "Artificial Intelligence Based Fake or Fraud Phone Calls Detection," *Dialnet Journal*, 2024. [Online]. Available: https://dialnet.unirioja.es/descarga/articulo/9906272.pdf

[12] B. A. Yehya and N. Salhab, "Telecommunications Fraud Machine Learning-Based Detection," *ResearchGate*, 2024. [Online]. Available: https://www.researchgate.net/publication/383177588_Telecommunications_Fraud_Machine_Learning-based_Detection

[13] M. Julme and P. Agarkar, "A Study on UPI Fraud Detection Using Blockchain," *Journal of Emerging Technologies and Innovative Research (JETIR)*, vol. 12, no. 4, pp. 245–250, Apr. 2025. [Online]. Available: https://www.jetir.org/papers/JETIR2504D43.pdf

[14] K. Hasan, M. N. Hosen, and K. Saha, "Federated Learning for Telecom Fraud Detection: A Privacy-Preserving Approach to Overcoming Data Fragmentation and Enhancing Security," *Semantic Scholar Preprint*, 2024. [Online]. Available: https://pdfs.semanticscholar.org/39af/3ceab9065897838df10edfadfce35516fa92.pdf

[15] L. D. Lakshmi, B. Sravanthi, A. S. P. Raju, and B. G. Sadvika, "Hybrid Deep Learning Model for UPI Fraud Detection Using CNN and Random Forest," *ResearchGate Preprint*, 2025. [Online]. Available: https://www.researchgate.net/publication/390227261_HYBRID_DEEP_LEARNING_MODEL_FOR_UPI_FRAUD_DETECTION_USING_CNN_AND_RANDOM_FOREST

[16] B. Sethi, S. Mhatre, S. Yadav, S. Das, and V. Jadhav, "Machine Learning-Based UPI Fraud Detection: A Comprehensive Approach Using Random Forest," *Atlantis Press Proceedings*, 2025. [Online]. Available: https://www.atlantis-press.com/article/126016569.pdf

[17] A. Deshmukh, P. E. de la Rosa, R. V. Rodriguez, and S. Dasari, "Enhancing Privacy in IoT-Enabled Digital Infrastructure: Evaluating Federated Learning for Intrusion and Fraud Detection," *Sensors*, vol. 25, no. 10, art. 3043, 2025. [Online]. Available: https://www.mdpi.com/1424-8220/25/10/3043

[18] Coinlaw.io, "UPI vs ATM Transactions Statistics — July 2025," 2025. [Online]. Available: https://coinlaw.io/upi-vs-atm-transactions-statistics/

[19] BFSI Elets Online, "UPI achieves record 613 million daily transactions in June 2025 as digital payments surge," 2025. [Online]. Available:

https://bfsi.eletsonline.com/upi-achieves-record-613-million-daily-transactions-in-june-2025/

[20] RMA India, "India's cyber fraud losses soar 206% to ₹22,845 crore from 2024," 2025. [Online]. Available: https://rmaindia.org/indias-cyber-fraud-losses-soar-206-to-%E2%82%B922845-crore-from-2024/

[21] Press Information Bureau (PIB), "Cybercrime cases and telecommunication-related frauds in India (2024 update)," Government of India, 2024. [Online]. Available: https://www.pib.gov.in/PressNoteDetails.aspx?ModuleId=3&NoteId=155384

[22] Reserve Bank of India (RBI), "Annual Report on Digital Payments and Cyber Security," 2024. [Online]. Available: https://rbi.org.in

This paper is an original synthesis of existing research. All external sources have been properly cited, and no part of this work has been copied verbatim.