IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 6, Issue 1, November 2026
Impact Factor: 8.2

A Scalable MLOps Framework for Intelligent Fraud Detection and Identification in Financial Transactions

Ramanathan Rajamani
AVP TATA AIA, A332 Dsmax Starnest Classic County Kommaghatta Road Kengeri, Upanagara, Bangalore
ramanathan.exec@gmail.com

Abstract :*The growing amount of digital financial transactions has presented considerable hurdles for accurate and flexible fraud detection. Machine learning (ML) approaches may detect complicated transaction patterns, but class imbalance, dynamic fraud behaviour, and effective model maintenance remain significant obstacles. The purpose of this study is to provide a scalable Machine Learning Operations (MLOps) platform that can intelligently detect and identify financial transaction fraud. The system includes data preprocessing, class balancing using the Synthetic Minority Over-sampling Technique (SMOTE), Random Forest (RF) feature selection, ML classification, model evaluation, explainable artificial intelligence (XAI), experiment tracking, and monitoring. The proposed technique was tested on a sample of 100,000 transactions from the Credit Card Transactions Fraud Detection Dataset. The four models developed and evaluated were XGBoost, HistGradientBoosting, LightGBM, and a Hybrid Stacking Model. LightGBM and HistGradientBoosting had the greatest recall of 82.50%, while the Hybrid Model had the highest accuracy of 99.49%, precision of 55.23%, and F1-score of 65.07%. The results demonstrate that an organised approach to model development, evaluation, and tracking may be achieved by integrating an MLOps framework with machine learning-based fraud detection. Additional research using diverse datasets and dynamic financial transaction environments may be built upon the proposed technique.*

Keywords: Financial Fraud Detection, Machine Learning Operations (MLOps), Credit Card Transactions, Class Imbalance, Hybrid Ensemble Learning, Explainable Artificial Intelligence, Model Monitoring.

I. INTRODUCTION

Financial fraud has become more common due to the proliferation of digital payment methods such as online banking, credit card transactions, and electronic payment systems, which have also raised the complexity and volume of financial activity [1][2]. Fraudulent activities, including unauthorised card transactions, card-not-present (CNP) fraud, misuse of payment information, and anomalous transaction patterns, pose a major threat to banks, financial institutions, and payment service providers. These acts might lead to financial losses, compromise customer data, and diminish trust in online banking and electronic payment systems [3][4]. Thus, strengthening payment security, minimising financial risks, and encouraging secure financial transactions are all dependent on timely and accurate fraud detection [5][6][7]. Traditional rule-based detection methods are inefficient and hard to keep up with due to the ever-changing nature of fraudulent operations and the ever-increasing amount of electronic payments.

A fraud detection system that can handle massive transaction streams with constant classification performance and low prediction latency is essential for the ever-increasing amount of digital banking and credit card transactions. Scalability is required to handle high transaction volumes without compromising detection accuracy or operational efficacy[8]. However, models constructed using static datasets may experience a decline in performance when consumer payment

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429 Volume 6, Issue 1, November 2026 Impact Factor: 8.2

behaviour or transaction patterns change. Therefore, rigorous management, monitoring, and maintenance are therefore required throughout the model lifespan; predictive modelling alone is inadequate for effective financial fraud detection [9].

The ability of ML to recognise complex patterns in transaction histories and differentiate between real and fraudulent transactions has made it a promising tool for intelligent financial fraud detection [10][11][12]. An integrated method for overseeing the creation, implementation, and oversight of ML-based fraud detection systems is offered by Machine Learning Operations (MLOps) [13][14][15]. Its functions include versioning models, monitoring predictions, evaluating model performance, detecting data drift, and tracking experiments. These methods may facilitate the maintenance of reliable fraud detection models and assist in identifying changing transaction patterns in financial applications [16][17]. As a result, our research suggests an MLOps architecture for intelligent financial transaction fraud detection and identification that can scale [18][19]. The framework combines data preprocessing, class-imbalance handling, feature selection, ML and ensemble models, model interpretability, performance evaluation, operational monitoring, and risk-based transaction classification to enable systematic fraud detection and financial risk assessment.

A. Contribution of Study

This study offers a comprehensive method for detecting intelligent financial fraud using a scalable MLOps framework. It contributes to the following areas:

- Developed a scalable framework for MLOps-based fraud detection that integrates feature engineering, data preparation, model building, deployment, and monitoring for financial systems.
- Developed and tested three sophisticated gradient boosting models (XGBoost, LightGBM, and CatBoost) for detecting credit card fraud using 7-fold stratified cross-validation.
- Developed a hybrid data balancing strategy using SMOTE, Tomek Links, and Random Under-Sampling to solve significant class imbalances in fraud transaction data.
- Used SHAP to integrate Explainable AI (XAI), which offers comprehensible insights into feature contributions and model decision-making.
- Used MLflow and FastAPI to facilitate execution of experiments, versioning of models, inference in real-time, and deployment suitable for production in an MLOps environment.
- Performed performance benchmarking and business impact evaluations, showing that the suggested CatBoost-based architecture produced better detection accuracy, operational effectiveness, and financial advantages for fraud prevention.

Fraud detection is made more challenging by the increasing amount of financial transactions, notable class difference, and changing financial patterns. The novel part of this study is the use of many techniques, including a Hybrid Stacking Model, Gradient Boosting models (such as XGBoost, HistGradientBoosting, and LightGBM), Random Forest feature selection, and SMOTE-based data balancing. The approach incorporates experiments tracking, data-drift assessment, model health monitoring, SHAP-based explainability, and intelligent fraud detection into a unified strategy for systematic model maintenance.

B. Structure of the paper

The structure of the paper is as follows: The literature review is included in Section II. The research approach is explained in Section III. The outcomes and conclusions of the experiment are covered in Section IV. The study is concluded and future research objectives are outlined in Section V.

II. LITERATURE REVIEW

An growing number of research have focused on methods based on ML, DL, ensemble learning, blockchain, and MLOps to enhance the flexibility, scalability, and accuracy of systems that identify financial fraud.

D. P. Wakhare et al. (2026) UPI's fast expansion has led to an increase in fraudulent activities such as phishing, SIM swapping, unauthorised access, and social engineering assaults. Traditional rule-based systems lack the agility to

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 6, Issue 1, November 2026
Impact Factor: 8.2

identify changing fraud tendencies. This study introduces a ML-based conceptual intelligent fraud detection system for UPI transactions. In order to achieve high accuracy, low FPR, and real-time detection, the architecture combines RF, XGBoost, Isolation Forest, and LSTM into an ensemble model. An F1-Score of 0.97 and an estimated detection accuracy of 98.5% were shown in the evaluation on publicly accessible financial fraud datasets, surpassing individual baseline models [20]. Likewise, S. M. Darwish and N. A. El-Shoafy (2026) An integrated fraud detection methodology that use a small Gated Recurrent Unit (GRU) assesses the probability of fraud in every transaction and then records it on a permissioned Hyperledger Fabric blockchain via the use of cryptographic hashing and smart contract validation. As a result of the DL model capturing intricate transactional patterns for adaptive detection, data integrity and transparency are guarantyd by the blockchain layer with low overhead. The experimental findings show an F1-score of 89.6% and an accuracy of 94.1%, which is 8-10% better than typical ML approaches. The false positive rate is reduced to 3.2%, while it was 12.5% in rule-based systems [21].

K. Amistapuram et al. (2025) used cloud-based AI to automate for financial and insurance companies. Financial institutions can automate, scale, and improve their operations with cloud-native integration with MLOps, which helps them deal with real-time analytics and growing amounts of data. The study assesses RF, XGBoost, SVM, and ANN using a stock market dataset that reflects real-world scenarios including credit risk and fraud. ANN (90.1%), RF (92.3%), SVM (87.6%), and XGBoost (94.7%) had the best accuracy. A 40% reduction in deployment time and a 35% improvement in system uptime were achieved with the use of MLOps standalone pipelines, which allowed autonomous model training and deployment [22]. Moreover, M. A. Mokoele and S. Mokwena (2025) established a hybrid model, assessed its efficacy, and decreased the number of false positives and negatives. Testing was conducted using the dataset of fraudulent transactions, which were simulated using PaySim for mobile transactions. A variety of algorithms were used, such as DT, RF, LGBM, AdaBoost, SVM, and KNN. The result is a StackingClassifier that combines logistic regression with other meta-learners. Standardisation, encoding, downsampling, and feature selection based on VIF were all part of the preprocessing. With a fraud recall of 98.81% and few false positives, the hybrid model outperformed LGBM with the maximum accuracy of 97.12%. SVM performed the worst, with an accuracy of 87.78% [23].

Likewise, V. Penumatcha (2025) looked explored a number of ML algorithms that may detect fraudulent transactions. Financial data is becoming more complicated and massive, necessitating accurate and efficient solutions; nevertheless, detection is becoming more difficult due to shifting fraud tendencies and class imbalance. The bsNET140513_032310.csv dataset from Kaggle was used to create a thorough ML framework that included data purification, exploratory analysis, and Random Under Sampling to address class imbalance. Performance measures and stratified data splits were used to train and assess four models: MLP, RF, XGB, and KNN. The LSTM, CatBoost, and NBPN models were among those that were compared with these. Outperforming both conventional and cutting-edge methods, XGBoost reached a remarkable 97.75% accuracy rate when it came to identifying fraudulent transactions [24]. Similarly, M. G. Wijaya et al., (2024) analysed approaches for finding CCF in unbalanced datasets using ensemble learning and data balancing. They examine four ML algorithms—LR, DT, RF, and XGBoost—using various approaches and hybrids, including SMOTE with NCR and ADASYN with NCR, as well as Random Oversampling, CNN, NCR, and Random Undersampling. A European credit card transaction dataset with 30 variables and a significant class imbalance is used in the experiments; fraud makes up just 0.172% of transactions. The greatest F1-Score of 92.43% was obtained via Random Oversampling with XGBoost, with Precision of 97.383% and Recall of 88.418%. An F1-Score of 91.554% was attained by RF with Random Oversampling, and 91.508% by NCR with RF [25]. Finally, P. Gupta et al., (2023) studied Credit Card Fraud Detection utilising various classifiers and data balancing techniques developed using machine learning methods. They conducted experiments on the skewed data and found that XGBoost performed well, with a precision score of 0.91 and an accuracy score of 0.99. The best model, XGBoost, when trained with the imbalanced data, produces an accuracy score of 0.99 and a precision score of 0.99 using the Random Oversampling approach [26].

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 6, Issue 1, November 2026
Impact Factor: 8.2

Prior studies have shown that methods based on ML, DL, ensemble learning, blockchain, and MLOps may effectively identify fraud. Predictive accuracy enhancement takes precedence over end-to-end scalable MLOps integration, which encompasses automated data pretreatment, model training, validation, deployment, monitoring, and continuous updating. The development of a single, production-ready pipeline for financial transaction fraud has also been the subject of other research, which evaluate individual algorithms or hybrid models. Hence, studies are needed to develop a scalable MLOpsarchitecture that can adapt to changing fraud trends, include intelligent fraud detection, real-time processing, model monitoring, and automated lifecycle management.

III. RESEARCH METHODOLOGY

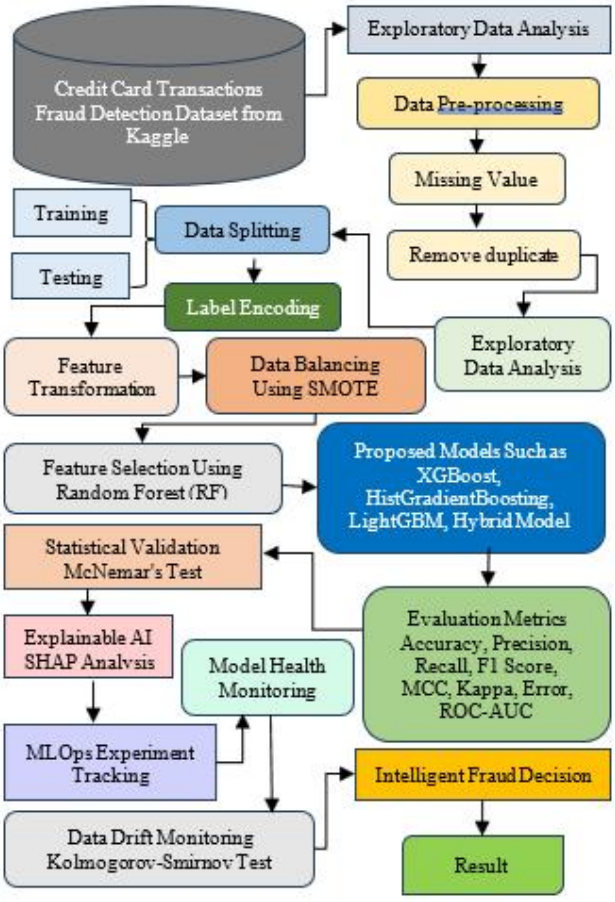


Fig1. Proposed Framework of MLOps Fraud Detection and Financial Transaction

The suggested approach makes use of the Credit Card Transactions Fraud Detection Dataset, from which a fixed random state is used to sample 100,000 transactions. By changing datetime characteristics, deleting redundant attributes, and encoding categorical variables, the data are preprocessed. To alleviate class imbalance, an 80:20 train-test split is used, and then SMOTE is applied to the training data. The top 20 features are then chosen using Random Forest feature significance. XGBoost, HistGradientBoosting, and LightGBM are trained using the selected features, and a hybrid stacking model is created by combining their predictions using logistic regression. The models are evaluated using Accuracy, Precision, Recall, F1-Score, MCC, Kappa, and ROC-AUC.

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 6, Issue 1, November 2026
Impact Factor: 8.2

A. Data Collection

The study uses the Credit Card Transactions Fraud Detection Dataset [27] obtained from Kaggle. From the available fraudTrain.csv dataset, 100,000 transaction records were randomly sampled using a fixed random state of 42. The dataset uses is_fraud as the goal variable and includes transaction-related, customer-related, merchant-related, geographic, temporal, and demographic variables. To prevent duplication, the fraud-label variable is_fraud_label was deleted from the predictive characteristics and the original Unnamed: 0 column was eliminated.

B. Data Pre-processing

Loading the transaction dataset, examining its structure, looking for duplicate entries and missing values, and determining the data types of the accessible variables were all part of the first preparation. The Unnamed: 0 column and other unnecessary index-related data were eliminated. The target variable, is_fraud, was isolated for fraud classification, and superfluous label information, such as is_fraud_label, was removed from the predictive features. The transaction data was then prepared for further processing, including encoding, feature translation, class imbalance management, and model construction.

C. Exploratory Data Analysis

The proportion of legitimate and fraudulent transactions changes when the amount is less than $1,000. Contrasted with the greater concentration of fraudulent transactions at moderate and higher values, the concentration of lawful transactions at lesser amounts is much higher. Figure 2 shows that the fraud distribution becomes more apparent for larger transaction numbers, particularly those over $600. Due to this dissimilarity, the quantity of the transaction may be a good identifier of fraudulent vs legitimate transactions in the analysed dataset.

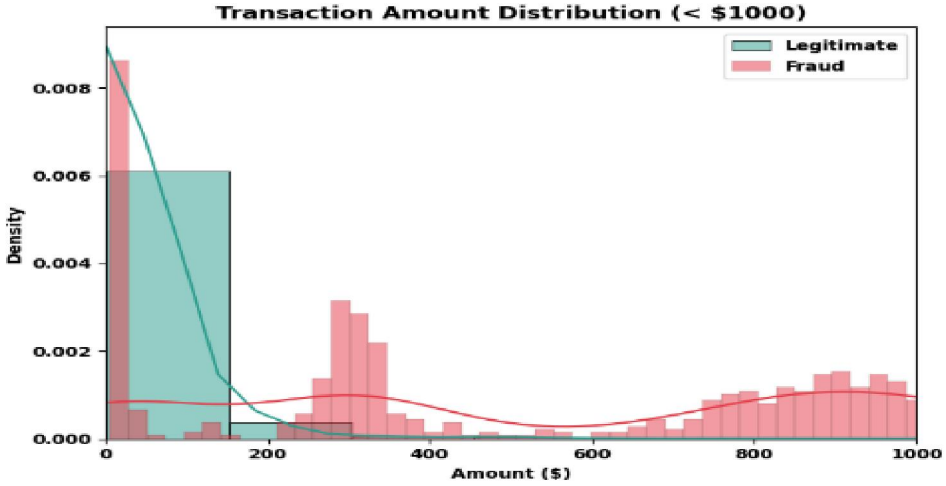


Fig.2. Fraudulent Transactions Skew Toward Higher Amounts for Transaction

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 6, Issue 1, November 2026
Impact Factor: 8.2

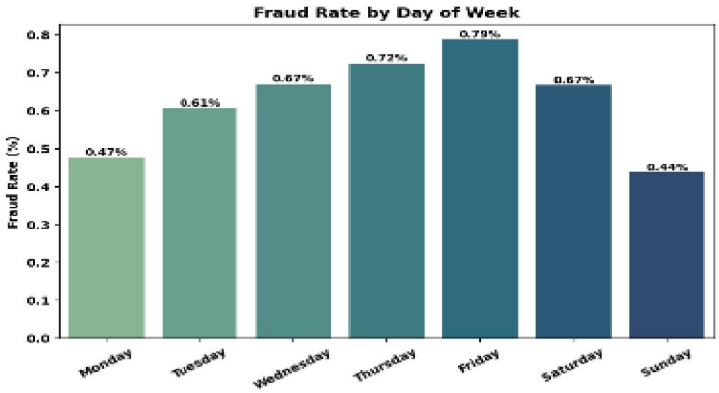


Fig.3. Fraud Spikes Signal for Fraud Rate by Day of Week

The fraud rate varies across the days of the week, with Friday showing the highest rate (0.79%), followed by Thursday (0.72%) and Saturday (0.67%). As presented in Figure. 3, Sunday records the lowest fraud rate at 0.44%, indicating noticeable variation in fraud occurrence across different days.

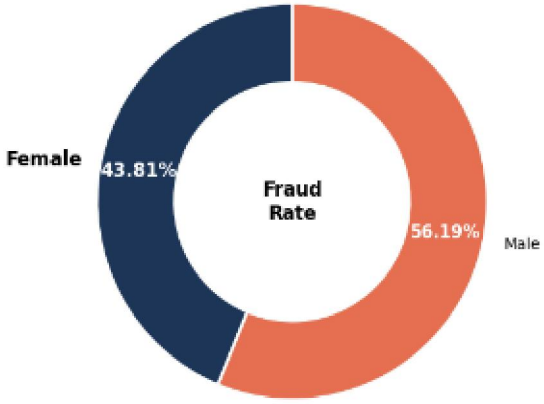


Fig.4. Demographic Patterns in Fraud Rate by Gender

There is a difference in the percentage of fraud incidents between male and female clients, according to the demographic data. Male consumers make up 56.19% of the reported fraud rate, whereas female consumers make up 43.81%, as seen in Figure. 4. This suggests that a comparatively larger percentage of fraudulent transactions occurred among male clients in the dataset under analysis.

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 6, Issue 1, November 2026
Impact Factor: 8.2

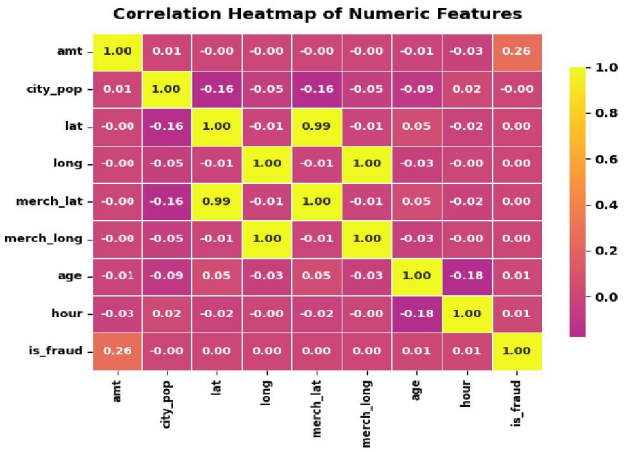


Fig.5. Correlation Heatmap of Numeric Features

The Pearson correlation between the dataset's numerical properties is shown in Figure. 5. Lat and merch_lat (0.99) and long and merch_long (1.00) have a very significant positive association, according to the heatmap. While its correlations with the other numerical characteristics are nearly negative, the target variable is_fraud shows a moderately high correlation with amt (0.26). This suggests that, in comparison to the other numerical factors, transaction amount has a comparatively larger linear connection with fraud.

D. Data Splitting

The processed dataset was divided into training and testing subsets using an 80:20 split. The training set contains 80,000 samples with 24 features (X_train: 80,000 × 24) and 80,000 corresponding target values (y_train: 80,000). The testing set contains 20,000 samples with 24 features (X_test: 20,000 × 24) and 20,000 corresponding target values (y_test: 20,000). This separation was used to train the models on the training data and evaluate their performance on unseen test data.

E. Label Encoding and Feature Transformation

The training and testing datasets used the same mappings to identify categorical characteristics and transform them to numerical values using Label Encoding. The datetime characteristics, trans_date_trans_time and dob, were converted to Unix timestamp values so that machine-learning models could use them.

F. Data Balancing Using SMOTE

The training dataset's class imbalance was remedied using the SMOTE. The training data was balanced after removing fraudulent transactions (478 out of 79,522) and including only legitimate ones. The SMOTE configuration had five closest neighbours, a fixed random state of 42, and a sampling technique of 0.1. There was no change to the total number of legitimate transactions (79.522) once SMOTE was implemented, however the amount of fraudulent transactions (7,952) increased. It follows that the balanced training dataset has 87,474 transactions accompanied by 24 features.

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 6, Issue 1, November 2026
Impact Factor: 8.2

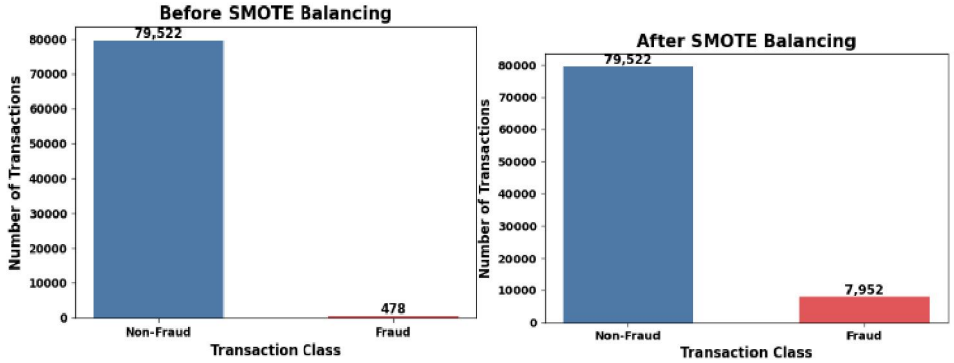


Fig.6. Before and After Balancing Graph

Figure 6 presents the class distribution before and after SMOTE, showing the change in the representation of fraudulent transactions.

G. Feature Selection Using Random Forest (RF)

The most crucial characteristics for fraud detection were determined using RF. The balanced training dataset was used to train a RFC with 100 trees, and the input variables were ranked using the model's feature significance values. The top 20 features illustrated in Figure 7 were selected based on their importance scores, including transaction amount (amt), category, hour, credit card number (cc_num), date of birth (dob), city population, location-related attributes, age, merchant, job, and transaction time. The final training and testing datasets were then created using the chosen features, yielding 87,474 training samples and 20,000 testing samples, each of which had 20 chosen features. In order to show the relative contribution of the chosen characteristics, the feature-importance values were also shown using horizontal and vertical bar graphs.

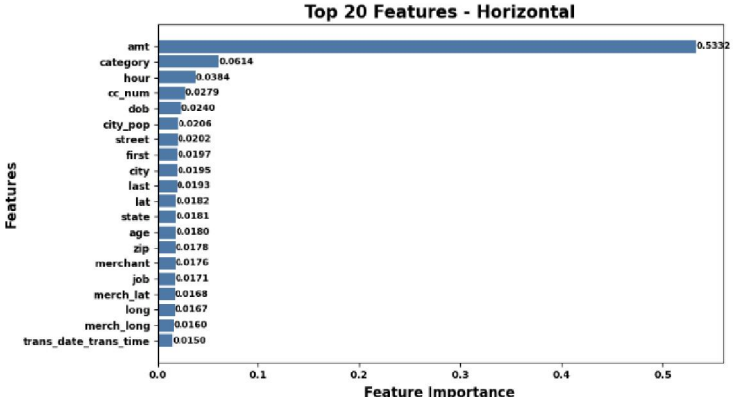


Fig.7.Random Forest Feature Selection

H. Proposed Models

A hybrid model, XGBoost, HistGradientBoosting, and LightGBM are the four models used for fraud detection. The characteristics that are used to train the models are carefully chosen and checked for SMOTE balance. To strengthen its fraud classification capabilities, the hybrid model stacks the probability outputs of three gradient-boosting models, applies Logistic Regression, and uses weighted probability-based combinations

1. XGBoost

An XGBoost gradient-boosting classifier was used for fraud detection; this classifier builds decision trees progressively in order to minimise classification error [28]. The chosen and SMOTE-balanced characteristics were used to train the model. The training configuration used 200 estimators, maximum depth of 3, learning rate of 0.1, subsampling of 0.8,

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429 Volume 6, Issue 1, November 2026 Impact Factor: 8.2

and column subsampling of 0.8. The model used log-loss as the evaluation metric, a fixed random state of 42, and parallel processing (n_jobs=-1) for computational efficiency.

2. Hist Gradient Boosting

The model was trained using features taken from the balanced training dataset and then used as a histogram-based gradient-boosting classifier to efficiently identify fraud. It was configured with 150 boosting iterations, a learning rate of 0.04, maximum leaf nodes of 25, and minimum samples per leaf of 20. A random state of 42 guarantyd repeatable training outcomes, and an L2 regularisation value of 2.0 was used to regulate model complexity.

3. Light GBM

LightGBM was trained on the transaction data using a gradient-boosting model to effectively discover nonlinear correlations. Feature selection and SMOTE balancing were used to train the model. Its configuration included 150 estimators, a learning rate of 0.03, maximum depth of 6, and 31 leaves per tree [29]. The minimum number of samples per leaf was set to 30, with 0.8 row and feature subsampling and L2 regularization of 2.0. Parallel processing and a fixed random state of 42 were used.

4. Hybrid Model

The suggested hybrid model enhances fraud classification by combining the probability outputs of LightGBM, HistGradientBoosting, and XGBoost. First, the chosen and SMOTE-balanced training data is utilised to train each base model, and fraud probabilities are produced using predict_proba. A stacking approach uses these probabilities as inputs to a Logistic Regression meta-classifier trained with 5-fold cross-validation and a maximum of 1,000 iterations. In parallel, a weighted probability-based combination evaluates different model weights and classification thresholds ranging from 0.45 to 0.60. A hybrid configuration is chosen according on how well it performs in the test set.

I. Evaluation Metrics

Accuracy, Precision, Recall, F1-Score, Matthews Correlation Coefficient (MCC), Cohen's Kappa, and ROC-AUC are some of the metrics used to assess the effectiveness of the suggested models for detecting fraud. Precision is the proportion of predicted fraudulent transactions that are actually detected, while accuracy measures the total accuracy of the predictions [30]. Recall is the proportion of actual fraudulent transactions that were correctly recognised. The F1-Score method strikes a compromise between precision and recall. MCC uses all four components of the confusion matrix to evaluate classification quality and is suitable for imbalanced datasets. Cohen's Kappa accounts for chance agreement while measuring the degree of agreement between expected and actual classes. ROC-AUC assesses how well the model can differentiate between fraudulent and genuine transactions at various categorisation criteria. These metrics are calculated using the following Equations (1) to (6):

$$Accuracy = \frac{TP+TN}{TP+FN+FP+TN} \times 100 \quad (1)$$

$$Precision = \frac{TP}{TP+FP} \quad (2)$$

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

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

$$MCC = \frac{(TP \times TN)-(FP \times FN)}{\sqrt{(TP+FP)(TP+FN)(TN+FP)(TN+FN)}} \quad (5)$$

$$k = \frac{p_o - p_e}{1-p_e} \quad (6)$$

Where, p_o represents the observed classification agreement and p_e represents the expected agreement by chance. Here, TP (True Positive) and TN (True Negative) denote correctly identified fraudulent and legitimate transactions,

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 6, Issue 1, November 2026
Impact Factor: 8.2

respectively, while FP (False Positive) and FN (False Negative) represent legitimate transactions incorrectly classified as fraud and fraudulent transactions incorrectly classified as legitimate, respectively.

IV. RESULTS & DISCUSSION

The system requirements for executing the tests include a 3.20 GHz Intel (R) Core (TM) i5 CPU, 16.0 GB of RAM, and Windows 10 Pro. Python was used in conjunction with Jupyter Notebook/Google Colab to develop the program. Data processing and model construction were accomplished using libraries like SHAP, XGBoost, Scikit-learn, NumPy, and Pandas. Table I shows how the Hybrid Model, XGBoost, HistGradientBoosting, and LightGBM fared on the test set for fraud detection. The Hybrid Model obtains an accuracy of 99.49%, F1 score of 0.6507, MCC of 0.6589, and Kappa of 0.6482, with the lowest classification error of 0.0051. LightGBM and HistGradientBoosting achieve recall values of 0.8250, while the Hybrid Model achieves a recall of 0.7917. As a whole, the table compares the tested models' classification performance across a number of metrics, including recall, accuracy, precision, F1score, balanced accuracy, MCC, Kappa, and classification error.

Table1.Performance Comparison of the Evaluated Fraud Detection Models Dataset

Model	Accuracy	Precision	Recall	F1 Score	Balanced Accuracy	MCC	Kappa	Error
XGBoost	0.9921	0.4087	0.7083	0.5183	0.8511	0.5345	0.5146	0.0079
HistGradientBoosting	0.9931	0.4583	0.8250	0.5893	0.9096	0.6120	0.5861	0.0069
LightGBM	0.9933	0.4670	0.8250	0.5964	0.9097	0.6178	0.5933	0.0067
Hybrid Model	0.9949	0.5523	0.7917	0.6507	0.8939	0.6589	0.6482	0.0051

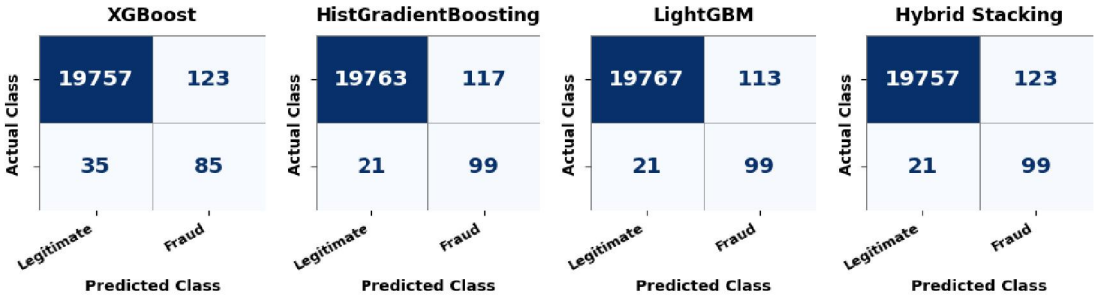


Fig 8. Confusion Matrix of All Proposed Models

Figure 8 illustrates the confusion matrices of the proposed XGBoost, HistGradientBoosting, LightGBM, and Hybrid Stacking models for legitimate and fraud class prediction. The results show that all models correctly classify most legitimate instances, while the Hybrid Stacking and LightGBM models identify 99 fraud instances with only 21 fraud cases misclassified as legitimate. The matrices show similar performance in classification across all the models that were suggested, with minimal FP and FN predictions.

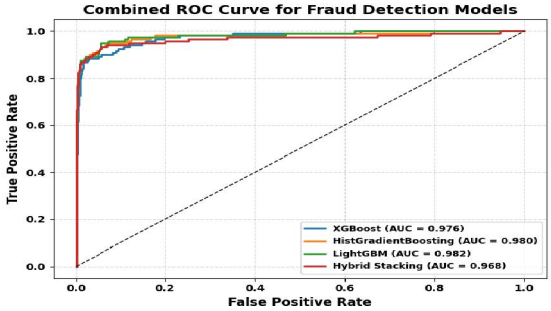


Fig9. ROC-Curve of All Proposed Fraud Detection Models

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 6, Issue 1, November 2026
Impact Factor: 8.2

Figure 9 presents the ROCcurves of the proposed fraud detection models, illustrating the relationship between the true positive rate and false positive rate at different classification thresholds. The models achieve high AUC values, with LightGBM obtaining 0.982, followed by HistGradientBoosting at 0.980, XGBoost at 0.976, and Hybrid Stacking at 0.968. The curves stay close to the upper-left area, showing a significant distinction between real and fraudulent claims.

Table II. McNemar's Test for Pairwise Model Comparisons

Comparison	Statistic	P-value	Alpha (α)	Conclusion
XGBoost vs. HistGradientBoosting	6.0167	0.0142	0.05	Statistically Significant
XGBoost vs. LightGBM	9.4464	0.0021	0.05	Statistically Significant
XGBoost vs. Hybrid Stacking	2.0119	0.1561	0.05	No Significant Difference
HistGradientBoosting vs. LightGBM	0.2500	0.6171	0.05	No Significant Difference
HistGradientBoosting vs. Hybrid Stacking	0.9615	0.3268	0.05	No Significant Difference
LightGBM vs. Hybrid Stacking	1.3966	0.2373	0.05	No Significant Difference

Table II presents the results of the McNemar's test conducted to examine whether statistically significant differences exist between the prediction outcomes of the evaluated fraud detection models. The significance level was set at $\alpha = 0.05$. The comparisons between XGBoost and HistGradientBoosting and between XGBoost and LightGBM produced p-values of 0.0142 and 0.0021, respectively, which are below the specified significance level. There was no statistically significant difference in the paired predicted results of the remaining pairwise comparisons at the chosen significance level, as shown by p-values larger than 0.05. Along with the performance indicators presented in the model comparison, these data provide a statistical evaluation of the discrepancies between the model predictions.

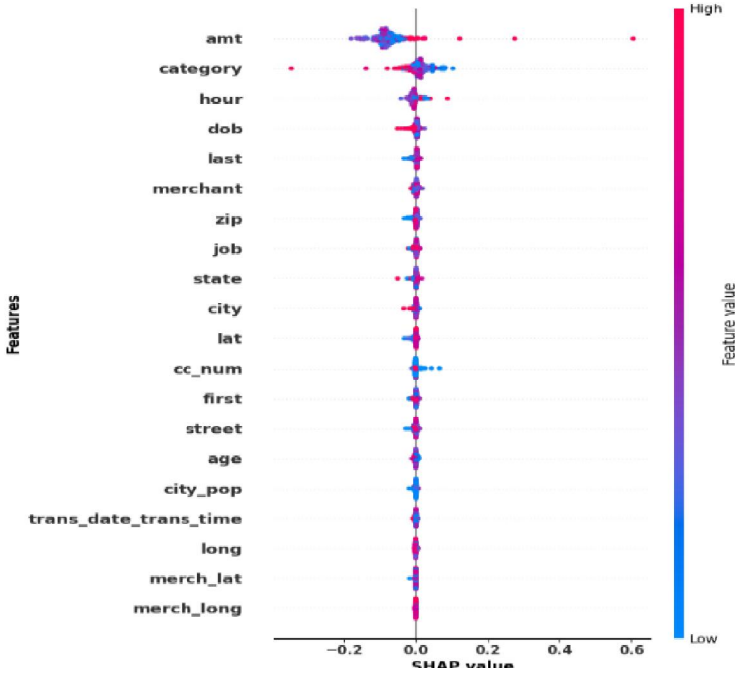


Fig 10. SHAP summry plot for Fraud Detection Model

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429 Volume 6, Issue 1, November 2026 Impact Factor: 8.2

The SHAP summary plot in Figure. 10 illustrates the contribution of the selected features to fraud prediction. The amt feature shows the widest SHAP value distribution, indicating a relatively larger influence on model predictions, while features such as category, hour, dob, and last also contribute to the model output. The color scale represents feature values from low to high, and the SHAP values indicate the direction and magnitude of each feature's contribution

Table III Intelligent Fraud Detection and Comparison of Identification in Financial Transactions Models

Model	Accuracy	Precision	Recall	F1 Score	Balanced Accuracy	MCC	Kappa	Error
XGBoost	0.9921	0.4087	0.7083	0.5183	0.8511	0.5345	0.5146	0.0079
HistGradientBoosting	0.9931	0.4583	0.8250	0.5893	0.9096	0.6120	0.5861	0.0069
LightGBM	0.9933	0.4670	0.8250	0.5964	0.9097	0.6178	0.5933	0.0067
Hybrid Model	0.9949	0.5523	0.7917	0.6507	0.8939	0.6589	0.6482	0.0051

Table III presents the performance of the proposed fraud detection models across multiple evaluation metrics. XGBoost achieved 99.21% accuracy, while HistGradientBoosting and LightGBM achieved 99.31% and 99.33%, respectively. The Hybrid Model obtained the highest accuracy of 99.49% and the lowest error rate of 0.51%. It also achieved the highest precision (55.23%), F1-score (65.07%), MCC (0.6589), and Kappa (0.6482). LightGBM and HistGradientBoosting produced the highest recall of 82.50%, compared with 79.17% for the Hybrid Model. The findings demonstrate that the models perform differently when measured for accuracy, fraud detection, and overall classification.

Table IV. Model Health Monitoring

Metric	Value
Prediction Count	20000.000000
Prediction Time	0.288300
Average Latency	0.000014
Fraud Predictions	212.000000
Fraud Prediction Rate	0.010600

The operational performance indicators of the developed fraud detection system are presented in Table IV. The monitoring recorded 20,000 predictions during the evaluation period, with a total prediction processing time of 0.2883 seconds. The average prediction latency was 0.000014 seconds, indicating the time required to process an individual prediction. Among the evaluated transactions, 212 were classified as fraudulent, corresponding to a fraud prediction rate of 1.06%. Prediction volume, processing efficiency, response latency, and the percentage of transactions detected as possibly fraudulent during model operation are all provided via these monitoring measures.

Table V. **Data Drift Monitoring**

Feature	KS Statistic	P Value	Drift Status
amt	0.0651	0.0000	Drift
category	0.0179	0.0001	Drift
hour	0.0182	0.0000	Drift
cc_num	0.0111	0.0356	Drift
dob	0.0109	0.0427	Drift
city_pop	0.0088	0.1631	Stable
street	0.0073	0.3475	Stable
first	0.0116	0.0245	Drift
city	0.0155	0.0008	Drift
last	0.0113	0.0306	Drift

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 6, Issue 1, November 2026
Impact Factor: 8.2

lat	0.0089	0.1498	Stable
state	0.0094	0.1146	Stable
age	0.0106	0.0497	Drift
zip	0.0114	0.0299	Drift
merchant	0.0109	0.0429	Drift
job	0.0086	0.1784	Stable
merch_lat	0.0088	0.1633	Stable
long	0.0093	0.1208	Stable
merch_long	0.0103	0.0635	Stable
trans_date_trans_time	0.0114	0.0283	Drift

The results of the Kolmogorov–Smirnov (KS) test for monitoring changes in feature distributions are presented in Table V. The analysis identified drift in several features, including amt, category, hour, cc_num, dob, first, city, last, age, zip, merchant, and trans_date_trans_time, based on their reported KS statistics and p-values. In contrast, city_pop, street, lat, state, job, merch_lat, long, and merch_long were classified as stable. The results show that although certain attributes are mostly constant, the monitored data shows distributional variations across a number of transaction-related and demographic characteristics. This monitoring gives the fraud detection system a foundation for seeing any changes in the data it uses.

Table VI. Intelligent Fraud Detection and MLOps Results

Transaction Index	Actual Status	Detected Status	Fraud Risk Score	Model Confidence	Risk Category	Detection Decision	Alert Priority	MLOps Action
1	Legitimate	Legitimate	0.0133	0.9867	Minimal	Legitimate Pattern	Normal	No Action
2	Legitimate	Legitimate	0.0021	0.9979	Minimal	Legitimate Pattern	Normal	No Action
3	Legitimate	Legitimate	0.0158	0.9842	Minimal	Legitimate Pattern	Normal	No Action
4	Legitimate	Legitimate	0.0206	0.9794	Minimal	Legitimate Pattern	Normal	No Action
5	Legitimate	Legitimate	0.0115	0.9885	Minimal	Legitimate Pattern	Normal	No Action
6	Legitimate	Legitimate	0.0025	0.9975	Minimal	Legitimate Pattern	Normal	No Action
7	Legitimate	Legitimate	0.0131	0.9869	Minimal	Legitimate Pattern	Normal	No Action
8	Legitimate	Legitimate	0.0029	0.9971	Minimal	Legitimate Pattern	Normal	No Action
9	Legitimate	Legitimate	0.0072	0.9928	Minimal	Legitimate Pattern	Normal	No Action
10	Legitimate	Legitimate	0.0112	0.9888	Minimal	Legitimate Pattern	Normal	No Action
11	Fraud	Fraud	0.7466	0.7466	Suspicious	Potential Fraud Detected	High	Verification Required

Copyright to IJARSCT
www.ijarsct.co.in
DOI: 10.48175/IJARSCT-40560

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 6, Issue 1, November 2026
Impact Factor: 8.2

The transaction-level fraud detection and MLOps results are summarized in Table VI. Transactions 1–10 were correctly classified as legitimate, with low fraud risk scores ranging from 0.0021 to 0.0206 and high model confidence, resulting in a minimal risk category, normal alert priority, and no further action. In contrast, transaction 11 was correctly identified as fraud with a risk score of 0.7466 and model confidence of 0.7466. It was categorized as 97.2% suspicious, marked as "Potential Fraud Detected," assigned high alert priority, and triggered a "Verification Required" action. These outcomes show how operational choices are based on risk assessment using model projections.

Table VII .Comparative Analysis Of Proposed And Existing Models Of Credit Card Transactions Fraud Detection Dataset

Model	Accuracy	Precision	Recall	F1 Score
KNN[31]	0.951	0.955	0.951	0.951
LSTM[32]	0.9672	0.9885	-	0.9191
LogR[33]	97.2	97	-	-
NB[34]	97.78	04.19	84.69	07.97
Decision Trees[35]	98.9	99.2	99.1	99.1
XGBoost Proposed	0.9921	0.4087	0.7083	0.5183
HistGradientBoosting Proposed	0.9931	0.4583	0.8250	0.5893
LightGBM Proposed	0.9933	0.4670	0.8250	0.5964
Hybrid Model Proposed	0.9949	0.5523	0.7917	0.6507

The experimental results shown in Table VII that the proposed models achieve high overall accuracy on the credit card fraud detection dataset. Among the proposed approaches, the Hybrid Model records the highest accuracy of 99.49%, followed by LightGBM (99.33%), HistGradientBoosting (99.31%), and XGBoost (99.21%). The proposed models also provide precision, recall, and F1-score values, allowing their fraud-detection performance to be examined beyond accuracy. The Hybrid Model obtains a precision of 55.23% and an F1-score of 65.07%, while LightGBM achieves a recall of 82.50% and an F1-score of 59.64%. Existing models report varying results, with Decision Trees showing 98.9% accuracy and 99.1% F1-score. The findings show that the suggested techniques perform comparably across all of the measures that were used for assessment.

V. CONCLUSION & FUTURE WORK

As the number and complexity of online financial transactions continue to rise, fraud detection has emerged as a critical study field. This research integrated data preprocessing, class-imbalance management, feature selection, ML models, model interpretation, and operational monitoring to provide a scalable MLOps-based system for intelligent fraud detection and identification. The proposed framework was evaluated using 100,000 transactions sampled from the Credit Card Transactions Fraud Detection Dataset. The class imbalance was addressed using SMOTE, and 20 relevant features were selected using Random Forest feature significance. Among the evaluated models, the Hybrid Stacking Model achieved an accuracy of 99.49%, precision of 55.23%, F1-score of 65.07%, MCC of 0.6589, and Kappa of 0.6482, with a classification error of 0.51%. LightGBM and HistGradientBoosting achieved a recall of 82.50%, while LightGBM obtained an ROC-AUC of 0.982. The SHAP analysis shed light on the major features' contributions, and the MLOps monitoring findings showed that for 20,000 test transactions, the average prediction latency was 0.000014 seconds. Distribution shifts in many characteristics were also detected via data-drift monitoring.

However, the research was reviewed based on a single publicly accessible historical dataset. Future research will look at numerous real-world datasets, constantly updated transaction streams, online learning, automated retraining, deep learning, and graph-based algorithms to better adaptability to changing fraud trends and large-scale financial settings.

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 6, Issue 1, November 2026
Impact Factor: 8.2

REFERENCES

[1]A. Cherif, A. Badhib, H. Ammar, S. Alshehri, M. Kalkatawi, and A. Imine, "Credit card fraud detection in the era of disruptive technologies: A systematic review," J. King Saud Univ. - Comput. Inf. Sci., vol. 35, no. 1, pp. 145–174, Jan. 2023, doi: 10.1016/j.jksuci.2022.11.008.

[2]R. Bin Sulaiman, V. Schetinin, and P. Sant, "Review of Machine Learning Approach on Credit Card Fraud Detection," Human-Centric Intell. Syst., vol. 2, no. 1, pp. 55–68, June, 2022, doi: 10.1007/s44230-022-00004-0.

[3]D. K. Jayswal and S. Gupta, "Intelligent Detection of Anomalous Online Payments Using Machine Learning for Financial Cyber Defense," in 2026 International Conference on Intelligent Digitization of Systems and Services (IDSS), IEEE, May 2026, pp. 102–108. doi: 10.1109/IDSS69894.2026.11691329.

[4]E. A. L. Marazqah Btoush, X. Zhou, R. Gururajan, K. C. Chan, R. Genrich, and P. Sankaran, "A systematic review of literature on credit card cyber fraud detection using machine and deep learning.," PeerJ. Comput. Sci., vol. 9, p. e1278, 2023, doi: 10.7717/peerj-cs.1278.

[5]A. Mittal and A. Gupta, "A Multi-Pipeline Deep Learning Based Framework for Fraud Detection in Financial Transactions," in 2026 International Conference on Computing, Communication, Control and Cyber-Physical Systems (I5CPS), IEEE, Feb. 2026, pp. 1–6. doi: 10.1109/I5CPS67958.2026.11452292.

[6]N. Adik, "Secure Dispute Resolution Framework for Online Financial Transactions," Eur. J. Electr. Eng. Comput. Sci., vol. 10, no. 3, pp. 21–25, june, Jun. 2026, doi: 10.24018/ejece.2026.10.3.70440.

[7]B. S. Shaikh, "Using LLM Chaining for Enhanced Fraud Detection in Credit Card Transactions: A Multi-Stage Approach," Int. J. Sci. Res. Eng. Dev., vol. 8, no. 4, pp. 1–7, Aug, 2025, doi: 10.5281/zenodo.16889309.

[8]N. D. Anh Luong and S. Xie, "Explainable ensemble machine learning for financial transaction fraud detection: Insights from XGBoost and deep neural networks," J. Financ. Data Sci., vol. 12, p. 100195, Dec. 2026, doi: 10.1016/j.jfds.2026.100195.

[9]S. Tarakampet, "Real-Time Fraud Detection at Scale: An Architectural Framework for FinTech Big Data Systems," J. FinTech Digit. Assets, vol. 1, no. 1, pp. 45–55, May, 2026, doi: 10.34900/jfda.v1i1.1380.

[10]L. H. Aros, L. X. B. Molano, F. Gutierrez-Portela, J. J. M. Hernandez, and M. S. R. Barrero, "Financial fraud detection through the application of machine learning techniques: a literature review," Humanit. Soc. Sci. Commun., vol. 11, no. 1, p. 1130, Sep. 2024, doi: 10.1057/s41599-024-03606-0.

[11]S. Motie and B. Raahemi, "Financial fraud detection using graph neural networks: A systematic review," Expert Syst. Appl., vol. 240, p. 122156, Apr. 2024, doi: 10.1016/j.eswa.2023.122156.

[12]E. Ileberi, Y. Sun, and Z. Wang, "A machine learning based credit card fraud detection using the GA algorithm for feature selection," J. Big Data, vol. 9, no. 1, p. 24, Dec. 2022, doi: 10.1186/s40537-022-00573-8.

[13]V. Plakandaras, P. Gogas, T. Papadimitriou, and I. Tsamardinos, "Credit Card Fraud Detection with Automated Machine Learning Systems," Appl. Artif. Intell., vol. 36, no. 1, pp. 1584–1599, August, Dec. 2022, doi: 10.1080/08839514.2022.2086354.

[14]S. A. Pushkala, "Deep Learning Techniques for Real-Time Fraud Detection in Financial Transactions," in 2025 International Conference on Computational Innovations and Sustainable Technologies (ICCIST), IEEE, Dec. 2025, pp. 1–7. doi: 10.1109/ICCIST67338.2025.11438826.

[15]B. A. Brahmandam, "MLOps in Finance: Automating Compliance & Fraud Detection," Int. J. Comput. Trends Technol., vol. 73, no. 4, pp. 35–41, Apr. 2025, doi: 10.14445/22312803/IJCTT-V73I4P105.

[16]F. D. Silbermann, "Cloud Native AI and Deep Learning Models for Real Time Fraud Detection and Cybersecurity in Financial Institutions," nternational J. Adv. Res. Comput. Sci. Technol., vol. 7, no. 1, pp. 17-25,May, 2024, doi: 10.15662/IJARCST.2024.0701803.

[17]S. Paul, C. Huyen, and A. Finnegan, "Scalable Payment Ecosystem Design Using Machine Learning–Based Platform Engineering and Automation," 2026.

[18]R. K. Alluri, "Context-Aware Fraud Detection Models Deployed via Feature Pipelines in ML Flow," Int. Sci. J. Eng. Manag., vol. 03, no. 10, pp. 1–6, Oct, 2024, doi: 10.55041/ISJEM02013.

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429 Volume 6, Issue 1, November 2026 Impact Factor: 8.2

[19]S. Barman, "Operationalizing Reliability in MLOps: A Case for Continuous Monitoring of Tabular ML Systems," in 2026 International Conference on Connected Intelligence for Industrial Applications (CI2A), IEEE, Apr. 2026, pp. 1–6. doi: 10.1109/CI2A69097.2026.11576911.
[20]D. Sreenivasulu, "Fraud Detection in UPI Transactions Using Machine Learning," Int. J. Res. Publ. Rev., vol. 7, no. 4, pp. 10394–10400, Apr. 2026, doi: 10.55248/gengpi.07.0426.a1105.
[21]S. M. Darwish and N. A. El-Shoafy, "A hybrid deep learning–blockchain framework for reliable fraud detection: a big data–driven approach for real-time FinTech systems," J. Big Data, vol. 13, no. 1, pp. 129, July, 2026, doi: 10.1186/s40537-026-01506-5.
[22]K. Amistapuram, L. Pandiri, V. R. Raju, S. Paleti, S. Singireddy, and G. K. Sheelam, "AI-Based Cloud Infrastructure and MLOps Frameworks for Scalable Data Engineering Across Banking and Insurance," in 2025 IEEE International Conference on Communication Networks and Computing (CNC), Sonbhadra, India: IEEE, 2025, pp. 186–192, December. doi: 10.1109/CNC68716.2025.11484532.
[23]M. A. Mokoele and S. Mokwena, "Enhancing Fraud Detection in Financial Transactions through a Hybrid Machine Learning Model," in 2025 IEEE AFRICON, IEEE, Dec. 2025, pp. 1–7. doi: 10.1109/AFRICON66545.2025.11533845.
[24]V. Penumatcha, "A Comprehensive Machine Learning Method for Identifying Frauds in Digital Transactions in Realtime Banking Settings: Difficulties and Consequences," in 2025 International Conference on Emerging Trends in Networks and Computer Communications (ETNCC), Windhoek, Namibia: IEEE, 2025, pp. 1–12, August. doi: 10.1109/ETNCC66224.2025.11299775.
[25]M. G. Wijaya, M. F. Pinaringgi, A. Y. Zakiyyah, and Meiliana, "Comparative Analysis of Machine Learning Algorithms and Data Balancing Techniques for Credit Card Fraud Detection," Procedia Comput. Sci., vol. 245, pp. 677–688, 2024, doi: 10.1016/j.procs.2024.10.294.
[26]P. Gupta, A. Varshney, M. R. Khan, R. Ahmed, M. Shuaib, and S. Alam, "Unbalanced Credit Card Fraud Detection Data: A Machine Learning-Oriented Comparative Study of Balancing Techniques," Procedia Comput. Sci., vol. 218, pp. 2575–2584, 2023, doi: 10.1016/j.procs.2023.01.231.
[27]K. Shenoy, "Credit Card Transactions Fraud Detection Dataset," kaggle.
[28]O. Lawal, A. Paschal, A. Okolie, C. Obunadike, M. O. Ikhifa, and S. I. Edozie, "An Explainable XGBoost Framework for Detecting Fraudulent Financial Transactions," J. Sci. Res. Reports, vol. 31, no. 12, pp. 244–255, Dec. 2025, doi: 10.9734/jsrr/2025/v31i123769.
[29]K. Huang, "An Optimized LightGBM Model for Fraud Detection," J. Phys. Conf. Ser., vol. 1651, no. 1, pp. 1–7, November, Nov. 2020, doi: 10.1088/1742-6596/1651/1/012111.
[30]A. Mniai, M. Tarik, and K. Jebari, "A Novel Framework for Credit Card Fraud Detection," IEEE Access, vol. 11, pp. 112776–112786, 2023, doi: 10.1109/ACCESS.2023.3323842.
[31]A. R. Khalid, N. Owoh, O. Uthmani, M. Ashawa, J. Osamor, and J. Adejoh, "Enhancing Credit Card Fraud Detection: An Ensemble Machine Learning Approach," Big Data Cogn. Comput., vol. 8, no. 1, pp. 1–6, Jan. 2024, doi: 10.3390/bdcc8010006.
[32]I. Benchaji, S. Douzi, B. El Ouahidi, and J. Jaafari, "Enhanced credit card fraud detection based on attention mechanism and LSTM deep model," J. Big Data, vol. 8, no. 1, pp. 151, December, 2021, doi: 10.1186/s40537-021-00541-8.
[33]H. Z. Alenzi and N. O, "Fraud Detection in Credit Cards using Logistic Regression," Int. J. Adv. Comput. Sci. Appl., vol. 11, no. 12, 2020, doi: 10.14569/IJACSA.2020.0111265.
[34]S. Gupta, "An Advanced Machine Learning Method Used for Identification of Banking Fraud in Financial Transactions," J. Glob. Trends Artif. Intell., vol. 1, no. 1, pp. 28–35, 2025.
[35]L. Theodorakopoulos, A. Theodoropoulou, A. Tsimakis, and C. Halkiopoulos, "Big Data-Driven Distributed Machine Learning for Scalable Credit Card Fraud Detection Using PySpark, XGBoost, and CatBoost," Electronics, vol. 14, no. 9, pp. 1–32, April, 2025, doi: 10.3390/electronics14091754.