Journal of Science Engineering Technology and Management Science *ISSN:3049-0952*
Volume 03, Issue 06, June 2026 *www.jsetms.com*

http://doi.org/10.64771/jsetms.2026.v03.i06.pp330-337

Adaptive Fraud Detection Based on Machine Learning Models for Financial Technology Applications

Vikas Kumar Pandey
Senior Software Engineer, PayPal India Pvt. Ltd.
Whitefield, Bangalore
vikaspandey.cdac@gmail.com

***Abstract*—Financial technology (FinTech) has become an urgent topic with regard to financial fraud, because of the numerous online transactions, and the changing patterns of online fraud. To improve the effectiveness and robustness of fraud detection, this research suggests an adaptive machine learning (ML)-based fraud detector. The methodology entails using the IEEE-CIS fraud detection dataset, which is preprocessed, feature-engineered, and treated with class-imbalance remedies like SMOTE. LightGBM and XGBoost are two modernized ensemble models, that are implemented and tested. Explainable AI models, such as SHAP and LIME, are used to increase the model transparency. Experimental results indicate that XGBoost is more effective than LightGBM, achieving higher accuracy (99.1%) as well as higher precision, recall and F1-score. The given structure is a good way to capture the intricate patterns of fraud without becoming uninterpretable. This study demonstrates the significance of real-time fraud detection using adaptive, scalable, explainable ML solutions in current financial systems.**

Keywords*—Financial Technology (FinTech), Fraud Detection, Machine Learning, Explainable AI, Financial Transactions.

I. INTRODUCTION

In recent years, the emergence of technologies and paradigms such as e-commerce and the financial technology (FinTech) industry has led to an increase in financial fraud[1]. Online transactions, mobile banking and systems of online payments have resulted in new opportunities that fraudsters can use to take advantage of vulnerabilities in systems. Additionally, fraudsters are constantly improving their methods and are using advanced means to cover their fraudulent actions as legitimate business transactions hence making it harder to detect fraud [2][3]. Consequently, financial fraud has turned into an active and multifaceted issue that must be constantly improved in terms of detection methods and technologies [4][5].

Financial fraud refers to the act of obtaining financial gain through illegal or deceptive means and can occur across various sectors, including banking, insurance, taxation, and corporate environment [6][7][8]. Financial transaction fraud, identity theft, and money laundering have, over the past few years, become a significant challenge for organizations and financial institutions. Although there have been many attempts to curb these threats, financial fraud continues to cause significant losses to the economy and erode trust in financial systems [9][10][11][12]. Conventional fraud-detection systems are mostly manual or rule-based, making them time-consuming, expensive to operate, and ineffective for large-scale, high-dimensional data. Moreover, such methods often fail to keep pace with evolving fraud trends, leading to more FP and missed fraud.

Fraud detection systems are becoming more efficient and automated because to advancements in data mining, machine learning (ML), and artificial intelligence (AI) [13][14]. In cases when the traditional approaches would have complicated the detection of concealed tendencies and irregularities, ML can consume large amounts of transaction data. These methods can be categorized as supervised and unsupervised learning [15]. The basis of supervised learning is using labeled data to classify a transaction as authentic or fraudulent, but unsupervised learning aims to classify an anomaly, in the absence of an existing label. Lately, supervised learning algorithms, such as decision trees (DT), support vector machines (SVM), and ensemble methods, have fared fairly well in detecting fraud [16][17]. However, the majority of advanced ML models, including ensemble and deep learning (DL) models, are black boxes and, therefore, difficult to understand and trust. Explainable Artificial Intelligence (XAI) addresses this by providing an explanation of model decisions in terms of such tools as SHAP and LIME [18]. It helps clarify things and build trust, resulting in improved decision-making. This increases the reliability of fraud detection systems and their usefulness in real-world FinTech contexts.

A. Motivation and Contribution of study

The increasing use of digital financial systems has greatly increased risk and complexity of fraud. The conventional fraud detection systems are often set and rule-based, making it impossible for them to adapt to the shifting patterns of fraud. The impact of this limitation is increased false positives and fraud omissions. The modern FinTechs have required smart, versatile and extensible solutions. ML is capable of dealing with vast amounts of transactions and discovering latent patterns. Therefore, the need to create a strong, real-time, and explainable fraud detection model to enhance accuracy and reliability has motivated this work. The key contributions are given below:

- Proposed an adaptive fraud detection framework using ML for FinTech applications.
- Implemented and compared advanced ensemble models: LightGBM and XGBoost.

- Applied data preprocessing, feature engineering, and imbalance handling techniques (SMOTE, SMOTE-ENN).
- Integrated Explainable AI techniques (SHAP and LIME) for model transparency.
- Addressed challenges of scalability, interpretability, and real-time fraud detection.

B. Justification and novelty

This paper deal with the key shortcomings of the existing fraud detecting mechanisms, including limited flexibility, high complexity and inability to be interpreted. Compared to the conventional approaches, the proposed framework integrates adaptive learning and high-performance ensemble models with explainable AI approaches in a smooth manner. LightGBM and XGBoost coupled with SHAP and LIME allow the framework to capture a strong trade-off between predictive and model transparency. Its practical importance is further boosted by the fact that it utilizes big real-world data and can handle imbalance. In this, the novelty lies in its coherent, adaptive and explainable architecture, providing a scalable, high-precision fraud detection system suited to dynamically changing FinTech environments.

C. Structure of paper

This paper is organized as follows: **Section II** presents a literature review on ML in fraud detection. In **Section III**, proposed methodology and model development are outlined. **Section IV** presents experimental findings, and compares and contrasts results. Lastly, the study ends with a **Section V** that summarizes the research and provides future research directions.

II. Literature Review

Existing literature has shown the high effectiveness of hybrid, ensemble, and DL models in fraud detection, but most approaches are not trained to adapt in real time and are computationally intensive and slow to respond to novel data.

H. S. Mohammed, Z. B. Sallow, and H. M. Zangana (2026) proposes a novel hybrid approach that blends DL models with algorithms for finding anomalies to make fraud detection in digital banking more accurate, powerful, and flexible. The suggested method combines autoencoder-based unsupervised anomaly detection with a DNN architecture instructed using supervised learning to discern complex transactional patterns and uncover previously unrecognized fraudulent strategies. The accuracy of the suggested hybrid AI model is 98.8% [19].

S. B. Shah (2025) applied the ML techniques to the Kaggle Financial Fraud Detection Dataset, which includes class balancing, feature engineering, and preprocessing. RF, AdaBoost, LightGBM, and a Voting Classifier were optimized with GridSearchCV with LightGBM having the maximum accuracy (90.20%). SHAP analysis also showed important features that ensured the effectiveness and interpretability of Ensemble Learning models in detection of fraud in financial institutions [20].

H. A. Al Dulaimi et al. (2025) introduce a real-time behavioral anomaly and fraud detection architecture tailored for FinTech systems by integrating temporal DL with federated and explainable AI (XAI) pipelines. The system was trained on more than 220 million mobile wallet, algorithmic trading, and cross-border remittance transactions, using LSTM Autoencoders, Graph Neural Networks (GNNs), and reinforcement-based feedback loops. The framework achieved classification accuracy of over 97.85%, F1-scores of 0.947% and latency of less than 110 milliseconds per transaction stream [21].

Silvia *et al.* (2024) examine use of behavioral analytics in combination with ML models to reduce FP and increase accuracy by analyzing the patterns of user behavior to detect fraud. To identify the main trends, 200 articles (2020-2024) were subjected to bibliometric analysis using VOS viewer and Google Scholar. Such words as user behavior and ML algorithms came to the forefront. CNNs were tested to have up to 95% accuracy and other ML models were studied, such as neural networks and DL [22].

Sharma *et al.* (2024) ML and DL models' relative efficacy in improving the fraud-detection capabilities of digital finance systems, using CNNs and RNNs as examples, is the focus of this research. To identify problematic transactions in a large collection of transactional data, both supervised and unsupervised learning techniques are employed. Such type of ML algorithms that are incorporated in this methodology include CNNs and RNNs to model, or automated methods to clean up data. The analysis indicates that the RNN model does even better, having a 95.8% accuracy rate, a 93.7% sensitivity rate, a 97.5% specificity rate, and an AUC of 0.972 [23].

G. Eswar Prasad et al. (2023) use a dataset of 284,807 transactions conducted by European cardholders in 2013, of which 492 were fraudulent, to investigate how ML algorithms may be used to detect credit card fraud. The dataset was preprocessed using Label Encoding, SMOTE, and PCA to minimize number of features. A training dataset was used to assess the effectiveness of ML-based classification models including DT, SVM, and ANNs. The models were evaluated using a variety of criteria, including rec, prec, and acc. When compared to DT and SVM, ANN model performed the best, with highest prec, acc, and rec of 98.41%, 98.69%, and 98.98%, respectively [24].

Despite the hybrid, ensemble, and DL models achieving high accuracy in detecting fraud, current research is not truly adaptable to changing fraud, trends in real-time financial conditions. The majority of solutions are based on fixed datasets, offer inadequate support for concept drift, and are highly computationally demanding, making them difficult to deploy in FinTech systems. Moreover, the interpretability in the context of DL models is inadequate, whereas the scalability of federated and real-time architectures is inadequate. Hence, it is urgently required to have lightweight, flexible and explainable ML systems that can learn continuously using streaming financial transaction data. Table I shows the comparison of recent studies on the basis of methodology, datasets, performance outcomes, limits, and future research objectives in use of financial technology.

TABLE I. Literature Review of Machine Learning Approaches for Fraud Detection in Financial Technology Applications

Authors	Approaches	Dataset	Key Findings	Research Gaps	Future Directions

H. S. M. Mohammed et al. (2026)	Hybrid Deep Neural Network (DNN) integrated with Autoencoder-based unsupervised anomaly detection under a hybrid learning paradigm	Digital banking transaction dataset (not fully specified)	Achieved 98.8% accuracy with improved robustness in fraud pattern recognition	High computational complexity; limited interpretability; weak cross-domain validation	Develop Explainable AI (XAI)-driven hybrid models and optimize lightweight architectures for real-time FinTech deployment
Shah (2025)	Ensemble learning framework (RF, AdaBoost, LightGBM, Voting Classifier) optimized using GridSearchCV with SHAP interpretability.	Kaggle Financial Fraud Detection Dataset	LightGBM achieved 90.20% accuracy; SHAP improved model transparency and feature importance interpretation.	Limited adaptability to evolving fraud patterns; static learning pipeline	Integrate temporal learning and adaptive retraining mechanisms for dynamic fraud detection
H. A. Al Dulaimi et al. (2025)	Real-time fraud detection architecture combining LSTM Autoencoders, GNNs, Reinforcement Learning, Federated Learning, and XAI	Large-scale FinTech dataset (220M+ transactions from wallets, trading, remittance systems)	Achieved 97.85% accuracy, F1-score of 0.947, and <110 ms latency per transaction	System complexity and communication overhead in federated setup; scalability challenges	Develop edge-optimized federated learning models and reduce computational overhead for real-time deployment.
Silvia et al. (2024)	Behavioral analytics-based fraud detection using ML and DL; bibliometric analysis with CNN evaluation	200 scholarly articles (2020–2024) from Google Scholar	Behavioral patterns significantly enhance detection; CNN achieved up to 95% accuracy	Lack of real-world deployment validation; literature-focused analysis	Implement real-time behavioral fraud detection using streaming financial datasets
Sharma et al. (2024)	Comparative DL framework using CNN and RNN models for fraud detection	Financial transaction dataset (large-scale unspecified dataset)	RNN outperformed CNN with 95.8% accuracy, AUC = 0.972	No adaptive learning mechanism; limited interpretability of deep models	Introduce adaptive deep learning with concept drift handling and explainable architectures
G. Eswar Prasad et al. (2023)	Classical ML models (Decision Tree, SVM, ANN) with SMOTE balancing and PCA dimensionality reduction	European credit card transaction dataset (284,807 records, 492 fraud cases)	ANN outperformed conventional models with 98.69% accuracy and 98.98% recall.	Static dataset; lack of real-time and evolving fraud detection capability	Develop incremental learning models for streaming fraud detection in dynamic environments

III. Methodology

The process flow of an adaptive fraud detection system implemented in financial technology using ML models, as illustrated in Fig. 1. The IEEE-CIS fraud detection dataset is used first, and then the data is cleaned and prepared to deal with missing values and other issues. The cleaned data are then encoded to convert categorical data into numerical data, which can be used in modeling. Then, dataset is split into two parts: a training set (80%) and a testing set (20%). To address class imbalance, proper data imbalance management strategies are implemented. The process of feature engineering is then undertaken to improve relevant features to achieve improved model performance. LightGBM and XGBoost are proposed ML models that are trained and evaluated. Model interpretability is achieved by the application of SHAP and LIME approaches. Lastly, performance metrics of system, evaluating acc, prec, rec, and F1, are used to evaluate effectiveness of system, and final results are produced.

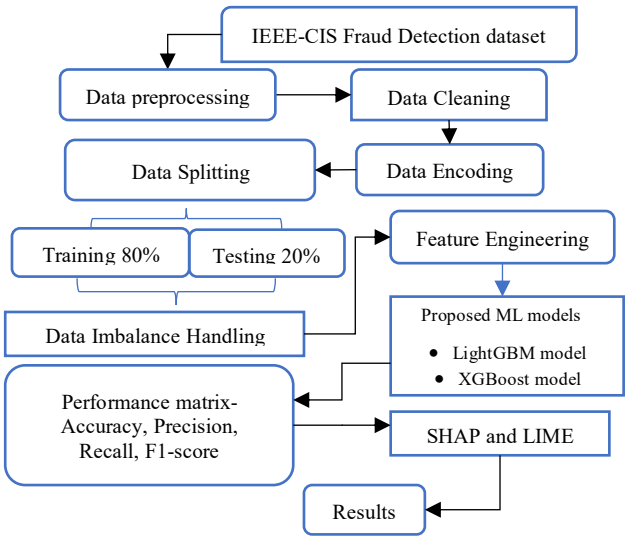

Fig. 1. Flowchart Depicting Proposed Framework for FinTech Security

The subsequent is described in the following sections, in accordance with methodology and with proposed flowchart.

A. Data Collection and Visualization

The IEEE Computational Intelligence Society (IEEE-CIS) Vesta Corporation launched IEEE-CIS Fraud Detection dataset, was employed in the current study. It includes real-world e-commerce transactions to develop ML-based fraud detection systems. It is a dataset that is a combination of transaction and identity data and can be accessed through the Kaggle.

The dataset consists of two main CSV files:

- **transaction.csv:** It has 590,540 transaction records and 394 features, including anonymized transactional

features, behavioral features, and engineered variables to detect fraud.

- **identity.csv:** It has 144,233 records containing 41 features, which include identity-related and device-level data, including browser type, device characteristics, and anonymized user identifiers.

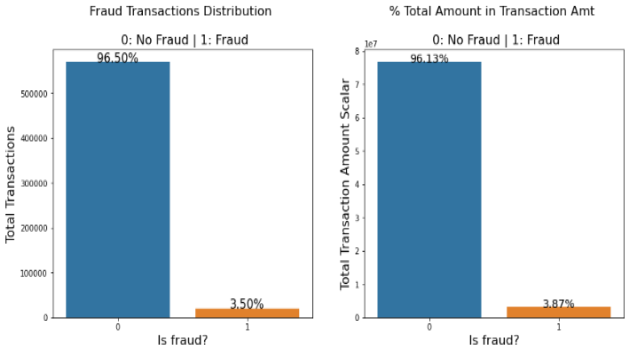

Fig. 2. Distribution of Financial Transactions

Fig. 2 presents allocation of financial transactions in two categories, with the pattern being very imbalanced. In the two subplots, most transactions are of the dominant class, with the figure standing at about 96.5% and 96.13%, respectively, and the minority class constitutes about 3.5% and 3.87%, respectively.

In Fig. 3, the distribution of transactions in various card networks and their relation to fraud are given. Visa is the leader (65.16%), MasterCard takes the second place (32.04%), Discover and American Express play a minor role. The plot of fraud distribution indicates that although transaction volumes are highest with Visa and MasterCard, percentage of fraud varies across card types.

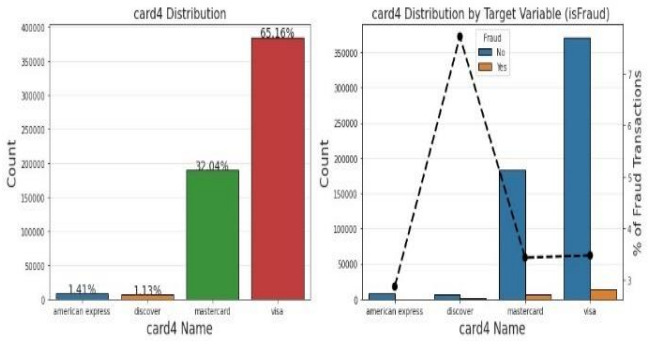

Fig. 3. Distribution of Card4

B. Data Preprocessing

This stage focuses at enhancing the quality of the data and how to prepare the dataset for training on an item ML model. It guarantees uniformity, minimizes noisy data and incomplete data in the dataset. Discussed below are the steps that follow:

- Features with very high ratios of missing values are eliminated in order to enhance data quality and minimize noise in data.
- For numerical characteristics, median imputation is used to manage the remaining missing values, while categorical features are filled using mode or assigned an "Unknown" category to preserve data integrity.

C. Data Encoding

Categorical variables are converted into numerical form to ensure model compatibility and improve performance. Label Encoding is applied for tree-based models such as LightGBM and XGBoost, while Frequency Encoding is used for high-cardinality features to preserve distribution patterns efficiently. The feature is cleaned, which involves removing zero-variance, highly correlated (>0.95), and low-information features to reduce redundancy and enhance model efficiency.

D. Data Splitting

The dataset is split into 20% for testing and 80% for training after preprocessing. To enhance model performance and lessen overfitting, some of training data is also utilized as a validation set. This approach ensures reliable evaluation on unseen data.

E. Data Imbalance Handling

To address highly imbalanced nature of dataset, SMOTE and SMOTE-ENN are applied only on training data to generate synthetic minority (fraud) samples and improve class distribution. Also, misclassification of cases of fraud is penalized by class weighting so that learning between fraud and non-fraud classes is balanced. Fig. 4 reveals the distribution of the classes with the balancing technique and the counts of the classes are the same non-fraud (0) and fraud (1) with 1000 instances in each. This means that the dataset is now balanced, removing the issues of class imbalance and allowing fair training of the model.

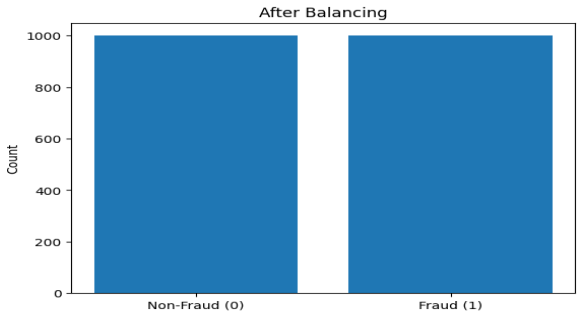

Fig. 4. Data Distribution after Balancing

F. Feature Engineering

Feature engineering is used to improve ML models' prediction power by identifying significant patterns in unprocessed transaction data. It involves transforming temporal attributes such as transaction time into meaningful cyclical or time-based features, constructing behavioral features that capture user transaction patterns, and generating aggregated indicators based on historical transaction behavior. These engineered features help in capturing hidden fraud patterns and significantly improve the effectiveness of fraud detection models.

G. Proposed LightGBMModel

LightGBM is a DT-based GB framework designed to be highly efficient for high-dimensional, sparse datasets [25]. At each iteration, LightGBM adds a new decision tree to minimize a loss function defined over true values and predictions, a technique that has demonstrated strong predictive performance (Equation 1).

$$\mathcal{L} = \sum_{i=1}^{n} l(y_i, \hat{y}_i) + \Omega(f_t) \quad (1)$$

Where: y_i is actual value of sample i. $\hat{y}_i$ is predicted value of sample i. $l(\cdot)$ is loss function, typically squared loss or logistic loss. f_t is regularization term that controls model complexity and prevents overfitting.

H. Proposed XGBoost Model

XGBoost is a type of ensemble learning that makes a lot of low learners, typically DT and integrates their predictions to develop a strong predictive model [26]. The approach improves overall model performance by iteratively minimizing a given loss function by successively adding weak learners. XGBoost is quite efficient for both regression and classification workloads because to this method. A mathematical structure that forecasts yi based on input features xi is referred to as a model in supervised learning. For example, Equation (2) expresses the prediction in a linear model:

$$\hat{y}_i = \sum_j \theta_j x_{ij} \tag{2}$$

Depending on the job, such as classification or regression, the prediction value might have different meanings. The parameters are the unknowns that must be learned from the data. The coefficients θ are the parameters in linear regression issues. Finding the optimal parameters θ that best suit training data x_i and labels y_i is task of training model. The gradient boosting method, in which decision trees are trained one after the other, is shown in Fig. 5. A robust final model is created by combining the predictions of each tree, which corrects the residual errors of the preceding ones.

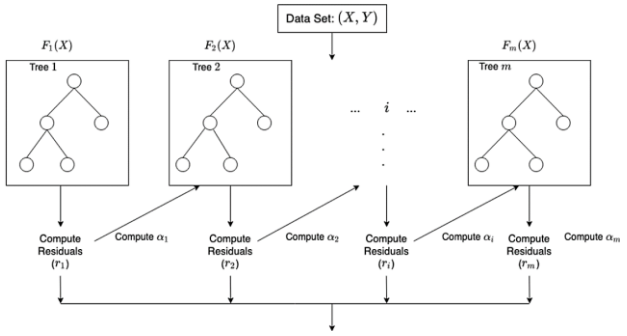

Fig. 5. Structure of the XGBoost model

I. Interpretability Analysis Using SHAP and LIME

The system uses explainable AI techniques such as SHAP and LIME to improve the clarity of fraud detection. SHAP offers global interpretability, it establishes if LIME offers local justifications for particular transactions, while value of each characteristic is used to model predictions. This combination is able to provide overall and instance-level insight into model behavior, improve trust, interpretability, and decision-making in financial fraud detection systems.

J. Performance Matrix

Classic methods for evaluating ML classifiers can use confusion metrics, which compare the true labels in the dataset with the model's predictions. In this context, true positive (TP), true negative (TN), false positive (FP), and false negative (FN), respectively.

Accuracy: used to gauge how well data processing and evidence domain recovery are progressing [27]. Equation (3) may be used to describe percentage of outcomes that are successfully categorized as follows:

$$\text{Accuracy} = \frac{TP+TN}{TP+FP+FN+TN} \tag{3}$$

Precision: Precision, a performance metric, measures percentage of positives that were correctly identified to all positives. This is shown in Equation (4) as follows:

$$\text{Precision} = \frac{TP}{TP+FP} \tag{4}$$

Recall: Equation (5) shows the recall, also called the sensitivity, is the ratio of linked instances recovered over the total number of retrieved instances.

$$\text{Recall} = \frac{TP}{TP+FN} \tag{5}$$

F1_Score: The f-measure takes into account both recall and precision. Equation (6) illustrates how the average weight of all values can be considered the f-measure.

$$F1 - score = 2 * \frac{(precision * recall}{(precision + recall)} \tag{6}$$

ROC: The ROC curve is used to compare TPR and FPR to evaluate model performance across different thresholds. In this experiment, five classifiers were not resampled, and their AUC-ROC scores were compared. The proposed model had the highest AUC, indicating it is the most suitable for differentiating between churned and non-churned customers.

IV. Results and Discussion

The proposed framework is implemented in Python 3.10 using Jupyter Notebook/Google Colab. It is tested on a system with an Intel Core i7 processor, 16 GB RAM, and an optional NVIDIA Tesla T4 graphics card, Windows 11 or Ubuntu 22.04. This environment facilitates effective computation, scalability, and reproducibility, thereby enabling the efficient processing of large volumes of financial transaction data for fraud detection applications. Table II presents a quantitative performance analysis of the proposed LightGBM and XGBoost models. LightGBM achieves 98.6% acc, 97.9% prec, 96.5% rec, and 97.2% F1, while XGBoost outperforms it with 99.1% acc, 98.8% prec, 98.6% rec, and 98.7% F1. These findings show that XGBoost performs better in fraud detection and achieves a better balance between precision and recall, leading to more reliable detection of fraudulent transactions.

TABLE II. Performance Evaluation of Fraud Detection Models Using Machine Learning for Financial Technology

Measures	LightGBM Model	XGBoost Model
Accuracy	98.6	99.1
Precision	97.9	98.8
Recall	96.5	98.6
F1-score	97.2	98.7

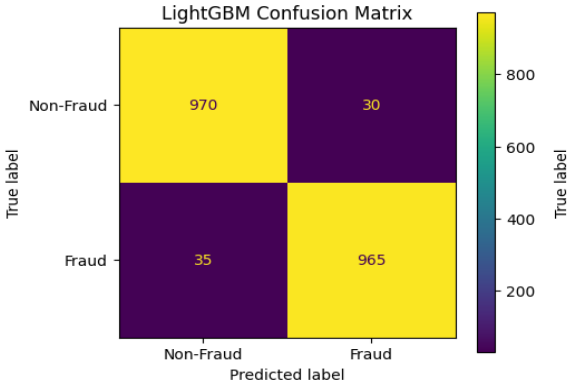

Fig. 6. Confusion matrix of the LightGBM Model

Fig. 6 illustrates LightGBM model's confusion matrix, and its classification performance is, with the correct identification of 970 cases that have not been committed to fraud and 965 cases that have been committed to fraud. The model has the lowest errors of 30 and 35, which indicate that it is very accurate and well-balanced in its forecasting for both classes.

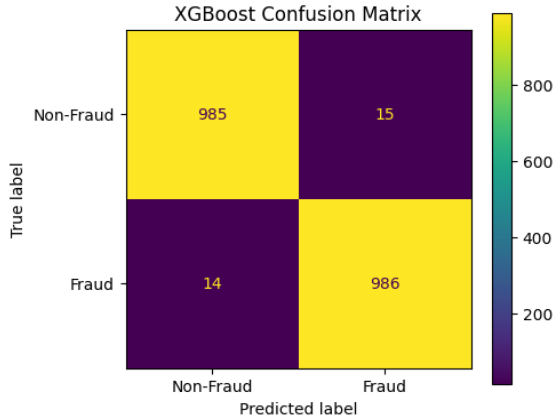

Fig. 7. Confusion matrix of the XGBoost model

Fig. 7 presents confusion of XGBoost model, where it is observed that model showed high accuracy and predicted well across both classes, with 985 non-fraud transactions and 986 fraud transactions being correctly identified.

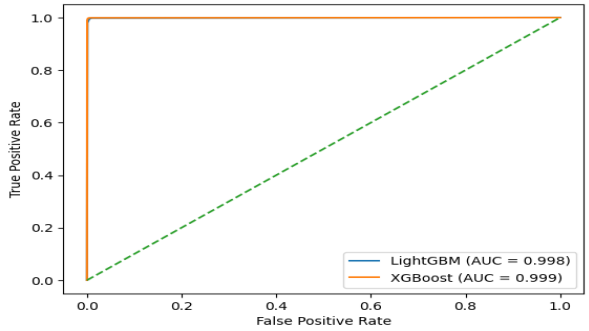

Fig. 8. ROC Curve of the proposed models

The comparison of ROC curves for LightGBM and XGBoost models in Fig. 8 indicates that both models have very good classification performance, with AUCs of 0.998 and 0.999, respectively. Near the upper-left corner, the curves are almost parallel, signifying high TP and low FP, implying that XGBoost is slightly superior to LightGBM.

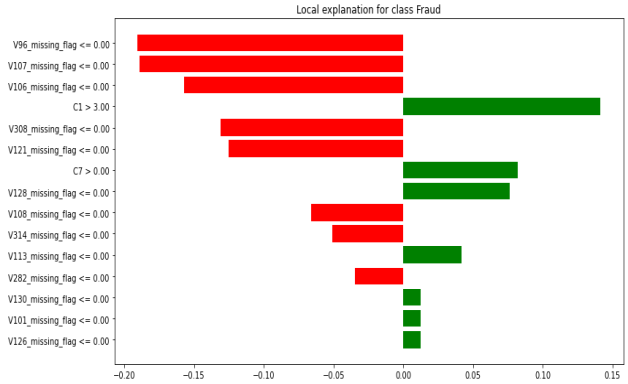

Fig. 9. LIME Local Explanation Plot

Fig. 9 presents feature importance analysis, which indicates most impactful variables in model predictions. The chart indicates a combination of positively (green) and negatively (red) contributing features, where a few dominant features have a much larger impact than others. This indicates that extraneous factors have minimal bearing on the forecast result and that the model is reliant on important traits.

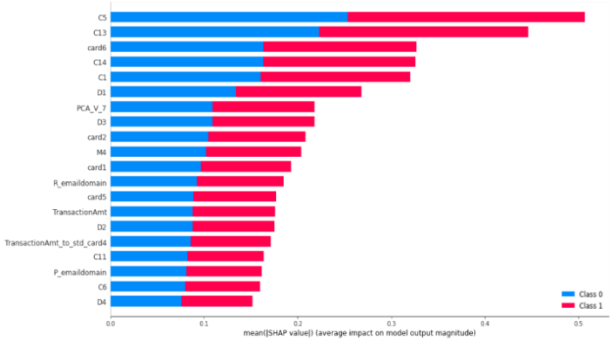

Fig. 10. SHAP Summary Plot

Fig. 10 demonstrates the comparison of feature contributions between Class 0 and Class 1, which underlines impact of various features on each of classes. The chart indicates that there are features that are more contributory to one of the classes than the other class and this implies that there are different patterns in classification. The contribution is dominated by a few features, and rest have minor roles, which is an indicator of the model being dependent on the defining variables to create a difference between the two classes.

A. Comparative Analysis and Discussion

Table III presents a detailed comparison of current ML models and the proposed models in detecting financial fraud in terms of key performance indicators. Although the traditional models, including the SVM and LR, have a high accuracy, their low precision and F1-scores suggest that they are limited in the ability to classify fraudulent transactions. Forest-based algorithms such as the RF and MLP exhibit better balance in terms of metrics. The proposed LightGBM and XGBoost models are, however, evidently better than all the baseline approaches, as they achieve higher acc, prec, rec and F1. It is also significant that XGBoost yields the best overall result, indicating that it can handle unequal financial data and identify complex fraud trends, resulting in a fraud detection system that is stronger and more dependable.

TABLE III. COMPARISON BETWEEN THE EXISTING AND PROPOSED MODEL PERFORMANCE FOR FINANCIAL FRAUD DETECTION

References	Models	Accuracy	Precision	Recall	F1-score
[28]	SVM	98	54	93	57
[29]	MLP	95.8	97.6	93.9	95.8
[30]	Logistic Regression	94	54	92	57
[31]	Decision Tree	90.9	90.9	91.9	92.8
[32]	Random Forest	94.9	94	94	94
Proposed	LightGBM	98.6	97.9	96.5	97.2
Proposed	XGBoost	99.1	98.8	98.6	98.7

The results of the experiment confirm the advantage of using ML models to predict financial fraud in FinTech systems. Both LightGBM and XGBoost are very predictive; however, XGBoost will always be superior in all measures of evaluation compared to LightGBM. A combination of feature engineering and imbalance handling greatly enhances the ability to detect, particularly the minority frauds. XAI techniques, such as SHAP and LIME, provide helpful features to comprehend significance of characteristics and model performance. These strengths are accompanied by the fact that there are still issues with real-time deployment and computational efficiency. The results demonstrate that the balance between accuracy, interpretability, and scalability is crucial in creating effective fraud detection systems.

V. Conclusion and Future Scope

A proposed adaptive ML-based framework is to identify financial fraud in FinTech applications to increase the accuracy, robustness, and scalability of the modern digital transaction environment. The ensemble models like LightGBM and XGBoost, extensive data preparation and feature engineering, and class imbalance control techniques enable the system to be a good predictor. Empirical evidence suggests that XGBoost is more effective than LightGBM, especially when dealing with highly skewed data, as well as capturing intricate and changing fraud dynamics. Moreover, explainable AI algorithms such as SHAP and LIME not only simplify understanding models but also increase their transparency, which is more believable and simplifies the application of models to the real world. With these positive results, there are some limitations. The framework is mainly based on batch learning, and might not be able to effectively process streaming data in real time or fraud behaviors that are quickly changing. In addition, small datasets can be replaced with large ones and more complex models can also make the computation more complex and thus influence the deployment efficiency. Future studies will be undertaken with regards to adding on-line adaptive features under online and incremental learning methods. DL will be researched based on federated learning, and edge computing to enhance scalability, efficiency and privacy. Increasing interpretability and reducing model complexity to make their deployment in dynamic and high-frequency FinTech environment practical and efficient will also be a subject of future research.

References

[1] S. Shivam, V. Nutalapati, T. P. Patel, A. K. Padhy, M. Kumari, and R. Purushothaman, "Transformer Based Framework for Imbalanced Transaction Fraud Detection in FinTech Systems," in *2026 14th International Symposium on Digital Forensics and Security (ISDFS)*, Boston, MA, USA: IEEE, 2026, pp. 1–6, March. [Online]. Available: https://ieeexplore.ieee.org/abstract/document/11459102

[2] S. K. Hashemi, S. L. Mirtaheri, and S. Greco, "Fraud Detection in Banking Data by Machine Learning Techniques," *IEEE Access*, vol. 11, pp. 3034–3043, 2023, doi: 10.1109/ACCESS.2022.3232287.

[3] J. W. Sajja and A. Nerella, "Enterprise Finance Reimagined: Harneessing ERP and Data Innovation for Next-Generation Value Creation," 2024. doi: 10.52710/cfs.743.

[4] A. Nerella, P. Badri, K. Sundravadivelu, and R. Murugesan, "Navigating Regulatory Hurdles in AI-Driven Credit Card Approvals: Balancing Innovation and Compliance," *J. Inf. Syst. Eng. Manag.*, vol. 8, no. 4, pp. 1–9, Nov. 2023.

[5] R. Palwe, "Three Layers of Trust in AI Interfaces: Interface, Behavior, and Organization," *Int. J. Sci. Res.*, vol. 15, no. 1, pp. 1152–1160, Jan. 2026, doi: 10.21275/SR26112072531.

[6] A. Parupalli, "The Evolution of Financial Decision Support Systems : From BI Dashboards to Predictive Analytics," *KOS J. Bus. Manag.*, vol. 1, no. 1, December, pp. 1–8, 2023, [Online]. Available: https://kelvinpublishers.com/Articles/jbm/JBM-1-01.pdf?jid=jbm

[7] S. Singamsetty, "Efficacy of Data Governance a Cutting Edge Approach to Ensuring Data Quality in Machine Learning for Banking Industry," in *2024 2nd International Conference on Signal Processing, Communication, Power and Embedded System (SCOPES)*, IEEE, Dec. 2024, pp. 1–7. doi: 10.1109/SCOPES64467.2024.10991944.

[8] P. Parida and N. Senguttuvan, "Responsible Utilization of Cloud in Retail Banking Ecosystem," *Int. J. Comput. Appl.*, vol. 187, no. 49, pp. 34–39, Oct, 2025, doi: 10.5120/ijca2025925835.

[9] E. M. Al-dahasi, R. K. Alsheikh, F. A. Khan, and G. Jeon, "Optimizing fraud detection in financial transactions with machine learning and imbalance mitigation," *Pure KFUPM*, vol. 42, no. 2, Feb, pp. 01–08, 2025, doi: https://doi.org/10.1111/exsy.13682.

[10] H. N. Patel, "AI-Driven Fraud Detection in Insurance Claims : A Deep Learning Framework for FinTech Risk Intelligence," *J. Glob. Res. Electron. Commun.*, vol. 1, no. 12, pp. 76–82, 2025.

[11] H. N. Dholariya, "GVIF: A Governed Vector Intelligence Framework for AI-Driven Cloud Data Modernization in Regulated Financial Systems," *Int. J. Comput. Exp. Sci. Eng.*, vol. 12, no. 1, pp. 371–386, January, 2026, doi: 10.22399/ijcesen.4797.

[12] B. Mohan, V. R. Surasani, and R. Kumar, "Autonomous Data Stewardship: Multi-Agent AI for Real-Time Master Data Management in Financial Services," in *2026 IEEE 16th Annual Computing and Communication Workshop and Conference (CCWC)*, Las Vegas, NV, USA: IEEE, 2026, pp. 0689–0698, February. doi: 10.1109/CCWC67433.2026.11393832.

[13] K. Ramkumar, R. R. Sagunthala, A. Professor, A. Nerella, S. Kilaru, and G. Gladson Battu, "A Temporal Graph Neural Network Approach for Deep Fraud Detection in Real-Time Financial Transactions," in *THE 16th INTERNATIONAL IEEE CONFERENCE ON COMPUTING, COMMUNICATION AND NETWORKING TECHNOLOGIES (ICCCNT)*, 2025, pp. 1–5.

[14] S. Irfan, "Identification of Financial Fraud Transactions: A Cybersecurity via Machine Learning Methods," in *2026 IEEE International Conference on AI Engineering and Innovations (AIEI)*, NIT Jamshedpur, India: IEEE, 2026, pp. 1–6, May. doi: 10.1109/AIEI69164.2026.11497127.

[15] A. Thennakoon, C. Bhagyani, S. Premadasa, S. Mihiranga, and N. Kuruwitaarachchi, "Real-time Credit Card Fraud Detection Using Machine Learning," in *2019 9th International Conference on Cloud Computing, Data Science & Engineering (Confluence)*, 2019, pp. 488–493. doi: 10.1109/CONFLUENCE.2019.8776942.

[16] D. Patel, "Explainable Risk Decision Systems Using Artificial Intelligence Models for Payment Fraud Identification with Mitigation," in *2026 14th International Symposium on Digital Forensics and Security (ISDFS)*, Boston, MA, USA: IEEE, 2026, pp. 01–06, April. doi: 10.1109/ISDFS69419.2026.11459006.

[17] N. Radhasharan, "Real-Time AI and Data Transparency in Financial Services: A New Era of Trust and Liquidity Optimization," *J. Comput. Anal. Appl.*, vol. 35, no. 1, pp. 205–215, Jan. 2026, doi: 10.48047/jocaaa.2026.35.01.18.

[18] S. A. Pushkala, "Explainable AI for Trustworthy Predictions: The

Comparative Analysis of SHAP, LIME, and Counterfactuals," in *2025 2nd International Conference on Recent Trends in Electrical, Electronics and Computing Technologies (ICRTEECT)*, IEEE, Oct. 2025, pp. 1–5. doi: 10.1109/ICRTEECT67512.2025.11448658.

[19] H. S. Mohammed, Z. B. Sallow, and H. M. Zangana, "AI - Driven Fraud Detection in Digital Banking : A Hybrid Approach using Deep Learning and Anomaly Detection," *Sist. J. Sist. Inf.*, vol. 15, no. 1, pp. 209–219, 2026.

[20] S. B. Shah, "Advancing Financial Security with Scalable AI: Explainable Machine Learning Models for Transaction Fraud Detection," in *2025 4th International Conference on Distributed Computing and Electrical Circuits and Electronics (ICDCECE)*, Ballari, India: IEEE, 2025, pp. 1-7 , April. doi: 10.1109/ICDCECE65353.2025.11034838.

[21] H. A. Al Dulaimi *et al.*, "AI-Driven Behavioral Anomaly and Fraud Detection Models for Real-Time High-Frequency Financial Transactions in FinTech Systems," in *2025 3rd International Conference on Cyber Resilience (ICCR)*, IEEE, Jul. 2025, pp. 1–7. doi: 10.1109/ICCR67387.2025.11292440.

[22] P. Silvia, Q. Aini, E. A. Nabila, Henderi, and H. Nusantoro, "The Role of User Behavior Patterns in Enhancing Fraud Detection in Online Banking: A Bibliometric Analysis," in *2024 2nd International Conference on Technology Innovation and Its Applications (ICTIIA)*, IEEE, Sep. 2024, pp. 1–6. doi: 10.1109/ICTIIA61827.2024.10761930.

[23] R. Sharma and A. Sharma, "Combatting Digital Financial Fraud through Strategic Deep Learning Approaches," in *2024 2nd International Conference on Sustainable Computing and Smart Systems (ICSCSS)*, 2024, pp. 824–828. doi: 10.1109/ICSCSS60660.2024.10625249.

[24] E. P. Galla *et al.*, "Enhancing Performance of Financial Fraud Detection Through Machine Learning Model," *J. Contemp. Educ. Theory Artif. Intell.*, pp. 1–7, 2023, doi: 10.47991/2996-4954/JCETAI-101.

[25] X. Zhao, Y. Liu, and Q. Zhao, "Improved LightGBM for extremely imbalanced data and application to credit card fraud detection," *IEEE Access*, vol. 12, pp. 159316–159335, 2024.

[26] M. Ononiwu, T. Azonuche, F. Onum, and O. Enyejo, "Machine Learning Approaches for Fraud Detection and Risk Assessment in Mobile Banking Applications and Fintech Solutions," *Int. J. Sci. Res. Scince, Eng. Technol.*, vol. 10, no. 4, pp. 371–390, 2023, doi: https://doi.org/10.32628/IJSRSET.

[27] S. Gupta, "Securing Unified Payments Interface : A Deep Learning Approach for Fraudulent Transaction Detection," *J. Glob. Res. Multidiscip. Stud.*, no. 1, pp. 1–8, 2025.

[28] M. Jabeen, S. Ramzan, A. Raza, N. L. Fitriyani, M. Syafrudin, and S. W. Lee, "Enhanced Credit Card Fraud Detection Using Deep Hybrid CLST Model," *Mathematics*, vol. 13, no. 12, p. 1950, Jun. 2025, doi: 10.3390/math13121950.

[29] K. Hayat and B. Magnier, "Data Leakage and Deceptive Performance: A Critical Examination of Credit Card Fraud Detection Methodologies," *Mathematics*, vol. 13, no. 16, p. 2563, Aug. 2025, doi: 10.3390/math13162563.

[30] P. Sharma, S. Banerjee, D. Tiwari, and J. C. Patni, "Machine Learning Model for Credit Card Fraud Detection- A Comparative Analysis," *Int. Arab J. Inf. Technol.*, vol. 18, no. 6, pp. 75–82, 2021, doi: 10.34028/iajit/18/6/6.

[31] N. K. Trivedi, S. Simaiya, U. K. Lilhore, and S. K. Sharma, "An Efficient Credit Card Fraud Detection Model Based on Machine Learning Methods," *Int. J. Adv. Sci. Technol.*, 2020.

[32] S. Rajora *et al.*, "A Comparative Study of Machine Learning Techniques for Credit Card Fraud Detection Based on Time Variance," 2018, pp. 1958–1963. doi: 10.1109/SSCI.2018.8628930.