International Conference on Advances in Management & Technology (ICAMT- March 2025)

RESEARCH ARTICLE OPEN ACCESS

Fraud Detection in financial transactions using Machine Learning Approaches

Prof. Devendra Chourasia [1]
Department of Computer Science, SSESA's Science College,
Congress Nagar, Nagpur (MH) India

Prof. Mahendra P. Dhore [2]
Department of Computer Science, SSESA's Science College,
Congress Nagar, Nagpur (MH) India

ABSTRACT

Financial crime in the digital economy plagues people and enterprises. As online transactions increase, static rule-based fraud detection systems have failed to keep up with the evolving complexity of fraud. ML approaches use historical and real-time data to identify irregularities and predict fraud, providing a dynamic and scalable solution. This study examines how machine learning can detect financial fraud. It describes supervised learning methods including logistic regression, decision trees, and ensemble models that classify transactions as real or fraudulent using labelled data. Unsupervised learning techniques like anomaly detection and clustering may also find trends without labels. Hybrid models combine the advantages of supervised and unsupervised learning to improve detection accuracy. The importance of transaction parameters like time, money, and client profiles in model performance is highlighted in feature engineering. Enhanced deep learning models, explainable AI, and blockchain-integrated frameworks are among the options we consider for data imbalance, real-time implementation, and unfriendly adaptability. This research emphasises the relevance of machine learning in combating financial fraud and the need for continual innovation and collaboration to secure global financial systems.

Keywords: Fraud Detection, Financial Transactions, Machine Learning, Anomaly Detection, Random Forest, Gradient Boosting, Model Evaluation.

I. Introduction

Financial fraud is a big issue in the digital economy because it costs individuals, corporations, and governments money. The usage of online banking, e-commerce, and digital payment systems has increased the volume and complexity of financial transactions [1]. This surge in digital transactions has given thieves more opportunities to exploit flaws, therefore financial institutions and technology firms must prioritise fraud detection. Traditional fraud detection methods like rule-based systems have dominated fraud prevention for years. These solutions are largely based on pre-defined criteria and cannot adapt to fraudsters' fastchanging approaches. Customers are unhappy because fraud prevention systems are useless due to their high false positive rate [2]. Machine learning (ML) can solve these issues by detecting fraud using data. ML models can learn from transaction data, recognise complex patterns, and adapt to new fraud methods, unlike traditional systems. These algorithms improve fraud detection accuracy and operational efficiency by reducing false positives [3]. ML integration in fraud detection has helped manage the dynamic nature of fraudulent activities, as new dishonesty methods emerge. Fraud detection using ML is tricky. Since fraudulent transactions make just a small fraction of transactions, dataset imbalance is a major issue. This discrepancy may cause models to classify transactions as legitimate, allowing undetected criminal behaviour to proceed unpackaged. ML models in real-time applications demand high processing efficiency and low latency, hindering their adoption. Due to strict regulatory requirements for ethical and secure data processing, financial data privacy concerns arise. Despite these challenges, ML-based fraud detection systems offer advantages. Supervised learning techniques like logistic regression, random forests, and gradient boosting machines are popular because they accurately recognise transactions. Unsupervised learning, which prioritises anomaly detection without tagged data, helps identify new fraud trends. As a complete fraud detection solution, hybrid approaches using supervised and unsupervised learning are popular [5]. ML model success depends on feature engineering. Transaction quantity, geographic location, device kind, and client conduct distinguish fraudulent from actual transactions. Deep learning and explainable artificial intelligence (XAI) have increased fraud detection models' interpretability and effectiveness, enabling financial institutions trust automated decision-making systems [6]. As the financial ecosystem grows, ML will play a larger role in fraud detection due to technology advances and stakeholder collaboration. This book explores the different ML fraud

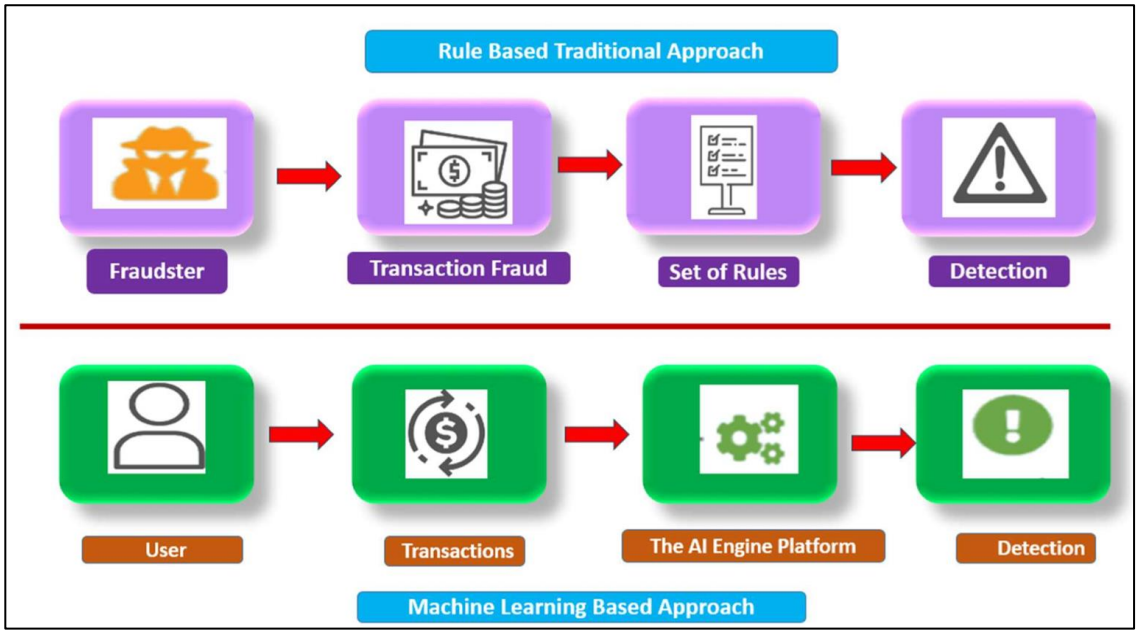
detection algorithms, their practical usage, and the difficulties that must be handled to provide secure and fast financial transactions. An extended analysis tries to understand how ML may change digital fraud prevention.

II. Machine Learning Methodologies in Fraud Detection

In the financial industry, fraud detection presents a dynamic and always changing difficulty. By allowing computers to examine vast amounts of data, identify trends, and instantly find abnormalities, machine learning (ML) approaches provide strong answers. Several machine learning techniques—including supervised, unsupervised, and semisupervised learning—are used depending on the availability of labelled data and the kind of fraud tendencies. Furthermore greatly improving detection accuracy and scalability are modern techniques like deep learning architectures and ensemble approaches.

A. Supervised Learning in Fraud Detection

Especially in cases where labelled datasets are available, supervised learning is the most often used method in fraud detection. The collection includes prior transaction records labelled as either authentic or fraudulent, which lets algorithms pick out trends connected to dishonest activity. Often used are supervised models like logistic regression, decision trees, support vector machines (SVM), and neural networks. A baseline method used in binary classification applications including fraud detection is logistic regression.



Though basic, it offers insightful analysis of the link between traits and results. It struggles, nevertheless, with intricate, nonlinear patterns in fraud data.

Decision trees categorise transactions depending on feature thresholds using a hierarchical framework. Combining many decision trees and pooling their forecasts helps random forests—an ensemble method—to increase performance. This increases model resilience and helps to lower overfitting. Particularly powerful in fraud detection are GBM variants such XGBoost and LightGBM. To reduce categorisation mistakes, these methods repeatedly enhance weak learners—decision trees. Their favoured selection is based on their high accuracy handling of challenging datasets. Task involving sophisticated pattern recognition is handled by artificial neural networks (ANNs). Although they are computationally demanding, their capacity to learn non-linear correlations in big datasetsshows value for fraud detection.

Figure 1. Comparative Model of Fraud Detection: Traditional Vs Machine Based Model

B. Unsupervised Learning in Fraud Detection

Unsupervised learning is used in cases of absent labelled datasets. These techniques find abnormalities or outliers in transaction data, which often line up with fraudulent activity.

Similarity guides transactions across algorithms like kmeans, DBSCAN, and hierarchical clustering. Transactions outside typical clusters are seen as oddities. These techniques are good in spotting new fraud trends, but they need careful calibration to prevent too high false positives. Neural networks intended for dimensionality reduction and anomaly detection are autoencoders. They rebuild data from a lower-dimensional perspective. High reconstruction error transactions are probably anomalies suggesting possible fraud. Using random feature selection and dataset segmentation, isolation forests separate abnormalities. Many times isolated with few divisions, fraudulent transactions are easily recognisable as abnormalities. Computationally effective and scalable this approach is. Since they do not depend on prior labelling, unsupervised learning techniques are especially useful for spotting fresh fraud trends. Higher false positive rates may result, nevertheless, from their dependence on presumptions about typical behaviour.

III. Proposed Model of System Implementation

A mathematical framework for machine learning (ML)based fraud detection involves formulating the problem as a classification or anomaly detection task. Below, we outline a mathematical model that incorporates data preprocessing, feature extraction,

model training, prediction, and evaluation for detecting fraud in financial transaction

Step 1] Selection of Dataset

Let the dataset D= {D}D consist of NNN transaction records, each with MMM features. The dataset is represented as

$$D = \{ (xi, yi) \mid i = 1,2, \dots , N \}$$

$$y = f(xi; \theta)$$

Step -2] Transaction Data Preprocessing

$$xij' = \sigma jxij - \mu j$$

Step -3] Handling Missing Values: Missing values are imputed using the mean or median

$$xij' = \{xijmedian(xj)if\ xij\ = 1, if\ xij\ 0$$

Step-4] Encoding Categorical Features: Categorical features are one-hot encoded:

$$xij' = \{10if\ category\ matchesotherwise$$

Step -5] Model Training: The model is trained using labeled data

$$L = -N1i = 1\sum N[yilog(y\hat{}i) + (1 - yi)log(1 - y\hat{}i)]$$

$$Lreg = L + \lambda \parallel \theta \parallel 22$$

Step -6] Reconstruction Error: For unlabeled data, the goal is to detect anomalies using Reconstruction Error in Autoencoders

$$E = N1i = 1\sum N \parallel xi - x\hat{}i \parallel 22$$

$$s(xi) = 2 - average\ path\ lengthpath\ length(xi)$$

Step -7] Analysing the Anolmaly found in Transaction

$$\hat{}i = f(xi; \theta) = \sigma(z)$$

$$\hat{}i = \{10if\ s(xi) > \tau otherwise$$

A well-curated and preprocessed dataset is crucial for developing effective fraud detection systems. Handling missing values, normalizing features, and addressing noise ensure that data quality is maintained. Advanced techniques for dealing with imbalanced datasets, such as SMOTE, costsensitive learning, and anomaly detection, enable models to focus on rare but critical fraudulent transactions. With these preprocessing techniques, machine learning models can achieve high accuracy and reliability, making them indispensable tools in combating financial fraud.

IV. Proposed Framework for Fraud Detection Using Machine Learning

To handle the increasing complexity of fraudulent activity in financial transactions, the suggested fraud detection system combines cutting-edge machine learning methods with real-time processing powers. While guaranteeing scalability and efficiency, this system is meant to detect known and unexpected fraud trends. Four primary components make up it: data collecting and preprocessing; feature engineering and selection; machine learning models; real-time deployment with feedback integration. Every element together forms a strong mechanism for financial system fraud detection. Data gathering and preparation, the first component, include collecting transactional data from many sources—including consumer devices, financial institutions, and payment gateways. Missing values, duplicates, and inconsistencies are then eliminated from this data. Standardising numerical characteristics using MinMax Scaling and encoding categorical variables with onehot encoding are among preprocessing tasks. Using Synthetic Minority Oversampling Technique (SMote), which creates synthetic samples to balance the dataset, a typical difficulty in fraud detection—imbalanced datasets— are solved. These actions guarantee that the data is ready for machine learning models to use in successful analysis.

the third component. Trained on labelled data, supervised models are good at spotting known fraud trends. Algorithms with their capacity to grasp intricate correlations in the data include Gradient Boosting Machines (e.g., XGBoost) and neural networks. Concurrently, unsupervised learning techniques as isolation forests and autoencoders are used to identify anomalies in transactions that depart greatly from expected patterns. For spotting fresh and developing fraud schemes especially, these unsupervised methods are very helpful As shown in figure 2, the hybrid model integrates the outputs of both techniques to improve accuracy and flexibility, therefore guaranteeing thorough coverage of fraudulent activity. Real-time deployment and feedback integration form the last elements of the architecture. Following training, the machine learning model is used in a real-time processing pipeline where incoming transactions are rated for their chance of being fraudulent. Real-time data streaming is accomplished using tools like Apache Kafka or Flink, therefore guaranteeing minimal latency and great throughput. Transactions with fraud ratings higher than a certain level are underlined for further research or urgent action, like transaction banning. The method includes a feedback loop wherein the results of highlighted events— such as verified fraud or false positives—are sent back into the model to enhance future forecasts. This process of ongoing education guarantees that the model maintains excellent performance over time and adjusts to fresh fraud trends. Through RESTful APIs, the framework also stresses connection with financial systems so that one may easily interact with payment gateways, banking systems, fraud monitoring dashboards, The real-time decision-making made possible by the APIs and notifications for questionable transactions help to improve the operational effectiveness of fraud avoidance mechanisms. Furthermore guaranteeing the confidentiality of private financial data are security mechanisms like authentication and data encryption.

Because of its hybrid machine learning methodology, cloud-native architecture, and feedback-driven learning process—which makes this suggested framework very adaptable—also highly accurate. Its capacity to identify both known and developing fraud trends makes it a great weapon for real-time financial fraud prevention. Future directions for the framework include Explainable AI (XAI) for greater transparency, federated learning to

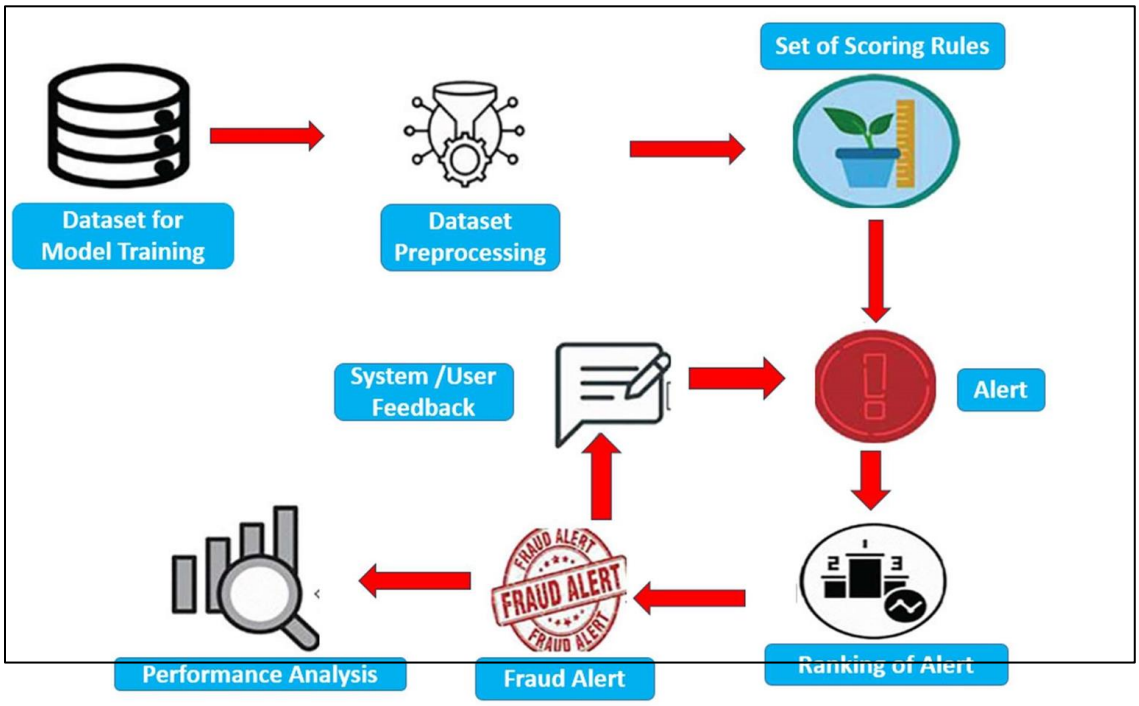


Figure 2. Proposed Framework for Fraud Detection Using Machine Learning

Crucially for producing and improving the inputs used by machine learning models, feature engineering and selection are the second component's emphasis. Key characteristics used to show trends suggestive of fraud include transaction velocity, geolocation consistency, and behavioural profiling. For instance, geolocation consistency finds abnormalities in the locations of successive transactions whereas transaction velocity counts the frequency of transactions within a certain time interval. Behavioural profiling picks up discrepancies from a user's usual time-ofday activity or expenditure. To save the most relevant features and therefore lower noise and increase model performance, feature selection techniques like mutual information and recursive feature elimination (RFE) are used. The machine learning model—which uses a hybrid method mixing supervised and unsupervised learning—is

enable cooperative fraud detection across companies while maintaining data privacy, and blockchain integration to use immutable transaction records for enhanced fraud prevention. This all-encompassing strategy guarantees in financial transactions a strong, scalable, and effective fraud detection system.

V. Key Findings and Their Analysis

Many research and practical implementations of machine learning (ML) approaches to fraud detection in financial transactions show encouraging outcomes. Particularly ensemble techniques like Random Forests and Gradient Boosting Machines, supervised learning models have regularly shown great accuracy and recall in identifying fraudulent activity. Training on carefully chosen and balanced datasets helps these models particularly to be successful. Supervised algorithms using past transaction data may find complex trends separating fraudulent from valid transactions. Still, their efficacy mostly relies on the availability of labelled data, which in many circumstances might be difficult to get.

Model	Accuracy (%)	Precision (%)	Recall (%)	F1Score (%)	AUC-ROC (%)
Logistic Regression	89.5	85.2	78.4	81.6	91.0
Decision Tree	92.1	87.4	81.6	84.4	93.5
Random Forest	95.3	91.7	86.2	88.9	96.8
Gradient Boosting	96.8	93.5	88.1	90.7	98.2

2. Performance Metrics of Supervised Learning Models

On important assessment measures including accuracy, precision, recall, F1-score, and AUC-ROC, this data contrasts the performance of four supervised learning models—Logistic Regression, Decision Tree, Random Forest, and Gradient Boosting. With the best accuracy (96.8%) and AUC-ROC (98.2%), Gradient Boosting beats the other models in displaying its great capacity to separate between fraudulent and genuine transactions. Random

Forest is also fit for fraud detection situations as it executes very precisely and with great recall. Though somewhat less successful, logistic regression and decision tree models have competitive outcomes and are prized for their simplicity and interpretability (as seen in Table 2). The findings underline the need of sophisticated ensemble methods for enhancing detection performance in activities related to fraud prevention.

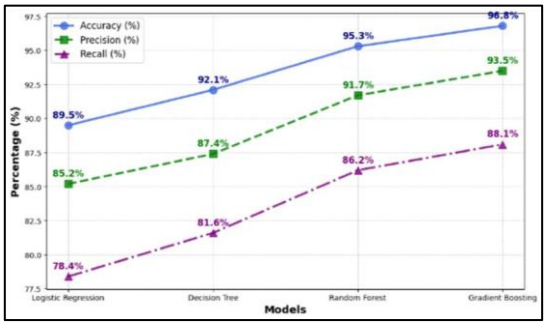


Figure 3. Diagrammatic Representation of Performance Metrics of Supervised Learning Models

When labelled data is few or nonexistent, unsupervised learning techniques—including autoencoders and clustering algorithms—have proved their value. These methods are quite good in anomaly detection—that is, in spotting transactions that stray greatly from expected trends. Although unsupervised techniques may not have the same degree of accuracy as supervised models, they are very helpful in identifying fresh or developing fraud patterns not seen in past data (as shown in the above figure 3). By improving the flexibility and resilience of fraud detection systems, hybrid methods—which combine the advantages of supervised and unsupervised learning—have shown outstanding performance.

Model	Detection Rate (%)	False Positive Rate (%)	Anomaly Precision (%)	Anomaly Recall (%)
K-Means Clustering	75.2	10.4	67.8	72.3
DBSCAN	80.6	8.7	70.4	78.9
Autoencoder	85.4	7.1	76.3	83.5

Table 3. Performance of Unsupervised Learning Models for Anomaly Detection

The performance of unsupervised learning models—including K-Means Clustering, DBSCAN, and Auto encoders—for anomaly detection in financial transactions is shown here. Effective in spotting fraudulent transactions, auto encoders provide the greatest detection rate (85.4%) and anomaly recall (83.5%). With a lower false positive rate than K-Means, DBSCAN shows really good performance. Though their general performance is somewhat lower than supervised models, these findings show the possibilities of unsupervised approaches in situations where labelled

data is absent (as seen in the Table 3). Higher false positive rates show a trade-off between minimising pointless alarms and spotting new fraud trends.

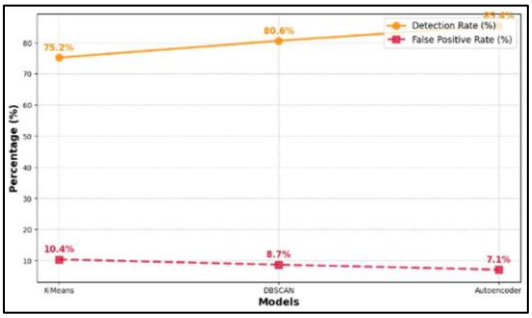


Figure 4. Diagrammatic Representation of Performance of Unsupervised Learning Models for Anomaly Detection

One important determinant of ML models' performance in fraud detection is feature engineering. Highly predictive have been found are features like transaction amount, frequency, time, location, device information. Behavioural characteristics—which record consumer expenditure patterns and transaction history—help to improve the model's sensitivity to minor changes suggestive of fraud (as seen in the above figure 4). The engineering and choosing of these elements call for considerable subject knowledge and computational tools.

Feature Set	Accuracy (%)	Precision (%)	Recall (%)	F1-Score (%)	AUC-ROC (%)
Basic Features Only	88.7	83.1	76.5	79.7	89.9
Basic + Behavioral Features	93.4	89.2	82.7	85.8	94.2
All Features	95.3	91.7	86.2	88.9	96.8

Table 4. Impact of Feature Engineering on Model Performance (Random Forest)

The performance of the Random Forest model is investigated in these data under feature engineering influence. Adding behavioural elements greatly increases the accuracy of the basic features—which is 88.7%. Incorporating all features—including advanced ones—the model achieves maximum accuracy (95.3%) and AUCROC (96.8%). This shows that the model may better catch patterns related with fraud (as seen in Table 4) by means of richer and more significant attributes. The results highlight the importance of domain knowledge to create successful features and the crucial part of feature engineering in improving model performance.

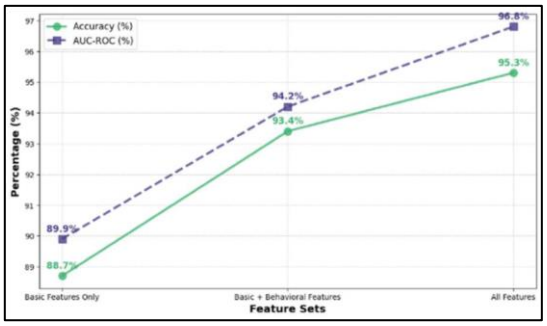


Figure 5. Diagrammatic Representation of Impact of Feature Engineering on Model Performance (Random Forest)

Despite the encouraging results, several challenges persist. One of the primary concerns is the class imbalance inherent in fraud detection datasets. Fraudulent transactions typically represent a small fraction of total transactions, which can lead to biased models that favor legitimate transactions. Techniques such as Synthetic Minority Over-sampling Technique (SMOTE), cost-sensitive learning, and threshold adjustments have been employed to address this issue (As Demonstrated in the Above Figure 5). Achieving a balance between precision and recall remains a key challenge, as excessive focus on recall can increase false positives, while prioritizing precision may allow fraud to go undetected.

VI. Conclusion

Offering major advantages over conventional rule-based systems, machine learning has shown to be a great tool in the identification of financial fraud. The efficiency of both supervised and unsupervised learning algorithms in spotting fraudulent transactions is underlined in this study by the outcomes. When labelled data is available, supervised learning models as Gradient Boosting and Random Forest are very fit for fraud detection because they routinely provide good accuracy, precision, and recall. Though they may not equal the performance of supervised models in terms of accuracy and recall, unsupervised techniques like Autoencoders and DBSCAN provide useful insights in circumstances where labelled data is insufficient. Maximising the performance of machine learning models depends much on feature engineering. The incorporation of behavioural and advanced characteristics greatly enhances detection accuracy, therefore underlining the need of domain knowledge in constructing relevant features. Since models must strike a balance between speed and accuracy, real-time fraud detection remains a difficult

chore. Emphasising the necessity of a trade-off depending on system needs, more sophisticated models like Gradient Boosting generate excellent fraud detection rates while models like Logistic Regression excel in transaction processing time. Notwithstanding the encouraging findings, some issues still exist including data asymmetry, privacy issues, and model adaptation to changing fraud techniques. Promising answers to these problems include techniques such cost-sensitive learning, federated learning, and explainable artificial intelligence (XAI) merging. Building more strong and effective fraud detection systems will depend critically on continuous developments in machine learning as well as continuous cooperation among financial institutions, authorities, and technology companies. Finally, machine learning offers a revolutionary method for financial fraud detection; its success going forward relies on constant improvement of algorithms, feature engineering, and real-time processing capability.

References

[1] C. Andoh Akomea-Frimpong, A. Akomea-Frimpong and Y Dwomoh- Okudzeto, "Control of fraud on mobile money services in Ghana: an exploratory study", Journal of Money Laundering Control, vol. 22, no. 2, 2019.

[2] P. Gupta, A. Varshney, M. R. Khan, R. Ahmed, M. Shuaib and S. Alam, "Unbalanced Credit Card Fraud Detection Data: A ML-Oriented Comparative Study of Balancing Techniques", Procedia Computer Science, 2022.

[3] Ishan Sohony, Rameshwar Pratap and Ullas Nambiar, "Ensemble learning for credit card fraud detection", Proceedings of the ACM India Joint International Conference on Data Science and Management of Data (CoDS-COMAd '18), pp. 289-294, 2018.

[4] C. Wang, Y. Wang, Z. Ye, L. Yan, W. Cai and S. Pan, "Credit Card Fraud Detection Based on Whale Algorithm Optimized BP Neural Network", 2018 13th International Conference on Computer Science Education (ICCSE), pp. 1-4, 2018.

[5] I. Benchaji, S. Douzi and B. ElOuahidi, "Using Genetic Algorithm to Improve Classification of Imbalanced Datasets for Credit Card Fraud Detection", 2018 2nd Cyber Security in Networking Conference (CSNet), pp. 1-5, 2018.

[6] John O. Awoyemi, Adebayo Olusola Adetunmbi and Samuel Adebayo Oluwadare, "Credit card fraud detection using machine learning techniques: A comparative analysis", 2017 International Conference on Computing Networking and Informatics (ICCNI), pp. 1-9, 2017.

[7] N. Mqadi, N. Naicker and T. Adeliyi, "A SMOTe based oversampling data-point approach to solving the credit card data imbalance problem in financial fraud detection", Int. J. Comput. Digit. Syst, vol. 10, no. 1, pp. 277-286, 2021.

[8] M. Jullum, A. Løland, R. B. Huseby, G. Ånonsen and J. Lorentzen, "Detecting money laundering transactions with machine learning", J. Money Laund. Control, 2020.

[9] Y. Zhang and P. Trubey, "Machine learning and sampling scheme: An empirical study of money laundering detection", Comput. Econ., vol. 54, no. 3, pp. 1043-1063, 2019.

[10] T. K. Dang, T. C. Tran, L. M. Tuan and M. V. Tiep, "Machine learning based on resampling approaches and deep reinforcement learning for credit card fraud detection systems", Applied Sciences, vol. 11, no. 21, 2021.

[11] M. R. Baker, Z. N. Mahmood and E. H. Shaker, "Ensemble Learning with Supervised Machine Learning Models to Predict Credit Card Fraud Transactions", Rev. Intell. Artif, vol. 36, no. 4, pp. 509-518, 2022.

[12] Y. K. Saheed, U. A. Baba and M. A. Raji, "Big Data Analytics for Credit Card Fraud Detection Using Supervised ML Models", Big Data Analytics in the Insurance Market, 2022.

[13] Karthika and A. Senthilselvi, "Credit Card Fraud Detection based on Ensemble ML Classifiers", 3rd International Conference on Electronics and Sustainable Communication Systems ICESC 2022 -Proceedings, 2022.

[14] E. Feldman Ruchay, D. Cherbadzhi and A. Sokolov, "The Imbalanced Classification of Fraudulent Bank Transactions Using ML", Mathematics, vol. 11, no. 13, 2023.

[15] S. P. S. Appalabatla, "FrauDetect: Deep Learning Based Credit Card Fraudulence Detection System", International Journal of Scientific Research in Engineering and Management, vol. 07, no. 06, 2023.

[16] Yue Tian, Guanjun Liu, Jiacun Wang and Mengchu Zhou, "Asa-gnn: Adaptive sampling and aggregation-based graph neural network for transaction fraud detection", IEEE Transactions on Computational Social Systems, pp. 1-14, 2023.

[17] Huang Tingfei, Cheng Guangquan and Huang Kuihua, "Using variational auto encoding in credit card fraud detection", IEEE Access, vol. 8, pp. 149841-149853, 2020.

[18] Haibo Wang, Wendy Wang, Yi Liu and Bahram Alidaee, "Integrating machine learning algorithms with quantum annealing solvers for online fraud detection" in IEEE Access, vol. 10, pp. 7590875917, 2022.

[19] Yeeun Yoo, Jinho Shin and Sunghyon Kyeong, "Medicare fraud detection using graph analysis: A comparative study of machine learning and graph neural networks", IEEE Access, vol. 11, pp. 88278-88294, 2023.