Enhancing Machine Learning for Credit Card Fraud Detection: A Comparison of Supervised Models

Amandeep Kaur[1], Arwinder Dhillon[2], Komal Singh Gill[3], Rajinder Vir[4]
{amandeepkaur.kaur06@gmail.com[1], arwinder.31660@lpu.co.in[2], gilldrkomal@gmail.com[3], rvir75@gmail.com[4] }

School of Computer Science and Engineering[1,2,3,4], Lovely Professional University, Punjab, 144411, India

Abstract. In the digital age, fraud involving credit cards has become a serious concern that impacts lots of people and companies worldwide. Reliable and intelligent fraud detection systems are becoming more and more important as e-commerce and online financial transactions grow in popularity. Conventional rule-based algorithms sometimes struggle with big, unbalanced datasets and are unable to identify new fraud tendencies. Using anonymized real-world data, this study explores the exploitation of Machine Learning models so that efficiently fraudulent credit card transactions could be identified. In this work three popular supervised learning models that are Random Forest, XGBoost, and Logistic Regression were used. Because of the extreme imbalance of the information, just 0.17% of transactions were deemed fraudulent, pre-processing techniques like normalization and the Synthetic Minority Oversampling Technique (SMOTE) is being utilized for balancing group in order to overcome this. Key performance indicators like Positive Predictive Value, sensitivity, f-measures, accuracy, and the region beneath the ROC curve (AUC-ROC) are used for evaluating the models' performance. With an AUC of 0.94, XGBoost outperforms the other models examined, demonstrating its resilience to differentiate between real and fake transactions. According to our findings, ensemble learning models in particular, XGBoost are very good at detecting fraud and can be included into real-time financial monitoring systems to improve security.

Keywords: XGBoost, Supervised Learning, Ensemble Models, SMOTE, Imbalanced dataset.

1 Introduction

The prevalence of fraud involving credit cards are now a major agitate for financial organizations throughout the world as digital transactions continue to develop as a result of the extensive use of e-wallet services, mobile banking, and online shopping. As per the Nilson Report, the fraud involving credit card has cost the world more than $28 billion in recent years, and estimates are still rising [1]. In addition to resulting in immediate financial loss, fraudulent credit card activity damages consumer confidence and necessitates expensive recovery procedures from service providers.

IICTDS 2025, December 19-20, Chandigarh, India
Copyright © 2026 EAI
DOI 10.4108/eai.19-12-2025.2362577

Conventional fraud detection systems, which are mostly rule-based and expert-driven, frequently don't have the adaptability to change with the quickly changing fraud strategies [2]. Additionally, these algorithms have trouble with extremely unbalanced datasets, which lead to large false negative rates. Because machine learning (ML) approaches can identify small anomalies in real-time and learn patterns from past data, they are being used more and more to overcome these restrictions[3]. The discipline of machine learning (ML) has developed as a strong and flexible remedy for these constraints. With little assistance from humans, ML models are able to recognize hidden patterns, learn from past transaction data, and generate predictions in real time. When labeled datasets are available for classification problems like fraud detection, supervised learning has demonstrated the most promise among the several machine learning techniques. Supervised models may effectively distinguish between suspicious and normal activity by training on features taken from transaction history, such as transaction amount, time, and anonymised PCA-transformed data. In particular, supervised learning algorithms present interesting solutions by distinguishing between fraudulent and authentic transaction data. In this work, we use the publicly accessible Kaggle dataset to compare three supervised ML models: XGBoost, Random Forest (RF), and Logistic Regression (LR) [4]. The models' efficacy in identifying fraudulent behavior in a highly imbalanced dataset is assessed using key assessment criteria like sensitivity, also referred to as recall and positive predictive value known as precision, f-measures, accuracy, and ROC-AUC. The Figure 1. demonstrates the connection between the rise of scam in credit card and the growth in digital transactions. Fraud detection systems have a number of difficulties as transaction volumes rise, such as extremely unbalanced datasets, the requirement for real-time processing, anonymised feature sets, and the possibility of false positives.

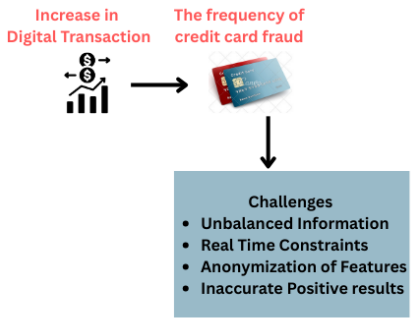


Fig. 1. Difficulties of identifying fraud in credit card within the framework of growing online transactions.

The remainder of this work is structured as follows: An overview is of pertinent research that has already been conducted and the most recent techniques for identifying credit card fraud provided in section 2. In section 3 the problem definition with respect to resarch is defined. The dataset, preprocessing methods, and study methodology are all covered in Section 4. The implementation is discussed in section 5 in which the findings and performance comparisons are shown. In section 6

the discussions related to this study is being discussed. The study is finally concluded and future trends are discussed in Section 7.

2 Related Work

Owing to the shortcomings of conventional rule-based systems and the rise in digital transactions, ML methods have become popular in fraud detection in recent years. Several research have investigated how well machine learning models use transaction-level characteristics and anomalous patterns to identify fraud related to credit card transactions. The role of ML techniques in detecting the fraud has raised substantially between 2013 and 2025, as shown in Figure 2, indicating that both academia and industry have recognized its potential.

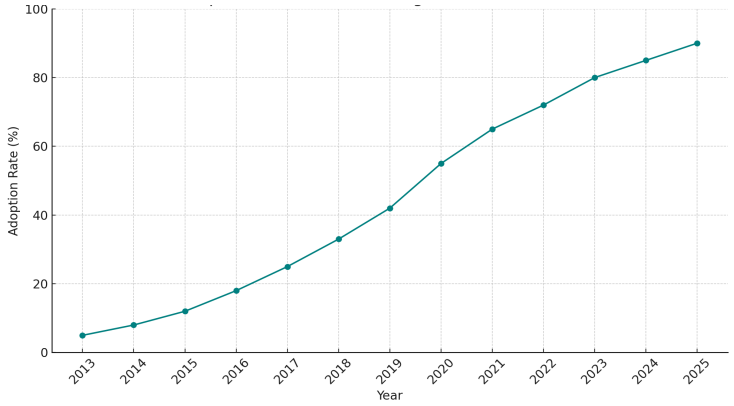


Fig. 2. Estimated Adoption of ML in Fraud Detection (2013-2025)

This section examines important research that has used ensemble approaches, data balancing strategies, and supervised learning algorithms to analyses fraud related to credit card in actual datasets. Khalid et al. [5] investigated how the identification of fraud in credit card could be enhanced by using ensemble ML techniques. A distinctive ensemble model that is a combination of KNN, SVM, RF, Bagging and Boosting are utilized to demonstrates the crucial issues such as imbalance data, false positives, and real-time processing. To deal with imbalanced data, methods like SMOTE and under-sampling are used. The study focuses on how ensemble techniques could be used to create in detecting the fraud that are more reliable.

The Isolation Forest model was used by Dal Pozzolo et al. [6] the irregularities were identified in credit card transactions. The precision that is achieved is only 6% and an achieved recall is 38%, The model didn't work well with fake transactions detection, despite achieving a high overall accuracy of 99%. This draws attention to a major problem that arises frequently in fraud detection tasks that is handling extremely imbalanced datasets.

The Credit Card Outlier Detection (CCOD) method, which integrate several ML models to recognize anomalies in credit card transactions, was given in research B Chugh et al. [7]. The main aim

of the concept was to improve the precision of detecting scam by using ensemble techniques, and it has attained a total accuracy of 90% with an F-measure of 0.91 and an AUC of 0.93.
V.G Helder et al. [8] analyzed the most pertinent input variables, researchers used the Variable Search for Neighborhoods heuristic to increase the detection in credit card. Nevertheless, the study lacked specific performance criteria that are essential for evaluating and comparing efficacy, such as the Area Under the Curve (AUC), accuracy, and precision.The direct comparability of their results with our research, where such parameters were measured and published directly, is limited by this omission. The XGBoost-based model used in this work presented a robust AUC of 0.94.
P Gupta et al. [9] presented ML classifier comparisons on unbalanced credit card fraud datasets. Data balancing procedures like Random oversampling were examined to determine how they affected classifier performance. With 0.91 precision and 99% accuracy, XGBoost outscored other models in the test. The precision increased to 0.99 with Random Oversampling, improving fraud detection. Since the study did not include AUC data, direct comparison versus our suggested XGBoost model, which had an AUC of 0.94, is limited.
Malik et al. [10] introduced a hybrid ML architecture for fraud detection in credit card. It uses Adaboost, LGBM, and XGBoost classifiers. Adaboost + LGBM outperformed the other hybrid setups with an AUROC of 0.82, high precision of 0.97, and recall of 0.64.
VN Dornadola et al. [11] exhibited a practical and dynamic methodology, however it failed to report crucial assessment parameters like AUC, which are needed to validate the model. The lack of these metrics prevents comparison with our proposed XGBoost model, which evaluates accuracy, precision, recall, F1-score, and AUC of 0.94. Table 1 lists key implementation-based studies, research gaps, results, and techniques.

3 Problem Statement

Fraud related to credit card is a big concern for customers as well as for banks because digital transactions are growing at an alarming rate and online banking and e-commerce sites are becoming more popular. Scammers always coming up with new ways to take advantage of weaknesses in payment systems, which makes it harder to recognize fraud. When it comes to detecting fraud, traditional rule-based systems frequently have high false-positive rates and can't adapt to new patterns of fraud. This makes customers unhappy and costs more to run the business. This imbalance makes it harder to work out models and lowers the accuracy of ordinary ML models when it comes to finding fraud. Intelligent, flexible, and scalable fraud detection solutions are needed to solve these issues. This study examines how deep learning and ensemble-based methods could be used for better accurately identify fraud in credit card. The suggested method seeks to increase detection accuracy while reducing false positives and guaranteeing prompt replies by utilizing synthetic oversampling (SMOTE), dimensionality reduction (PCA), and potent classification algorithms like XGBoost and Random Forest. Contributing to better secure financial systems that can manage contemporary fraud detection difficulties in real-time is the ultimate objective.

Table 1: Summary of Related Work

Reference	Technique Used	Result	Limitation
[6]	For Anomaly Detection Isolation Forest was used	99% accuracy, 6% precision, 38% recall, and 0.22 AUPRC	Unbalanced data leads to poor fraud detection; accuracy is high but precision and recall are low.
[7]	CCOD model using ensemble techniques	90% accuracy, 0.91 F1-score, and 0.93 AUC	AUC is marginally lower than our XGBoost, and the scope of improvement with the ensemble alone is constrained.
[8]	Seven machine learning classifiers are used in Variable Neighborhood Search (VNS) to pick features.	Enhanced capacity for discrimination with high-dimensional data	lacks conventional evaluation metrics and has neither published accuracy nor AUC.
[9]	Random Oversampling in XGBoost	Precision: 0.91, Accuracy: 99% (0.99 with oversampling)	Comparability is limited by partial metric reporting; AUC is not reported.
[10]	Architecture for hybrid machine learning: Adaboost + LGBM + XGBoost	AUROC: 0.82, Recall: 0.64, Precision: 0.97	Our model's AUROC is lower; the complex ensemble requires adjustment.
[11]	Streaming fraud detection using clustering	A new adaptive method using classifiers for clustering	AUC and other important parameters are not disclosed; there is little comparison with XGBoost.

4 Methodology

Using a publicly accessible dataset, we use supervised machine learning in this study to identify fraud related to credit card. Our approach includes dimensionality reduction, thorough pre-

processing, efficient management of data imbalance, and testing of several categorization models.

4.1 Dataset Description

The dataset is collected from Kaggle on fraud detection in credit card. This dataset consists of 284,807 transactions related to credit card from September 2013 across Europe are included. The 492 of these transactions are classified as fake, which results in a severely unbalanced dataset (0.172%). The main characterstics of the dataset is mentioned below.

- Time: The average amount of interval between transactions.
- Amount: The sum of the transaction.
- Class: 1 for fraud and 0 for legitimacy.
- V1 to V28: PCA-transformed anonymized characteristics.

The dataset's notable class imbalance is depicted Figure 3(a), where fraudulent transactions make up a very little percentage of all transactions. It highlights the challenge of identifying rare fraud occurrences in highly skewed databases.

The dissemination of quantities of transactions for both legitimate and illegal transactions is contrasted in Figure 3(b). The tendency of fraudulent transactions to cluster at smaller quantities raises the possibility that fraudsters favor smaller values in order to evade discovery.

4.2 Implementation

A thorough experimental pipeline that included data collection, preprocessing, dimensionality reduction, classification, and performance evaluation was created in order for detecting fraudulent credit card transactions. In order to correct class imbalance and derive valuable insights from anonymized transactional features, each step is essential.

4.2.1 Gathering and Preparing Data

- Data was collected in which only 492 of the 284,807 transactions in it have been classified as fraudulent. In order to comprehend the feature distributions and data quality, preliminary investigation was carried out using descriptive statistics and visualization [12].
- Normalisation: StandardScaler was used to standardize the features Time and Amount because they are not scaled. This changes the data is shown in equation 1.

$$X = \frac{z - \mu}{\sigma} \tag{1}$$

where the input value is denoted by z, the average by μ, and SD, or standard deviation by σ. This guarantees that every feature contributing equally to the model's training.

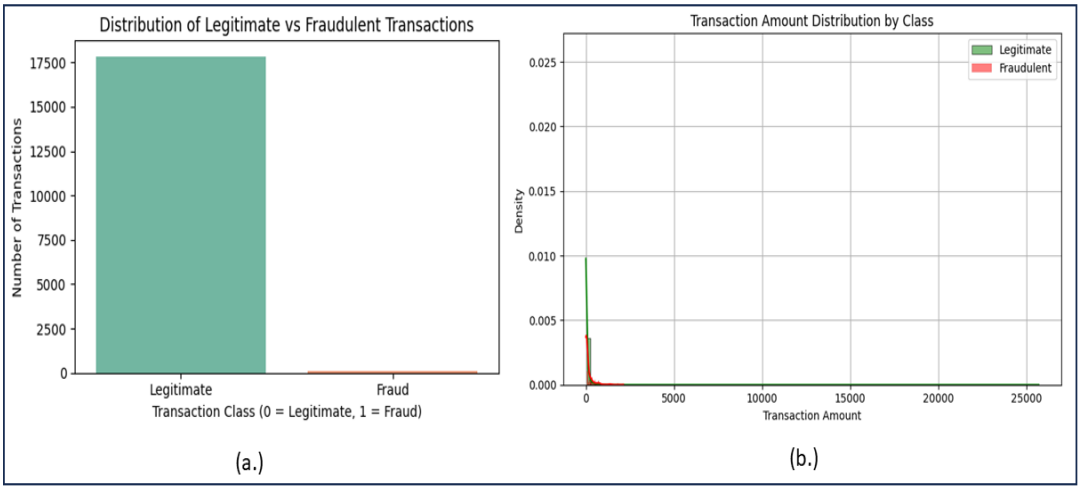


Fig. 3. (a.)Distribution of Legitimate vs Fraudulent Transactions; (b.)Transaction Amount Distribution by Class

- Managing Class Imbalance: In the dataset a considerable disparity in class exists(fraud $\approx$ 0.17%). In order to counteract this, SMOTE which creates examples of synthetic minority classes by comparing feature spaces of preexisting samples. This enhances the model's generalization to cases of minority classes.

4.2.2 Dimensionality Reduction Using PCA

In order to protect privacy, the dataset authors conducted a Principal Component Analysis (PCA) modification includes V1 through V28 features. Without trying to reverse-engineer the original qualities, we kept this alteration. A statistical method called principal component analysis (PCA) preserves the maximum variance while reducing correlated features to a smaller number of uncorrelated components as shown in equation 2.

$$Z = XW \tag{2}$$

Here, W is the covariance matrix's matrix of eigenvectors of X, Z is the modified feature matrix, and X is the input data matrix. PCA improves the effectiveness of model training by lowering overfitting [13]. The thorough experimental methodology used in the study is depicted in the Figure 4. Every stage, from gathering data to assessing the model, is essential to creating a trustworthy and efficient fraud detection system.

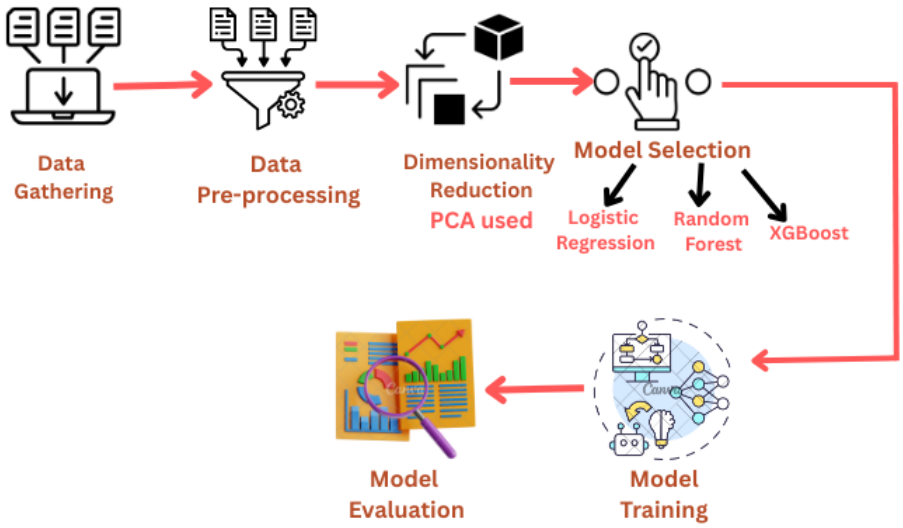


Fig. 4. Methodology

4.2.3 Model Selection

Three supervised classification methods were put into practice: XGBoost Classifier, Random Forest, and Logistic Regressionm [14]. Models are trained on labeled datasets in supervised learning, a machine learning technique in which each instance in training is accompanied by an output label. This makes the method suitable for issues with binary categorization, such as detecting fraud since it enables it to assimilate the connection between input features and the right output. These models were chosen because they are robust when handling highly unbalanced datasets, such as credit card fraud data, and they have a history of success in binary classification problems.

- **Logistic Regression(LR)** A linear framework that uses the logistic function to assess the likelihood of a binary outcome as in equation 3 [15].

$$P(Z = 1 \mid X) = \frac{1}{1 + e^{-(\alpha_0 + \alpha_1 X_1 + \cdots + \alpha_n X_n)}} \tag{3}$$

- **Random Forest(RF)** A group of decision trees (DT) that produce class predictions after being trained on a bootstrap sample. A majority vote determines the final product. It manages nonlinear data effectively and is resistant to overfitting [16].

- **XGBoost Classifier** An improved gradient boosting framework that builds trees more precisely by utilizing second-order derivatives[17]. Regularization and a loss function are com-

Table 2: Summarizes the Performance Metrics

Technique	Accuracy	Precision	Recall	F-measures	ROC-AUC
LR	0.97	0.83	0.66	0.73	0.90
RF	0.99	0.92	0.84	0.88	0.93
XGBoost	0.99	0.94	0.89	0.91	0.94

bined in the objective function. where Ω is the regularization term to regulate model complexity and l is the loss function (such as log-loss).

4.2.4 Key Performance Indicators and Analysis

Once the model is selected, the model's execution is analyzed by precision, recall, f-measure, accuracy, and the ROC-AUC curve [18, 19]. The True Positive Rate (TPR) and False Positive Rate (FPR) are plotted against the ROC curve. Area Under the Curve (AUC), where 1.0 represents perfect separation and 0.5 represents random guessing.

5 Results

In this research five primary performance metrics are used to analyze the classification's model performance, these performance metrics are being mentioned in figure 5. These metrics plays the vital part in evaluating the model's efficacy, specifically when dealing with unstable datasets such as fraud detection related to credit card. While precision and recall are more informative in detecting the fraud, precision shows the percentage of projected fraud cases that are real fraud, and recall demonstrates the model's ability to spot fake situations accuracy offers a broad measure of correctness. While the F-measures used for balancing precision and recall, the ROC-AUC offers a model's overall performance as a single measure. Table 2 demonstrates the assessment metrics for the three classification models that were assessed in this study: XGBoost, RF, and LR: Precision, Recall, F1-score, Accuracy, and ROC-AUC [20]. These assessment offer a comparative analysis of the way each model can recognize fraudulent transactions.

A visual depiction of the three models' pros and cons is made possible by the graphical comparison of these measures shown in Figure 5(b). This graphic makes it easy to see which model is superior to others in terms of several performance metrics.

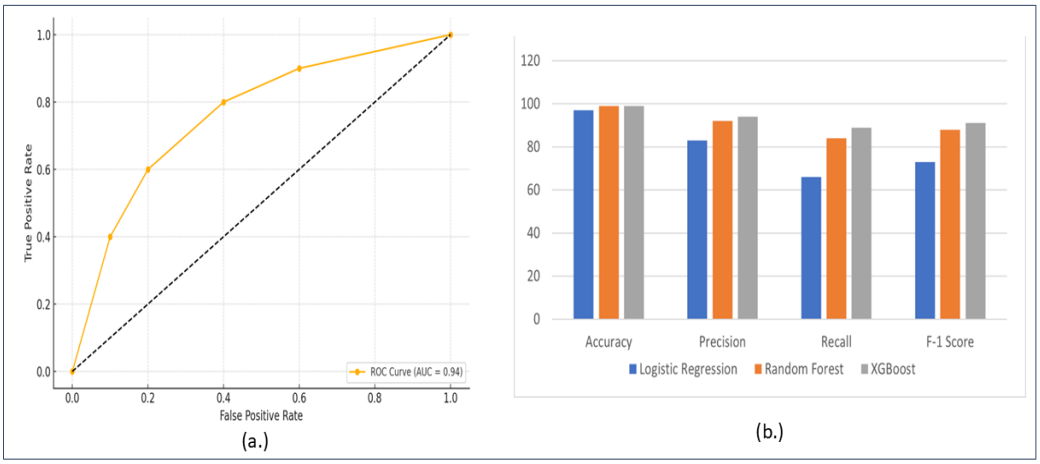


Fig. 5. (a.)ROC Curve for Identifying Credit Card Fraud; (b.)Comparison of assessment metrics among models

Plotting the Recall against the False Positive Rate at different threshold levels, the Receiver Operating Characteristic (ROC) curve shows how diagnostically capable a binary classification model is. Better model performance is shown by a larger area under the ROC curve (AUC). The XGBoost model had a remarkable capacity to differentiate between fake and real transactions in our study, as evidenced by its amazing AUC of 0.94. In comparison to previous models, the curve has a sharper increase, indicating fewer false positives and more accurate forecasts in Figure 5(a.).

6 Discussion

When supervised ML models are compared for detecting the credit card fraud, XGBoost routinely performs better than RF and LR on important assessment measures like accuracy, precision, sensitivity, F-measure, and AUC. XGBoost's methodology, which improves accuracy through iterative learning, enables it to identify between fake and legitimate transactions involving a high level of effectiveness (AUC of 0.94). Despite its ease of use and speed of execution, particularly in real-time systems, logistic regression has a very low recall (0.66) because it has trouble identifying intricate non-linear correlations in the data. False negatives, or fraudulent transactions that go unnoticed, are crucial in financial security applications, and this could result in more of them. Conversely, RF achieves a compromise by outperforming LR in terms of recall (0.84) and AUC (0.93) because of its ensemble-based architecture, which lowers variance and overfitting. Class imbalance and feature anonymization in the dataset pose problems for Random Forest and Logistic Regression, notwithstanding their robustness. PCA was used to control dimensionality and to artificially oversample the minority class SMOTE was utilized.

7 Conclusion

This research work presents the practical implications of fraud detection by exploiting ML techniques to this high financial loss risk, fraud risk, and fraud detection class imbalanced cases. In this research the comparisons of three different supervised classification models RF, XGBoost and Logistic Regression are used to greatly outperform the rest through their ensembles masted via XGBoost in the hyperparameter search with respect to financial loss risk fraud classified cases accuracy, precision, sensitivity, F-measure and AUC. Its complexity of the fraud risk detection problem is primarily through retaining the capacity to model non-linearities and self-correct adaptively. The resilience of the models has improved dramatically due the pre-processing techniques, primarily SMOTE and PCA, which worked to fully populate and balance the feature space streamlining critical to retaining useful patterns. XGBoost set the upper bound of the performance in the tests, but showcased significant potential room for advancement. Future research on the class imbalanced fraud risk cases would extend to retaining the financial loss risk fraud detection problem complexity, streamlining and integrating deep learning (DL) models. Recurrent Neural Networks (RNN) and CNN (Convolutional Neural Networks) for better of time and space of the transaction patterns. On the other hand, real time fraud loss detection in sequential stream is about possible with online learning which would greatly lessen the time delays of fraud strategies response without. Investigating explainable AI (XAI) methods may also boost the automated decision-making's transparency and trustworthiness, specifically in financial institutions where accountability is essential.

References

[1] Beju DG, Făt CM. Frauds in banking system: Frauds with cards and their associated services. In: Economic and financial crime, sustainability and good governance. Springer; 2023. p. 31-52.

[2] Bhattacharyya S, Jha S, Tharakunnel K, Westland JC. Data mining for credit card fraud: A comparative study. Decision support systems. 2011;50(3):602-13.

[3] Dal Pozzolo A, Boracchi G, Caelen O, Alippi C, Bontempi G. Credit card fraud detection: a realistic modeling and a novel learning strategy. IEEE transactions on neural networks and learning systems. 2017;29(8):3784-97.

[4] Kaur D, Saini A, Gupta D. Credit card fraud detection using machine learning, deep learning, and ensemble of the both. In: 2022 Seventh International Conference on Parallel, Distributed and Grid Computing (PDGC). IEEE; 2022. p. 484-9.

[5] Khalid AR, Owoh N, Uthmani O, Ashawa M, Osamor J, Adejoh J. Enhancing credit card fraud detection: an ensemble machine learning approach. Big Data and Cognitive Computing. 2024;8(1):6.

[6] Dal Pozzolo A, Caelen O, Le Borgne YA, Waterschoot S, Bontempi G. Learned lessons in credit card fraud detection from a practitioner perspective. Expert Systems with Applications. 2014;41(10):4915-28. Available from: https://www.sciencedirect.com/science/article/pii/S095741741400089X.

[7] Chugh B, Malik N, Gupta D, Alkahtani BS. A probabilistic approach driven credit card anomaly detection with CBLOF and isolation forest models. Alexandria Engineering Journal. 2025;114:231-42.

[8] Helder VG, Filomena TP, Ferreira L, Kirch G. Application of the VNS heuristic for feature selection in credit scoring problems. Machine Learning with Applications. 2022;9:100349.

[9] Gupta P, Varshney A, Khan MR, Ahmed R, Shuaib M, Alam S. Unbalanced credit card fraud detection data: A machine learning-oriented comparative study of balancing techniques. Procedia Computer Science. 2023;218:2575-84.

[10] Malik EF, Khaw KW, Belaton B, Wong WP, Chew X. Credit card fraud detection using a new hybrid machine learning architecture. Mathematics. 2022;10(9):1480.

[11] Dornadula VN, Geetha S. Credit card fraud detection using machine learning algorithms. Procedia computer science. 2019;165:631-41.

[12] Maharana K, Mondal S, Nemade B. A review: Data pre-processing and data augmentation techniques. Global Transitions Proceedings. 2022;3(1):91-9.

[13] Jia W, Sun M, Lian J, Hou S. Feature dimensionality reduction: a review. Complex & Intelligent Systems. 2022;8(3):2663-93.

[14] Camacho L, Douzas G, Bacao F. Geometric SMOTE for regression. Expert Systems with Applications. 2022;193:116387.

[15] Das A. Logistic regression. In: Encyclopedia of quality of life and well-being research. Springer; 2024. p. 3985-6.

[16] Sun Z, Wang G, Li P, Wang H, Zhang M, Liang X. An improved random forest based on the classification accuracy and correlation measurement of decision trees. Expert Systems with Applications. 2024;237:121549.

[17] Gündoğdu S. Efficient prediction of early-stage diabetes using XGBoost classifier with random forest feature selection technique. Multimedia Tools and Applications. 2023;82(22):34163-81.

[18] Miao J, Zhu W. Precision–recall curve (PRC) classification trees. Evolutionary intelligence. 2022;15(3):1545-69.

[19] Fang Y, Zhang Y, Huang C. Credit card fraud detection based on machine learning. In: Proceedings of the 2024 2nd International Conference on Image, Algorithms and Artificial Intelligence. vol. 115. Springer Nature; 2024. p. 35.

[20] Kaur A, Gupta D. A hybrid classification model for prediction of academic performance of students: An edm application. In: International Conference on Emergent Converging Technologies and Biomedical Systems. Springer; 2022. p. 59-71.