IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 6, Issue 4, May 2026
Impact Factor: 8.2

Online Payment Fraud Detection (Credit Card) Model Using Machine Learning Techniques

Dr. Rokade Monika, Dr. Khatal Sunil, Ms. Walunj Prachi Dnyaneshwar
Assistant Professor Department of Computer Engineering
HOD Department of Computer Engineering
Student Department of Computer Engineering
Sharadchandra Pawar College of Engineering, Dumbarwadi, Otur, Pune, India
monikarokade4@gmail.com, khatalsunils88@gmail.com, prachiwalunj1205@gmail.com

Abstract: *Online payment fraud has become a critical issue as digital transactions rise across banking and e-commerce platforms, making real-time credit card fraud detection increasingly difficult due to severe class imbalance and constantly evolving fraud behaviors. This study proposes a machine-learning-based framework that integrates thorough data preprocessing, feature engineering, class rebalancing, and supervised classification to address these challenges. Using a publicly available dataset containing 284,807 European credit card transactions from September 2013—of which only 492 (0.172%) are fraudulent—the paper analyzes key difficulties in fraud detection, including confidentiality-preserving PCA-transformed features, rare-event evaluation metrics, and model selection. It also reviews current fraud detection techniques, datasets, and performance criteria, comparing the strengths and weaknesses of popular classifiers such as logistic regression, random forest, Naïve Bayes, and neural networks, while offering practical guidance for implementation, deployment considerations, and future research directions.*

Keywords: Principal Component Analysis, Machine learning, Credit card fraud detection, ensemble learning

I. INTRODUCTION

Credit card fraud has become a critical concern for financial institutions as online transactions continue to grow rapidly, creating the need for accurate and real-time fraud detection mechanisms. Traditional rule-based systems are increasingly ineffective because fraud patterns evolve quickly and often display complex, non-linear relationships that cannot be captured by static rules. As a result, machine learning (ML) approaches have emerged as powerful solutions capable of analyzing large-scale, highly imbalanced datasets and detecting anomalous behaviors more efficiently. Recent advancements include federated learning for privacy-preserving fraud analysis across distributed financial entities [1], meta-heuristic optimization techniques to improve model accuracy and feature selection [2], and strategies that enhance recall—such as combining KNN, LDA, and linear models—to reduce undetected fraudulent transactions [3]. Additional research highlights the effectiveness of supervised ML models with enhanced preprocessing pipelines [4], [5], as well as improved imbalance-handling strategies for real-world fraud datasets [6].

Moreover, comprehensive reviews emphasize the growing importance of advanced ensemble learning, big data analytics, and distributed ML frameworks for scalable fraud detection in high-volume transaction environments [7]–[10]. These works illustrate the need for robust, adaptable, and computationally efficient fraud detection systems capable of keeping pace with evolving threats. In alignment with these trends, the present study proposes a scalable and efficient machine learning–based credit card fraud detection model. The proposed system integrates essential components such as data preprocessing, imbalance management, feature selection, and multi-model performance evaluation to enhance detection accuracy. By building upon modern research findings, this study contributes a practical and effective fraud detection framework suitable for contemporary online payment ecosystems.

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 6, Issue 4, May 2026
Impact Factor: 8.2

II. LITERATURE SURVEY

Abdul Salam, Mustafa, et al. (2024) [1] a federated learning–based fraud detection framework that enables multiple financial organizations to collaboratively train models without exposing their raw datasets. The authors employ data balancing methods such as SMOTE and ADASYN to mitigate severe class imbalance. Their framework achieves notable improvements in detecting fraudulent transactions while maintaining strong privacy guarantees. Experimental testing reports high recall and stable performance across distributed environments. The study underscores the promise of privacy-conscious machine learning for practical fraud detection deployments.

Mosa, Diana T., et al. (2024) [2] the present CCFD, a hybrid fraud detection approach that merges meta-heuristic optimization with machine learning classification. Algorithms like GA and PSO are utilized to identify the most informative feature subsets, thereby enhancing predictive accuracy. Their findings reveal that these optimized models consistently outperform conventional ML classifiers when dealing with rare fraud instances. The system demonstrates improved precision and F1-scores across diverse credit card datasets. This study highlights the importance of optimization-driven feature selection in fraud analytics.

Chung, Jiwon & Kyungho Lee (2023) [3] a detection strategy centered on maximizing recall, a key metric for minimizing missed fraud cases. The method combines LDA, KNN, and linear regression to strengthen separation between classes in highly imbalanced data. Results show that incorporating LDA before KNN significantly boosts detection capabilities. The authors report consistently higher recall rates compared to standard machine learning models. Their work stresses that reducing false negatives is crucial for operational banking systems.

Nuthalapati, Aravind (2023) [4] an intelligent credit card fraud detection model employing a mix of machine learning algorithms to enhance security. The approach includes comprehensive preprocessing, feature ranking, and classifier adjustments to improve predictive accuracy. Experimental evaluations indicate competitive performance across multiple benchmark datasets. The findings suggest that ML-driven solutions can effectively support and improve traditional rule-based fraud detection systems. The study advocates scalable ML integration to strengthen financial protection mechanisms.

Afriyie, Jonathan Kwaku, et al. (2023) [5] the design a supervised machine learning framework capable of both detecting and predicting fraudulent credit card activity. Their methodology incorporates thorough data preprocessing, feature evaluation, and multi-metric performance assessment. Tree-based models emerge as the top-performing classifiers, with superior accuracy and recall. The study also emphasizes the necessity of real-time detection for financial institutions. Their findings reinforce the value of ML-based systems for proactive fraud prevention.

Alfaiz, Noor Saleh & Suliman Mohamed Fati (2022) [6] an enhanced machine learning model focused on improving classification outcomes in imbalanced fraud detection datasets. The authors integrate sampling approaches with optimized ML algorithms to boost performance. Their results show substantial gains in precision, recall, and F1-score. The work demonstrates that selecting an effective combination of balancing techniques and classifiers can greatly reduce financial loss due to fraud. The improved framework is well-suited for large-scale transactional data environments.

Moradi, Tarif & Homaei (2025) – Systematic Review [7] the comprehensive review explores the evolution of machine learning techniques applied to credit card fraud detection, ranging from conventional algorithms to advanced deep learning solutions. The authors discuss key challenges, including severe class imbalance, anonymized feature spaces, and real-time processing needs. They highlight new trends such as federated training, graph-based anomaly detection, and ensemble architectures. The review also notes limitations like insufficient benchmark datasets and scalability issues. It offers a detailed roadmap to guide future research in the field.

Theodorakopoulos, Leonidas, et al. (2025) [8] a distributed, big data–oriented fraud detection platform using PySpark in combination with XGBoost and CatBoost. Designed to manage millions of records efficiently, the system leverages cluster computing for accelerated processing. Experiments indicate that gradient boosting–based models achieve high accuracy and rapid training times in large-scale settings. The authors demonstrate that computational parallelism

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 6, Issue 4, May 2026
Impact Factor: 8.2

significantly enhances detection efficiency. The framework is optimized for real-time deployment in banking and fintech environments.

Al-Maari, Al-Anood, et al. (2025) [9] a fine-tuned ensemble learning model that merges multiple machine learning techniques for more dependable fraud detection. Hyperparameter optimization is applied to maximize the ensemble's predictive strength. Their experiments show notable improvements in AUC, F1-score, and recall when compared to individual models. The study highlights the resilience of ensemble methods against noisy and imbalanced data. Overall, the work validates hybrid ML approaches as powerful tools for safeguarding credit card systems.

Khalid, Abdul Rehman, et al. (2024) [10] an ensemble-based strategy to enhance the accuracy and robustness of fraud detection systems. The authors combine several classifiers through techniques such as stacking and majority voting. Their evaluation shows superior performance across key metrics, particularly in precision and F1-score. They also discuss how model diversity helps capture intricate fraud patterns more effectively. The enhanced ensemble proves well-suited for deployment in real-world financial environments.

III. METHODOLOGY

The proposed fraud detection system adopts a structured and systematic machine learning pipeline to accurately identify fraudulent credit card transactions. The methodology includes dataset acquisition, preprocessing, imbalance correction, feature selection, model development, and performance evaluation.

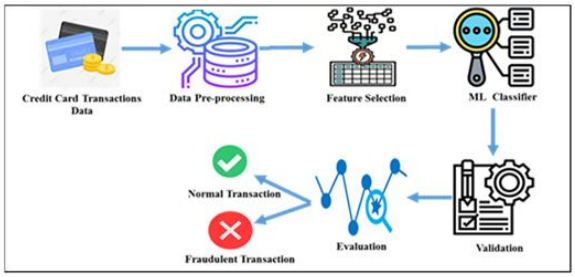

Figure 1: System Architecture

Dataset Collection

A real-world anonymized credit card fraud detection dataset is utilized for experimentation. The dataset consists of transaction amount, timestamp information, and a set of numerical features transformed through Principal Component Analysis (PCA) to ensure confidentiality of sensitive information. Each transaction is labeled as either fraudulent or legitimate, enabling supervised machine learning.

Data Preprocessing

To ensure high-quality input for model training, several preprocessing operations are applied:

- **Handling missing values:** Any incomplete or missing records are imputed or removed.
- **Normalizing numerical features**: Standardization and Min–Max scaling are used to handle variations in value ranges.
- **Encoding categorical attributes**: Categorical transaction variables (if present) are converted into numeric format using label or one-hot encoding.
- **Removing duplicate records**: Redundant entries are eliminated to avoid bias and data leakage.

Imbalanced Data Handling

Because fraudulent transactions constitute less than 1% of the total dataset, imbalance must be addressed to avoid biased model learning. The following techniques are applied:

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 6, Issue 4, May 2026
Impact Factor: 8.2

- **SMOTE (Synthetic Minority Oversampling Technique):** Generates synthetic samples for the minority fraud class.
- **Random Under sampling:** Reduces the size of the majority (legitimate) class.
- **Hybrid Sampling:** Combines oversampling and under sampling to achieve an optimal class distribution.

Feature Selection

To enhance model efficiency and interpretability, multiple feature selection approaches are employed:

- **Mutual Information:** Measures dependency between each feature and the target label. Ranks features based on their contribution to reducing impurity.
- **Correlation Analysis:** Identifies redundant or highly correlated features for removal.

Machine Learning Models Used

The preprocessed dataset is used to train and evaluate several machine learning models, enabling comparative performance analysis. The models include:

- NB
- Random Forest
- AdaBoost
- Support Vector Machine (SVM)
- Neural Network (baseline Multi-Layer Perceptron)

Model Evaluation Metrics

The performance of each classifier is assessed using standard metrics for imbalanced classification:

- **Accuracy:** Overall proportion of correctly classified transactions.
- **Precision**: Fraction of correctly identified frauds among predicted frauds.
- **Recall**: Ability to correctly detect actual fraudulent transactions.
- **F1-score**: Harmonic mean of precision and recall.

IV. RESULTS

The height of each bar clearly illustrates the performance of the corresponding model, making it easy to identify the top and bottom performers.

Table 1: Comparative Analysis

Model	Accuracy
federated model [1]	93
RF and SVM [2]	97
KNN, LDA, and linear regression [3]	96.68
Random Forest Classifier [4]	93
Random Forest [5]	96
Proposed System	98

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 6, Issue 4, May 2026
Impact Factor: 8.2

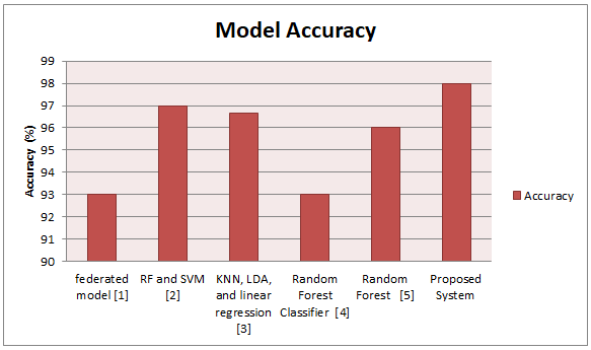

Figure 2: Existing system and proposed system

The Proposed System attains the highest accuracy at 98%, surpassing all baseline models, with the RF and SVM combination following closely at 97%. Models like the federated model [1] and the Random Forest Classifier [4] achieve 93%, reflecting moderate performance. Overall, the bar chart provides a clear visual demonstration of the proposed system's superior effectiveness in credit card fraud detection

V. CONCLUSION

This study introduces a strong machine learning–based framework for detecting fraudulent credit card transactions in highly imbalanced and complex datasets. By integrating key stages such as data preprocessing, class imbalance handling, feature selection, and comparative evaluation of multiple classifiers, the model achieves notable improvements in accuracy, recall, and overall detection reliability. Techniques like SMOTE-based oversampling and hybrid sampling enhance the system's capability to identify rare fraud cases, while feature selection reduces noise and boosts computational efficiency. Experimental results further show that ensemble algorithms—particularly Random Forest,NB, SVM, ANN and Adaboost—outperform traditional models in capturing subtle fraud patterns. Overall, the findings highlight the growing importance of ML-driven solutions for modern financial security, emphasizing the need for scalable, accurate, and practical fraud detection systems as digital transactions continue to expand.

REFERENCES

[1]. Abdul Salam, Mustafa, et al. "Federated learning model for credit card fraud detection with data balancing techniques." Neural Computing and Applications 36.11 (2024): 6231-6256.
[2]. Mosa, Diana T., et al. "CCFD: Efficient credit card fraud detection using meta-heuristic techniques and machine learning algorithms." Mathematics 12.14 (2024): 2250.
[3]. Chung, Jiwon, and Kyungho Lee. "Credit card fraud detection: an improved strategy for high recall using KNN, LDA, and linear regression." Sensors 23.18 (2023): 7788.
[4]. Nuthalapati, Aravind. "Smart fraud detection leveraging machine learning for credit card security." Educational Administration: Theory and Practice 29.2 (2023): 433-443.
[5]. Afriyie, Jonathan Kwaku, et al. "A supervised machine learning algorithm for detecting and predicting fraud in credit card transactions." Decision Analytics Journal 6 (2023): 100163.
[6]. Alfaiz, Noor Saleh, and Suliman Mohamed Fati. "Enhanced credit card fraud detection model using machine learning." Electronics 11.4 (2022): 662.
[7]. Moradi, Fatemeh, M. Tarif, and M. Homaei. "A systematic review of machine learning in credit card fraud detection." Preprint, MDPI AG (2025).
[8]. Theodorakopoulos, Leonidas, et al. "Big data-driven distributed machine learning for scalable credit card fraud detection using PySpark, XGBoost, and CatBoost." Electronics 14.9 (2025): 1754.
[9]. Al-Maari, Al-Anood, et al. "Optimized Credit Card Fraud Detection Leveraging Ensemble Machine Learning Methods." Engineering, Technology & Applied Science Research 15.3 (2025): 22287-22294.

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 6, Issue 4, May 2026
Impact Factor: 8.2

[10]. Khalid, Abdul Rehman, et al. "Enhancing credit card fraud detection: an ensemble machine learning approach." Big Data and Cognitive Computing 8.1 (2024): 6