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, June 2026
Impact Factor: 8.2

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

Walunj Prachi Dnyaneshwar, Dr. Monika Rokade, Dr. Sunil Khatal
Student, Dept. of Computer Engineering
Sharadchandra Pawar College of Engineering, Otur, India

Abstract: *Credit card fraud has become one of the largest issues facing businesses and banks in the digital age, as online banking, online shopping, and digital transactions have all increased the likelihood of fraud, causing monetary loss for both the bank and customer. Fraud detection is especially difficult due to most transactions being legitimate. Traditional systems are often ruled-based, and as fraud systems become more sophisticated, they will inevitably produce high rates of fraud under the detection threshold, and high rates of missed fraud cases. This work aims to provide a solution using a framework, (MFCC), based on machine learning. The framework includes data preprocessing and the use of Principal Component Analysis (PCA) to handle class imbalance through the Synthetic Minority Oversampling Technique (SMOTE). The framework also considers Feature Selection using the Information Gain approach. The framework is focused on the Random Forest (RF) and Support Vector Machine (SVM) algorithms, both of which are popular in classification and anomaly detection. Both models used in this experiment are good at spotting fraudulent transactions. The Random Forest model shows that accuracy is 96.36%, precision is at 100%, recall is 75.51%, and F-measure is 86.05%. On the other hand, the Support Vector Machine model (or SVM) does an even better job than Random Forest by reaching an accuracy of 97.55%, precision of 97.2%, recall of 100%, and an F-measure of 98.58% with much fewer fraudulent transactions being flagged as good ones (or false negatives). The Smart SVM Fraud Detection Framework is a great way of spotting fraudulent transactions since it combines several good preprocessing and machine learning techniques. Of all the models we used, SVM is the model of choice to use in real-world credit card transaction frameworks to bolster transactional digital security*

Keywords: Credit Card Fraud Detection, Machine Learning, SMOTE, Random Forest, Support Vector Machine, Hybrid Model.

I. INTRODUCTION

Digital banking has transformed the financial nation on a global scale with its ease, speed, and security. With these transformations in the payment system, credit cards have become one of the most predominant payment systems. But problems have emerged with the growing popularity in digital payment systems with an increase in credit card payment system frauds. Credit card fraud has become one of the most significant financial crimes with an even greater impact on the customer and the banking system with billions of dollars of loss reporting annually. [1][2].

Fraudulent transactions occur when a credit card payment system is exploited with unauthorized access. Methods vary, with phishing, skimming, identity fraud, theft, and even data breaches. Fraud is an ever-evolving crime and has rendered many modern systems out of date. Traditional systems focus on a few basic rules; with the evolved techniques, these systems fail to detect such fraud and end up with a high rate of false positives. Such rates result in frustration for valid end users and loss of trust in digital banking systems [3][4].

A major issue with fraud detection in the credit card payment system is in the disparity of data sets. Majority of the data in a real system is comprised of valid transactions with only a small amount of data pertaining to fraud. Such an extreme system imbalance impacts the capability of machine learning models to accurately detect fraud. Most

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, June 2026 Impact Factor: 8.2

traditional systems end up learning on the majority class data (valid transactions) and as a consequence detection of fraud is very weak. [5][6].

Machine learning has become popular because it can solve problems that previous methods can't. Since fraud can be detected by identifying anomalous patterns in transaction data, machine learning has been useful in this area. In fraud detection, random forest and support vector machine (SVM) play important roles, as do artificial neural networks (ANN). Breakthroughs have been achieved in fraud detection and anomaly detection thanks to advancements in some deep learning techniques, as well as hybrid models that combine various algorithms [9][10].

Data preprocessing techniques also have important roles to play when it comes to improving model performance. Other techniques such as Principal Component Analysis (PCA) and the Synthetic Minority Oversampling Technique (SMOTE) can also be useful. While PCA is useful for the reduction of large amounts of data, SMOTE can handle imbalances in data classification in such a way that models can be trained more effectively on the data [1][5].

The objective of this study is to provide a foundation on which credit card fraud detection can be efficiently implemented. This study focuses on machine learning, data preprocessing, and the use of various classification techniques. The key concern is to enhance the accuracy of the detection of fraud, in addition to limiting the number of false positives. This can increase the security and dependability of online payment systems..

II. LITERATURE SURVEY

Alatawi et al. [1] designed an ML-based system that uses credit card transaction data from IoT devices to detect fraudulent activities. They stressed the need to quickly analyze data in order to identify fraudulent activities. They used various classification methods to detect abnormal behaviors in transaction data. They observed that the developed system is able to detect fraudulent activities in the data with a high level of accuracy when big data is used. They indicated that the use of IoT devices for transaction data collection is becoming increasingly important in detecting financial fraud.

In [2], Alrasheedi et al. reviewed and compared different ML models for credit card fraud detection. They investigated models like Decision Trees, and Random and Support Vector Machines. It was observed that the models that use an ensemble of several classifiers outperform the accuracy and the measure of interest in fraud detection over the single classifiers included in the comparison. The results of the study also highlighted that when dealing with a class-imbalance dataset, the choice of evaluation measures is of great importance. The work in this study is of great importance for the choice of models for fraud detection tasks.

Al-Maari et.al [3] proposed a new fraud detection system that uses ML with ensemble techniques. This study showed that the addition of multiple classifiers to a model increases the accuracy and the robustness of the prediction of the model, and that the developed method eliminates the issue of the high number of false positives that were observed in previous fraud detection models. The authors were also able to improve the speed of the model while continuing to provide a high level of accuracy to the users of the model. This study mainly supports the application of the ensemble approach to multiple fraud detection tasks.

Theodorakopoulos et al. [4] display their architecture for a highly scalable fraud detection system that employs big data technologies such as PySpark, XGBoost and CatBoost. Their approach is designed to manipulate massive transactional data in distributed computing. The system demonstrated quantifiable improvements in fraud detection with positive impact to speed, size and accuracy. The team stressed the need to couple big data technologies with machine learning. This work is especially helpful to financial organizations that work with big data.

Wang et al. [5] assessed the operating performance of a framework combining techniques for data balancing and learning via ensembles. The fraud detection challenge of unbalanced classes was tackled using the SMOTE approach. The authors noted higher detection success for fewer fraudulent cases as well as increased performance of the model with a balanced dataset. Their work also emphasizes the value of data cleansing for fraud detection.

Unogwu and Filali [6] examined several algorithms for detecting fraud such as the Decision Trees and Support Vector Machines to identify fraudulent patterns in credit card transactions. The authors noted that while traditional algorithms

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, June 2026 Impact Factor: 8.2

have high success rates, they perform poorly on data that is highly imbalanced. This work aids in recognizing the limits of fraud detection.

Afriyie et al. [7] created a supervised learning model that predicts fraudulent credit card transactions. This model is based on constructing labeled datasets to train various classification algorithms. This study explains that supervised learning models require sufficient training data in order to achieve sufficient accuracy. They also stressed that, even with a small amount of training data, the performance of the learning model can be greatly improved by including a well-defined set of features. Their study recommends the use of supervised learning in fraud detection systems.

Khalid et al. [8] presented an ensemble machine learning model that attempts to improve the accuracy of fraud detection systems. This model combines a collection of various machine learning models to improve the system's precision and recall. This study explains that, when dealing with complex fraud detection, an ensemble model will outperform a single model. This study also considered the challenges that were associated with the under-fitting and over-fitting of the model. This study also explains that, when creating a fraud detection model, combining various machine learning algorithms can be very beneficial.

For the first time, Alarfaj and Shahzadi [9] suggested the use of advanced Graph Neural Networks and autoencoders within the realm of deep learning to further the advancements of fraud detection technologies. This model is able to analyze the relationships between various transactions, thereby identifying complex fraudulent behavior. For this study, the analysis of deep learning advanced the real-time detection of fraudulent activities. This study also considered the challenges that were associated with creating deep learning models in order to meet the goal of containing the fast and ever-evolving tactics employed by fraudsters. This study has provided advancements for the technologies employed in the detection of fraud.

Qayoom et al. [10] offered the first deep learning-based fraud detection system based on the concepts of deep learning and advanced artificial intelligence technologies. This study further developed the concept of an adaptive system by utilizing real-time data that is readily available for deep learning. This study suggested that fraud detection technologies can be both intelligent and self-sufficient. The study also provided evidence that the use of deep learning in the realm of advanced artificial intelligence can further the development of fraud detection technologies.

III. METHODOLOGY

The developed system is called the Machine Learning-based Framework for Credit Card Fraud Detection (MFCC). It has the ability to identify whether a particular transaction is legitimate or fraudulent. It employs a multi-step, elaborate, and sophisticated methodology. This includes data collection, modification, preprocessing, handling class imbalance, selection of pertinent attributes, training of the model, and assessment of the model. All of these phases contribute to a better-performing system.

Data Collection

This study uses data from actual credit card transactions. In order to maintain privacy, some sensitive components are modified using Principal Component Analysis (PCA). The dataset contains both genuine and fraudulent transactions. In this case, fraudulent transactions make up a tiny portion of the dataset and the dataset is very imbalanced.

Pre-Processing Data

To get data ready for a machine learning model, we have to do a lot of data prep. This involves cleaning the data which includes removing duplicate entries and filling or removing missing values. Raw data can also have weirdly large or small numbers. We can use normalization to fix that. To make sure that all the data features contribute equally, we can scale and standardize the features. PCA or, principal component analysis, is a technique that helps us reduce the number of features that we are working with. It helps to keep the essential data we need

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, June 2026
Impact Factor: 8.2

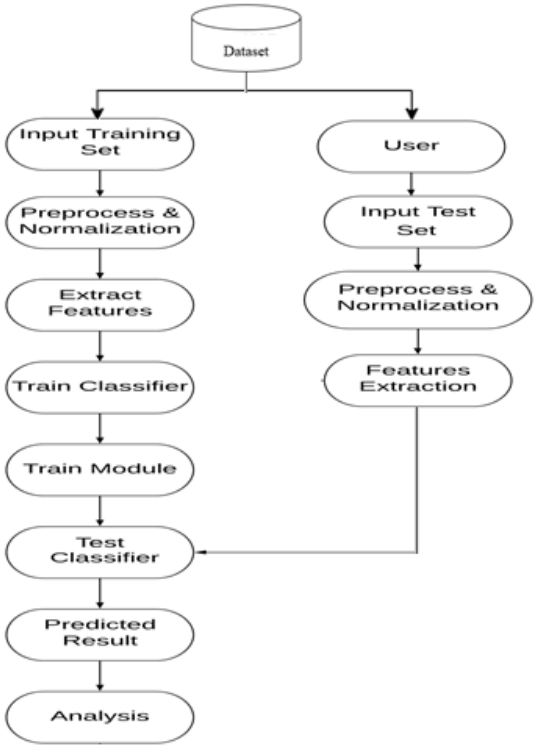

Figure 1: System Architecture

Managing Class Imbalance (SMOTE)

Fraud detection faces multiple problems, with class imbalance being one of the most significant. In this research, the balancing of datasets through SMOTE (Synthetic Minority Oversampling Technique) is implemented. Bias towards the majority class is reduced, and the model learns, as SMOTE uses the generation of synthetic representation of the minority class (fraudulent transactions).

Feature Selection (Information gain)

Feature selection is used to find important attributes for detecting fraud. The Information Gain method ranks important features. With the chosen significant features, the model minimizes computational complexity and enhances accuracy for better predictions.

Model Training and Testing

For this step, the dataset was split into training and testing sets with an 80:20 ratio. Training data is the subset used to teach the model how to make predictions. Test data is the subset that has never been seen or manipulated by the model. Their purpose is to evaluate how well the model generalizes to new data. Techniques like cross-validation are also applied. The model utilizes the strengths of SVM and RF. An ensemble learning method combines several decision trees to improve accuracy and robustness. SVM is a strong classifier that separates the data with the best hyperplane and operates especially well with high dimensional data.

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, June 2026
Impact Factor: 8.2

Model Analysis

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

- **Accuracy:** The percentage of transactions that were accurately classified**.**
- **Precision:** the number of frauds that were accurately identified out of the total number of anticipated frauds**.**
- **Recall:** the ability to correctly identify real fraudulent transactions**.**
- **F1-score:** the harmonic mean of memory and precision.

Algorithm

Hybrid SVM–ANN Fraud Detection Model

Input: Credit card transaction dataset D

Output: Fraud prediction results with evaluation metrics (Accuracy, Precision, Recall, F1 Score)

Step 1: Load Dataset

Load dataset D from CSV file

Convert D into WEKA Instances format

Step 2: Set Class Attribute

Set class attribute as last attribute in D

Step 3: Convert Class Type

IF class attribute is numeric THEN

Convert class attribute to nominal

ENDIF

Step 4: Randomize and Split Dataset

Randomize dataset D

Split D into:

D_train (70%)

D_test (30%)

Step 5: Train SVM Model

Initialize SVM classifier

Train SVM using D_train

Step 6: Extract SVM Probabilities

FOR each instance x in D DO

Compute probability distribution using SVM

Extract fraud probability (class = 1)

Add new attribute svm_prob_class1 to x

ENDFOR

Step 7: Create Extended Dataset

Form extended dataset D_ext with new feature

Step 8: Train RF Model

Initialize RF classifier

Train RF using D_ext_train

Step 9: Testing Phase

Apply RF model on D_ext_test

Predict class label

Step 10: Performance Evaluation

Calculate:

Accuracy

Precision

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, June 2026
Impact Factor: 8.2

Recall
F1 Score

Step 11: Output Results
Display evaluation metrics and predictions
END

IV. RESULTS

For the implementation of the proposed fraud detection system, Java accompanied by the WEKA machine learning framework was used. Java was leveraged in building the application, in the implementation of the hybrid model, and in the integration of several preprocessing and classification techniques. About the construction and assessment of the models for fraud detection, the WEKA library, which offers multifarious machine learning algorithms and data preprocessing methods, was utilized.

Table 1: Model Performance Comparison

Model	Accuracy	Precision	Recall	F1 Score
Random Forest	96.36	100	75.51	86.05
SVM	97.55	97.2	100	98.58

Table 1 comparison shows that all three models achieved high accuracy in detecting fraudulent transactions. Random Forest achieved an accuracy of 96.36% with perfect precision (100%), but its recall was lower at 75.51%, meaning some fraud cases were not detected. The Support Vector Machine model performed better with 97.55% accuracy and achieved a recall of 100%, successfully identifying all fraudulent transactions.

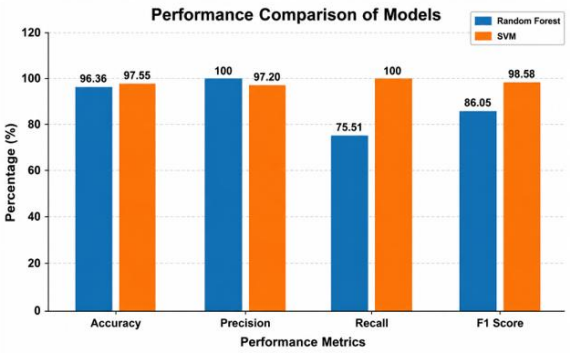

Figure 2: Performance Comparison of Random Forest and SVM

Figure 2 provides information about Random Forest (RF) and Support Vector Machine (SVM) models in terms of evaluating them based on Accuracy, Precision, Recall, and F1 Score. SVM performs better compared to Random Forest in figure 2. SVM has the highest accuracy of 97.55%. A good SVM model will attain a high recall and thus, a 100% recall SVM will detect all cases of fraud. This is the ideal model for fraud detection as it is very costly for fraud cases to go undetected. Random Forest achieves 100% Precision or in other words, very few false positives. However, its recall is only 75%, meaning it detects 75 out of 100 cases and SVM detects the other 25. Hence, Random Forest will be of limited value in everyday fraud detection. SVM (98.58%) achieves the highest F1 Score compared to Random Forest (86.05%). F1 score measures the balance between precision and recall and thus the model which provides a good trade-off between minimization of fraud and errors will have the highest F1 Score. Based on the chart, SVM is the best and most efficient model to use to detect credit card fraud. SVM is better in cases of class imbalance and fraud detection is best to be identified in fraud cases.

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, June 2026
Impact Factor: 8.2

V. CONCLUSION

This study created a framework to detect credit card fraud using machine learning. Three of the major roadblocks of this system were data imbalance, high dimensionality, and the adaptivity of fraud. Three methods, Principal Component Analysis (PCA), the Synthetic Minority Oversampling Technique (SMOTE), and Information Gain, were used for feature selection in this study. These preprocessing steps also significantly improved the quality of the dataset as well as the learning capability of the models. The purpose of this study was to evaluate the performance of two learning machine methods, Random Forest (RF) and Support Vector Machine (SVM). This study used the measures of the standard metrics of Accuracy, Precision, Recall, and F1 Score to perform the evaluation. Experimental results proved that while Random Forest assessed perfect Precision (100%), it assessed a lower Recall (75.51%) and thus that some fraudulent transactions were not detected. Comparatively, the SVM method assessed an improvement over Random Forest in all the standard metric measures. SVM assessed an Accuracy of (97.55%), a perfect Recall of (100%), and the highest Score of all methods, (98.58%). The results of this study also proved that SVM is a more reliable model for credit card fraud detection. In the real world, most companies need to be able to detect all fraudulent activities, thus SVM is well-suited for financial applications. To sum it up, this framework shows an impressive improvement to fraud detection. This is largely due to its advanced preprocessing methods and quick handling of machine learning algorithms. Systems like these are likely to help reduce financial losses, make transactions more secure, and help build a greater level of trust among customers in modern online payment methods. To help increase the reliability and flexibility of this system, additional work should involve the use of deep learning models along with real time detection features.

REFERENCES

[1]. Alatawi, Mohammed Naif. "Detection of fraud in IoT based credit card collected dataset using machine learning." Machine Learning with Applications 19 (2025): 100603.
[2]. Alrasheedi, Masad A. "Enhancing Fraud Detection in Credit Card Transactions: A Comparative Study of Machine Learning Models." Computational Economics (2025): 1-27.
[3]. 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.
[4]. 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.
[5]. Wang, Yuhan. "A data balancing and ensemble learning approach for credit card fraud detection." 2025 4th International Symposium on Computer Applications and Information Technology (ISCAIT). IEEE, 2025.
[6]. Unogwu, Omega John, and Youssef Filali. "Fraud detection and identification in credit card based on machine learning techniques." Wasit Journal of Computer and Mathematics Science 2.3 (2023): 16-22.
[7]. 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.
[8]. Khalid, Abdul Rehman, et al. "Enhancing credit card fraud detection: an ensemble machine learning approach." Big Data and Cognitive Computing 8.1 (2024): 6.
[9]. Alarfaj, Fawaz Khaled, and Shabnam Shahzadi. "Enhancing Fraud detection in banking with deep learning: Graph neural networks and autoencoders for real-time credit card fraud prevention." IEEE Access 13 (2024): 20633-20646.
[10]. Qayoom, Abdul, et al. "A novel approach for credit card fraud transaction detection using deep reinforcement learning scheme." PeerJ Computer Science 10 (2024): e1998..