VOL. 13 NO. 7,
AUGUST, 2026
E-ISSN 3026-9679
P-ISSN 3026-9423

African Scholar Publications & Research International
www.africanscholarpub.com

Journal of Business Dev. and Management Res. (JBDMR)

Effectiveness of the Random Forest Classifier Algorithm in Credit Card Fraud Detection in Financial Institutions

Bukola Olaosebikan Ogunribido[1]; Tolulope Gbenga Olaoluwa[2]; & David Akinwumi[3]

[1]Department of Business Administration, Faculty of Administration and Management, Adekunle Ajasin University, Akungba Akoko, Ondo State, Nigeria. [2]ALAT Engineering, Wema Bank Plc, Lagos, Nigeria. [3]Department of Computer Science, Faculty of Computer Science, Adekunle Ajasin University, Akungba Akoko, Ondo State, Nigeria.
Corresponding Author: bukola.ogunribido@aaua.edu.ng
DOI: https://doi.org/10.70382/ajbdmr.v13i7.0119

Abstract

The increase in the number of credit card users and transactions has caused fraudulent activities to increase significantly. As a result, it has become crucial for financial institutions to develop robust and accurate fraud detection systems to minimise financial losses. One of the popular ways of achieving this is through machine learning. The objective of this research is to develop a model that can detect whether a credit card used for a transaction has been compromised. In this study, a detection system was developed using the random forest classifier. The classifier was used to analyse one million, two hundred and ninety-six thousand, six hundred and seventy-five (1,296,675) synthetic credit card transactions generated with the Sparkov Data Generator and obtained from Kaggle. The dataset contained one million, two hundred and eighty-nine thousand, one hundred and sixty-nine (1,289,169) non-fraudulent transactions and seven thousand five hundred and six (7,506) fraudulent transactions. The data was split with 70% used for training and 30% for testing before being passed into the random forest classifier. The training data was used to fit the model while the testing data was used to evaluate it, and a confusion matrix was plotted. A feature importance plot showed the factors with the highest importance in the dataset. The performance of the algorithm was analysed using accuracy, precision, recall, and F1-score, with fraudulent transactions treated as the positive class. The software used was the Python programming language and the scikit-learn module. The model achieved an overall accuracy of 99.76%, a precision of 91.40%, a recall of 64.20%, and an F1-score of 75.43%, correctly identifying 64.20% of fraudulent transactions and 99.96% of non-fraudulent transactions. The results demonstrate that the random forest classifier is highly precise but that recall remains

constrained by severe class imbalance, indicating that resampling or cost-sensitive learning is required before deployment. The model developed contributes to the management of credit card fraud in financial institutions.

Keywords: Credit Card Fraud, Financial Institutions, Random Forest, Class Imbalance, Machine Learning, Transactions.

Introduction

Digital banking has eased financial transactions but has also contributed to a significant rise in fraudulent activities (Hussain et al., 2024). Credit card companies must identify fraudulent transactions to prevent customers from being charged for unauthorised purchases. Financial fraud poses a great threat to governments, business organisations, and financial institutions, with far-reaching effects on productivity. Credit card transactions have risen as a result of heavy reliance on internet technology in recent times, but credit card fraud has also increased as a result of both online and offline transactions.

In the United Kingdom, losses from unauthorised financial fraud involving credit cards and remote banking amounted to £844.8 million in 2018, while banks and card issuers prevented £1.66 billion in unauthorised fraud in the same year, equivalent to preventing £2 of every £3 of attempted fraud. In Nigeria, the third quarter industry fraud report released by the Nigeria Inter-Bank Settlement System (NIBSS) revealed that fraudsters attempted 46,126 attacks in 2020 and were successful on 41,979 occasions, a success rate of approximately 91 per cent. Institutions suffered losses of up to ₦3.36 billion between July and September 2020, representing a substantial increase relative to the same period in 2019. Balogun et al. (2024) similarly report that recorded instances of credit card theft in Nigeria rose from over 5,000 in 2018 to over 10,000 in 2019, based on Central Bank of Nigeria figures. Credit card fraud therefore remains one of the most significant threats to financial institutions and businesses today.

The losses caused by credit card fraud harm customers, retailers, and issuing banks alike; for this reason, developing tools for identifying and combating credit card fraud is critical. Fraud detection involves tracking the behaviour of large populations of users in order to predict and avert unacceptable behaviour such as fraud, intrusion, or default. As credit card transactions grow, computational approaches for dealing with credit card fraud have attracted considerable attention. Nevertheless, the identification of credit card fraud remains a common and challenging problem that has yet to be adequately addressed.

Prior studies have established that machine learning is an effective approach to fraud detection (Varmedja et al., 2019; Ileberi et al., 2021; Balogun et al., 2024). However, limitations remain in the reported methods. Many rely on the widely used European cardholder dataset, on which principal component analysis has already been performed, obscuring the interpretability of individual features. Balogun et al. (2024) explicitly note this constraint, observing that the anonymised variables V1 to V28 exhibited low correlation with the fraud label in their own analysis. Others report headline accuracy figures on severely imbalanced data without disclosing recall on the minority fraud class, which is the metric of practical interest. This study addresses these gaps by applying a random forest classifier to a large, non-PCA-transformed transaction dataset in which the original features remain interpretable, and by reporting class-specific performance metrics with the fraudulent class explicitly treated as the positive class.

The dataset employed for this purpose is a large corpus of synthetic transactions produced by the Sparkov Data Generator, which simulates cardholder spending from realistic demographic, merchant, and geographic profiles and reproduces the behavioural and temporal patterns of genuine banking activity, including the extreme rarity of fraud. Synthetic data of this kind serves as a valid proxy for real-world transactions because it preserves these statistical properties while remaining free of the privacy and regulatory constraints that compel the anonymisation, and consequent loss of feature interpretability, of authentic cardholder datasets.

Aim and Objectives

The aim of this study is to evaluate the effectiveness of the random forest classifier algorithm in detecting fraudulent credit card transactions in financial institutions. The specific objectives are to: (i) develop a random forest-based fraud detection model on a large-scale, interpretable transaction dataset; (ii) evaluate the model using accuracy, precision, recall, and F1-score, with fraud as the positive class; (iii) identify the transaction features that contribute most to fraud detection; and (iv) compare the performance of the model with results reported in the existing literature.

Literature Review

Overview of Credit Card Fraud

Credit card fraud is a form of financial crime involving the unauthorised use of another person's credit card to make purchases or withdraw funds. It may be committed online, over the telephone, or in person, and can cause substantial financial loss as well as damage to the victim's credit standing and reputation (Balogun et al., 2024). Unauthorised credit card fraud occurs when the account holder does not permit the payment to proceed and a third party completes the transaction, whereas authorised credit card fraud occurs when a legitimate customer is deceived into processing a payment to an account controlled by a criminal.

Among the most common mechanisms of credit card fraud is skimming, in which a device is attached to a card reader such as an automated teller machine or fuel pump in order to capture the information encoded on the magnetic strip of a card. The captured information can then be used to make unauthorised transactions, produce counterfeit cards, or be sold illicitly. A second common mechanism is phishing, in which individuals are induced to disclose their card details through fraudulent emails or websites that imitate legitimate sources (Balogun et al., 2024). Fraud may accordingly take the form of physical card fraud or virtual card fraud, the latter involving the theft and reuse of the card number, expiry date, and card verification value entered during an online purchase.

Credit Card Fraud Detection

Fraud detection refers to the methods and analytics that enable firms to detect and prevent unlawful financial activity, including fraudulent credit card transactions, identity theft, cyber intrusion, and insurance fraud. Credit card fraud detection is specifically the process of identifying fraudulent purchase attempts and rejecting them rather than completing the sale. Payment cards are simple to use, requiring only a few digits to be transmitted to the issuing bank to identify the account and authorise the transaction. This simplicity is also the source of their vulnerability, since strict data security is difficult to enforce on a small set of numbers that must be shared with every counterparty.

Traditional manual methods of detecting fraudulent transactions are time-consuming and inefficient, and developments in big data have rendered such methods impracticable at scale; consequently, computational methods have become the principal focus of financial institutions seeking to address

credit card fraud (Balogun et al., 2024). Contemporary solutions employ machine learning to manage data analysis, predictive modelling, decision-making, fraud alerts, and remediation. Such systems benefit multiple stakeholders: financial institutions reduce losses and improve customer satisfaction, cardholders benefit from increased security, and merchants avoid chargeback losses while maintaining their reputation.

A recurring methodological difficulty in this domain is class imbalance. Fraudulent transactions typically constitute well under one per cent of all records. Balogun et al. (2024) observe that this imbalance poses a direct challenge to machine learning algorithms, which may tend to classify most transactions as normal and thereby fail to detect the fraudulent ones. Accuracy alone is consequently an inadequate measure of performance, and precision, recall, and the F1-score computed with respect to the fraudulent class are the appropriate evaluation criteria.

Empirical Review

Varmedja, Karanovic, Sladojevic, Arsenovic and Anderla (2019) compared several machine learning algorithms for credit card fraud detection using the European cardholder dataset. Because the dataset was highly imbalanced, the Synthetic Minority Over-sampling Technique (SMOTE) was applied, followed by feature selection and a train–test split. The algorithms evaluated were logistic regression, random forest, naive Bayes, and the multilayer perceptron. Random forest achieved the strongest performance, with an accuracy of 99.96%, a precision of 96.38%, and a recall of 81.63%. The study demonstrated the value of oversampling for imbalanced fraud data, although its reliance on a PCA-transformed dataset limited any interpretation of which underlying transaction attributes drove the predictions.

Taha and Malebary (2020) proposed an intelligent approach to credit card fraud detection using an optimised light gradient boosting machine (OLightGBM), in which a Bayesian-based hyperparameter optimisation algorithm was integrated to tune the parameters of a LightGBM classifier. Experiments were performed on two real-world public credit card transaction datasets. On the first dataset the approach achieved an accuracy of 98.40%, an area under the curve of 90.94%, a precision of 97.34%, a recall of 40.59%, and an F1-score of 56.95%. The study established the importance of systematic hyperparameter optimisation, a step frequently omitted in comparable work. It is notable that despite the high accuracy and precision reported, recall on the fraudulent class remained below 41%, illustrating the difficulty of achieving balanced performance on severely skewed data.

Makolo and Adeboye (2021), working within the Nigerian context at the University of Ibadan, proposed a Hybrid Credit Card Fraud Detection (HCCFD) system for financial institutions. Rather than applying a conventional supervised classifier, the authors adopted an anomaly detection strategy that combined a genetic algorithm with a multivariate normal distribution to identify fraudulent transactions on an imbalanced dataset. The hybrid model achieved an accuracy of 93.3% and an F-score of 93.5%, outperforming an artificial neural network, a decision tree, and a support vector machine trained on the same data, which recorded F-scores of 84.2%, 80.0%, and 68.5% respectively. The study is significant for demonstrating that anomaly-based formulations offer a viable alternative to purely supervised approaches when fraudulent examples are scarce, and for situating the fraud detection problem within the operational realities of Nigerian financial institutions. Its principal limitation is that the reported gains were established against a comparatively narrow set of baseline classifiers.

Ileberi, Sun and Wang (2021) implemented a machine learning framework for credit card fraud detection using a real-world imbalanced dataset generated from European cardholders, in which 0.172% of transactions were fraudulent. The class imbalance was addressed by resampling with SMOTE, and

the framework was evaluated using support vector machine, logistic regression, random forest, extreme gradient boosting, decision tree, and extra tree classifiers, each additionally paired with the AdaBoost algorithm. In the first phase, random forest performed optimally in terms of classification quality with a Matthews correlation coefficient of 0.88, while random forest and extra tree achieved the highest accuracy at 99.5%. Pairing the classifiers with AdaBoost produced further improvements, with the decision tree registering a Matthews correlation coefficient increase of 0.20 and extreme gradient boosting an increase of 0.28. The study confirmed that ensemble methods combined with resampling outperform single classifiers, though the reliance on the anonymised European dataset again precluded interpretation of the underlying transaction features.

Alarfaj, Malik, Khan, Almusallam, Ramzan and Ahmed (2022) conducted a comprehensive evaluation of state-of-the-art machine learning and deep learning algorithms for credit card fraud detection. The authors compared conventional classifiers, including decision tree, random forest, logistic regression, and extreme gradient boosting, against deep learning architectures based on convolutional neural networks with varying layer configurations. A convolutional neural network with twenty layers achieved the strongest results, obtaining an accuracy of 99.9%, an F1-score of 85.71%, a precision of 93%, and an area under the curve of 98%. The authors additionally performed experiments with balanced data in order to minimise the false negative rate. The study provided a valuable consolidated comparison across model families, although the computational cost and reduced interpretability of the deep architectures remain practical constraints for deployment in financial institutions.

Esenogho, Mienye, Swart, Aruleba and Obaido (2022) addressed the imbalance problem through a combination of feature engineering and ensemble learning. Their approach employed a long short-term memory neural network as the base learner within the adaptive boosting (AdaBoost) technique, while the hybrid synthetic minority oversampling technique and edited nearest neighbour (SMOTE-ENN) method was used to resample the training data. The proposed method was benchmarked against a support vector machine, a multilayer perceptron, a decision tree, conventional AdaBoost, and a standalone long short-term memory network. The authors reported that all classifiers performed better when trained on resampled data, and that the proposed ensemble obtained the highest sensitivity and specificity at 0.996 and 0.998 respectively. The study reinforces the finding that hybrid resampling combined with boosting materially improves detection of the minority class, though the sequential architecture entails greater training cost than tree-based alternatives.

Mienye and Sun (2023) proposed a deep learning ensemble incorporating data resampling for credit card fraud detection. Their approach combined long short-term memory and gated recurrent unit neural networks as base learners within a stacking ensemble framework, with a multilayer perceptron serving as the meta-learner, while the hybrid synthetic minority oversampling technique and edited nearest neighbour (SMOTE-ENN) method was employed to balance the class distribution. The experimental results showed that combining the deep learning ensemble with SMOTE-ENN achieved a sensitivity of 1.000 and a specificity of 0.997, superior to other widely used classifiers reported in the literature. The study demonstrates that combining architectural diversity with explicit imbalance handling substantially improves minority class detection, although the resulting model is considerably more complex to train and maintain than a single tree-based classifier.

Balogun, Kupolusi and Akomolafe (2024) examined the application of machine learning algorithms to credit card fraud detection using the European cardholder dataset of 284,807 transactions, of which 492 were fraudulent. Random forest, logistic regression, k-nearest neighbours, decision trees, and naive Bayes were implemented in the R programming language, with SMOTE applied to address the class imbalance and recursive feature elimination used for feature selection. Random forest on the original

data achieved the strongest overall performance, with an accuracy of 0.9996, a precision of 0.9487, a recall of 0.7959, an F1-score of 0.8658, and an area under the receiver operating characteristic curve of 0.9796. On the resampled data, random forest attained a recall of 0.8469 with a reduced precision of 0.8529, illustrating the precision–recall trade-off introduced by oversampling. The authors concluded that transaction amount was the most important variable for detecting fraud, and observed that the anonymised variables V1 to V28 showed low correlation with fraudulent transactions, indicating that these may be of limited value for detection.

Oluwatoyin and Akinola (2024) developed a real-time credit card fraud detection and reporting system intended to identify fraudulent transactions promptly and alert users. The study employed random forest and decision tree classifiers together with several sampling techniques, generating multiple model configurations for comparison. The random forest model combined with SMOTE and hyperparameter tuning proved most effective, achieving an accuracy of 97%, an F1-score of 95%, and a precision of 98%. The model was integrated with a messaging service to deliver real-time fraud alerts. Notably, the study employed the same Sparkov-generated transaction dataset used in the present research, and the authors excluded the geographic and population features on the grounds that they contributed little predictive value. The work is significant for prioritising real-time alerting rather than offline classification alone.

Hussain, Dhanawat, Aslam, Iqbal and Tripura (2024) addressed the imbalanced data problem directly by systematically comparing over-sampling and under-sampling methods for credit card fraud detection. Their results showed that the choice of resampling strategy materially affects recall on the fraudulent class, and that the selection between over-sampling and under-sampling involves a measurable trade-off between recall and precision. The study is directly relevant to the present work, as it quantifies the performance consequences of leaving severe class imbalance unaddressed.

Taken together, these ten studies establish that machine learning is effective for credit card fraud detection, and that ensemble methods, whether tree-based or built upon sequential neural architectures, consistently outperform single classifiers. A clear consensus also emerges regarding class imbalance: every study that applied a resampling technique, whether SMOTE (Varmedja et al., 2019; Ileberi et al., 2021; Oluwatoyin & Akinola, 2024) or the hybrid SMOTE-ENN method (Esenogho et al., 2022; Mienye & Sun, 2023), reported materially higher recall on the fraudulent class than comparable models trained without it.

Two limitations nonetheless recur across the literature. First, the majority of studies rely on the European cardholder dataset, whose features have been transformed by principal component analysis and are therefore uninterpretable; Balogun et al. (2024) note directly that these anonymised variables showed low correlation with the fraud label. Second, several studies report headline accuracy figures on severely imbalanced data without disclosing recall on the fraudulent class. The results of Taha and Malebary (2020) are instructive in this regard: an accuracy of 98.40% and a precision of 97.34% were accompanied by a recall of only 40.59%. It is also notable that studies grounded in the Nigerian context remain comparatively few, with Makolo and Adeboye (2021) and Balogun et al. (2024) among the limited number addressing the problem from that perspective, despite the scale of credit card fraud losses recorded domestically. The present study addresses these gaps by using a large dataset in which the original transaction features are retained and interpretable, by reporting class-specific metrics with fraud explicitly designated as the positive class, and by contributing further evidence from a Nigerian research setting.

Methodology

Dataset

The dataset used in this study consists of synthetic credit card transactions generated using the Sparkov Data Generator and obtained from Kaggle. Unlike the widely used European cardholder dataset, the features in this dataset have not been transformed by principal component analysis and therefore remain directly interpretable. The dataset contains attributes including transaction date and time, credit card number, merchant, category of payment, amount, first name, last name, gender, address, job, date of birth, and transaction identifier. It comprises 1,296,675 transactions in total, of which 7,506 (0.58%) are fraudulent and 1,289,169 (99.42%) are non-fraudulent. The use of synthetic data is a deliberate methodological choice, permitting a realistically large transaction volume with interpretable features while avoiding the privacy constraints that necessitate anonymisation in real cardholder datasets. The same dataset was employed by Oluwatoyin and Akinola (2024), which permits a degree of direct comparison between the two studies.

Data Preprocessing

The transaction dataset was read into the system using the read_csv function of the Pandas library. Label encoding was applied to convert categorical text attributes into numerical form. This approach was preferred over one-hot encoding because several of the nominal attributes are of very high cardinality, the merchant and job fields alone containing several hundred distinct values each, so that one-hot encoding would have expanded the feature space into thousands of sparse binary columns and substantially increased memory consumption and training time without a commensurate benefit for a tree-based learner. Random forests are, moreover, comparatively robust to the arbitrary integer orderings that label encoding introduces, since each tree partitions a feature through repeated binary splits rather than by interpreting the magnitude of the encoded value. Target encoding and frequency encoding, which avoid imposing a spurious ordinal scale, were considered as alternatives; these were not adopted at this baseline stage but represent a natural refinement for subsequent work. It is nonetheless acknowledged that label encoding imposes an artificial ordinal relationship on nominal attributes of high cardinality, such as merchant and job, and this is a recognised limitation of the present preprocessing pipeline that is discussed further among the limitations of the study. The dataset was subsequently partitioned into a training set comprising 70% of records and a testing set comprising the remaining 30%. The training set was used to fit the classifier, while the testing set was reserved for evaluation of predictive performance.

The Random Forest Algorithm

The random forest algorithm may be summarised in the following steps:

a. Randomly select N samples from the training set with replacement (bootstrap sampling).

b. Build a decision tree using the selected samples and a random subset of the available features.

c. Repeat steps (a) and (b) to construct T trees.

d. Aggregate the predictions of the T trees by majority vote to obtain the final prediction.

The classification problem is formalised as follows. Let X denote a d-dimensional input space and let Y denote a set of K classes, as shown in equation (3.1):

$$Y = \{1, 2, ..., K\} \qquad (3.1)$$

A decision tree is a hierarchical structure that partitions X into disjoint regions R_j, for j = 1, 2, ..., J. Each leaf node of the tree is associated with a class label, which is the mode of the training samples falling within the corresponding region. Given a test sample x, the algorithm traverses the tree from the

root to a leaf node according to the feature values of x, and outputs the class label associated with that leaf node. The decision rule is expressed in equation (3.2):

$$C(x) = \text{argmax}_k (n_k / N_j) \quad (3.2)$$

where $C(x)$ is the predicted class label for sample x, N_j is the number of training samples in region R_j, and n_k is the number of training samples of class k within R_j.

The random forest algorithm extends the decision tree by constructing T trees using different bootstrap subsets of the training data and different random subsets of the features. The classification rule for a random forest is given in equation (3.3):

$$C(x) = \text{argmax}_k \sum^{T}_{t=1} I(h_t(x) = k) \quad (3.3)$$

where $h_t(x)$ denotes the prediction of the t-th tree for input x, and $I(\cdot)$ is an indicator function that returns 1 when the condition holds and 0 otherwise. The predicted class is therefore the class receiving the greatest number of votes across the T trees.

Model Configuration

The model was implemented in the Python programming language using the RandomForestClassifier of the scikit-learn library. The ensemble was configured with 98 estimators (n_estimators = 98), and a fixed random seed of 60 (random_state = 60) was set to ensure that the train–test partition and the bootstrap sampling are reproducible. All remaining hyperparameters were left at their scikit-learn default settings: the Gini impurity was used as the splitting criterion, no maximum tree depth was imposed so that nodes were expanded until all leaves were pure or contained fewer than two samples, a minimum of two samples was required to split an internal node, a minimum of one sample was required at each leaf node, and the number of features considered at each split was set to the square root of the total number of features. Bootstrap sampling was enabled, as is the default.

No resampling technique, class weighting, or decision threshold adjustment was applied at this stage. This was a deliberate methodological choice, intended to establish the baseline behaviour of the random forest classifier under severe class imbalance so that the magnitude of the resulting recall penalty could be measured directly. The implications of this decision are examined in the results and among the limitations of the study.

Evaluation Metrics

Model performance was assessed using a confusion matrix, from which true positives (TP), true negatives (TN), false positives (FP), and false negatives (FN) were derived. In accordance with standard practice in fraud detection research, the fraudulent class is designated as the positive class throughout, since correct identification of fraudulent transactions is the objective of the system. Accuracy, precision, recall, and the F1-score were computed as defined in equations (4.1) to (4.4) in the following section.

Results and Discussion

This study developed a credit card fraud detection system using a random forest classifier. Of the 388,994 transactions in the test partition, 2,268 were fraudulent and 386,726 were non-fraudulent. The classifier correctly identified 1,456 of the 2,268 fraudulent transactions and 386,589 of the 386,726 non-fraudulent transactions, producing 137 false positives and 812 false negatives. Treating fraud as the positive class, this yields TP = 1,456, TN = 386,589, FP = 137, and FN = 812.

The performance metrics were calculated from the confusion matrix as follows:

Accuracy = (TP + TN) / (TP + TN + FP + FN) = 388,045 / 388,994 = 0.9976 or 99.76% (4.1)

Precision = TP / (TP + FP) = 1,456 / 1,593 = 0.9140 or 91.40% (4.2)

Recall = TP / (TP + FN) = 1,456 / 2,268 = 0.6420 or 64.20% (4.3)

F1-score = 2 × (Precision × Recall) / (Precision + Recall) = 0.7543 or 75.43% (4.4)

The overall accuracy of 99.76% must be interpreted with caution. Because only 0.58% of transactions in the dataset are fraudulent, a trivial classifier that labelled every transaction as legitimate would achieve an accuracy of approximately 99.42%. The accuracy figure therefore conveys little information about the practical utility of the model, and the class-specific metrics are the meaningful measures of performance. This concern is precisely the one raised by Balogun et al. (2024), who noted that severe imbalance may lead algorithms to classify most transactions as normal while missing the fraudulent ones.

The precision of 91.40% indicates that when the model flags a transaction as fraudulent, it is correct in slightly more than nine cases out of ten. This is operationally valuable, as it implies a low burden of false alerts on fraud investigation teams and minimal disruption to legitimate customers. The false positive rate was 0.035%, meaning that fewer than four in every ten thousand legitimate transactions were incorrectly flagged.

The recall of 64.20%, by contrast, reveals the principal limitation of the model: approximately 36% of fraudulent transactions passed undetected. This outcome is characteristic of tree-based classifiers trained on severely imbalanced data without resampling or cost-sensitive adjustment. Because the majority class dominates the training signal, the classifier adopts a conservative decision boundary that favours precision over recall. The resulting F1-score of 75.43% reflects this asymmetry between the two measures.

This pattern is consistent with results reported elsewhere in the literature. Taha and Malebary (2020) obtained an accuracy of 98.40% and a precision of 97.34% alongside a recall of only 40.59%, a still more pronounced instance of the same asymmetry. Balogun et al. (2024) reported that their random forest model achieved a recall of 0.7959 on the original imbalanced data, rising to 0.8469 after SMOTE resampling while precision declined from 0.9487 to 0.8529. Varmedja et al. (2019) and Ileberi et al. (2021), both of whom applied SMOTE before training, likewise reported higher recall than is obtained here. The comparison indicates that the recall deficit observed in the present study is attributable to the absence of imbalance handling rather than to any inherent limitation of the random forest algorithm, and that the precision–recall trade-off documented by Hussain et al. (2024) is the mechanism at work.

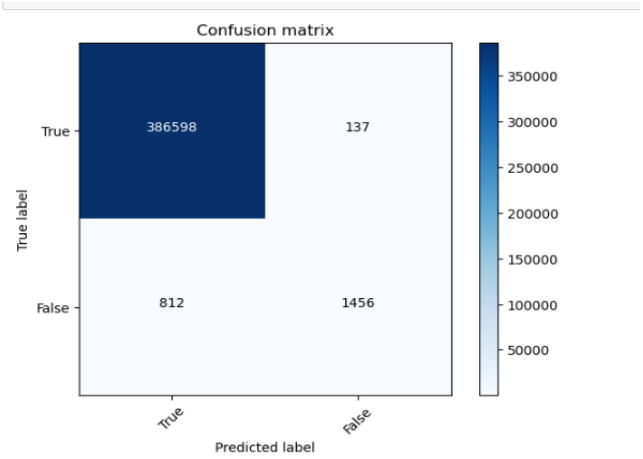


Figure 1: Confusion Matrix for the Random Forest Classifier

Source: Authors' computation (2026)

```
feat_importances = pd.Series(clf_B.feature_importances_, index=data.columns)
feat_importances.nlargest(8).plot(kind='barh')
```

```
<Axes: >
```

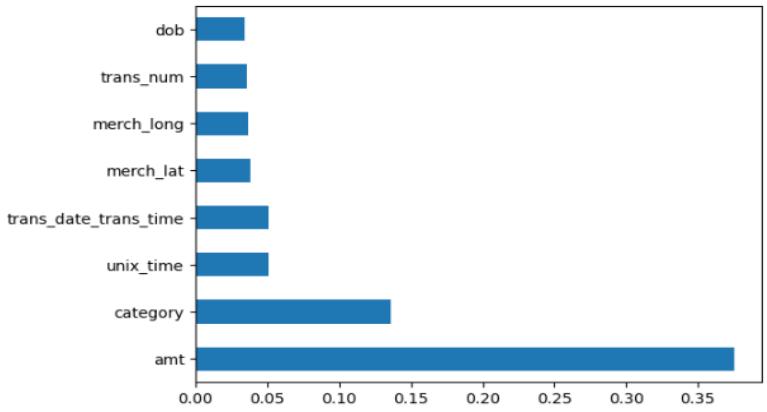


Figure 2: Feature Importance Plot for the Random Forest Classifier
Source: Authors' computation (2026).

Figure 2 presents the feature importance plot for the random forest classifier. Transaction amount emerged as the most important feature for fraud detection, followed by transaction category, unix_time, and trans_date_trans_time. This finding corresponds closely with that of Balogun et al. (2024), who likewise identified transaction amount as the most important variable for detecting credit card fraud. The prominence of category is intuitive, as fraudulent transactions frequently cluster within particular merchant categories. The importance of the two temporal features indicates that the timing of transactions carries substantial discriminative signal. Because the dataset retains its original, untransformed features, these importances are directly interpretable and can inform rule-based screening alongside the model, an advantage not available in studies based on PCA-transformed data.

Comparison with Existing Studies

Table 1 presents the performance of the present model alongside results reported in the reviewed literature. All comparison figures are drawn from the respective published studies and are reproduced as reported by their authors; a dash indicates that the metric in question was not reported in the source study. It should be noted that these studies employ different datasets, preprocessing pipelines, and resampling strategies, and the comparison is therefore indicative rather than a controlled benchmark.

Table 1: Comparison of Reported Performance with Existing Studies

Author(s) and Year	Best Model	Accuracy	Precision	Recall	F1-Score
Varmedja et al. (2019)	Random Forest + SMOTE	99.96%	96.38%	81.63%	—
Taha & Malebary (2020)	OLightGBM	98.40%	97.34%	40.59%	56.95%
Makolo & Adeboye (2021)	HCCFD (genetic algorithm)	93.30%	—	—	93.50%
Ileberi et al. (2021)	Random Forest + SMOTE	99.50%	—	—	—
Alarfaj et al. (2022)	CNN (20 layers)	99.90%	93.00%	—	85.71%

Author(s) and Year	Best Model	Accuracy	Precision	Recall	F1-Score
Esenogho et al. (2022)	LSTM–AdaBoost + SMOTE-ENN	—	—	99.60%	—
Mienye & Sun (2023)	LSTM–GRU ensemble	—	—	100.00%	—
Balogun et al. (2024)	Random Forest (original)	99.96%	94.87%	79.59%	86.58%
Oluwatoyin & Akinola (2024)	Random Forest + SMOTE (tuned)	97.00%	98.00%	—	95.00%
Current Study (2026)	Random Forest (no resampling)	99.76%	91.40%	64.20%	75.43%

Source: Authors' compilation from the cited studies (2026).

Figures for Taha and Malebary (2020) relate to the first of the two datasets they evaluated. Figures for Balogun et al. (2024) relate to their random forest model on the original, non-resampled data. Esenogho et al. (2022) and Mienye and Sun (2023) report sensitivity and specificity rather than accuracy, precision, and F1-score; their sensitivity values of 0.996 and 1.000 respectively are shown in the recall column.

The comparison shows that the accuracy achieved in the present study is broadly consistent with the literature, and that its precision is competitive with several published models. Its recall, however, is lower than that reported by most studies applying resampling techniques, though it exceeds the 40.59% recall reported by Taha and Malebary (2020) for an optimised gradient boosting model. The models reporting the highest recall in Table 1 are precisely those that applied SMOTE, SMOTE-ENN, or an equivalent resampling method prior to training. The present study, which deliberately omitted resampling in order to establish a baseline, illustrates the magnitude of the recall penalty that severe class imbalance imposes when left unaddressed.

Summary, Conclusion and Recommendations

Summary

This study examined credit card fraud detection using a random forest classifier, a machine learning algorithm well suited to classification tasks of this nature. The work was motivated by the need for fraud detection methods capable of keeping pace with evolving fraud tactics as the volume of online transactions increases. The study reviewed existing literature on credit card fraud detection using machine learning algorithms, identifying two recurring limitations: reliance on PCA-transformed datasets that preclude feature interpretation, and the reporting of aggregate accuracy figures that obscure poor recall on the minority fraudulent class.

A random forest classifier was implemented on a dataset of 1,296,675 transactions with interpretable features. The data was preprocessed, partitioned into training and testing sets, and evaluated using accuracy, precision, recall, and F1-score with fraud designated as the positive class. The model achieved an accuracy of 99.76%, a precision of 91.40%, a recall of 64.20%, and an F1-score of 75.43%. Transaction amount, category, and temporal attributes were identified as the most important predictive features.

Conclusion

The study demonstrates that the random forest classifier achieves high precision in identifying fraudulent credit card transactions, correctly classifying more than nine in ten flagged transactions while

maintaining a false positive rate below 0.04%. This makes the model operationally attractive from the perspective of investigative workload and customer experience. However, the recall of 64.20% indicates that approximately one third of fraudulent transactions escape detection when the classifier is trained on severely imbalanced data without corrective measures. In operational terms, a recall of 64.20% means that 35.80% of fraudulent transactions are not intercepted, and every undetected case exposes the institution to direct financial liability under card-scheme chargeback rules, under which the disputed amount is ultimately borne by the issuer or the merchant. This reinforces why recall, rather than headline accuracy, is the metric of primary operational concern. The model in its present configuration is therefore best understood as a high-precision screening component rather than a complete fraud detection solution, and should be deployed in conjunction with complementary controls.

A broader conclusion follows from this result. Accuracy alone is an inadequate and potentially misleading measure of performance in fraud detection, since the extreme class imbalance inherent to the problem allows very high accuracy to coexist with poor detection of the cases of interest. Financial institutions evaluating such systems should require the disclosure of class-specific recall and precision rather than aggregate accuracy figures.

Contribution to Knowledge

This study contributes to the existing body of knowledge in three respects. First, it evaluates random forest performance on a large, non-PCA-transformed transaction dataset, permitting direct interpretation of which transaction attributes drive fraud predictions, an analysis that is not possible with the anonymised datasets predominantly used in the literature. Second, it reports a transparent baseline for random forest performance under severe class imbalance without resampling, quantifying the recall penalty that such imbalance imposes and providing a reference point against which the benefit of resampling techniques can be measured. Third, it provides practical guidance to financial institutions on the appropriate evaluation criteria for fraud detection systems.

Limitations of the Study

Several limitations should be acknowledged. The dataset employed is synthetic and generated by the Sparkov Data Generator; while it provides realistic transaction volumes and interpretable features, findings may not transfer directly to live transaction data. No resampling technique or class weighting was applied, which accounts for the constrained recall reported above. Label encoding was applied to nominal attributes of high cardinality, imposing an artificial ordinal relationship that may affect the model's treatment of those features. Evaluation was based on a single train–test partition rather than cross-validation, and threshold-independent metrics such as the area under the precision–recall curve were not computed. Finally, the comparison presented in Table 1 draws on studies using different datasets, preprocessing pipelines, and evaluation protocols, and is therefore indicative rather than a controlled benchmark.

Recommendations for Future Research

Future research should address the limitations identified above. The most immediate priority is the incorporation of class imbalance handling, whether through SMOTE, adaptive synthetic sampling, class weighting, or cost-sensitive learning, together with a systematic evaluation of the resulting trade-off between precision and recall; the results of Balogun et al. (2024), in which recall rose from 0.7959 to 0.8469 while precision fell from 0.9487 to 0.8529, indicate the shape this trade-off is likely to take. Decision threshold tuning should be undertaken in conjunction with this, since the operational cost of a

missed fraudulent transaction typically exceeds that of a false alert. Hyperparameter optimisation, following the Bayesian approach demonstrated by Taha and Malebary (2020), may yield further gains. Subsequent studies should also adopt stratified k-fold cross-validation and report the area under the precision–recall curve, which is more informative than accuracy for imbalanced classification. Validation on real rather than synthetic transaction data, ideally in partnership with a financial institution, would strengthen the external validity of the findings. Finally, the applicability of these methods to other categories of financial transaction, including wire transfers, bank transfers, and mobile payments, merits investigation.

References

Alarfaj, F. K., Malik, I., Khan, H. U., Almusallam, N., Ramzan, M., & Ahmed, M. (2022). Credit card fraud detection using state-of-the-art machine learning and deep learning algorithms. *IEEE Access, 10*, 39700–39715. https://doi.org/10.1109/ACCESS.2022.3166891

Balogun, O. O., Kupolusi, J. A., & Akomolafe, A. A. (2024). Credit card fraud detection using machine learning algorithms. *British Journal of Computer, Networking and Information Technology, 7*(3), 1–35. https://doi.org/10.52589/BJCNIT-YDIJNXG2

Esenogho, E., Mienye, I. D., Swart, T. G., Aruleba, K., & Obaido, G. (2022). A neural network ensemble with feature engineering for improved credit card fraud detection. *IEEE Access, 10*, 16400–16407. https://doi.org/10.1109/ACCESS.2022.3148298

Hussain, A., Dhanawat, V., Aslam, A., Iqbal, N., & Tripura, S. (2024). Credit card fraud detection using machine learning techniques: Dealing with imbalanced data using over-sampling and under-sampling methods. *Proceedings of the 2024 Beyond Technology Summit on Informatics International Conference (BTS-I2C)*, East Java, Indonesia, 676–681. https://doi.org/10.1109/BTS-I2C63534.2024.10942037

Ileberi, E., Sun, Y., & Wang, Z. (2021). Performance evaluation of machine learning methods for credit card fraud detection using SMOTE and AdaBoost. *IEEE Access, 9*, 165286–165294. https://doi.org/10.1109/ACCESS.2021.3134330

Makolo, A., & Adeboye, T. (2021). Credit card fraud detection system using machine learning. *International Journal of Information Technology and Computer Science, 13*(4), 24–37. https://doi.org/10.5815/ijitcs.2021.04.03

Mienye, I. D., & Sun, Y. (2023). A deep learning ensemble with data resampling for credit card fraud detection. *IEEE Access, 11*, 30628–30638. https://doi.org/10.1109/ACCESS.2023.3262020

Oluwatoyin, J. A., & Akinola, S. (2024). Real time credit card fraud detection and reporting system using machine learning. *European Journal of Computer Science and Information Technology, 12*(4), 36–56. https://doi.org/10.37745/ejcsit.2013/vol12n43656

Taha, A. A., & Malebary, S. J. (2020). An intelligent approach to credit card fraud detection using an optimized light gradient boosting machine. *IEEE Access, 8*, 25579–25587. https://doi.org/10.1109/ACCESS.2020.2971354

Varmedja, D., Karanovic, M., Sladojevic, S., Arsenovic, M., & Anderla, A. (2019). Credit card fraud detection – Machine learning methods. *Proceedings of the 18th International Symposium INFOTEH-JAHORINA (INFOTEH)*, East Sarajevo, Bosnia and Herzegovina, 1–5. https://doi.org/10.1109/INFOTEH.2019.8717766