LeadSci Journal of Management, Innovation and Social Sciences (JMISS)

RESEARCH ARTICLE ISSN(e): 3105-4935, ISSN(p): 3136-2040 OPEN ACCESS

Enhancing Credit Card Fraud Detection through Data Preprocessing and SMOTE-Based Class Balancing: A Comparative Evaluation of Machine Learning Models

Nafiu Yahuza[1]*, Ahmad Baita Garko[2], Abubakar Atiku Muslim[3], Abdullahi Usman Gulumbe[1]

[1]Department of Computer Science, Federal University Birnin Kebbi, Kebbi State, Nigeria
[2]Department of Computer Science, Federal University Dutse (FUD), Jigawa State, Nigeria
[3]Department of Electric/Electronics Engineering Abdullahi Fodio University of Science and Technology Aliero, Kebbi State Nigeria

Abstract

Credit card fraud remains a major challenge for financial institutions, both financially and operationally, as digital transactions continue to grow and fraud datasets remain highly imbalanced. This study compares the performance of several supervised machine learning models for fraud detection, using a unified data preprocessing pipeline. The approach includes removing duplicates, applying RobustScaler normalization, engineering features and using the Synthetic Minority Oversampling Technique (SMOTE) to balance classes before training. Four models were developed and tested Logistic Regression, Decision Tree, Random Forest and Artificial Neural Network (ANN) using the publicly available Kaggle Credit Card Fraud Detection dataset. Their performance was measured with Accuracy, Precision, Recall, F1-score and ROC-AUC metrics. Results showed that thorough preprocessing combined with SMOTE significantly improved the models ability to detect fraudulent transactions. Among them, the Random Forest model delivered the strongest overall performance, proving especially effective at handling highly imbalanced financial data. The comparative analysis also highlighted that ensemble learning methods generally outperform single classifiers in both accuracy and minority-class recognition. These findings emphasize the importance of pairing robust preprocessing strategies with machine learning techniques to boost fraud detection in real-world financial systems. The proposed system offers institutions a scalable and practical solution for building intelligent fraud detection systems, while laying the groundwork for future integration of Explainable AI (XAI) and real-time detection tools.

Keywords: *Credit Card Fraud Detection, Machine Learning, Random Forest, SMOTE, Data Preprocessing*

Introduction

The rapid expansion of electronic payment systems and online transactions has heightened the risk of financial fraud, particularly credit card fraud. With millions of transactions processed daily, manual detection is no longer feasible, making intelligent automated systems essential. Machine learning has proven highly effective in this area because it can uncover complex behavioral patterns from historical financial data (Abdallah et al., 2016; Hernandez Aros et al., 2024).

*Corresponding Author: yahuza.nafiu@fubk.edu.ng
Received: 14 July 2026; Received in revised form 17 August 2026; Accepted: 27 August 2026;
Available Online: 31 August, 2026

A major challenge, however, lies in the severe imbalance between legitimate and fraudulent transactions. Since legitimate cases vastly outnumber fraud, models often lean toward the majority class, reducing their ability to detect fraudulent activity. To address this, researchers have explored techniques such as SMOTE, feature engineering and cost-sensitive learning to improve minority-class recognition (Chawla et al., 2002; He & Garcia, 2009).

Recent studies have compared different machine learning algorithms for fraud detection. Preciado Martínez et al. (2025) evaluated multiple machine learning models for banking fraud detection and reported that ensemble learning methods generally achieve superior predictive performance. Similarly, Almalki and Masud (2025) integrated Explainable Artificial Intelligence (XAI) with stacking ensemble learning to improve both predictive accuracy and model transparency. Furthermore, Thivaios et al. (2026) highlighted that future fraud detection systems should combine robust preprocessing, effective class balancing, explainability and scalable machine learning architectures to improve real-world deployment.

Recent studies continue to advance this field, proposing novel ML architectures tailored for real-time deployment and scalability. Shakeabubakor (2024) demonstrates how ML models can be integrated with cloud-based data warehousing to enable real-time fraud detection in high-throughput environments.

Although previous studies have demonstrated the effectiveness of individual machine learning algorithms, comparatively fewer studies have investigated the combined impact of comprehensive data preprocessing, feature engineering, SMOTE based class balancing and comparative evaluation of multiple machine learning models within a unified experimental framework. Such an integrated evaluation is essential for identifying the most suitable algorithms for highly imbalanced financial transaction datasets.

Therefore, this study presents a comparative performance evaluation of four supervised machine learning algorithms Logistic Regression, Decision Tree, Random Forest and Artificial Neural Network for credit card fraud detection after applying data preprocessing, feature engineering and SMOTE based class balancing. The models are evaluated using Accuracy, Precision, Recall, F1-score and ROC-AUC to determine the most effective approach for intelligent fraud detection.

Contributions of this study

This study makes the following contributions:

1. It develops a comprehensive preprocessing pipeline incorporating data cleaning, feature engineering, feature scaling and SMOTE-based class balancing.
2. It evaluates the performance of four widely used supervised machine learning algorithms under identical experimental conditions.
3. It compares model performance using multiple evaluation metrics, including Accuracy, Precision, Recall, F1-score and ROC-AUC.
4. It provides recommendations for selecting suitable machine learning models for practical financial fraud detection systems.

Related Work

Financial fraud detection has attracted considerable research attention due to the rapid growth of digital banking, electronic payment systems and online financial transactions. Artificial intelligence and machine learning techniques have increasingly replaced traditional rule-based fraud detection systems because of their ability to learn complex transaction patterns and adapt to evolving fraudulent behaviours (Abdallah et al., 2016; West & Bhattacharya, 2016).

Several studies have investigated the application of supervised machine learning algorithms for financial fraud detection. Liu et al. (2022) developed a hybrid machine learning model for credit card fraud detection and demonstrated that ensemble learning techniques achieved higher predictive accuracy than individual classifiers. Similarly, Nama and Obaid (2024) employed deep learning techniques for financial fraud identification and reported that neural network models effectively captured complex nonlinear relationships within financial transaction datasets. However, both studies emphasized predictive performance without extensively addressing the challenges associated with severe class imbalance before model development.

Recent literature has increasingly focused on improving fraud detection through advanced preprocessing techniques and comprehensive comparative evaluation. Hernandez Aros et al. (2024) reviewed contemporary machine learning approaches for financial fraud detection and concluded that effective preprocessing, feature engineering and appropriate evaluation strategies are fundamental for improving fraud detection accuracy. Likewise, Rao and Mandhala (2024) surveyed modern machine learning and data mining techniques and identified data imbalance, scalability and practical deployment as major challenges affecting fraud detection systems.

With the increasing emphasis on trustworthy artificial intelligence, explainable and interpretable fraud detection models have also received significant attention. Almalki and Masud (2025) proposed an explainable artificial intelligence system that combines stacking ensemble learning with Explainable AI (XAI) techniques to improve both prediction accuracy and model interpretability. Their findings demonstrated that integrating explainability with ensemble learning enhances user confidence in fraud detection systems. Nevertheless, the study primarily focused on explainability rather than investigating the influence of comprehensive preprocessing and class balancing techniques on overall model performance.

Comparative evaluation studies have also become increasingly common in recent years. Preciado Martínez et al. (2025) compared several machine learning algorithms for detecting fraudulent banking transactions and reported that ensemble learning methods consistently outperformed traditional single classifiers. However, their investigation concentrated mainly on algorithm comparison and did not evaluate the contribution of integrated preprocessing stages such as duplicate removal, feature engineering, feature scaling and SMOTE-based class balancing prior to classification.

More recently, Thivaios et al. (2026) presented a comprehensive survey of machine learning and deep learning techniques for financial fraud detection, emphasizing the growing importance of graph neural networks, explainable artificial intelligence, multimodal financial data and scalable real-world deployment architectures. The authors concluded that future fraud detection systems should integrate

robust preprocessing pipelines, comprehensive evaluation metrics and practical deployment considerations to improve reliability in operational financial environments.

Research Gaps

Although previous studies have demonstrated the effectiveness of various machine learning algorithms for fraud detection, several research gaps remain. Many studies primarily focus on comparing algorithm performance without systematically evaluating the impact of comprehensive preprocessing pipelines that combine duplicate removal, feature scaling, feature engineering and SMOTE based class balancing. Furthermore, relatively few studies evaluate multiple supervised machine learning algorithms using a unified experimental system and a comprehensive set of evaluation metrics, including Accuracy, Precision, Recall, F1-score and ROC-AUC. Addressing these gaps is essential for identifying machine learning models capable of achieving reliable and balanced fraud detection performance on highly imbalanced financial transaction datasets.

To address these limitations, the present study implements a comprehensive preprocessing pipeline incorporating data cleaning, RobustScaler normalization, hour-based feature engineering and SMOTE-based class balancing before training four supervised machine learning models: Logistic Regression, Decision Tree, Random Forest and Artificial Neural Network. Their predictive performance is comparatively evaluated using multiple classification metrics to determine the most suitable model for intelligent financial fraud detection.

Table 1: Summary of Related Studies

Author(s)	Year	Method	Major Finding	Limitation
Liu et al.	2022	Hybrid Machine Learning	Ensemble methods improved fraud detection accuracy	Limited discussion of preprocessing
Hernandez Aros et al.	2024	Literature Review	ML techniques are effective for fraud detection	No experimental comparison
Rao & Mandhala	2024	Survey	Identified imbalance and deployment challenges	Review only
Almalki & Masud	2025	XAI + Stacking Ensemble	Improved transparency and prediction	Limited preprocessing analysis
Preciado Martínez et al.	2025	Comparative ML Models	Ensemble methods outperformed single models	No integrated preprocessing pipeline
Thivaios et al.	2026	Comprehensive Survey	Highlighted GNNs, XAI and deployment challenges	Recommended further practical validation
Sadineni	2020	Comparative ML Models	Artificial Neural Network achieved the highest prediction accuracy	Limited evaluation metrics and no real-time implementation.

Table 1 summarizes recent studies on machine learning-based financial fraud detection. Although previous researchers have reported promising classification performance using various machine

learning algorithms, most studies focused primarily on algorithm comparison or explainable AI techniques. Limited attention has been given to evaluating the combined effect of comprehensive preprocessing, feature engineering and SMOTE-based class balancing. This research addresses these limitations through an integrated experimental framework.

Materials and Methods

Research Methodology

The workflow begins with acquiring the dataset, which is then carefully prepared through preprocessing steps. Once the data is cleaned and standardized, the severe class imbalance is addressed using the Synthetic Minority Oversampling Technique (SMOTE). This ensures that fraudulent transactions are adequately represented before model development.

Following preprocessing, the dataset is split into training and testing subsets. The training portion is used to build and fine tune the machine learning models, while the testing subset serves as an independent benchmark to evaluate how well the models classify transactions they have never seen before. This separation is critical, as it allows researchers to measure the true generalization ability of each algorithm in detecting fraud.

The process continues with model training, where algorithms learn patterns from the balanced training data. Afterward, the models are tested against the unseen data and their performance is measured using established evaluation metrics. This structured pipeline from acquisition to preprocessing, balancing, training and evaluation ensures that the models are not only accurate but also reliable in real-world fraud detection scenarios

Four supervised machine learning algorithms are selected for comparison:

1. Logistic Regression
2. Decision Tree
3. Random Forest
4. Artificial Neural Network

These algorithms were selected because they represent different learning approaches, ranging from traditional statistical classification models to advance nonlinear learning models.

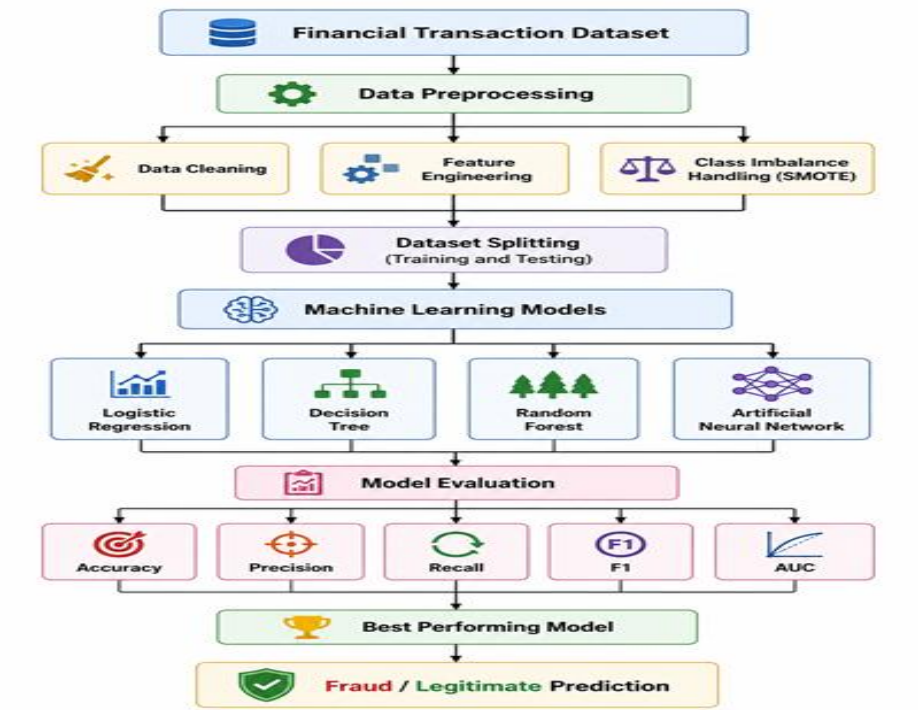

Figure 1: Architecture of the Proposed Financial Fraud Detection System

Figure 1 presents the architecture of the proposed financial fraud detection system. The architecture illustrates how transaction data flow through preprocessing, feature engineering, machine learning, prediction and output modules to produce fraud detection decisions.

Dataset Description

This study used the Kaggle Credit Card Fraud Detection dataset, which contains 284,807 European transactions represented by 31 attributes. The dataset is highly imbalanced: 99.83% legitimate transactions versus only 0.17% fraudulent ones. The target variable, Class**,** indicates whether a transaction is legitimate (Class = 0) or fraudulent (Class = 1).

The dataset is characterized by a severe class imbalance, Such imbalance presents a significant challenge for machine learning algorithms because models tend to favour the majority class, resulting in poor fraud detection performance. Consequently, appropriate preprocessing and class balancing techniques were applied before model development.

Table 2: Characteristics of the Credit Card Fraud Dataset

Attribute	Description
Dataset Source	Kaggle Credit Card Fraud Detection Dataset
Total Transactions	284,807
Legitimate Transactions	284,315
Fraudulent Transactions	492
Features	30 Predictor Variables
Target Variable	Class (0 = Legitimate, 1 = Fraud)

Table 2 presents the characteristics of the dataset used in this study. The dataset contains 284,807 credit card transactions, of which only 492 are fraudulent. This severe class imbalance highlights the need for appropriate preprocessing and class balancing techniques before model training.

Data Preprocessing

1. **Data cleaning:** Removed 1,081 duplicate records, leaving 283,726 unique transactions.
2. **Feature scaling:** Applied RobustScaler to Time and Amount attributes to reduce the influence of outliers.
3. **Feature engineering:** Extracted an "Hour" variable from the Time attribute to capture temporal fraud patterns

Table 3: Data Preprocessing Activities

Preprocessing Step	Purpose
Duplicate Removal	Eliminate repeated transaction records
Missing Value Check	Ensure data completeness
RobustScaler	Normalize Time and Amount features
Feature Engineering	Extract Hour feature
SMOTE	Balance minority fraud class

These operations in Table 3 were carried out sequentially to improve data quality, eliminate redundancy, normalize feature distributions and prepare the dataset for machine learning model development.

Feature Engineering

Feature engineering was carried out to strengthen the predictive power of the machine learning models. From the original *Time* attribute, a new temporal feature called *Hour* was derived to indicate the specific hour in which each transaction took place. By introducing this feature, the models were able to capture time-based patterns often linked to fraudulent activity, while still preserving the anonymized nature of the dataset.

Class Balancing Using SMOTE

To address the severe class imbalance in the training dataset, the Synthetic Minority Oversampling Technique (SMOTE) was applied. Unlike simple duplication, SMOTE generates synthetic minority class samples by interpolating between existing neighboring minority observations (Chawla et al., 2002). This approach enriches the representation of fraudulent transactions while avoiding redundancy. Importantly, SMOTE was applied only to the training data, which helps reduce bias toward legitimate transactions while preserving the integrity of the independent testing set.

Before oversampling, the training dataset consisted of 198,277 legitimate transactions and just 331 fraudulent ones. After applying SMOTE, both classes were expanded to contain 198,277 records each, resulting in a balanced training dataset of 396,554 transactions. This balanced distribution provided

the models with a fairer learning environment, enabling them to better capture fraud-related patterns and improving their ability to generalize to unseen data.

Table 4: Class Distribution Before and After SMOTE

Training Dataset	Legitimate	Fraudulent	Total
Before SMOTE	198,277	331	198,608
After SMOTE	198,277	198,277	396,554

As shown in Table 4, the training dataset was initially highly imbalanced, with 198,277 legitimate transactions compared to only 331 fraudulent ones. To correct this imbalance, the minority class was synthetically oversampled using SMOTE, resulting in an equal distribution of 198,277 observations per class. This adjustment produced a balanced dataset of 396,554 transactions, which reduced the tendency of the models to favor the majority class and enhanced their ability to learn meaningful fraud-related patterns. By ensuring equal representation of both classes, the models were better positioned to achieve reliable and effective fraud detection performance.

Train Test Partitioning

The cleaned dataset was divided into training and testing subsets using a 70:30 stratified sampling approach. Stratification ensured that the original class distribution was preserved across both subsets, allowing each model to be evaluated under representative conditions. This method provided a fair basis for comparison by maintaining consistency in the proportion of legitimate and fraudulent transactions. The final partition produced 198,608 samples for training and 85,118 samples for testing, establishing a balanced framework for model development and performance evaluation.

Table 5: Train-Test Dataset Partition

Dataset	Number of Samples	Percentage
Training	198,608	70%
Testing	85,118	30%

Table 5 illustrates the dataset distribution following stratified train test partitioning. In this process, seventy percent of the data was allocated for model training, while the remaining thirty percent was reserved for independent evaluation. This partitioning strategy ensured that the models were trained on a sufficiently large sample while still being tested against an unbiased subset, thereby providing a reliable measure of their ability to generalize to unseen transactions.

Machine Learning Models

Four supervised machine learning algorithms were implemented and systematically evaluated in this study.

1. **Logistic Regression (LR):** A linear probabilistic classifier commonly used as a baseline for binary classification tasks.

2. **Decision Tree (DT):** A tree-based model that recursively partitions the feature space into homogeneous subsets, offering interpretability but prone to overfitting.
3. **Random Forest (RF):** An ensemble learning method that aggregates multiple decision trees through bootstrap sampling and random feature selection, thereby improving predictive accuracy and reducing overfitting.
4. **Artificial Neural Network (ANN):** A multilayer feed-forward neural network capable of capturing complex nonlinear relationships within financial transaction data.

Performance Evaluation Metrics

The predictive performance of the machine learning models was assessed using five widely recognized classification metrics: Accuracy, Precision, Recall, F1-score and Receiver Operating Characteristic - Area Under the Curve (ROC-AUC).

1. Accuracy reflects the overall proportion of correctly classified transactions.
2. Precision measures the proportion of transactions predicted as fraudulent that were truly fraudulent, thereby indicating the reliability of positive predictions.
3. Recall evaluates the model's ability to correctly identify fraudulent transactions, making it particularly important in highly imbalanced datasets where minority-class detection is critical.
4. F1-score represents the harmonic mean of Precision and Recall, providing a balanced measure that is especially suitable for imbalanced data scenarios.
5. ROC-AUC quantifies the discriminatory capability of each classifier across varying thresholds, offering a comprehensive view of model performance beyond a single decision boundary (Naidu et al., 2023).

Together, these metrics provide a robust framework for evaluating the effectiveness of fraud detection models, ensuring that both majority and minority classes are adequately considered in performance assessment.

Experimental Environment

The proposed system was implemented in Python within the Jupyter Notebook environment, providing an interactive platform for data analysis and model development. Data preprocessing and manipulation were carried out using Pandas and NumPy, which offered efficient handling of large-scale transaction records. The supervised machine learning models were developed with the Scikit-learn library, ensuring consistency and reproducibility across experiments.

For the Artificial Neural Network, TensorFlow/Keras was employed to construct and train multilayer feed-forward architectures capable of modeling complex nonlinear relationships. Finally, graphical visualizations and performance plots were generated using Matplotlib**,** enabling clear presentation of results and comparative analysis of model performance.

Results and Discussion

Model Performance

The predictive performance of the four supervised machine learning models was assessed using Accuracy, Precision, Recall, F1-score and ROC-AUC**.** Together, these metrics provide a comprehensive evaluation of classifier effectiveness, particularly in the context of highly imbalanced datasets where accuracy alone can be misleading. While Accuracy reflects the overall proportion of correctly classified transactions, Precision and Recall highlight the model's ability to correctly identify fraudulent cases without being overwhelmed by legitimate ones. The F1-score balances these two measures, making it especially relevant for fraud detection tasks. Finally, ROC-AUC offers an aggregate view of each model's discriminatory capability across varying thresholds, ensuring a robust comparison of performance.

Table 6: Comparative Performance Comparison of Machine Learning Models

Model	Accuracy (%)	Precision (%)	Recall (%)	F1-score (%)	ROC-AUC (%)
Logistic Regression	97.18	4.98	88.03	9.43	97.34
Decision Tree	99.17	14.29	71.83	24.22	85.84
Random Forest	99.95	91.45	75.35	82.63	96.84
Artificial Neural Network	99.89	65.20	80.28	72.15	98.75

The results presented in Table 6 demonstrate that the four machine learning algorithms exhibited varying capabilities in detecting fraudulent credit card transactions. Although all classifiers achieved relatively high overall accuracy, noticeable differences were observed in Precision, Recall, F1-score and ROC-AUC values. This finding reinforces the argument that relying solely on accuracy is insufficient for evaluating fraud detection models, since highly imbalanced datasets can produce deceptively high accuracy while failing to identify fraudulent transactions effectively.

1. Logistic Regression

The Logistic Regression classifier achieved an overall accuracy of 97.18%**,** the lowest among the evaluated models. While the model obtained a relatively high Recall of 88.03%**,** its Precision was only 4.98%**,** resulting in a very low F1-score of 9.43%**.** These results indicate that although Logistic Regression successfully detected many fraudulent transactions, it also generated a large number of false positives. This behavior is expected, as Logistic Regression assumes a linear decision boundary, limiting its ability to capture the complex nonlinear relationships that characterize financial fraud patterns.

Decision Tree

The Decision Tree model produced a higher overall accuracy **(**99.17%**)** than Logistic Regression and demonstrated improved fraud detection capability, with a Recall of **7**1.83% and an F1-score of 24.22%**.** However, its relatively low Precision (14.29%) suggests that the classifier remained susceptible to false-positive predictions. While Decision Trees are valued for their interpretability and

computational efficiency, their tendency to overfit complex datasets reduces their ability to generalize effectively, particularly in highly imbalanced financial transaction scenarios.

Random Forest

Among the evaluated classifiers, the Random Forest model achieved the best overall performance. It recorded an accuracy of 99.95%, Precision of 91.45%, Recall of 75.35%, F1-score of 82.63% and ROC-AUC of 96.84%. These results indicate that Random Forest effectively balanced fraud detection capability with a low false-positive rate. Its superior performance can be attributed to the ensemble learning mechanism, which combines multiple decision trees through bootstrap aggregation and random feature selection. This approach enhances robustness, minimizes overfitting and improves the ability to capture complex nonlinear relationships in financial transaction data.

Artificial Neural Network (ANN)

The Artificial Neural Network also demonstrated excellent predictive performance, achieving an overall accuracy of **99.89%** and the highest ROC-AUC value **(98.75%)** among all evaluated models. Furthermore, the model achieved a Recall of **80.28%**, indicating strong capability in identifying fraudulent transactions. However, its Precision (**65.20%**) was lower than that of Random Forest, resulting in a slightly lower F1-score (**72.15%**). This suggests that although ANN successfully detected a larger proportion of fraudulent transactions, it also produced more false positives compared to Random Forest. Nevertheless, the results confirm that neural networks remain highly effective for modeling the nonlinear characteristics of financial transaction data.

Comparative Insights

The superior performance achieved by the Random Forest classifier is consistent with the findings of Liu et al. (2022), who reported that ensemble learning techniques outperform individual machine learning algorithms by combining multiple classifiers to improve predictive accuracy and robustness. Similarly, Preciado Martínez et al. (2025) observed that ensemble-based approaches consistently achieved better fraud detection performance than traditional single classifiers when evaluated on banking transaction datasets.

The present study also complements the work of Almalki and Masud (2025), who demonstrated that integrating ensemble learning with Explainable Artificial Intelligence (XAI) improves both prediction accuracy and model interpretability. Although this study focused primarily on preprocessing and comparative evaluation rather than explainability, the strong performance of Random Forest suggests that it could serve as an effective foundation for future integration with XAI techniques such as SHAP or LIME to enhance transparency and user trust.

Furthermore, the overall conclusions align with the comprehensive review conducted by Thivaios et al. (2026), who emphasized that effective fraud detection systems should combine robust preprocessing techniques, appropriate class balancing strategies, comprehensive evaluation metrics and advanced machine learning algorithms to achieve reliable real-world performance. The present study supports these recommendations by demonstrating that the integration of duplicate removal, feature engineering, RobustScaler normalization and SMOTE-based class balancing significantly enhanced the effectiveness of the evaluated machine learning models.

Overall Conclusion of Results

Collectively, these findings indicate that combining comprehensive data preprocessing with ensemble learning provides a reliable and practical solution for intelligent credit card fraud detection in modern financial systems. Random Forest emerged as the most effective model, striking the best balance between Precision and Recall, while ANN showed strong potential for capturing complex nonlinear patterns. Logistic Regression and Decision Tree, though useful as baselines, were less effective in handling the challenges posed by highly imbalanced datasets.

Accuracy Comparison

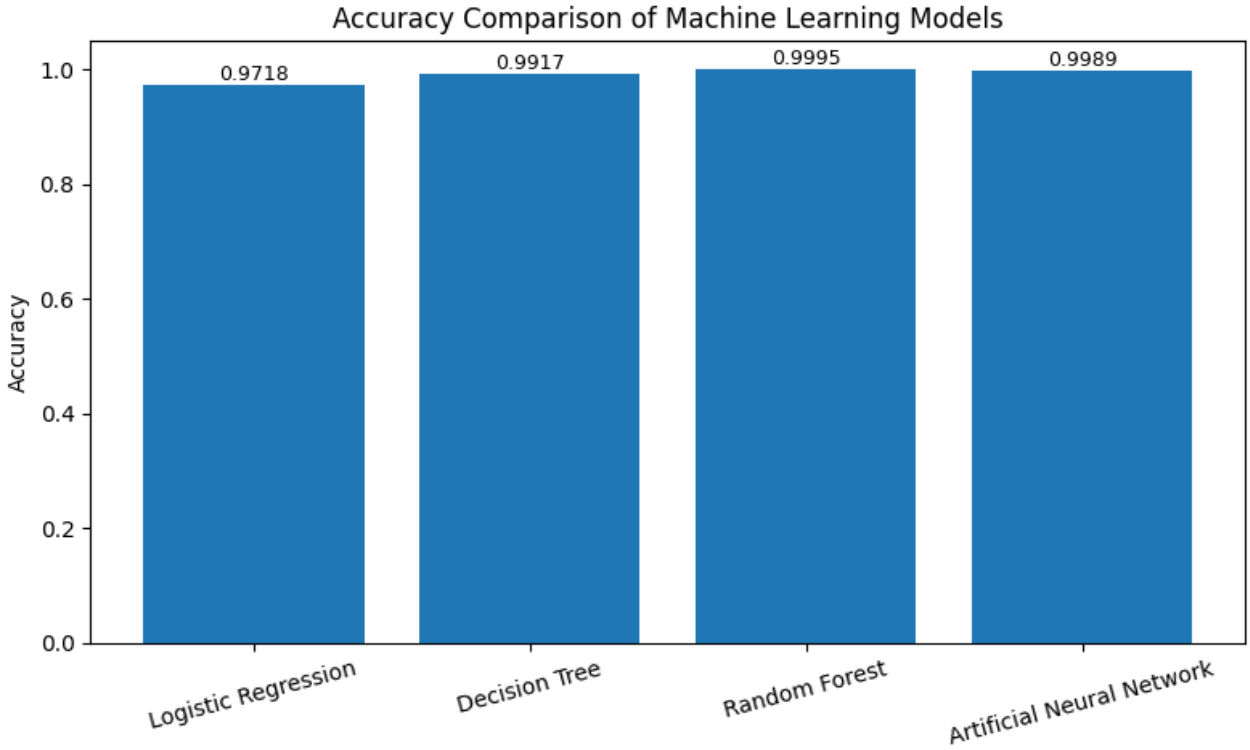

Figure 2: Accuracy Comparison of Machine Learning Models

Figure 2 illustrates the comparative classification accuracy of the four supervised machine learning algorithms evaluated in this study. The Random Forest model achieved the highest accuracy, closely followed by the Artificial Neural Network, while the Decision Tree and Logistic Regression classifiers produced comparatively lower accuracies. These results confirm that ensemble learning approaches, such as Random Forest, deliver superior predictive performance in the context of financial fraud detection, highlighting their robustness in handling complex and imbalanced transaction datasets.

Cross-Validation Analysis

Three-fold Cross-Validation was performed to evaluate the stability and generalization capability of the Random Forest model.

Table 7: Cross-Validation Results

Fold	Accuracy
Fold 1	99.90%
Fold 2	99.91%
Fold 3	99.87%
Average Accuracy	**99.89%**
Standard Deviation	**0.01%**

Table 7 presents the results of the 3-fold cross-validation performed on the Random Forest model. The model achieved an average validation accuracy of 99.89% with a standard deviation of 0.01%**,** indicating highly consistent performance across all validation folds. The extremely small standard deviation demonstrates the robustness and stability of the model, suggesting that it generalizes effectively to unseen data and is not overly sensitive to variations in the training subsets.

Real-Time Fraud Transaction Predictions

The proposed fraud detection approach successfully simulated a real-time transaction monitoring environment. Transactions from the testing dataset were processed sequentially, with each transaction independently classified by the trained Random Forest model. The simulation results confirmed that the system was capable of analyzing live transaction streams and generating immediate fraud predictions, accompanied by associated fraud probability scores.

Real-time fraud detection requires models that can process transaction information rapidly while adapting to evolving fraud patterns (West & Bhattacharya, 2016). The successful simulation demonstrates that the Random Forest model not only achieves high predictive accuracy in static evaluation but also performs reliably in dynamic, real-time scenarios. This highlights its potential for deployment in modern financial systems where timely detection is critical to minimizing losses and maintaining customer trust.

Table 8: Sample Real-Time Transaction Predictions

Transaction ID	Actual	Prediction	Probability
90161	Legitimate	Legitimate	0.00
16415	Fraudulent	Fraudulent	1.00
41458	Legitimate	Legitimate	0.00
274842	Legitimate	Legitimate	0.00
12369	Fraudulent	Fraudulent	0.99

The real-time fraud detection simulation demonstrated that the trained Random Forest model successfully processed transactions sequentially and produced immediate fraud predictions accompanied by fraud probability scores. The model correctly identified both legitimate and fraudulent transactions in the simulated transaction stream, confirming its suitability for deployment in real-time financial transaction monitoring systems. These findings further highlight the practical applicability of the proposed machine learning approach for operational fraud detection.

Discussion of Classification Performance

The comparative evaluation revealed that the four supervised machine learning algorithms exhibited varying levels of effectiveness in detecting fraudulent credit card transactions. Although all models achieved relatively high classification accuracy, their performance differed considerably when assessed using Precision, Recall, F1-score and ROC-AUC. This observation confirms that accuracy alone is insufficient for evaluating fraud detection systems, given the highly imbalanced nature of financial transaction datasets. Consequently, multiple evaluation metrics were employed to provide a more comprehensive assessment of classifier performance.

1. **Random Forest** achieved the best overall performance, recording an accuracy of 99.95%, Precision of 91.45%, Recall of 75.35%, F1-score of 82.63% and ROC-AUC of 96.84%. These results demonstrate its effectiveness in distinguishing fraudulent transactions while maintaining a low false-positive rate. Its superior performance is attributed to ensemble learning, which combines multiple decision trees through bootstrap aggregation, thereby minimizing overfitting and capturing complex nonlinear relationships.
2. **Artificial Neural Network (ANN)** also demonstrated strong predictive capability, achieving an accuracy of 99.89%, Recall of 80.28% and F1-score of 72.15%. Its relatively high Recall indicates strong detection of fraudulent transactions, making it suitable for applications where minimizing missed fraud cases is critical. However, its lower Precision compared with Random Forest suggests a higher false-positive rate, highlighting the trade-off between maximizing fraud detection and minimizing unnecessary alerts.
3. **Decision Tree** achieved an accuracy of 99.17%, but its Precision (14.29%) and F1-score (24.22%) were considerably lower than those of Random Forest and ANN. Although Decision Trees are valued for interpretability and efficiency, their susceptibility to overfitting limits generalization in complex, imbalanced datasets.
4. **Logistic Regression** produced the lowest overall performance. Despite achieving an accuracy of 97.18% and a Recall of 88.03%, its Precision was only 4.98%, resulting in a very low F1-score. This indicates that while Logistic Regression detected many fraudulent transactions, it also generated a large number of false positives. This limitation stems from its assumption of linear relationships, which restricts its ability to capture the nonlinear patterns typical of financial fraud.

The effectiveness of the preprocessing pipeline also contributed significantly to the observed performance. Duplicate removal improved data quality, RobustScaler reduced the influence of extreme values and feature engineering extracted temporal information from transaction timestamps. Most importantly, SMOTE addressed severe class imbalance by generating synthetic minority-class samples, enabling all classifiers to learn fraud-related patterns more effectively.

SMOTE had a particularly positive impact on Recall and F1-score, which are critical in fraud detection. Identifying fraudulent transactions is generally more important than marginal improvements in overall accuracy, since undetected fraud can result in substantial financial losses. The balanced training dataset therefore allowed the models to improve minority-class recognition while maintaining satisfactory overall performance.

These findings are consistent with Liu et al. (2022**)**, who reported that ensemble learning techniques outperform conventional algorithms in fraud detection tasks. Similarly, **Preciado** Martínez et al.

(2025**)** observed that ensemble-based classifiers consistently achieved superior performance compared with individual models. The strong performance of Random Forest in this study reinforces the growing evidence supporting ensemble learning for financial fraud detection.

The results also complement the work of Almalki and Masud (2025**)**, who demonstrated that integrating advanced machine learning with Explainable AI improves both predictive accuracy and transparency. Although explainability was not incorporated here, Random Forest provides a robust foundation for future integration with XAI techniques such as SHAP or LIME to enhance interpretability and support decision-making by fraud analysts.

Furthermore, the conclusions align with Thivaios et al. (2026), who emphasized that effective fraud detection systems should integrate preprocessing, robust algorithms, comprehensive evaluation metrics and practical deployment strategies. By combining duplicate removal, feature engineering, RobustScaler normalization, SMOTE-based balancing and comparative evaluation, this study addresses several deployment challenges identified in recent literature.

The cross-validation analysis further demonstrated the robustness of the approach. Random Forest achieved an average cross-validation accuracy of 99.89% with a standard deviation of 0.01%, indicating highly consistent performance across folds. This stability suggests strong generalization to unseen data, which is essential in fraud detection where transaction patterns vary over time.

Finally, the real-time simulation confirmed the practical applicability of the proposed methodology. The Random Forest model successfully processed sequential transactions and generated immediate fraud predictions with probability scores. These results demonstrate that the approach is not only effective under experimental conditions but also suitable for real-time deployment in banking and fintech environments. Collectively, the study provides both theoretical and practical evidence that integrating comprehensive preprocessing with ensemble learning offers a reliable solution for intelligent credit card fraud detection.

Practical Implications

The findings of this study demonstrate that integrating comprehensive preprocessing techniques with machine learning substantially enhances fraud detection performance. The proposed preprocessing pipeline comprising duplicate removal, RobustScaler normalization, feature engineering and SMOTE-based class balancing provides a reliable foundation for developing intelligent fraud detection systems. By improving data quality and addressing class imbalance, this pipeline enables machine learning models to more effectively capture fraud-related patterns and deliver robust predictive performance.

Financial institutions and fintech organizations can adopt this pipeline to strengthen the reliability of AI-driven fraud detection systems in operational environments. Specifically, preprocessing ensures that transaction data is clean, balanced and representative, which reduces bias and improves model generalization. This not only enhances fraud detection accuracy but also minimizes false positives, thereby reducing unnecessary alerts and improving customer trust.

In practice, the adoption of such a pipeline can support real-time fraud monitoring systems, allowing institutions to process large volumes of transactions efficiently while adapting to evolving fraud patterns. Consequently, the study provides actionable insights for the deployment of machine learning based fraud detection solutions in modern financial systems, bridging the gap between theoretical research and operational implementation.

Conclusion

This study presented a comparative evaluation of four supervised machine learning algorithms for credit card fraud detection, supported by a comprehensive data preprocessing pipeline. The methodology integrated duplicate removal, RobustScaler normalization, feature engineering and SMOTE based class balancing prior to training Logistic Regression, Decision Tree, Random Forest and Artificial Neural Network models. These preprocessing stages significantly improved the quality and balance of the training data, enabling the models to better distinguish between legitimate and fraudulent transactions.

The experimental results demonstrated that all evaluated models were capable of detecting fraudulent transactions with high classification performance. However, the Random Forest classifier consistently achieved the best overall results across all evaluation metrics, confirming its superior ability to handle highly imbalanced financial transaction data. The findings further emphasized that effective preprocessing and class balancing are critical for improving fraud detection performance, regardless of the classification algorithm employed.

Overall, the study provides empirical evidence that integrating data preprocessing with machine learning offers a practical and reliable approach for intelligent credit card fraud detection. The proposed methodology can serve as a valuable reference for researchers and financial institutions seeking to develop robust fraud detection systems capable of operating in highly imbalanced transaction environments.

Future Work

Future research should extend the evaluation of the proposed methodology using multiple financial transaction datasets from diverse domains to enhance its generalizability. In addition, advanced artificial intelligence techniques should be explored to further improve transparency, scalability and deployment in real-world financial systems. Promising directions include:

1. **Explainable Artificial Intelligence (XAI):** Integrating methods such as SHAP or LIME to improve interpretability and support decision-making by fraud analysts.
2. **Graph Neural Networks (GNNs):** Leveraging relational structures in transaction networks to capture hidden fraud patterns.
3. **Transformer-based deep learning models:** Applying attention mechanisms to model sequential transaction data more effectively.
4. **Federated learning:** Enabling collaborative fraud detection across institutions without compromising data privacy.
5. **Real-time streaming fraud detection:** Developing scalable architectures capable of processing high-volume transaction streams with minimal latency.

By pursuing these directions, future work can build upon the strong foundation established in this study to advance the development of intelligent, transparent and scalable fraud detection systems for modern financial environments.

References

1. Abdallah, A., Maarof, M. A., & Zainal, A. (2016). Fraud detection system: A survey. *Journal of Network and Computer Applications, 68*, 90-113. https://doi.org/10.1016/j.jnca.2016.04.007
2. Ali, A., et al. (2022). *Financial fraud detection based on machine learning: A systematic literature review*. Applied Sciences, 12(19), 9637. https://doi.org/10.3390/app12199637
3. Almalki, F., & Masud, M. (2025). *Financial fraud detection using explainable AI and stacking ensemble methods* (arXiv:2505.10050). arXiv. https://doi.org/10.48550/arXiv.2505.10050
4. Chawla, N. V., et al. (2002). *SMOTE: Synthetic minority over-sampling technique*. Journal of Artificial Intelligence Research, 16, 321-357.
5. He, & Garcia, E. A. (2009). Learning from imbalanced data. *IEEE Transactions on Knowledge and Data Engineering, 21*(9), 1263–1284. https://doi.org/10.1109/TKDE.2008.239
6. Hernandez Aros, L., Bustamante Molano, L. X., Gutierrez-Portela, F., et al. (2024). *Financial fraud detection through the application of machine learning techniques: A literature review*. Humanities and Social Sciences Communications, 11**,** Article 1130. https://doi.org/10.1057/s41599-024-03670-0
7. Liu, Y., Xia, Y., Yu, C., & Dou, Y. (2022). A hybrid machine learning model for credit card fraud detection. *Information Sciences, 585*, 67–82. https://doi.org/10.1016/j.ins.2021.07.002
8. Naidu, G., Zuva, T., & Sibanda, E. (2023). *A Review of Evaluation Metrics in Machine Learning Algorithms*. In *Artificial Intelligence Application in Networks and Systems* (pp. 15–25). This review highlights that accuracy, precision, recall, F1-score, and ROC-AUC are standard tools for classifier performance evaluation. http://dx.doi.org/10.1007/978-3-031-35314-7_2
9. Nama, F. A., & Obaid, A. J. (2024). *Financial fraud identification using deep learning techniques*. *Al-Salam Journal for Engineering and Technology*, *3*(1), 141–147. https://doi.org/10.55145/ajest.2024.03.01.012
10. Ngai, E. W. T., Hu, Y., Wong, Y. H., Chen, Y., & Sun, X. (2011). The application of data mining techniques in financial fraud detection: A classification framework and academic review of literature. *Decision Support Systems, 50*(3), 559-569. https://doi.org/10.1016/j.dss.2010.08.006
11. Preciado Martínez, P. M., López, A. M., & García, J. R. (2025). Comparative analysis of machine learning models for the detection of fraudulent banking transactions. *Cogent Business & Management, 12*(1), Article 2474209. https://doi.org/10.1080/23311975.2025.2474209
12. Rao, R. K., & Mandhala, V. N. (2024). Unveiling financial fraud: A comprehensive review of machine learning and data mining techniques. *Informatica, 29*(6), 1-25. https://doi.org/10.18280/isi.290620
13. Sadineni, P. K. (2020). *Detection of fraudulent transactions in credit card using machine learning algorithms*. Proceedings of the Fourth International Conference on I-SMAC (pp. 878-883). IEEE. https://doi.org/10.1109/I-SMAC49090.2020.9243545
14. Shakeabubakor, A. A. (2024). Real-time fraud detection in financial transactions: An integrated approach using machine learning and cloud-based data warehousing. *International*

Journal of Intelligent Systems and Applications in Engineering, 12(23s), 89-97. https://www.ijisae.org/index.php/IJISAE/article/view/7038

15. Thivaios, S., Kostopoulos, G., Stefani, A., & Kotsiantis, S. (2026). *A survey of machine learning and deep learning for financial fraud detection: Architectures, data modalities and real-world deployment challenges***.** *Algorithms, 19*(5), 354. https://doi.org/10.3390/a19050354
16. West, J., & Bhattacharya, M. (2016). Intelligent financial fraud detection: A comprehensive review. *Computers & Security, 57*, 47-56. https://doi.org/10.1016/j.cose.2015.10.005