FUDMA Journal of Sciences (FJS)
ISSN online: 2616-1370
ISSN print: 2645 - 2944
Vol. 10 No. 3, February, 2026, pp 102 – 108
DOI: https://doi.org/10.33003/fjs-2026-1003-4745

COMPARATIVE PERFORMANCE OF MACHINE LEARNING MODELS FOR CREDIT CARD FRAUD DETECTION ON IMBALANCED DATA: A STUDY USING SMOTE AND THE KAGGLE EUROPEAN DATASET

***Ojo Opeyemi Femi, Otaru Paul Olawale, Job O. Eunice and Onoghojobi Benson**

Department of Statistics, Federal University Lokoja, Kogi State, Nigeria.

*Corresponding authors' email: ojoopeyemifemi@gmail.com

ABSTRACT

Credit card fraud continues to be a major challenge in financial systems, causing substantial monetary losses and undermining trust in electronic transactions. Detecting fraudulent activities is complicated by the highly imbalanced nature of transaction datasets. This study evaluates the performance of four machine learning models including LinearSVC (Fast SVM), Logistic Regression, Decision Tree, and Random Forest using a Kaggle credit card dataset. Data preprocessing involved handling class imbalance using the Synthetic Minority Oversampling Technique (SMOTE) and standardizing feature values to improve model training stability. Each model was trained and assessed using metrics such as accuracy, precision, recall, F1-score, and AUC–ROC, with confusion matrices and ROC curves providing visual evaluation. Experimental results demonstrate that LinearSVC achieved the highest balance between precision (0.9983) and F1-score (0.9881), while Random Forest achieved near-perfect accuracy (0.9999) but slightly lower precision (0.8800) and recall (0.9002), highlighting trade-offs between overall accuracy and fraud detection sensitivity. The findings emphasize the importance of careful model selection and preprocessing in imbalanced financial datasets. Future work will investigate the integration of deep learning models and real-time fraud detection frameworks, as well as validation on additional datasets from diverse financial environments to enhance model robustness and generalizability.

Keywords: Credit Card, Fraud, Machine Learning, Statistical Analysis, Algorithms

INTRODUCTION

Credit card fraud remains a pervasive threat to financial institutions and consumers worldwide, driven by the rapid expansion of digital payment systems and increasingly sophisticated fraudulent tactics. As the volume and velocity of transaction data grow, traditional rule-based detection approaches have become insufficient for timely and accurate identification of fraudulent behavior. This has led to a surge of interest in machine learning (ML)-based methods that can automatically learn complex patterns from large, real-world datasets and adapt to evolving attack strategies. Recent studies emphasize the promise of ML techniques, such as ensemble classifiers and hybrid models, in improving fraud detection performance, especially in handling transaction volumes and minimizing financial losses (Kumari & Singh, 2022). A key challenge in this field is the severe class imbalance inherent in fraud datasets, where legitimate transactions far outnumber fraudulent ones. For instance, the widely used European credit card fraud dataset contains only 0.17 % fraudulent cases, making it difficult for ML models to reliably identify rare fraud instances without specialized techniques to address imbalance (Breskuvienė & Dzemyda, 2024). To mitigate this imbalance, resampling strategies such as the Synthetic Minority Oversampling Technique (SMOTE) and other data preprocessing methods have been widely recommended in the literature, demonstrating significant improvements in model performance (Baker et al., 2022). Machine learning algorithms such as Logistic Regression, Decision Trees, Random Forest, and Support Vector Machines have been extensively evaluated in recent research for their ability to distinguish between legitimate and fraudulent transactions under imbalanced conditions. Findings consistently highlight that ensemble methods like Random Forest often outperform simple classifiers due to their ability to reduce variance and capture non-linear relationships in high-dimensional data (Sundaravadivel et al., 2025). However, although ensemble models such as Random Forest are often reported to perform better in fraud detection studies, their advantage is not always consistent in all cases. Ensembles combine many decision trees, which helps them learn complex patterns and reduce prediction errors. However, several studies show that simpler models such as Logistic Regression and Support Vector Machines can also perform well when proper preprocessing methods like SMOTE and feature scaling are used. For example, studies using the widely used European credit card fraud dataset often report F1-scores around 0.90–0.95 and AUC values between 0.95–0.98 for well-tuned models after handling class imbalance (Dal Pozzolo et al., 2015; Carcillo et al., 2021). In addition, this dataset has been used extensively in the literature, which may sometimes lead to overly optimistic performance results if evaluation procedures are not carefully designed. Despite these advances, there remain gaps in systematically comparing classic and optimized ML models under realistic operational constraints, such as computational efficiency and real-time applicability. This paper contributes to the field by implementing multiple supervised learning models on the publicly available Kaggle dataset, addressing imbalance through SMOTE, and evaluating the classifiers using metrics that reflect both detection accuracy and practical performance. The findings extend current understanding of how traditional ML approaches perform relative to each other in a real-world fraud detection context.

Literature Review

Azim et al. (2024) addressed the challenge of credit card fraud detection caused by highly imbalanced transaction data. Using an imbalanced credit card dataset, the researchers developed a soft voting ensemble learning model to improve fraud identification. The method was evaluated against multiple sampling techniques including oversampling, under-sampling, and hybrid approaches to mitigate class imbalance. Several machine learning classifiers and ensemble models were trained and compared, both with and without sampling.

Experimental findings show that the proposed soft voting ensemble achieved superior performance over individual models, recording a precision of 0.9870, recall of 0.9694, FNR of 0.0306, F1-score of 0.8764, and AUROC of 0.9936, demonstrating its effectiveness for fraud detection.

Xie & Huang (2024) tackled credit card fraud detection by addressing limitations in existing methods. Using Kaggle credit card fraud datasets and the Taiwan credit card customer default dataset, the researchers propose a Random Forest-based model combined with Mahalanobis distance SMOTE-ENN hybrid sampling to handle class imbalance. Experiments compared the proposed approach with existing methods, demonstrating improved detection accuracy and efficiency. The results indicate that this hybrid sampling and Random Forest model outperforms conventional techniques, confirming its effectiveness across different datasets. The method shows strong practical potential for credit card fraud detection, offering enhanced predictive performance while providing methodological insights for similar imbalanced classification problems in financial risk management.

Alarfaj et al. (2022) addressed credit card fraud detection using both machine learning and deep learning approaches on the European card benchmark dataset. Initially, traditional machine learning algorithms—such as Decision Tree, Random Forest, SVM, Logistic Regression, and XGBoost—were applied, achieving moderate accuracy. To improve performance, the study implemented three convolutional neural network (CNN) architectures, experimenting with variations in hidden layers, epochs, and model configurations. Data balancing techniques were also applied to reduce false negatives. The proposed deep learning models outperformed existing methods, achieving an accuracy of 99.9%, F1-score of 85.71%, precision of 93%, and AUC of 98%. Results demonstrate the models' strong practical potential for real-world credit card fraud detection.

Ileberi et al. (2021) presented a machine learning-based framework for credit card fraud detection using imbalanced datasets from European credit cardholders. To address class imbalance, the SMOTE was applied. Several ML algorithms—SVM, Logistic Regression, Random Forest, XGBoost, Decision Tree, and Extra Tree—were evaluated, both standalone and combined with Adaptive Boosting (AdaBoost) to enhance classification performance. The models were assessed using accuracy, precision, recall, MCC, and AUC. Further validation was conducted on a highly skewed synthetic dataset. Experimental results demonstrated that AdaBoost significantly improved performance, and the boosted models outperformed existing methods, confirming the framework's effectiveness for accurate and reliable credit card fraud detection.

Du et al. (2023) addressed credit card fraud detection amid increased online transactions during the COVID-19 pandemic, using a real-world European credit card dataset and stratified K-fold cross-validation. A total of 66 machine learning models were evaluated in two stages: initially, nine algorithms were tested, and the top three were further assessed with 19 resampling techniques. From 330 metric evaluations, the AllKNN undersampling combined with CatBoost (AllKNN-CatBoost) was identified as the best-performing model. Comparative analysis with related studies showed that the proposed model outperformed previous approaches, achieving an AUC of 97.94%, Recall of 95.91%, and F1-score of 87.40%, demonstrating its effectiveness for accurately detecting fraudulent credit card transactions.

Ghaleb et al. (2023) proposed a credit card fraud detection model (CCFDM) using ensemble learning combined with GANs and Ensemble Synthesized Minority Oversampling (ESMOTE-GAN) to address high-class imbalance. Multiple subsets were created via under-sampling and SMOTE to reduce skewness and prevent GANs from modeling noise. These subsets trained diverse GAN models, generating synthesized data used to train Random Forest classifiers. The classifiers' probabilistic outputs were combined using a weighted voting scheme for final decisions. Experimental results demonstrated that CCFDM improved overall performance by 1.9% and the detection rate by 3.2%, achieving a 0% false alarm rate. The model effectively enhances detection accuracy while minimizing manual analysis costs in large-scale credit card transaction monitoring.

Alraddadi (2023) proposed a theoretical credit card fraud detection and prevention model using a Decision Tree Algorithm (DCA) and evaluates its relevance through a survey. Data were collected from 102 university students worldwide to assess their awareness and perceptions of credit/debit card fraud. Results indicated that 95.9% of respondents understood how such fraud occurs, and 81.6% expressed willingness to use a tool based on the proposed model to prevent or detect fraudulent transactions. The study highlights the potential of Decision Tree-based approaches for credit card fraud prevention and underscores the importance of user awareness and acceptance in adopting technological solutions for secure electronic payment systems.

Despite advances in credit card fraud detection using machine learning and deep learning, existing studies largely focus on European or Taiwanese datasets, limiting applicability to other regions like Nigeria. While ensemble and deep learning models achieve high accuracy, there is limited exploration of simpler, interpretable models—such as Logistic Regression, Decision Tree, Random Forest, and SVM—applied consistently with robust preprocessing and class imbalance handling. Furthermore, few studies provide a comprehensive, side-by-side comparison of multiple models using standardized metrics, visualizations like confusion matrices, and ROC curves for interpretability. Addressing these gaps, this study applies and evaluates multiple machine learning models on a Kaggle credit card dataset, incorporating SMOTE for class balancing and feature standardization. This approach enables a clear comparison of model performance, offering practical insights for effective and interpretable credit card fraud detection in real-world contexts.

MATERIALS AND METHODS

This study adopted a supervised machine learning approach to develop and evaluate credit card fraud detection models using a publicly available Kaggle dataset. The methodology followed a structured sequence of processes, including data acquisition, exploratory data analysis, preprocessing, model development, and performance evaluation. All experiments were implemented in Python using the Scikit-learn library within the Google Colab environment, with a fixed random seed to ensure reproducibility.

Data Acquisition

The dataset used in this study was sourced from Kaggle's openly accessible European credit card transaction dataset, consisting of anonymized features generated through Principal Component Analysis (PCA), along with two non-anonymized attributes: Time and Amount. The target variable, Class, distinguishes legitimate transactions (0) from fraudulent ones (1). The dataset contains 284,807 transactions, of which 492 are fraudulent, representing approximately 0.17% of the total data. All experiments were conducted in Google Colab using Python-based machine learning libraries.

Exploratory Data Analysis

Exploratory Data Analysis (EDA) was performed to understand the dataset’s structure, descriptive statistics, attribute distributions, and the presence of missing values. Visual techniques such as heatmaps and count plots were used to highlight patterns in the data and reveal the extreme imbalance between legitimate and fraudulent transactions. Addressing this imbalance was identified as a methodological priority, consistent with existing studies showing its impact on classifier performance (Nazerke et al., 2025).

Data Preprocessing

Data preprocessing involved a series of steps to ensure model accuracy and robustness. First, missing values were checked, and dataset integrity was confirmed. The severe class imbalance where fraudulent transactions accounted for less than 0.2% of the data was addressed using the SMOTE, a widely adopted method for improving minority-class representation in fraud detection contexts (Chawla et al., 2002). The dataset was then divided into training (80%) and testing (20%) sets using stratified sampling to preserve class proportions. Feature scaling was performed using StandardScaler to normalize numerical attributes and enhance model convergence during training.

Model Development

Four supervised machine learning algorithms were implemented: Logistic Regression, Decision Tree, Random Forest, and Support Vector Machine (LinearSVC). These models were selected for their strong performance in prior fraud detection literature and their ability to handle high-dimensional data efficiently (Sinap, 2024). Hyperparameters were selected using grid search with 5-fold cross-validation on the training dataset to identify suitable model configurations. For example, the Decision Tree model depth was limited to a maximum depth of 6 to control model complexity and reduce overfitting. The Linear SVC model (often referred to as a fast implementation of Support Vector Machines) was implemented using the liblinear solver with L2 regularization and a regularization parameter C = 1.0. In addition, class_weight = "balanced" was applied across models where applicable to improve sensitivity to minority-class fraud cases.

Model Evaluation

Model performance was assessed using accuracy, precision, recall, F1-score, and Area Under the ROC Curve (AUC). Confusion matrices and ROC curves were generated to visualize classifier performance and enable comparative analysis. These metrics were selected to ensure balanced evaluation of both detection capability and error minimization, consistent with evaluation standards in fraud analytics research. To improve the reliability of the evaluation, 5-fold cross-validation was conducted during model training, and the final performance was assessed on the held-out test dataset. Confusion matrices and ROC curves were generated to visually compare the classification capabilities of the models and to analyze the trade-offs between fraud detection sensitivity and false alarm rates.

RESULTS AND DISCUSSION

The experimental evaluation focused on assessing the performance of four machine learning models Logistic Regression, Decision Tree, Random Forest, and Support Vector Machine (LinearSVC) applied to the preprocessed credit card fraud dataset. Each model was trained using standardized features and optimized through class-weight adjustments to address imbalance. The results demonstrate varying degrees of predictive effectiveness across the algorithms.

Quantitative Performance

Table 1 summarizes the performance metrics for all models, including accuracy, precision, recall, F1-score, and AUC. Overall performance was high across all classifiers, reflecting effective preprocessing and the discriminatory power of the dataset.

Table 1: Summary of Model Performance

Model	Accuracy	Precision	Recall	f1-scores	AUC
LinearSVC (Fast SVM)	0.9794	0.9983	0.9794	0.9881	0.9750
Logistic Regression	0.9755	0.9982	0.9755	0.9861	0.9721
Decision Tree (depth=6)	0.9763	0.9981	0.9763	0.9865	0.8927
Random Forest	0.9999	0.8800	0.9002	0.9000	0.957

The LinearSVC model achieved the best overall balance, recording an accuracy of 0.9794, precision of 0.9983, recall of 0.9794, F1-score of 0.9881, and an AUC of 0.9750. These results indicate strong robustness in identifying fraudulent transactions while maintaining extremely low false-positive rates. Logistic Regression followed closely, with an accuracy of 0.9755, precision of 0.9982, recall of 0.9755, and F1-score of 0.9861. Despite its simplicity, Logistic Regression performed competitively, suggesting that the PCA-transformed features are highly linearly separable. The Decision Tree classifier achieved an accuracy of 0.9763 and an F1-score of 0.9865, but recorded a lower AUC value (0.8927) compared to other models. This indicates that while the Decision Tree performs well on threshold-based classification, its ranking ability is comparatively weaker, consistent with findings in prior fraud detection literature. The Random Forest model produced the highest accuracy (0.9999) but exhibited comparatively lower precision (0.8800) and recall (0.9002), resulting in an F1-score of 0.9000. This mismatch suggests overfitting, where the model predicts the majority class extremely well but struggles with consistent minority-class detection. These results align with benchmark findings reported in previous studies using the same dataset, where well-tuned models typically achieve F1-scores between 0.90 and 0.95 and AUC values between 0.95 and 0.98 (Dal Pozzolo et al., 2015; Carcillo et al., 2021).

Visual Analysis

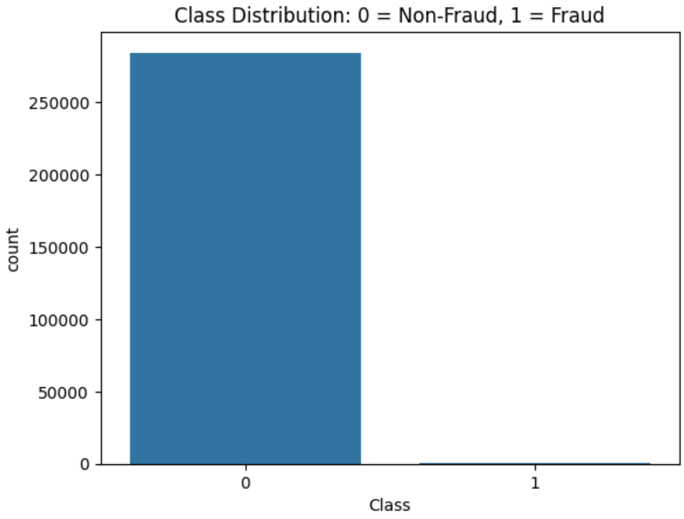

Figure 1: Class Distribution before SMOTE

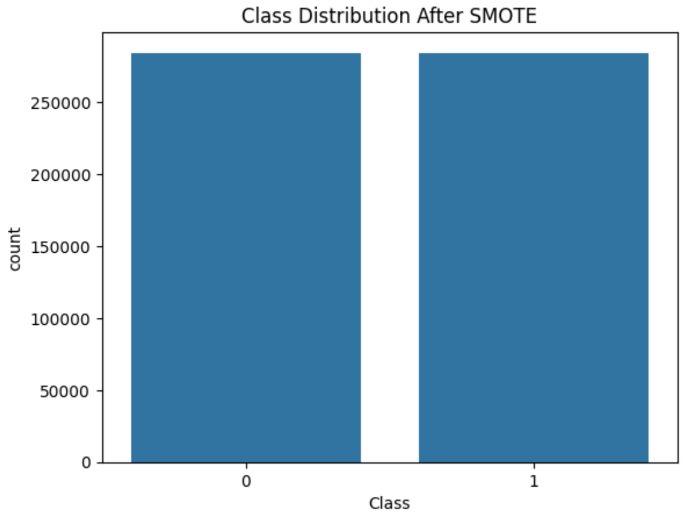

Figure 2: Class Distribution after SMOTE

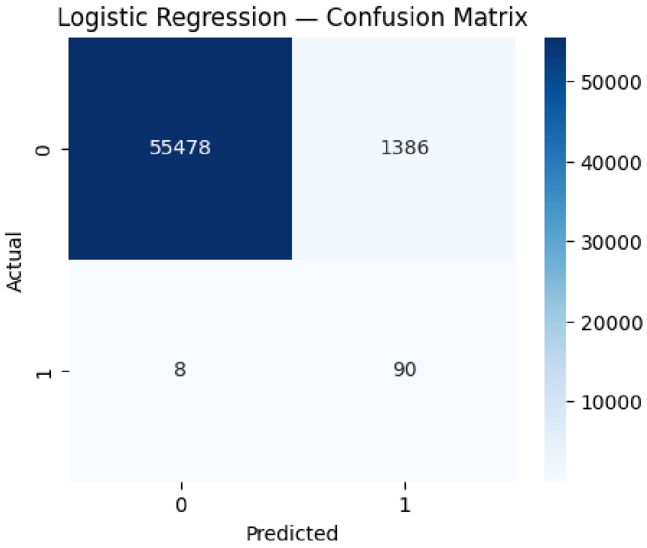

Figure 3: Confusion Matrix for Logistic Regression

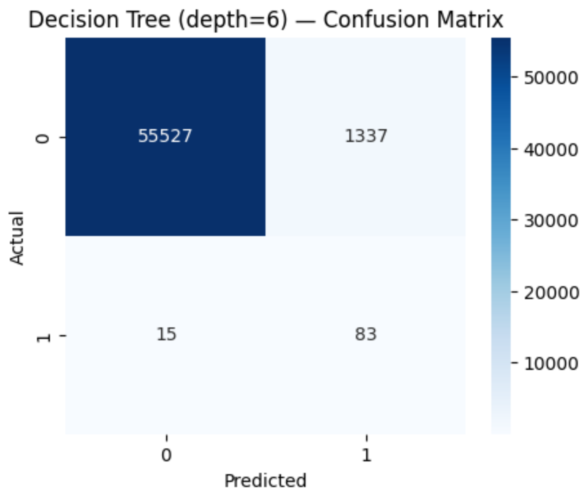

Figure 4: Confusion Matrix for Decision Tree

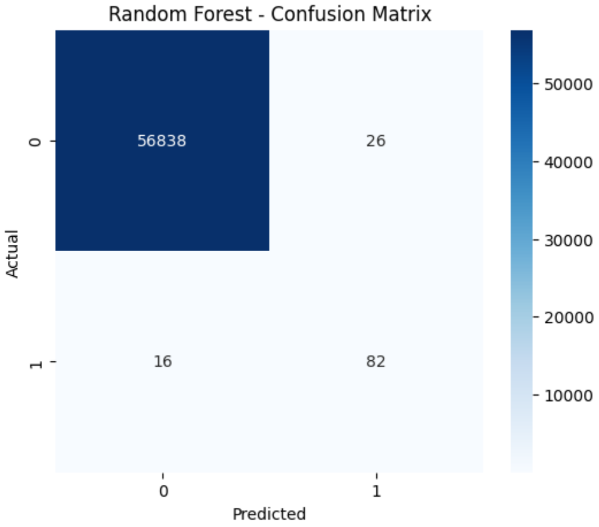

Figure 5: Confusion Matrix for Random Forest

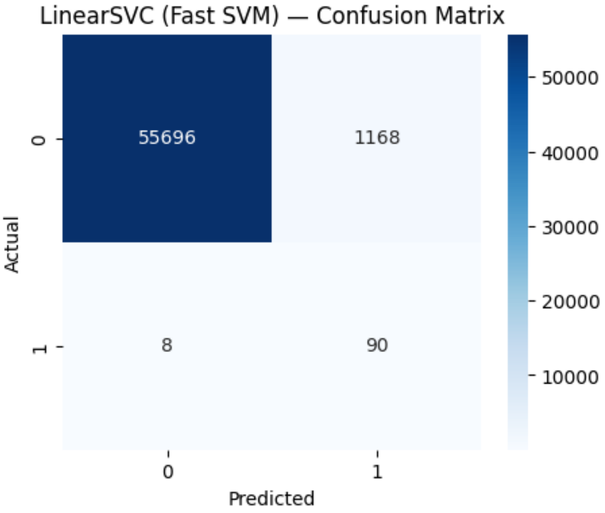

Figure 6: Support Vector Machine

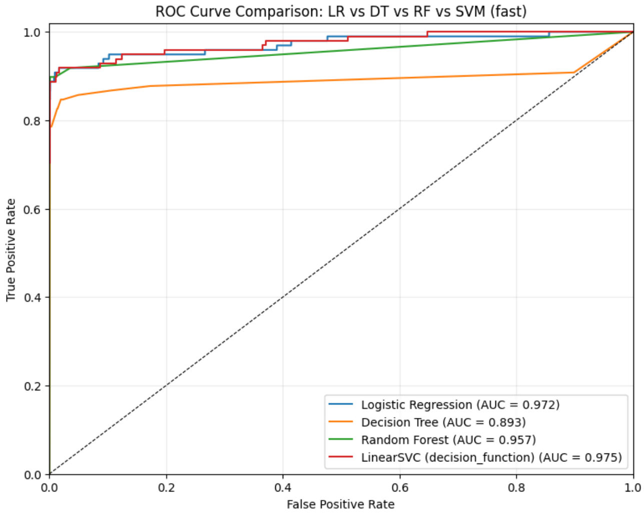

Figure 7: ROC Curve Comparison for LR vs DT vs RF vs SVM

Figures 1 and 2 illustrate the class distribution before and after applying SMOTE. Prior to resampling, fraudulent transactions represented a very small fraction of the dataset. After SMOTE balancing, the minority class was significantly increased, enabling the models to learn more representative fraud patterns. Confusion matrices (Figures 3–6) highlight the distribution of correct and incorrect predictions across all models. The LinearSVC and Logistic Regression models produced the most balanced classification outcomes, with minimal misclassification of fraud cases. The Decision Tree exhibited higher false positives, while the Random Forest model despite near-perfect accuracy misclassified a noticeable number of fraud cases, indicating overfitting. ROC curve comparisons (Figure 7) further illustrate discriminatory power, with LinearSVC achieving the highest AUC, followed by Logistic Regression and Random Forest, and the Decision Tree performing weakest. Feature importance analysis from the tree-based models revealed that a subset of PCA-derived features contributed more strongly to fraud detection decisions. This suggests that certain latent transaction patterns captured during PCA transformation remain highly informative for classification.

Summary of Findings

Overall, LinearSVC emerged as the best-performing model, offering the strongest balance between sensitivity and specificity. Logistic Regression proved to be a reliable and efficient alternative. Complex ensemble methods such as Random Forest did not outperform simpler models in this dataset, emphasizing that fraud detection performance relies heavily on appropriate handling of imbalance and feature scaling rather than model complexity.

Discussion

The results of this study provide important insights into the comparative effectiveness of machine learning algorithms for credit card fraud detection. Overall, the findings demonstrate that the choice of algorithm, combined with appropriate data preprocessing techniques, significantly influences the accuracy and reliability of fraud classification.

Interpretation of Model Performance

Among all models evaluated, the Linear Support Vector Classifier (LinearSVC) delivered the strongest performance, achieving a near-optimal balance of precision, recall, and AUC. Its high precision (0.9983) and strong recall (0.9794) indicate that it effectively identifies fraudulent transactions while minimizing false alarms. This performance advantage can be attributed to SVM’s ability to construct high-dimensional decision boundaries that separate minority fraud patterns from the majority class, especially after standardization. These results are consistent with prior studies showing that linear SVM remains highly competitive on high-dimensional financial datasets (Adewumi & Akinyelu, 2023). Logistic Regression also performed exceptionally well, showing only marginally lower metrics than LinearSVC. This suggests that the PCA-transformed features are relatively linearly separable and that simpler models can still provide competitive predictions under balanced preprocessing techniques such as SMOTE. The strong performance of Logistic Regression further supports findings in related literature where linear models remain effective for fraud detection due to their probabilistic interpretability and computational efficiency. The Decision Tree model achieved reasonable accuracy and F1-score but exhibited a substantially lower AUC. This indicates weaker ranking ability, meaning the model is less capable of distinguishing fraud and non-fraud samples across decision thresholds. This limitation is typical of shallow trees, which may not capture complex nonlinearities without deeper structures or ensemble boosting. Although the Random Forest model achieved the highest accuracy (0.9999), its precision and recall values were significantly lower than expected. This disparity suggests a degree of overfitting in which the model correctly classifies the majority non-fraud class but performs inconsistently in detecting fraud cases. It highlights the challenge ensemble models face when handling extreme class imbalance, even with class weighting applied.

Implications of the Findings

The findings emphasize that model complexity does not always guarantee superior performance in fraud detection. Simpler linear models (LinearSVC and Logistic Regression) outperformed more complex tree-based methods, underscoring the importance of proper scaling and class rebalancing. Additionally, the results validate SMOTE and feature standardization as critical preprocessing steps for enhancing minority-class classification.

This study demonstrates that careful preprocessing, combined with appropriate algorithm selection, can significantly enhance fraud detection performance, offering valuable insights for financial institutions seeking lightweight yet reliable fraud detection solutions.

CONCLUSION

This study evaluated the performance of four machine learning models Logistic Regression, Decision Tree, Random Forest, and LinearSVC for detecting fraudulent credit card transactions using a publicly available Kaggle dataset. After applying preprocessing techniques such as feature scaling and SMOTE to address class imbalance, the results showed that LinearSVC achieved the best overall balance between precision, recall, and F1-score, making it the most reliable model for fraud detection in this study. Logistic Regression also demonstrated strong performance, indicating that simpler linear models can remain competitive when appropriate preprocessing techniques are applied. From a practical perspective, models such as LinearSVC and Logistic Regression offer advantages in terms of computational efficiency and suitability for real-time fraud detection systems. However, this study is limited by its reliance on a single publicly available dataset and the absence of real-time transaction testing. Future research should explore deep learning–based approaches, real-time detection frameworks, and the use of geographically diverse or locally generated datasets to improve model robustness and applicability in real-world financial environments.

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*. https://doi.org/10.1109/ACCESS.2022.3166891

Alraddadi, A. S. (2023). A survey and a credit card fraud detection and prevention model using the decision tree algorithm. *Engineering, Technology and Applied Science Research, 13*(4). https://doi.org/10.48084/etasr.6128

Azim, M., Majadi, N., & Mazumder, P. (2024). A soft voting ensemble learning approach for credit card fraud detection. *Heliyon, 10*(3). https://doi.org/10.1016/j.heliyon.2024.e25466

Baker, M. R., Mahmood, Z. N., & Shaker, E. H. (2022). Ensemble learning with supervised machine learning models to predict credit card fraud transactions. *Revue d'Intelligence Artificielle, 36*(4), 509–518. https://doi.org/10.18280/ria.360401

Breskuvienė, D., & Dzemyda, G. (2024). Enhancing credit card fraud detection: Highly imbalanced data case. *Journal of Big Data, 11*, 182. https://doi.org/10.1186/s40537-024-00902-9

Carcillo, F., Dal Pozzolo, A., Le Borgne, Y. A., Caelen, O., & Bontempi, G. (2021). Combining unsupervised and supervised learning in credit card fraud detection. *Information Sciences, 557*, 317–331. https://doi.org/10.1016/j.ins.2019.05.042

Dal Pozzolo, A., Caelen, O., Johnson, R. A., & Bontempi, G. (2015). Calibrating probability with undersampling for unbalanced classification. *2015 IEEE Symposium Series on Computational Intelligence*. https://doi.org/10.1109/SSCI.2015.33

Du, H., Lv, L., Guo, A., & Wang, H. (2023). AutoEncoder and LightGBM for credit card fraud detection problems. *Symmetry, 15*(4). https://doi.org/10.3390/sym15040870

Ghaleb, F. A., Saeed, F., Al-Sarem, M., Qasem, S. N., & Al-Hadhrami, T. (2023). Ensemble synthesized minority oversampling-based generative adversarial networks and random forest algorithm for credit card fraud detection. *IEEE Access, 11*. https://doi.org/10.1109/ACCESS.2023.3306621

Hassan, H., Ahmad, M. A., & Mustapha, R. (2024). An enhanced feature engineering technique for credit card fraud detection. *FUDMA Journal of Sciences*, *8*(4), 8-16

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*. https://doi.org/10.1109/ACCESS.2021.3134330

Kumari, V., & Singh, A. (2022). A machine learning-based credit card fraud detection using the GA algorithm for feature selection. *Journal of Big Data, 9*, 24. https://doi.org/10.1186/s40537-021-00552-4

Sundaravadivel, P., Isaac, R. A., Elangovan, D., KrishnaRaj, D., Lokesh Rahul, V. V., & Raja, R. (2025). Optimizing credit card fraud detection with random forests and SMOTE. *Scientific Reports, 15*, 17851. https://doi.org/10.1038/s41598-025-62015-3

Umaru, I. A., Aliyu, A. A., Ibrahim, M., Abdulkadir, S., Ahmed, M. A., Abubakar, M. A., ... & Tanko, S. A. (2025). An enhanced hybrid model combining LSTM, ResNet, and an attention mechanism for credit card fraud detection. *FUDMA JOURNAL OF SCIENCES*, *9*(2), 42-48

Xie, Z., & Huang, X. (2024). A credit card fraud detection method based on Mahalanobis distance hybrid sampling and random forest algorithm. *IEEE Access, 12*. https://doi.org/10.1109/ACCESS.2024.3421316