FUDMA Journal of Sciences (FJS)
ISSN online: 2616-1370
ISSN print: 2645 - 2944
Vol. 10 No. 14, July, 2026, pp 210 – 217
DOI: https://doi.org/10.33003/fjs-2026-1014-5398

Mitigating Financial Fraud: A Hybrid SMOTE-Tomek and Stacked Ensemble Model Approach

Uduh Israel Akakoh and *Glory Nosawaru Edegbe

Department of Computer Science, Edo State University, Iyamho, Edo State, Nigeria.

*Corresponding authors' email: edegbe.glory@edouniversity.edu.ng

Received: 29 Jun. 2026
Accepted: 07 Aug. 2026
Published: 21 Aug. 2026

ABSTRACT
Credit card fraud is a menace to financial institutions, but detection is compromised by highly imbalanced transaction datasets. This study proposes an advanced machine learning framework optimized for fraud detection. To address the issue of data imbalance, SMOTE-Tomek Links is applied to synthetically generate minority fraud cases while removing noisy, overlapping majority-class instances. Recursive Feature Elimination (RFE) is deployed to identify the optimal features, and RandomizedSearchCV automates hyperparameter optimization. The study introduces a Stacked Logistic Regression ensemble to combine the predictive capacity of optimized Random Forest and XGBoost base classifiers. The model's effectiveness is assessed using seven evaluation methods: accuracy, recall, precision, confusion matrix, F1-score, Receiver Operating Characteristic Area Under the Curve (ROC-AUC) score and the Area Under the Precision-Recall Curve (AUC-PR) score. Findings reveal that the proposed stacked model performance surpasses both individual base models. While achieving deceptively high baseline accuracy across all models, the stacked ensemble delivers a superior AUC-PR score of 0.8207 and an F1-score of 0.93. This minimizes the confusion matrix misclassifications to just 20 False Negatives and 5 False Positives. The framework provides a cost-optimized operational engine that aggressively mitigates bank fraud losses while successfully shielding legitimate cardholders from accidental checkout declines.

Keywords: Credit Card Fraud; Imbalanced Data; SMOTE-Tomek Links; Recursive Feature Elimination; Stacked Ensemble Learning; AUC-PR

INTRODUCTION

Global credit card users grew from 1.1 billion in 2018 to 1.25billion in 2023, which represents a 2.79% growth, thereby driving an increase in fraud cases (Clearly Payments, 2024). Fawaz et al. (2022) and Asha and Suresh (2021) categorize major fraud types into four primary categories. The most common is Card-Not-Present (CNP) fraud.It involves data theft of credentials like the PAN and CVV, which occur through phishing or breaches. It accounts for 93% of US fraudulent charges and exceeds $6.2 billion annually (Cruz, 2025). Other types include card-present fraud, where criminals use stolen or cloned cards to make payments; application fraud, where fraudsters use a victim's personal data for fraudulent account creation; and friendly fraud, where legitimate cardholders falsely dispute authorized purchases with their financial institutions after receiving the goods or services.

Contemporary fraud prevention frameworks must look beyond traditional transactional monitoring and account for the entry channels through which financial data is initially compromised. Onyibe et al. (2024) established that identity theft and malware attacks remain prevalent cybercrimes that exploit individual user vulnerabilities. Within the domain of financial systems, such entry channels act as a primary gateway to credit card fraud, enabling fraudsters with the stolen personally identifiable information (PII) required to initiate unauthorized card-not-present (CNP) transactions or application fraud. While individual-based mitigation strategies such as multi-factor authentication (MFA) and password managers are vital to secure digital accounts and limit credential breaches, robust backend machine learning pipelines are simultaneously required to detect fraudulent patterns once those mitigating strategies are bypassed.

According to Castillo (2024), approximately 60% of credit card holders in 2023 were faced with some kind of attempted fraud. Nilson report, a payment industry research firm had stated that fraud losses globally linked to cards grew to $33.45 billion dollars in 2022 and went further to forecast this loss getting to $425.32billion by 2032 (Mullen, 2021). These reports motivated the reason for this work, which is to identify credit card fraud transactions using the proposed framework in this study.

This study attempts to improve existing systems by using the 2013 European credit card dataset. Traditional methods are replaced with a hybrid data-balancing technique applied strictly to the training data to avoid data leakage. Furthermore, it enhances model performance by using a stacking ensemble that combines bagging and boosting algorithms as base models, while optimizing parameters via RandomizedSearchCV instead of the more computationally expensive GridSearchCV.

The study seeks to develop a credit card fraud detection model using a hybrid approach in balancing the dataset and building the model.

The objectives of the study are to:

i. Develop an enhanced credit card fraud detection model by taking advantage of the strength of stacking.
ii. Compare the classification of the SMOTE-Tomek balanced dataset against the SMOTE balanced dataset using selected machine learning algorithms on the 2013 European credit card dataset.
iii. Compare the predictive performance of selected machine learning algorithms tuned using RandomizedSearchCV against those tuned using GridSearchCV.

iv. Review the performance improvement of the stacked model over the optimized base learners using seven evaluation methods.

Credit card fraud causes severe financial loss and poor credit ratings for cardholders. It also reduces institutional income and potential job losses (Balogun et al., 2024). Faced with an estimated future global loss of $425.32 billion and advanced technology threats, according to Mullen (2021), this study introduces a hybrid machine learning approach for dataset balancing and fraud detection. The results aim to enrich academic research and help the financial sector improve credit card transaction security.

Credit card fraud detection utilizes various pattern-matching and artificial intelligence techniques traditionally grouped into rule-based, statistical, and machine learning methods (Bakhtiari et al., 2023; Ahmad, 2024; HaiChao et al., 2024; Patel, 2023). This work focuses specifically on leveraging a machine learning approach to exploit its strengths in identifying fraudulent transactions.

Igor et al. (2021), Mienye and Jere (2024), Naresh et al. (2020), AlsharifHassan and Huthaifa (2023), Singh et al. (2022), Almazroi and Ayub (2023) identified some major challenges in credit card fraud detection. The challenges identified include data availability, feature engineering, scalability, unbalanced datasets, which cause bias towards majority samples, concept drift, performance measures and model algorithm selection (Fatokun et al., 2025). The team also proposed various techniques for handling these challenges.

The importance of data preprocessing cannot be overemphasized, as it improves the quality of the data by reducing noise, addressing redundancy in the dataset, and eliminating unnecessary data (Mujahid et al.,2024; Yang, 2024). Data preprocessing methods are classified under data cleaning, data transformation, and data reduction (Celtin & Yildiz, 2022).

Introduced by Wolpert (1992), stacking minimizes generalization errors caused by inherent algorithmic bias. This multi-layered architecture trains diverse base models (Level 0) on the initial dataset and then uses a meta-model (Level 1) trained on their predictions to learn, correct, and optimize final outputs.

Introduced by Chawla et al. (2002), SMOTE (Synthetic Minority Oversampling Technique) fixes model overfitting caused by imbalanced datasets. In comparison to traditional sampling methods that create narrow decision boundaries, SMOTE creates synthetic minority samples along line segments joining its k-nearest neighbors. The study concluded that combining SMOTE with majority class undersampling delivers the best performance.

Batista et al. (2004) showed that class overlapping rather than class imbalance alone negatively affects model performance. Their study established that oversampling outperforms undersampling under the ROC curve, with simple random oversampling exceeding complex techniques despite increasing model complexity. To resolve class overlapping, they proposed hybrid data-cleaning pipelines, specifically SMOTE-ENN and SMOTE + Tomek Links, which remove boundary noise and oversample the minority class.

Class imbalance degrades classifier performance, yet standalone oversampling causes overfitting and overlapping, while undersampling introduces noise and removes critical features. Consequently, a hybrid resampling approach provides an optimal solution to balance data and maximize model performance (Alamri & Ykhlef, 2024; Abdul Salam et al., 2024; Milli et al., 2024).

Pes (2021) investigated the best approach for handling a dataset with many features and highly imbalanced. The study investigates the effectiveness of combining feature selection and imbalance learning, using random forest as the machine learning model. The study showed that combining both techniques in building the hybrid pipeline was more efficient than using each technique alone.

Bahl et al. (2019) used RFE with random forest as a classifier in their study to systematically clean the model. The study was able to show that combining RFE with random forest helps improve model performance by the removal of unrelated features, as the output of a random forest model can be affected negatively if too many unrelated features are present. The study also shows that RFE is useful in mitigating issues caused by highly correlated variables.

Hyperparameter tuning is the process of identifying the optimum value of hyperparameters needed to optimize the functionality of the model (Sumaya et al., 2023). Applying hyperparameter tuning to a model increases the detection accuracy and reduces the risk of overfitting (Shantanu, 2024; Yang, 2024).

Bergstra and Bengio (2012) demonstrated that random search is more efficient than grid search for hyperparameter optimization. By navigating larger configuration spaces, random search identifies superior neural networks and deep belief network models using less computation time.

Also, random search provides operational advantages, including fault tolerance and the ability for asynchronous execution.

Bharti et al. (2024) developed a credit card fraud detection model using an imbalanced German credit card dataset. Their methodology combined SMOTE and TomekLinks for class balancing, optimized three base models (Random Forest, LDA, and Ridge) through multiple tuning techniques, and aggregated them using voting and stacking ensembles. Random Forest had the best individual F1-score (0.85) and the top stacking score (0.833). However, the study is limited by a small sample size (1,000 transactions), affecting generalizability, high computational costs, and an unspecified meta-learner.

Nurafni and Chuan-Ming (2025) proposed a stacking ensemble model combining SMOTE-ENN with Random Forest and Decision Tree base learners, with Random Forest being the meta-learner to enhance variance reduction, stability, and generalizability on the synthetic Paysim credit card dataset. The architecture achieved robust performance metrics, including a 1.00 ROC-AUC and a 0.92 F1-score. However, the model faced limitations such as high computational costs, potential real-world generalizability issues inherent to synthetic data, and a risk of data leakage indicated by the perfect ROC-AUC score.

Balogun et al. (2024) enhanced fraud detection on the dataset used in this study, using exploratory data analysis and SMOTE resampling across five machine learning algorithms. The analysis showed a severe class imbalance, identified the 'Amount' and specific PCA features as highly relevant, and deemed the 'Time' feature as not influential. Random Forest performed best both with SMOTE (0.85 F1-score) and without it (0.866 F1-score). Drawback of this study includes high possibility of noise introduction from relying solely on SMOTE and high computational costs from using GridSearchCV for hyperparameter tuning.

Emmanuel et al. (2022) developed a fraud detection model using a Genetic Algorithm with a Random Forest fitness function for feature selection, combined with SMOTE for dataset balancing. Evaluating using the dataset used in this study, Random Forest achieved the highest accuracy score of

0.9998, which was further validated on a synthetic dataset. Critical limitations of this study include relying solely on SMOTE for data balancing and using accuracy as the primary evaluation metric, which can mask poor minority class performance in highly skewed datasets.

Pratyush et al. (2021) evaluated SVM, Random Forest, ANN and Logistic Regression classifiers on the dataset used in this study, using SMOTE for data balancing. Recognizing the limitations of accuracy for skewed data, recall, precision, and F1-score was used, with ANN achieving the best F1-score of 0.91. A major drawback of this study is its sole reliance on SMOTE, which risks introducing noise into the dataset.

Meera (2022) applied the CRISP-DM framework to evaluate the performance of four machine learning algorithms on the dataset used in this study. Support Vector Machine performed best with 0.9994 accuracy, 0.711 AUC-ROC, and 51 misclassifications. However, the study is critically limited by leaving out data balancing techniques and relying heavily on accuracy as a primary evaluation metric for a highly imbalanced dataset.

Noor et al. (2022) bypassed data preprocessing to evaluate nine machine learning algorithms across nineteen resampling techniques on the dataset used in this study, using stratified cross-validation. AllKNN undersampling combined with CatBoost performed best with a 0.8740 F1-score and an AUC score of 0.9794. Critical limitations of the study include not paying attention to duplicate values during data preprocessing, omitting hyperparameter tuning, and relying solely on heatmap for feature selection.

MATERIALS AND METHODS

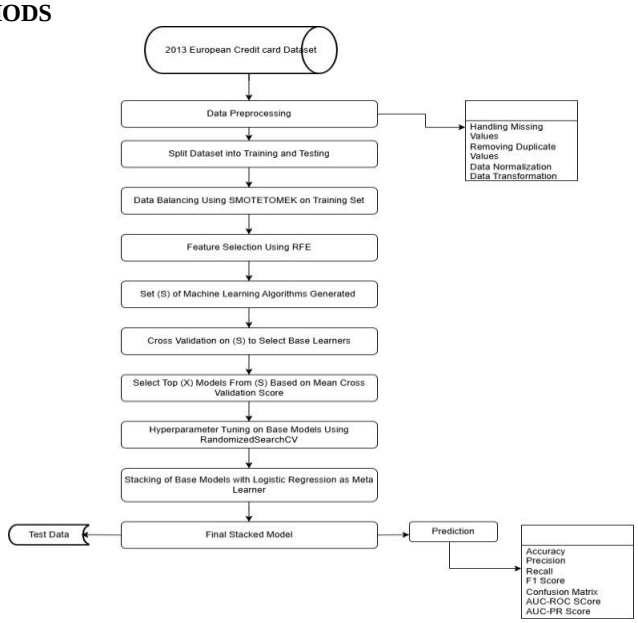

Figure 1: Architecture of Proposed Stacked Ensemble Model

The proposed machine learning methodology is designed to handle the highly imbalanced classification tasks that are evaluated using the selected dataset as illustrated in Figure 1. The proposed methodology incorporates data preprocessing, class balancing, feature selection, base model exploration, hyperparameter optimization, and ensemble stacking to maximize predictive accuracy and robustness.

Data Source and Preprocessing

The framework initiates with the chosen dataset, a benchmark dataset widely recognized for its severe class imbalance (fraudulent vs. legitimate transactions).

Data Preprocessing

Initial preprocessing is applied to the raw data. This step encompasses feature scaling, handling any missing or anomalous entries to ensure numerical stability for downstream estimators, and removing duplicate values.

Data Partitioning and Class Balancing

To ensure strict validation and prevent data leakage, data partitioning is executed before any synthetic oversampling.

Train and Test Split

The preprocessed dataset is stratified and divided into independent training and testing sets in the ratio of (80:20), respectively. The test partition is completely isolated and reserved exclusively for final model evaluation.

Class Balancing (SMOTE-TOMEK): Credit card fraud datasets are usually severely lopsided. To reduce this, SMOTE-TOMEK, a hybrid technique, is applied only to the training set. SMOTE synthesizes new minority class samples to increase sample proportion, while Tomek Links identify and remove noisy, overlapping data pairs along the class boundary. This creates a balanced, cleanly defined decision boundary.

Feature Selection

Recursive Feature Elimination (RFE)

To improve computational efficiency and eliminate redundant features, RFE is utilized. By recursively training a baseline estimator and removing the least significant features based on feature weightings/coefficients, RFE extracts the optimal features. This reduces the risk of overfitting and improves model generalizability.

Base Model Exploration and Selection

Instead of relying on a single classifier, this methodology uses an experimental selection pool strategy.

Algorithm Pool Generation

A comprehensive set (S) consisting of different machine learning algorithms, which includes Logistic regression, XGBoost, Random Forest, Support Vector Classifier, and Gradient boosting, is generated.

Cross-Validation Evaluation
A stratified K-fold cross-validation is implemented across all models in set (S) using the balanced training data.

Base Model Selection
The models are ranked according to their mean cross-validation scores using F1-score as the evaluation metric. The top (X) performing models are selected from pool (S) to serve as the base estimators for the final ensemble, and the least performing model is selected as the meta-learner.

Hyperparameter Optimization
RandomizedSearchCV
To enhance the predictive capacity of the selected top (X) base models, hyperparameter tuning is carried out using randomized search cross-validation. This approach picks a set number of parameter values from a defined distribution space, which is a computationally efficient alternative to exhaustive grid searches while securing near-optimal hyperparameters.

Ensemble Stacking and Final Prediction
The final phase leverages stacking to integrate the strengths of the individual tuned classifiers.
Stacked Generalization
The optimized top (X) base models are aggregated into a heterogeneous stacking architecture. The base models generate out-of-fold predictions during training.

Meta-Model Aggregation
Logistic Regression classifier is deployed as the meta-learner. The meta-learner takes the prediction probabilities/outputs of the base models as its Meta features and learns the optimal weights required to generate the final prediction.

Inference
The Final Stacked Model evaluates the previously isolated Test Data to produce the definitive fraud predictions, ensuring a realistic assessment of the architecture's real-world deployability.
Obtained from Kaggle, the selected dataset consists of 284,807 transactions that cut across a 48-hour span in September. The dataset is highly skewed with a fraud-to-non-fraud ratio of 1:579, and its features have been modified using PCA to reduce dimensionality and secure cardholder privacy. The dataset has 31 features: 28 anonymized features (V1 to V28) obtained through PCA, and three raw variables: Time (duration in seconds from the first transaction), Amount (payment value), and Class (binary target variable, where 0 represents genuine, and 1 represents fraudulent transactions). The dataset is severely imbalanced, comprising 283,253 legitimate records and only 473 fraud cases.
To evaluate the connection between the features in the dataset and the target variable, this study will apply a heatmap.
In Figure 2, V7, V10, V12, V14, V16, V17 shows negative correlation. V2, V4, and V11 show a positive correlation with the Class feature. The correlation among the V1– V28 features show zero correlation, which is expected for PCA-transformed features. This shows that the PCA transformation was properly done. The Amount and Time features have no relationship with the target variable. However, they have a relationship with some of the PCA features.
Features such as V12, V14, and V17 that have a strong negative correlation mean that as fraud decreases, these values increase. For features like V4 and V11 with moderate positive correlation, as fraud increases, the value of these features increases.

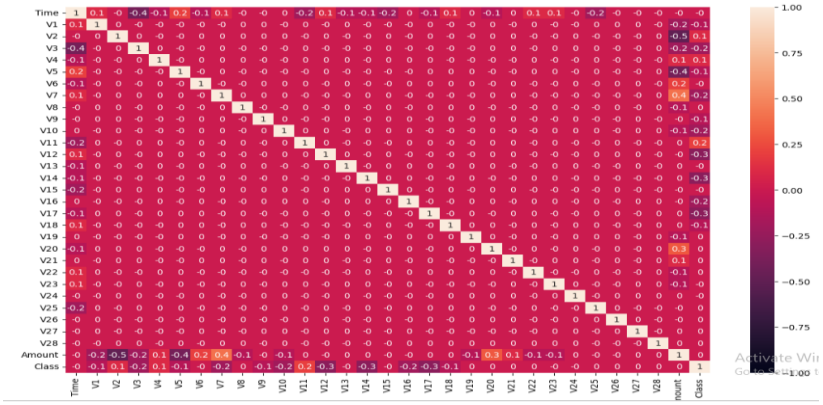

Figure 2: Heatmap Showing the Correlation between Data Features

Confusion matrix and macro average from Sklearn classification report due to the skewness of the dataset would be used to assess the efficiency of the proposed stack model. The metrics and their corresponding mathematical equations are presented below.
Accuracy: The accuracy score measures the proportion of both the fraudulent and genuine transactions that the model classifies correctly. It can be mathematically represented as

$$Accuracy\frac{=(TP+TN)}{TP+FP+TN+FN} \quad (1)$$

Precision: It is the ratio of the real fraud transactions to all the transactions reported as fraud by the model.

$$Precision=\frac{TP}{TP+FP} \quad (2)$$

Recall: It is the percentage of the real fraud transactions the model accurately classified from the total actual fraud cases.

$$Recall=\frac{TP}{TP+FN} \quad (3)$$

F1 score: It is the measure of the harmonic mean of precision and recall. It is a single score that balances the false negatives and false positives.

$$F1\text{-}Score=2\ x\frac{precisionxrecall}{precision+recall} \quad (4)$$

Following data balancing, RFE using a Random Forest estimator was applied to identify the best feature subset. Using an accuracy metric, the pipeline selected 29 optimal features (illustrated in Figure 3), which yielded the peak accuracy score of 0.9995.

```
Index (['Time', 'V1', 'V2', 'V3', 'V4', 'V5', 'V6', 'V7', 'V8', 'V9', 'V10',
       'V11', 'V12', 'V13', 'V14', 'V15', 'V16', 'V17', 'V18', 'V19', 'V20',
       'V21', 'V22', 'V24', 'V25', 'V26', 'V27', 'V28', 'Amount'],
      dtype='object')
```

Figure 3: Selected Features from Feature Selection Using RFE

To select the base models, this study used the best two average mean cross-validation scores of the machine learning models.

Table 1: Mean Cross-validation Score of Selected Machine Learning Models

Logistic Regression	XGBoost	Random Forest	Gradient Boosting	Support Vector Machine
0.9812	0.9999	0.9999	0.9899	0.9975

In Table 1, the mean cross-validation scores of the machine learning models using F1 score as the evaluation metric shows that Random Forest and XGBoost algorithms have the highest mean cross-validation score of 0.9999 each and were thereby selected as the base learners. Logistic regression is the least performing model and would be used as the meta-learner, which justifies the aim of stacking.

The selected base learners undergo hyperparameter tuning using RandomizedSearchCV and GridSearchCV. The specific hyperparameters targeted for tuning in each model are as follows.

Random Forest (n_estimators, max_depth, min_samples_split) and XGBoost (n_estimators, learning_rate, gamma, max_depth).

RESULTS AND DISCUSSION

In Table 2, the evaluation metrics of the base estimators across the selected hyperparameter tuning techniques show that RandomizedSearchCV produces slightly better results than GridSearchCV for both base models with less time. RandomizedSearchCV would be used as the hyperparameter tuning technique in building the proposed stacked model.

Table 2: Summary of Performance of Base Models across Hyperparameter Tuning Techniques

Model	Tuning technique	Accuracy	Precision	Recall	F1 Score	AUC-ROC
Random Forest	GridSearchCV	0.9995	0.9600	0.8900	0.9300	0.9423
	RandomizedSearchCV	0.9995	0.9700	0.8900	0.9300	0.9423
XGBoost	GridSearchCV	0.9994	0.9200	0.9000	0.9100	0.9637
	RandomizedSearchCV	0.9994	0.9200	0.9000	0.9100	0.9672

In Table 3, the evaluation metrics of base models balanced with SMOTE-Tomek Links, in comparison with the base models balanced using the SMOTE-only approach, with optimized parameters for both instances, show that the Random Forest algorithm balanced using SMOTE-Tomek Links performs slightly better than when it was balanced using SMOTE, with an F1 score of 0.93 as against 0.92. This implies that the model balanced with SMOTE-Tomek would have fewer misclassifications. We would now evaluate the performance of the base learners balanced with SMOTE-Tomek links against the proposed stacked model using seven methods.

Table 3: Summary of Proposed Stacked Model and Base Models

Model	Balancing Technique	Accuracy	Precision	Recall	F1 Score
XGBoost	SMOTE	0.9994	0.9200	0.9000	0.9100
Random Forest	SMOTE	0.9995	0.9600	0.8900	0.9200
XGBoost	SMOTE-Tomek Links	0.9994	0.9200	0.9000	0.9100
Random Forest	SMOTE-Tomek Links	0.9995	0.9700	0.8900	0.9300

From the results shown in Table 4, the suggested model has shown better performance over the individual models involved in developing the stacked model. This shows the model's capacity to balance between false negatives and false positives as it combines the strength of the XGBoost algorithm, which best identifies fraudulent transactions and the Random Forest algorithm, which is better at identifying legitimate transactions. While the AUC-ROC score may sometimes be misleading because of the imbalance between fraudulent and genuine transactions, the AUC-PR score of the proposed model strictly centres on the relationship between precision and recall, which reduces loss of funds and customer satisfaction, which are primary issues with false positives and false negatives.

Table 4: Summary of Performance of Proposed Stacked Model and individual Models

Model	Accuracy	Precision	Recall	F1 Score	AUC-ROC	AUC-PR
XGBoost	0.9994	0.9200	0.9000	0.9100	0.9579	0.8187
Random Forest	0.9995	0.9700	0.8900	0.9300	0.9423	0.8188
Proposed Stacked Model	0.9996	0.9700	0.8900	0.9300	0.9622	0.8207

As shown in Figures 4, 5 and 6, the proposed stacked model has 25 misclassifications, Random Forest had 26 misclassifications, and XGBoost had 34 misclassifications. However, the XGBoost algorithm has 19 false negatives as against the proposed model, which has 20 false negatives.

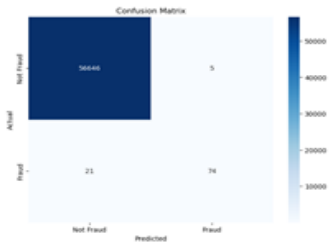

Figure 4: Confusion matrix of Random Forest

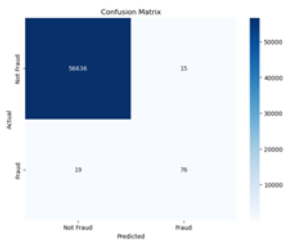

Figure 5: Confusion matrix of XGBoost

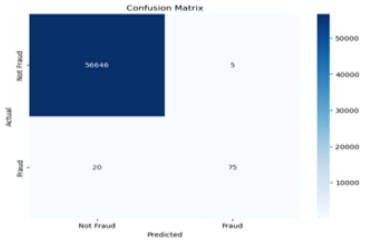

Figure 6: Confusion matrix of stacked model

The XGBoost model has shown strong competition with the proposed stacked model. However, when the models performance is compared using the three global architectural metrics (AUC-PR, AUC-ROC, F1 score), the proposed stacked model is a superior model. However, the fact that the model has 20 false negatives, as against the 19 recorded by the XGBoost model in the confusion matrix, is merely a threshold configuration issue and not a model capability issue. Future studies should run a threshold sweep on the proposed stacked model to lower its probability cutoff.

When compared to existing baseline literature, the model yields structural performance gains:

Handling Precision-Recall Trade-offs: Compared to Balogun et al. (2024) (Precision: 0.85, Recall: 0.85) and Meera (2022) (Precision: 0.93, Recall: 0.76), our framework establishes a steep increase to **0.97 Precision** while preserving **0.89 Recall**. This ensures that flagged anomalies are highly accurate alerts without missing actual fraud incidents.

The Robustness of the AUC-PR Vector: While Noor et al. (2022) achieved a higher nominal Recall (0.9591) and a slightly superior AUC-ROC (0.9794), their framework yielded a lower F1-Score (0.8740). This gap indicates a deficiency in the model. The most comparative literature Balogun et al. (2024) and Noor et al. (2022), omitted the **AUC-PR** metric entirely, relying on AUC-ROC, which is highly over-optimistic on skewed data owing to the large number of True Negatives.

State-of-the-Art Validation: Compared to Meera (2022), who did document an AUC-PR score of 0.7110, the recommended stacked model enhances this important metric to **0.8207**. This shows that the proposed model effectively separates the overlapping feature spaces of legitimate and fraudulent transactions at tighter operating thresholds.

CONCLUSION

To evaluate the proposed stacked model, its performance was benchmarked using the highly imbalanced, referenced dataset obtained from Kaggle. Due to the extreme class imbalance inherent to fraud detection, evaluation relied heavily on robust metrics like the AUC-ROC, F1-Score, and specifically AUC-PR to prevent class-imbalance distortions.

The proposed model showed strong predictive capabilities, with an accuracy of 0.9996, an F1-Score of 0.9300, and an AUC-PR score of 0.8207. Within the testing partition, the model produced only 25 total misclassifications on the confusion matrix. Minimizing these misclassifications directly reduces financial friction, as false negatives result in unrecovered stolen funds, while false positives result in unnecessary decline of transactions, leading to customer dissatisfaction.

Although the proposed framework showed superior performance in credit card fraud detection, the evaluation was limited to a single benchmark dataset. Nevertheless, future studies will focus on validating the model across additional public datasets, while conducting statistical and computational efficiency profiling to support real-world deployment.

REFERENCES

Abdul Salam, M., Fouad, K.M., Elbably, D.L., & Salah, M. E.(2024). Federated learning model for credit card fraud detection with data balancing techniques. *Neural Comput & Applic (2024)* (**36**), 6231–6256. https://doi.org/10.1007/s00521-023-09410-2

Ahmad, A. (2024). Adaptive Fraud Detection Systems: Real-Time Learning from Credit Card Transaction Data. *Advances in Computer Sciences (2024).* (7).

Alamri, M., & Ykhlef, M. (2024). Hybrid undersampling and oversampling for handling imbalanced credit card data. *IEEE Access*, *12*, 14050–14060. https://doi.org/10.1109/ACCESS.2024.3357091

Almazroi, A. A., & Ayub, N. (2023). Online payment fraud detection model using machine learning techniques. *IEEE Access*, *11*, 137188–137203. https://doi.org/10.1109/ACCESS.2023.3339226

AlsharifHassan, M. A., & Huthaifa, I. A. (2023). Credit Card Fraud Detection Using Enhanced Random Forest Classifier for Imbalanced Data. *International Conference on Advances in Computing Research (2023).* (700), 605-616.

Asha, R. B., & Suresh, K.(2021). Credit Card Fraud Detection Using Artificial Neural Network. *Global Transitions Proceedings (2021)*, 2(1), 35-41. https://doi.org/10.1016/j.gltp.2021.01.006

Balogun, O. O., Kupolosi, J. A., & Akomolafe, A. A.(2024). Credit Card Fraud Detection Using Machine Learning Algorithms. *British Journal of Computing, Networking and Information Technology 7(3) 1-35.* https://DOI.10.52589/BJCNIT-YDIJNXG2

Batista, G. E. A. P. A., Prati, R. C., & Monard, M. C. (2004). A study of the behavior of several methods for balancing machine learning training data. *ACM SIGKDD Explorations Newsletter*, *6*(1), 20–29. https://doi.org/10.1145/1007730.1007735

Bahl, A., Hellack, B., Balas, M., Dinischiotu, A., Wiemann, M., Brinkmann, J., Luch, A., Renard, B. Y., & Haase, A. (2019). Recursive feature elimination in random forest classification supports nanomaterial grouping. *NanoImpact*, *15*, Article 100179. https://doi.org/10.1016/j.impact.2019.100179

Bakhtiari, S., Nasiri, Z. & Vahidi, J. (2023). Credit card fraud detection using ensemble data mining methods. *Multimed Tools Appl (2023)* (**82)**, 29057–29075. https://doi.org/10.1007/s11042-023-14698-2

Bergstra, J., & Bengio, Y. (2012). Random search for hyper-parameter optimization. *Journal of Machine Learning Research*, *13*(2), 281–305. https://www.jmlr.org/papers/volume13/bergstra12a/bergstra12a.pdf

Bharti, C., Preeti, G., &Karnika, D. (2024). A Comprehensive Ensemble Approach Using Blending and Stacking for Credit Card Fraud Detection. International Journal on Smart & Sustainable Intelligent Computing (2024). 1(2) 19-33.

Castillo, M. (2024), Why Credit Card Fraud Alerts are Rising and how Worried You Should Be About Them. *CNBC.* https://www.cnbc.com/2024/09/12/why-credit-card-fraud-alerts-are-rising.html/

Celtin, V., & Yildiz, O. (2022). A comprehensive review on data preprocessing techniques in data analysis. *Pamukkale University Journal of Engineering Sciences*, *28*(2). https://doi.org/10.5505/pajes.2021.62687

Chawla, N. V., Bowyer, K. W., Hall, L. O., & Kegelmeyer, W. P. (2002). SMOTE: Synthetic Minority Over-sampling Technique. *Journal of Artificial Intelligence Research*, *16*, 321–357. https://doi.org/10.1613/jair.953

Clearly Payments (2024).*Statistics for Cash and Credit Card Use for Payments in 2024.*https://www.clearlypayments.com/blog/statistics-for-cash-and-credit-card-use-for-payments-in-2024/

Cruz, B.(2025). "62 Million Americans Experienced Credit Card Fraud Last Year. 2025 Credit Card Fraud Report and Statistics," [Online] Available: https://www.security.org/digital-safety/credit-card-fraud-report/.

Emmanuel, I., Yanxia, S., & Zenghui, W.(2022). A Machine Learning Based Credit Card Fraud Detection Using GA Algorithm for Feature Selection. *Journal of Big Data (2022), 9(24).*https://doi.org/10.1186/s40537-022-00573-8

Fatokun, J. O., Sikiru, M., Balogun, F., & Okorie, D. D. (2025). Fraud detection in Nigerian investment advisory sector using machine learning algorithms. *FUDMA Journal of Sciences*, *9*(10), 5–11. https://doi.org/10.33003/fjs-2025-0910-4004

Fawaz, A., Iqra, M., Hikmat K., Naif A., Muhammed R., & Muzamil A.(2022). Credit Card Fraud Detection Using State of the Art Machine Learning and Deep Learning Algorithms.*IEEE Access*. Vol 10 (202) 39700-39715.

HaiChao, D., Li, L., Wang, H., & Guo, A.(2024). A novel method for detecting credit card fraud problems. *PLoS ONE (2024)* 19(3): e0294537. https://doi.org/10.1371/journal.pone.0294537

Igor, M., Mladen, K., Damir, P., & Ljiljana, B. (2021). Credit Card Fraud Detection in Card-Not-Present Transactions: Where to Invest. *Appl. Sci (2021),* 11(6766) https://doi.org/10.3389/app11156766

Meera, A. (2022). Credit Card Fraud Detection Using Machine Learning.Thesis. Rochester Institute of Technology. https://repository.rit.edu/theses

Mienye, I. D., & Jere, N. (2024). Deep learning for credit card fraud detection: A review of algorithms, challenges, and solutions. *IEEE Access*, Advance online publication. https://doi.org/10.1109/ACCESS.2024.3426955

Milli, M. E. F., Aras, S., & Deveci Kocakoç, İ. (2024). Investigating the Effect of Class Balancing Methods on the Performance of Machine Learning Techniques*: Credit Risk Application. İzmir Yönetim Dergisi (2024),* 5(1), 55-70. https://doi.org/10.56203/iyd.1436742

Mujahid, M., Kına, E., Rustam, F., Villar, M. G., Alvarado, E. S., De La Torre Diez, I., & Ashraf, I. (2024). Data oversampling and imbalanced datasets: An investigation of performance for machine learning and feature engineering. *Journal of Big Data*, *11*(1), Article 87. https://doi.org/10.1186/s40537-024-00943-4

Mullen, C. (December, 2021). Card Industry Faces $400B in Fraud Losses Over Next Decade, Nilson Says. *Paymentsdrive*. https://www.paymentsdive.com/news/card-industry-faces-400b-in-fraud-losses-over-next-decade-nilson-says/611521/

Naresh, K. T., Sarita, S., Umesh, K. L., & Sanjeev, K. S. (2020). An Efficient Credit Card Fraud Detection Model Based On Machine Learning Methods. *International Journal of Advanced Science and Technology,* Vol 25 (5), 3414-3424.

Noor, S. A., Suliman, M. F. (2022). Enhanced Credit Card Fraud Detection Using Machine Learning. *Electronics 2022*, 11(66). https://doi.org/10.3390/electronics11040662.

Nurafni, D., &Chuang-Ming, L. (2025). Advanced Fraud Detection: Leveraging K-SMOTEENN and Stacking Ensemble to Tackle Data Imbalance and Extract Insights. IEEE *Access (2025),* (13), 10356-10370. https://doi.org/10.1109/ACCESS.2025.3528079

Onyibe, F. A. A., Edegbe, G. N., Omaji, S., & Olayinka, A. S. (2024). Combating cybercrime perpetrated via social media channels using individual resilience techniques. *Covenant Journal of Informatics and Communication Technology*, *11*(2), 1–15. https://journals.covenantuniversity.edu.ng/index.php/cjict/article/view/4065

Patel, K. (2023). Credit card analytics: A review of fraud detection and risk assessment techniques. *International Journal of Computer Trends and Technology*, *71*(10), 69–79. https://doi.org/10.14445/22312803/IJCTT-V71I10P109

Pratyush, S., Souradeep, B., Devyanshi, T., & Jagdish Chandra, P. (2021). Machine Learning Model For Credit Card Fraud Detection - A Comparative Analysis Model Using. *International Arab Journal of Information and Technology (2021)* 18(6). 789-796. https://doi.org/10.34028/iajit/18/6/6

Singh, A., Ranjan, R. K., & Tiwari, A. (2022). Credit card fraud detection under extreme imbalanced data: A comparative study of data-level algorithms. *Journal of Experimental & Theoretical Artificial Intelligence*, *34*(4), 571–598. https://doi.org/10.1080/0952813X.2021.1907795

Shantanu, K. (2024). Enhanced Fraud Detection in Financial Transactions Using

Hyperparameter-Tuned Random Forests, 2024. *15th International Conference on Computing Communication and Networking Technologies (ICCCNT), Kamand, India (2024),* (15), 1-7, https://doi:10.1109/ICCCNT61001.2024.10725958

Sumaya, S. S., Ibraheem, N., & Sarab, M. H.(2023). Credit Card Fraud Detection Using Improved Deep Learning Models. *Computers, Materials & Continua, (2024)* 78(1). https://doi.org/10.32604/cmc.2023.046051

Wolpert, D. H. (1992). Stacked Generalization, *Neural Networks*, 5(2). 241-259. https://doi.org/10.1016/S0893-6080(05)80023-1

Yang, G. (2024). Credit Card Fraud Detection Based on Machine Learning Prediction. *Proceedings of the 2024 2nd International Conference on Image, Algorithm, and Artificial Intelligence (ICIAAI) (2024), Advances in Computer Science Research* (115), pp 35-45. https://doi.org/10.2991/978-94-6463-540-9_5