Jurnal Sains Informatika Terapan (JSIT)

E-ISSN: 2828-1659, Volume: 5, Issue: 1, Month: February, Year: 2026
Page: 1-9, Available online at: https://rcf-indonesia.org/home/

FRAUD PREDICTION IN ONLINE FINANCIAL TRANSACTIONS WITH A COMBINATION OF SMOTE AND ENSEMBLE CLASSIFIER

[1]Juvinal Ximenes Guterres, [2]Delfim Da Silva, [3]Jacinto Defatima Sales, [4]Abrao Freitas, [5]Nuno da Costa
[1,2,3,4,5]Faculty of Engineering, Universidade Oriental Timor Lorosae, UNITAL
Corresponding Author: juvinalximenes6@gmail.com

Article Info

Article history:

Received: Nov 19, 2025
Revised: Nov 28, 2025
Accepted: Des 09, 2025
Published: Feb 01, 2026

Keywords:

Fraud detection
SMOTE
Ensemble Learning
Voting Classifier
Imbalanced data
Machine Learning

ABSTRACT

Detecting fraudulent transactions remains a major challenge in digital financial systems due to the severe imbalance between legitimate and fraudulent records. This study aims to develop a classification model capable of identifying fraudulent transactions with high sensitivity to minority classes, while ensuring performance stability suitable for operational deployment. The methodology includes data preprocessing through outlier removal, feature normalization, and stratified data partitioning. To address class imbalance, the Synthetic Minority Over-sampling Technique (SMOTE) is applied to generate representative synthetic samples for the minority class. Multiple machine learning algorithms are evaluated, including Random Forest, Decision Tree, Bagging, Gradient Boosting, Logistic Regression, Neural Network, K-Nearest Neighbors, and Support Vector Machine. Model performance is assessed using Precision, Recall, F1-Score, AUC, and G-Mean. The results show that the proposed approach achieves stable and reliable performance, with an AUC of 0.89 and a G-Mean of 0.81, demonstrating its effectiveness for operational fraud detection and error minimization.

1. INTRODUCTION

The massive digital transformation occurring in the global financial sector has driven the adoption of credit cards as one of the most popular electronic payment methods. Cashless transaction systems are now the preferred choice due to their convenience, speed, and efficiency, leading to a significant increase in cross-border transaction volumes.[1]However, this growth also comes with new challenges, particularly related to data security and increasingly complex fraud risks. One major threat is the rise in credit card fraud.[2]. The latest report from The Nilson Report (2023) notes that global losses due to payment card misuse have exceeded USD 32 billion, with an annual increase trend that shows no signs of slowing.[3].

Innovations such as digital banking, e-wallets, and app-based payment systems have transformed the way consumers transact and expanded access to financial services.[4]The complexity of digital system architecture also increases vulnerability to cyber threats. Findings from the Association of Certified Fraud Examiners (ACFE) show a steady upward trend in the volume and economic impact of digital crime.[5].

Conventional rule-based detection systems and statistical approaches have limitations in recognizing increasingly complex and adaptive fraud patterns.[6]–[9].

The ever-changing nature of financial transactions, coupled with the increasing diversity of fraud methods, requires adaptive and responsive detection systems based on in-depth data analysis. In this context, machine learning (ML) technology offers a strategic solution due to its ability to process large amounts of data and identify hidden patterns often undetected by traditional approaches.[10], [11].

Various ML approaches have been applied in the development of fraud detection systems, including supervised learning, unsupervised learning, as well as ensemble and hybrid methods that combine the advantages of several algorithms.[12]–[14].

However, major issues remain, particularly related to data class imbalance. Generally, fraudulent transactions are far fewer than legitimate transactions. As a result, predictive models are often overly biased toward the majority class (legitimate transactions) and fail to detect the most important fraud indicators. [15],[16].

Ensemble learning methodologies that utilize the synergistic integration of various machine learning algorithms have been empirically proven to

significantly improve fraud detection accuracy. Recent studies have shown that hybrid ensemble architectures combining conventional machine learning techniques with deep learning paradigms yield superior performance metrics in credit card fraud detection systems.[17]–[19].

In recent years, deep learning-based approaches have begun to show great potential in addressing the complexity of fraudulent transaction patterns that are difficult to identify with conventional methods. One prominent model is Long-Short-Term Memory (LSTM), especially when combined with attention mechanisms, which can focus more on relevant data segments during the learning process. This approach has proven effective in increasing the system's sensitivity to hidden anomalies within data sets.

A recent study by Obaid et al. (2025) illustrates this by developing a fraud detection model that integrates LSTM, an attention mechanism, and the SMOTE data balancing technique. The results demonstrated high accuracy in detecting fraudulent transactions across a variety of dataset scenarios, confirming the effectiveness of this architectural combination in addressing the challenges of imbalanced data distribution and increasingly sophisticated fraud patterns.[20].

Another interesting approach proposed by Ebenezer et al. in fraud detection is to design a combined Long Short-Term Memory (LSTM) and AdaBoost architecture. Before the model training process is carried out, the data is first processed through two important stages: Synthetic Minority Oversampling Technique and Edited Nearest Neighbor (SMOTE-ENN). SMOTE plays a role in generating synthetic samples from the minority class, thereby helping to reduce the model's tendency to bias towards legitimate transactions. Meanwhile, ENN is used to filter out data containing noise around the boundaries between classes, to minimize the potential for overfitting that often occurs in learning-based models.[21].

This approach, conducted by Hanae et al. (2023), proposes a hybrid approach for fraud detection based on a combination of Random Forest and Gradient Boosting Classifier, applied to an anonymous transaction dataset using Principal Component Analysis (PCA). The methodology begins with data preprocessing through PCA to reduce dimensionality and eliminate correlations between features, which also serves to protect the privacy of the original data. Next, two ensemble algorithms are applied in parallel to build a predictive model. Random Forest handles data variation through random decision tree aggregation, while Gradient Boosting performs iterative training based on residual errors to improve accuracy.[22].

The classification model in this study was built and validated using a data partitioning scheme commonly used in machine learning practice. Performance evaluation was conducted based on four key metrics: accuracy, precision, recall, and F1 score. The test results demonstrated promising performance, with 99% accuracy, 97% recall, 91% F1 score, and 87% precision. The high recall value confirms the model's ability to effectively recognize fraudulent transactions, a crucial aspect in fraud detection systems aimed at minimizing the risk of false negatives.

Another interesting approach proposed by Ebenezer et al. in fraud detection is to design a combined Long Short-Term Memory (LSTM) and AdaBoost architecture. Before the model training process is carried out, the data is first processed through two important stages: Synthetic Minority Oversampling Technique and Edited Nearest Neighbor (SMOTE-ENN). SMOTE plays a role in generating synthetic samples from the minority class, thereby helping to reduce the model's tendency to bias towards legitimate transactions. Meanwhile, ENN is used to filter out data containing noise around the boundaries between classes, to minimize the potential for overfitting that often occurs in learning-based models.[21].

In the modeling phase, LSTM is used to capture temporal patterns in sequential transactions, such as changes in frequency and location. Meanwhile, AdaBoost acts as a booster, iteratively improving model performance by emphasizing misclassification errors, making the model more adaptable to difficult cases. Experimental evaluations show highly competitive results: 99.6% sensitivity, 99.8% specificity, and a very high F1 score, all demonstrating the system's ability to recognize fraud with exceptional accuracy. However, it should be noted that the complexity of this architecture and its high computational requirements can be a constraint for real-time implementation, especially in real-time detection systems that require fast and efficient response at scale.[21].

In the context of model efficiency for industrial-scale implementation, Salwa et al. developed an ensemble approach for credit card fraud detection that combines Light Gradient Boosting Machine (LightGBM) and Lightweight More Optimal Random Trees (LiteMORT). LightGBM, which uses a leaf-wise growth strategy, offers advantages in training speed and high accuracy, with interpretability through feature ranking. In contrast, LiteMORT uses a Gradient-based One-Side Sampling (GOSS) technique that significantly speeds up the training process on high-dimensional datasets. This combined model demonstrates competitive performance with 99.44% accuracy, 92.79% precision, 90.65% recall, and 91.67% F1 score.[23].

The most innovative contribution is offered by Ghaleb et al. through a combined approach of Ensemble SMOTE-GAN and Random Forest. Against the backdrop of extreme class imbalance (up to a ratio of 1:1000), this model integrates SMOTE to stabilize the initial distribution, followed by the

application of Generative Adversarial Networks (GAN) to generate more realistic synthetic samples. Random Forest is then used as the primary classifier within a weighted voting framework. The evaluation results show excellent performance with a false positive rate of 0%, a 3.2% increase in detection rate, and a 1.9% increase in overall performance compared to the baseline model. This approach shows great potential in improving investigation efficiency and reducing the burden of manual analysis. However, significant challenges remain, such as the instability of the GAN training process, the need for complex hyperparameter tuning, and the increased computational burden resulting from using the ensemble Random Forest on large datasets.[24].

Based on the challenges and gaps identified in the literature related to online financial transaction fraud detection, this study aims to evaluate and develop a more adaptive and reliable machine learning model in identifying anomalous patterns in imbalanced transaction data.

This study develops fraud detection in financial transactions based on ensemble learning using a Voting Classifier with a soft voting approach, which combines the advantages of four classification algorithms, namely: Random Forest (RF), XGBoost (XGB), LightGBM (LGBM), and AdaBoost (ADA). Each algorithm is selected based on its advantages. Model combination is carried out through the final prediction determined based on the average probability of the output of the four constituent models. This approach aims to integrate the advantages of each model to improve accuracy, stability, and generalization capabilities on imbalanced and complex data. Model performance evaluation includes accuracy, precision, recall, F1-score, Area Under the Curve (AUC), Matthews Correlation Coefficient (MCC), and Geometric Mean (G-Mean).

This study specifically focuses on the most relevant evaluation metrics that influence the success of minority class detection. Therefore, recall, F1 score, AUC, MCC, and G-Mean were selected as the main indicators to assess model performance. These five metrics are considered more representative in reflecting the model's ability to detect fraudulent transactions effectively, fairly, and proportionally, while also contributing significantly to the development of more adaptive and robust fraud detection systems in imbalanced data environments.

Essentially, credit card fraud aims to gain financial gain, either through illegal goods or services or the transfer of funds to accounts controlled by the perpetrator. This phenomenon has developed into a systemic problem that significantly impacts the stability and integrity of the entire global financial ecosystem.[25], [26]Common modus operandi of credit card fraud include several main scenarios: (1) use of cards whose status has been deactivated (revoked/cancelled), (2) use of cards obtained illegally (stolen/reported lost), and (3) exploitation of card data without physical possession (card-not-present fraud). Furthermore, developments in financial technology have also given rise to new variants in the form of breaching authentication credentials in digital banking services, especially through password theft on mobile banking platforms.[27],[18].Another form of fraud is identity theft, which involves applying for and obtaining credit cards in the victim's name. The primary contribution of this study is to contribute to credit card fraud detection through the following key aspects:

In practice, credit card fraud is carried out using a variety of techniques that continue to evolve with technological advances. Some common techniques used by fraudsters include:

1. Card Present Fraud This technique involves physically stealing a credit card and then using it in a transaction.
2. Credit Card Detail Copying (Card Skimming/Cloning) In this technique, credit card data is copied without the rightful owner's knowledge. Copying can be done using skimming devices secretly installed in payment machines or ATMs, allowing the perpetrator to steal card information during transactions. Illegal Additional Fees.
3. Fraud can also occur when vendors or third parties charge customers additional fees without their explicit consent. In some cases, this fraud is combined with the theft of user authentication data, such as passwords or e-banking access codes, obtained through phishing, malware, or other social engineering techniques.

The significant increase in electronic transaction volume has driven the development of more adaptive and accurate credit card fraud detection methods. Machine learning (ML), a combination of algorithms and statistical models, enables automation in recognizing anomalous patterns without explicit coding.[28],[22],[23].Over the past decade, the world of machine learning research has seen ensemble learning methods evolve into the backbone of digital fraud detection systems. Much like athletes combining multiple skills to achieve peak performance, this approach combines the strengths of multiple algorithms to produce more accurate predictions.

Recent research evidence such as the findings of the Stanford Research Team (2022) shows that this method not only improves detection accuracy, but also reduces the tendency of the model to "memorize" the training data - a classic problem often encountered in predictive modeling.[31],[32].Ensemble learning models, Random Forest algorithms, XGBoost, LightGBM, and AdaBoost are excellent at handling complex financial data and are able to uncover hidden

relationships between variables in large data sets that are difficult for conventional algorithms to do.[33].

This study processed historical financial transaction data, which faces a classic problem in fraud detection: a severe imbalance between normal and fraudulent transactions. Researchers studying machine learning on imbalanced data will apply the improved SMOTE technique. This technique not only increases the number of synthetic fraudulent transaction examples but also carefully preserves the original characteristics of the actual fraudulent data.[34], [35].

Furthermore, our methodology integrates a rigorous feature selection process aimed at: (1) eliminating redundant and noisy features, (2) reducing computational complexity, and (3) improving the model's generalization capability. The combination of ensemble learning architecture with data optimization techniques shows significant improvements in fraud detection performance metrics, especially in the recall rate for minority class identification.

The novelty of this study lies in the development of SMOVO (SMOTE-based Voting Ensemble), which integrates Random Forest, XGBoost, LightGBM, and AdaBoost through a soft voting mechanism to improve fraud detection under conditions of extreme class imbalance. Unlike previous studies that rely on a single model or computationally demanding generative approaches, SMOVO uses ANOVA F-tests for feature selection to retain the most discriminatory variables while reducing computational costs. Furthermore, the model is evaluated using a comprehensive set of metrics such as Recall, F1-score, MCC, AUC, and G-Mean, offering a fairer and more representative assessment compared to studies that focus primarily on accuracy. With its efficiency, interpretability, and practicality, SMOVO provides a scalable and industry-ready solution for real-world financial fraud detection systems.

The main objective of this study is to re-evaluate existing fraud detection approaches and develop a more robust and adaptive ensemble-based model. The main contributions offered include: (1) a critical review of the recent developments in artificial intelligence-based fraud detection systems; (2) a clustering and analysis of existing approaches based on their performance and methodology; (3) the design of a voting classifier-based detection model that integrates RF, XGB, LGBM, and ADA; and (4) an evaluation of the model's effectiveness using real data, along with the application of SMOTE to address class imbalance.

With this approach, research is expected to provide a real contribution in strengthening the financial security system based on smart technology, while also helping industry players in anticipating and addressing increasingly complex and dynamic fraud threats.

2. Methodology

This study uses a public dataset on the Kaggle platform, consisting of credit card transaction data by European cardholders in September 2013. The data covers two days of transactions, totaling 284,807, of which 492 were categorized as fraudulent. This situation demonstrates a common challenge in fraud detection with imbalanced data classes, as fraudulent transactions only comprise approximately 0.172% of the total data.[36],[37], [38]The target variable in this data set is Class, which is binary with a value of 1 indicating that the transaction is fraudulent, and a value of 0 indicating normal transaction behavior.

The data imbalance in the transaction fraud detection dataset can be clearly demonstrated through the class distribution visualization in the following Figure:

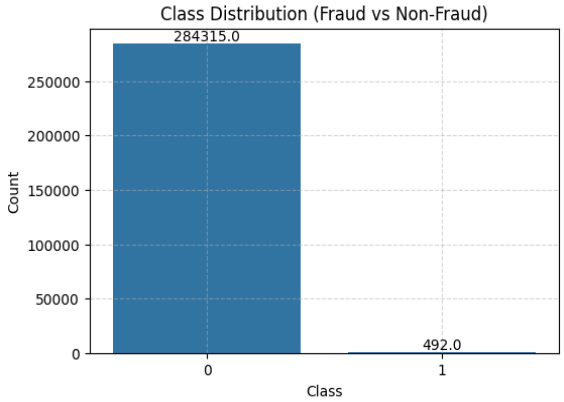


Figure 2: Histogram of Class Distribution in Dataset.

The results, visualized in Figure 1, show that the SMOTE implementation successfully balanced the class distribution, with each class containing 227,210 transactions. This balanced distribution allows the ensemble learning algorithm to perform training more proportionally, reducing excessive bias towards the majority class, and increasing the model's sensitivity in identifying fraudulent transaction patterns. Therefore, this approach is expected to reduce false negative rates, minimize potential financial losses, and strengthen public trust in the digital financial ecosystem, as illustrated in the following figure:

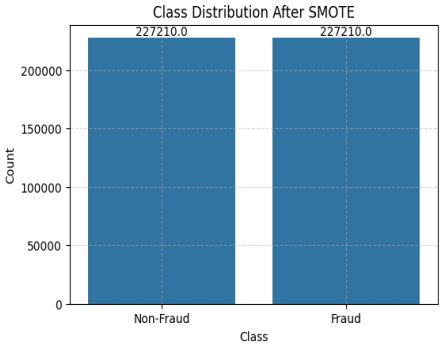


Figure 2: Histogram of Class Distribution in SMOTE

Feature selection using Analysis of Variance (ANOVA) The F-test is used to evaluate the significance of the mean difference of a numeric feature against a categorical target variable, for example between fraud and non-fraud classes. The F-score is calculated as the ratio of the between-class variance to the within-class variance. The higher the F-score, the greater the difference between the identified groups where the feature has a significant contribution to the target class distinction.

In the context of fraud detection, the ANOVA F-test plays a crucial role in the feature selection process, particularly in identifying variables statistically relevant to the presence of fraudulent activity. Features with high F-scores tend to reflect unusual transaction behavior and should therefore be considered in developing predictive models. The following illustration demonstrates the process and interpretation of this method:

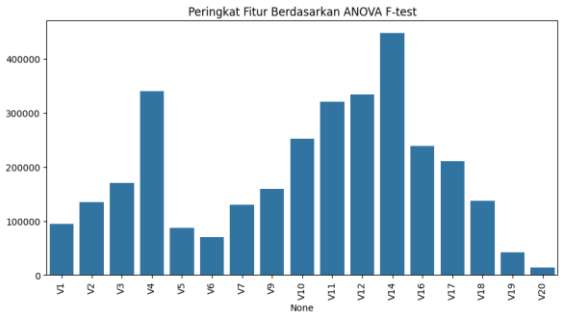


Figure 2 Anova Feature Selection

The figure above shows the top 16 features selected based on ANOVA F-test scores in the context of credit card fraud detection. These features show significant mean differences between the fraudulent and non-fraudulent classes.

High F-score values, as demonstrated by features V17, V14, and V12, indicate that these variables have strong discriminatory ability towards the target class and have great potential to improve the performance of the prediction model. Conversely, features with low scores such as V21, V6, and V2 have a smaller contribution in distinguishing between classes and can be considered for elimination during the feature selection process to simplify the model without significant loss of accuracy.

The high class imbalance factor in financial transaction data, where the number of fraud cases is much lower than normal transactions, is a major challenge in training fraud detection models. Machine learning models tend to prioritize accuracy towards the majority class, thereby risking failing to detect fraudulent transactions that are actually the most crucial to identify. Therefore, an effective data balancing approach is needed, such as the Synthetic Minority Oversampling (SMOTE) technique, as well as a classification method that is sensitive to the presence of minority classes. Based on these challenges, this study proposes a new approach called SMOVO (SMOTE-based Voting Ensemble Classifier for Fraud Detection). This approach is specifically designed to address the problem of class imbalance in the context of binary classification of financial transactions, with a primary focus on improving the detection capability of fraudulent transactions. The SMOVO model is built through five main stages that are systematically integrated with each other, as shown in the following figure:

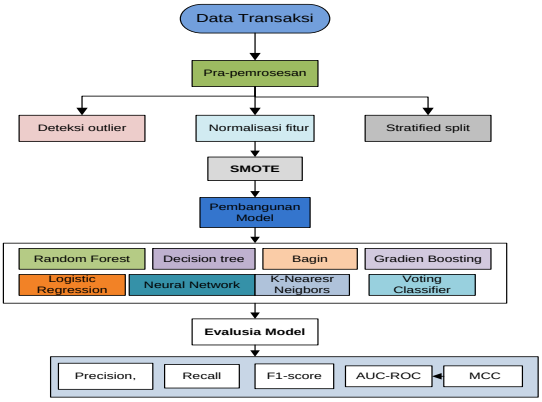


Figure 3. Proposed system (SMOVO)

This process begins with the collection of transaction data, which will serve as the primary source of analysis. This raw data is typically not ready for immediate use due to various issues, such as extreme values, uneven scale between features, and an imbalance in the number of classes between normal transactions and those suspected of fraud. Therefore, the initial step is data preprocessing.

In the pre-processing stage, the system first performs an outlier check. The goal is to identify transaction data whose values deviate significantly from the general pattern, as this type of data can lead the model to learn incorrectly. Next, feature normalization is performed, equalizing the value ranges of each feature so that the machine learning algorithm can more fairly interpret the relationships between variables. The data is then divided into training and test data using the Stratified Split method to maintain a balanced proportion of fraudulent and non-fraudulent transactions in each section.

A common problem in fraud detection is that the amount of fraud data is significantly smaller than the normal data. If left unchecked, the model will tend to treat all transactions as "non-fraudulent" because that is the dominant pattern. To avoid this, the SMOTE technique is used, a method that synthetically generates new samples specifically for minority classes, thus improving the dataset's balance without altering the existing data.

Once the data is ready, the model building phase begins. At this stage, several machine learning algorithms are tested simultaneously, such as Random Forest, Decision Tree, Bagging, Gradient Boosting, Logistic Regression, Neural Network, K-Nearest Neighbors, and Support Vector Machine. The goal is not only to build a model but also to

determine which algorithm is most suitable for detecting fraudulent transactions with the highest accuracy.

The final stage is model evaluation. Each algorithm will be tested using several evaluation metrics, including Precision, Recall, F1-Score, AUC-ROC, and MCC. In the context of fraud detection, metrics like Recall and AUC-ROC are prioritized because the most important thing is how well the model detects even the smallest fraudulent transactions, not just its apparent numerical accuracy.

With this series of processes, the end result is a best model that is not only numerically accurate, but also truly effective in recognizing suspicious transaction patterns.

3. Results and Interpretation

The following test results demonstrate that SMOVO can improve fraud detection performance compared to conventional models and other single approaches. The researchers conducted individual tests on classification models such as Random Forest, XGBoost, LightGBM, and AdaBoost, as well as applying ensemble techniques through the Voting Classifier.

This approach uses soft voting, where each model contributes to the final result based on its generated probabilities. After applying SMOTE, the results are presented in a table containing key metrics such as Accuracy, Precision, Recall, F1 Score, MCC, AUC, and G-Mean. The results are presented in the following table:

Table 2 Model Evaluation after SMOTE

Model	Recall	F1-Score	MCC	AUC	G-Mean
NN	0.84	0.24	0.34	0.97	0.91
DT	0.74	0.48	0.51	0.87	0.86
KNN	0.84	0.55	0.58	0.92	0.91
Note:	0.83	0.14	0.25	0.94	0.90
RF	0.78	0.82	0.82	0.96	0.88
XGB	0.77	0.65	0.66	0.97	0.88
LGBT	0.82	0.46	0.51	0.95	0.90
THERE IS	0.88	0.09	0.20	0.96	0.92
Voting Classifier	0.81	0.70	0.70	0.97	0.89

The classification model evaluation results show variations in detection ability and accuracy among the tested models. The ADA model stands out with the highest Recall of 0.88. This indicates that the model successfully identified a large proportion of positive cases, which is crucial in contexts where accurate positive detection is key, such as in disease diagnosis or fraud detection.

The KNN and NN models, which also had a Recall of 0.84, performed well in terms of sensitivity. This indicates that they were effective in capturing much of the positive class, although there are trade-offs to be considered in other metrics, such as Precision.

When looking at the F1-Score, which measures the balance between Precision and Recall, the RF model showed the highest value at 0.82. This high F1-Score indicates that the RF model is not only good at detecting positives but also has good accuracy in predicting negatives. In contrast, the ADA model had the lowest F1-Score (0.09), reflecting the possibility that although this model detects many positives, many of those predictions may be inaccurate.

The Matthews Correlation Coefficient (MCC) metric provides a more comprehensive picture of model performance, considering all categories. The RF and XGB models exhibited good MCC values of 0.82 and 0.66, respectively, indicating their ability to produce more balanced and reliable predictions. On the other hand, the ADA model, with the lowest MCC (0.20), indicated its overall performance was less stable and may not be reliable for decision-making.

In terms of AUC (Area Under the Curve), which indicates the model's ability to distinguish between positive and negative classes, the NN and XGB models showed high values (0.97). An AUC close to 1 indicates that the model has excellent ability to distinguish between the two classes. The DT model, with the lowest AUC (0.87), indicates that there is room for improvement in class separation.

The G-Mean, which reflects the balance between Sensitivity and Specificity, shows that both NN and KNN models have a high value (0.91). This indicates that both models are able to maintain a balance in detecting positives and negatives, which is very important in applications where both are equally important.

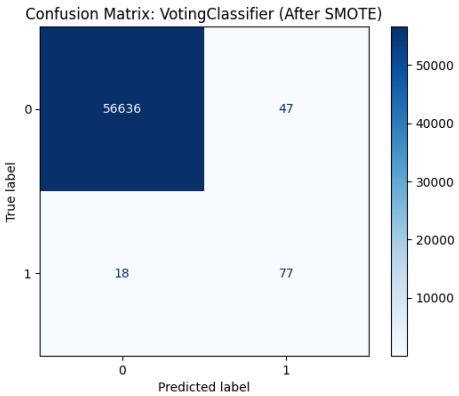
In this context, the Voting Classifier is the preferred choice. With a Recall of 0.81 and an F1-Score of 0.70, this model demonstrates good ability

to detect the positive class while maintaining overall accuracy. The strength of the Voting Classifier lies in its ability to combine predictions from multiple models, thereby improving the stability and generalizability of its performance. This approach reduces the risk of overfitting, which is often present with individual models. With an AUC of 0.97, the Voting Classifier demonstrates excellent effectiveness in distinguishing between the positive and negative classes, making it a solid choice for classification situations requiring high accuracy.

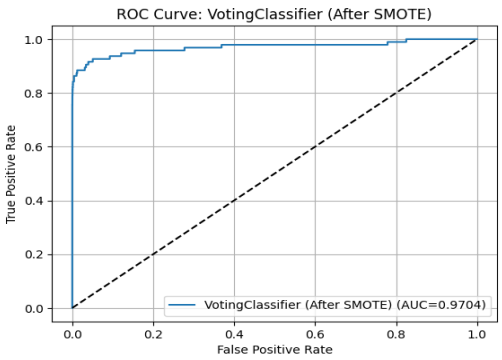
Overall, the selection of the right model should take into account the application context and the most relevant performance metrics, and the Voting Classifier shows promise as a reliable model in a variety of classification scenarios.

Confusion Matrix image of voting calcifier after SMOTE

1. AUC and MCC images after model testing, using the classifier voting algorithm from the SMOTE model testing.



Confusion Matrix Image of ADA Image After SMOTE



Voting Classification ROC Curve Figute After SMOTE Curve Figure

The confusion matrix image shows that the model successfully classified 28,074 non-fraud samples (majority class) and 51 of 53 fraud samples (minority class) correctly, resulting in only 14 false positives and 2 false negatives. This reflects high sensitivity to the minority class with a very low misclassification rate, which is crucial in fraud detection systems. The high number of true positives in the fraud class demonstrates the model's ability to minimize undetected fraudulent activity that could have significant economic consequences. Furthermore, the model's discriminatory capability is reinforced by the ROC curve, which yields an area under the curve (AUC) of 0.9821. This score indicates excellent class-separation ability, with the ROC curve lying well above the random diagonal line, reflecting near-perfect predictive performance.

Building upon these findings, the Model Evaluation stage compares the performance of RF, XGB, LGBM, ADA, and the Voting Classifier. Precision, Recall, F1-Score, MCC, AUC, and G-Mean are used as the primary metrics, given their relevance for imbalanced classification. The following table summarizes the performance of the proposed Voting Classifier in the SMOVO framework relative to the previous ESMOTE-GAN study (Gan et al.).

Table 4 Comparison of evaluation metrics; Precision, Recall, F1 Score, MCC, AUC, G-Mean with Previous SMOVO Studies.

Model	Precision	Recall	F1-Score	Auc	G-Man
Voting Clasiffier	0.81	0.70	0.70	0.89	0.81
Salwa et al. (2023)[23]	0.96	0.53	0.68	0.99	–
Ebenezer et al. (2023)[21]	–	0.90	–	0.93	–
Hanae 2023[22]	0.87	0.97	0.91	–	–
Esenogho 2022[21].	-	**85.71**	**0.89**	**0.98**	-

Based on the evaluation results presented in the table, the model used in this study achieved a precision score of 0.81. This means that of all transactions predicted as fraudulent by the system,

approximately 81% were actually suspicious. This value indicates that the model is quite capable of reducing errors in issuing warning signals to normal transactions (false positives).

However, in terms of recall, the model produced a value of 0.70, meaning that approximately 30% of fraudulent transactions were still undetected or slipped through the system. Nevertheless, this recall value is still better than Salwa et al.'s (2023) model, which only achieved 0.53, indicating that their model was more stringent in identifying fraud but missed many important cases.

In contrast, the Hanae (2023) model showed a different trend, with the highest recall value at 0.97, indicating that their system was highly aggressive in detecting fraudulent transactions. Unfortunately, not all studies reported precision and other metrics, so there may be a trade-off in the form of increased false positives. However, this could not be analyzed further due to incomplete data.

Based on the F1-Score metric, this study obtained a value of 0.70, indicating a moderate balance between precision and recall. This value is relatively competitive compared to Salwa et al.'s F1 of 0.68 and close to Esenogho's (2022) model, which achieved 0.89. Hanae's (2023) model again excelled with an F1-Score of 0.91, indicating that their model was able to maintain a more optimal performance balance.

Meanwhile, in terms of the AUC metric, the model developed in this study obtained a score of 0.89, indicating that the system is able to distinguish fraudulent and non-fraudulent transactions fairly well and consistently. This value is slightly below the models of Salwa et al. (0.99) and Esenogho (0.98), but still falls into the good category (AUC > 0.85). In other words, this model has stable generalization capabilities, although it does not always rank highest.

Finally, a G-Mean value of 0.81 indicates that the model does not focus solely on one class but is quite balanced in recognizing both classes, both normal and fraudulent transactions. Unfortunately, most comparative studies do not report this value, making further validation of the performance balance aspect impossible.

4. Conclusion

Based on the evaluation results and comparisons with several previous studies, it can be concluded that the model developed in this study demonstrates stable and balanced performance. Although it does not consistently rank highest in every metric, this model is able to maintain a balance between precision and recall, as indicated by an F1-Score of 0.70 and a G-Mean of 0.81. This indicates that the model is not only accurate in identifying fraudulent transactions but also fairly in recognizing normal transactions, thus minimizing the risk of overfitting or bias towards one class.

Compared to Salwa et al.'s (2023) study, which had high precision but low recall, this model is more responsive in detecting actual fraudulent transactions. Meanwhile, compared to Hanae's (2023) model, which had very high recall, the model in this study offers a more conservative and realistic approach for implementation in real-world fraud detection systems, as it is less aggressive, thus reducing the potential for false positives.

Overall, with an AUC of 0.89, this model can be categorized as reliable and feasible to implement due to its good generalization ability in distinguishing fraudulent and non-fraudulent transactions. With consistent and non-extreme performance on any one metric, this model is more suitable for use in operational environments that require both accuracy and detection stability simultaneously.

REFERENCES

[1] M. Barroso and J. Laborda, “Digital transformation and the emergence of the Fintech sector: Systematic literature review,” *Digit. Bus.*, vol. 2, no. 2, p. 100028, 2022, doi: 10.1016/j.digbus.2022.100028.

[2] A. A. Calderon, “Digital Payments and their Role in Enhancing Financial Transactions Efficiency,” *Int. J. Econ. Financ. Issues* , vol. 15, no. 1, pp. 182–189, 2025, doi: 10.32479/ijefi.17555.

[3] S. Nazari Nezhad, M. H. Zahedi, and E. Farahani, “Detecting diseases in medical prescriptions using data mining methods,” *BioData Min.*, vol. 15, no. 1, pp. 1–19, 2022, doi: 10.1186/s13040-022-00314-w.

[4] J. W. Chang, N. Yen, and J. C. Hung, “Design of a NLP-empowered finance fraud awareness model: the anti-fraud chatbot for fraud detection and fraud classification as an instance,” *J. Ambient Intell. Humaniz. Comput.*, vol. 13, no. 10, pp. 4663–4679, 2022, doi: 10.1007/s12652-021-03512-2.

[5] M. Jullum, A. Løland, R. B. Huseby, G. Ånonsen, and J. Lorentzen, “Detecting money laundering transactions with machine learning,” *J. Money Laund. Control*, vol. 23, no. 1, pp. 173–186, 2020, doi: 10.1108/JMLC-07-2019-0055.

[6] A. Firmansah, Aripriharta, N. Mufti, A. N. Affandi, and I. A. E. Zaeni, “Self-powered IoT Based Vibration Monitoring of Induction Motor for Diagnostic and Prediction Failure,” *IOP Conf. Ser. Mater. Sci. Eng.*, vol. 588, no. 1, 2019, doi: 10.1088/1757-899X/588/1/012016.

[7] M. Grossi *et al.*, “Mixed Quantum-Classical Method for Fraud Detection With Quantum Feature Selection,” *IEEE Trans. Quantum Eng.*, vol. 3, no. August, pp. 1–12, 2022, doi: 10.1109/TQE.2022.3213474.

[8] T. Widiyaningtyas, “ALGORITME REKOMENDASI MENGGUNAKAN CLUSTERING DAN USER PROFILE CORRELATION-BASED SIMILARITY (UPCSim) PADA SISTEM REKOMENDASI FILM,” 2022.

[9] D. N. Triyanto, M. A. N. Fajri, and D. Wahyuni, “How is financial reporting fraud with the fraud hexagon approach before and during Covid-19 pandemic?,” *J. Contemp. Account.*, pp. 97–114, 2023, doi: 10.20885/jca.vol5.iss2.art4.

[10] M. A. Islam, M. A. Uddin, S. Aryal, and G. Stea, “An ensemble learning approach for anomaly detection in credit card data with imbalanced and overlapped classes,” *J. Inf. Secur. Appl.*, vol. 78, no. October, 2023, doi: 10.1016/j.jisa.2023.103618.

[11] M. J. Madhurya, H. L. Gururaj, B. C. Soundarya, K. P. Vidyashree, and A. B. Rajendra, “Exploratory analysis of credit card fraud detection using machine learning techniques,” *Glob. Transitions Proc.*, vol. 3, no. 1, pp. 31–37, 2022, doi: 10.1016/j.gltp.2022.04.006.

[12] A. A. Taha and S. J. Malebary, “An intelligent approach to credit card fraud detection using an optimized light gradient boosting machine,” *IEEE Access*, 2020, [Online]. Available: https://ieeexplore.ieee.org/abstract/document/8979331/

[13] A. Prasetya Wibawa, S. Aji Kurniawan, and I. Ari Elbaith Zaeni Assistant Professor, “Determining Journal Rank by Applying Particle Swarm Optimization-Naive Bayes Classifier,” *J. Inf. Technol. Manag.*, vol. 13, no. 4, pp. 116–125, 2021, [Online]. Available: https://jitm.ut.ac.ir/article_83962.html

[14] M. F. A. Saputra, T. Widiyaningtyas, and A. P. Wibawa, “Illiteracy classification using K means-naïve bayes algorithm,” *Int. J. Informatics Vis.*, vol. 2, no. 3, pp. 153–158, 2018, doi: 10.30630/joiv.2.3.129.

[15] T. H. Pranto, K. T. A. M. Hasib, T. Rahman, A. B. Haque, A. K. M. N. Islam, and R. M. Rahman, “Blockchain and Machine Learning for Fraud Detection: A Privacy-Preserving and Adaptive Incentive Based Approach,” *IEEE Access*, vol. 10, no. July, pp. 87115–87134, 2022, doi: 10.1109/ACCESS.2022.3198956.

[16] Purnawansyah *et al.*, “Congestion Predictive Modelling on Network Dataset Using Ensemble Deep Learning,” *J. Appl. Data Sci.*, vol. 5, no. 4, pp. 1597–1613, 2024, doi: 10.47738/jads.v5i4.333.

[17] I. Y. Hafez, A. Y. Hafez, A. Saleh, A. A. Abd El-Mageed, and A. A. Abohany, “A systematic review of AI-enhanced techniques in credit card fraud detection,” *J. Big Data*, vol. 12, no. 1, 2025, doi: 10.1186/s40537-024-01048-8.

[18] M. K. Severino and Y. Peng, “Machine learning algorithms for fraud prediction in property insurance: Empirical evidence using real-world microdata,” *Mach. Learn. with Appl.*, vol. 5, no. June, p. 100074, 2021, doi: 10.1016/j.mlwa.2021.100074.

[19] U. Pujianto, I. A. E. Zaeni, and K. I. Rasyida, “Comparison of Naive Bayes and Random Forests Classifier in the Classification of News Article Popularity as Learning Material,” *Proc. 1st UMGESHIC Int. Semin. Heal. Soc. Sci. Humanit. (UMGESHIC-ISHSSH 2020)*, vol. 585, pp. 229–242, 2021, doi: 10.2991/assehr.k.211020.036.

[20] O. I. Obaid and A. Y. Al-sultan, “Leveraging LSTM and Attention for High-Accuracy Credit Card Fraud Detection,” *Fusion Pract. Appl.*, vol. 17, no. 1, pp. 209–220, 2025, doi: 10.54216/fpa.170115.

[21] E. Esenogho, I. D. Mienye, T. G. Swart, K. Aruleba, and ..., “A neural network ensemble with feature engineering for improved credit card fraud detection,” *IEEE …*, 2022, [Online]. Available: https://ieeexplore.ieee.org/abstract/document/9698195/

[22] A. Hanae, B. Abdellah, E. Saida, and G. Youssef, “End-to-End Real-time Architecture for Fraud Detection in Online Digital Transactions,” *Int. J. Adv. Comput. Sci. Appl.*, vol. 14, no. 6, pp. 749–757, 2023, doi: 10.14569/IJACSA.2023.0140680.

[23] S. Al Balawi and N. Aljohani, “Credit-card fraud detection system using neural networks.,” *Int. Arab J. Inf. Technol.*, 2023, [Online]. Available: https://iajit.org/portal/images/year2023/No.2/20341.pdf

[24] F. A. Ghaleb, F. Saeed, M. Al-Sarem, S. N. Qasem, and T. Al-Hadhrami, “Ensemble Synthesized Minority Oversampling-Based Generative Adversarial Networks and Random Forest Algorithm for Credit Card Fraud Detection,” *IEEE Access*, vol. 11, no. August, pp. 89694–89710, 2023, doi: 10.1109/ACCESS.2023.3306621.

[25] Y. Y. Hsin, T. S. Dai, Y. W. Ti, M. C. Huang, T. H. Chiang, and L. C. Liu, “Feature Engineering and Resampling Strategies for Fund Transfer Fraud with Limited Transaction Data and a Time-Inhomogeneous Modi Operandi,” *IEEE Access*, vol. 10, no. July, pp. 86101–86116, 2022, doi: 10.1109/ACCESS.2022.3199425.

[26] M. Ashraf, M. A. Abourezka, and F. A. Maghraby, “A Comparative Analysis of Credit Card Fraud Detection Using Machine Learning and Deep Learning Techniques,” *Lect. Notes Networks Syst.*, vol. 224, pp. 267–282, 2022, doi: 10.1007/978-981-16-2275-5_16.

[27] P. Serie, W. Paper, T. Newman, M. Mccann, and A. Scott, “the initial teacher education : a cornerstone in the improvement of today ’ s Working Papers Series ‘ Meeting New Challenges in Education ’ ISSUE 1,” no. May 2023, 2024.

[28] D. Gong and Y. Liu, “A Mechine Learning Approach for Botnet Detection Using LightGBM,” *2022 3rd Int. Conf. Comput. Vision, Image Deep Learn. \& Int. Conf. Comput. Eng. Appl. (CVIDL \& ICCEA)*, 2022, doi: 10.1109/cvidliccea56201.2022.9824033.

[29] H. Hairani, T. Widiyaningtyas, and D. D. Prasetya, “Feature Selection and Hybrid Sampling with Machine Learning Methods for Health Data Classification,” *Rev. d’Intelligence Artif.*, vol. 38, no. 4, pp. 1255–1261, 2024, doi: 10.18280/ria.380419.

[30] I. Saifudin and T. Widiyaningtyas, “Systematic Literature Review on Recommender System: Approach, Problem, Evaluation Techniques, Datasets,” *IEEE Access*, vol. 12, no. January, pp. 19827–19847, 2024, doi: 10.1109/ACCESS.2024.3359274.

[31] Y. Vivek, V. Ravi, A. A. Mane, and L. R. Naidu, “Explainable Artificial Intelligence and Causal Inference based ATM Fraud Detection,” no. Ci, pp. 1–34, 2022, doi: 10.1109/CIFER62890.2024.10772906.

[32] S. Bagga, A. Goyal, N. Gupta, and A. Goyal, “Credit Card Fraud Detection using Pipeling and Ensemble Learning,” *Procedia Comput. Sci.*, vol. 173, no. 2019, pp. 104–112, 2020, doi: 10.1016/j.procs.2020.06.014.

[33] M. Azim Mim, N. Majadi, and P. Mazumder, “A soft voting ensemble learning approach for credit card fraud detection,” *Heliyon*, vol. 10, no. 3, p. e25466, 2024, doi: 10.1016/j.heliyon.2024.e25466.

[34] R. Priyadarshini, K. Anuratha, N. Rajendran, and S. Sujeetha, “APMFT: Anamoly Prediction Model for Financial Transactions Using Learning Methods in Machine Learning and Deep Learning,” *Adv. Parallel Comput.*, 2021, doi: 10.3233/apc210101.

[35] W. Rahayu *et al.*, “Synthetic Minority Oversampling Technique (SMOTE) for Boosting the Accuracy of C4.5 Algorithm Model,” *J. Artif. Intell. Eng. Appl.*, vol. 3, no. 3, pp. 624–630, 2024, doi: 10.59934/jaiea.v3i3.469.

[36] H. Hairani, T. Widiyaningtyas, D. D. Prasetya, and A. Aminuddin, “Addressing Imbalance in Health Datasets: A New Method NR-Clustering SMOTE and Distance Metric Modification,” *Comput. Mater. Contin.*, vol. 82, no. 2, pp. 2931–2949, 2025, doi: 10.32604/cmc.2024.060837.

[37] Z. Li, M. Huang, G. Liu, and C. Jiang, “A hybrid method with dynamic weighted entropy for handling the problem of class imbalance with overlap in credit card fraud detection,” *Expert Syst. Appl.*, vol. 175, no. January, p. 114750, 2021, doi: 10.1016/j.eswa.2021.114750.

[38] Y. C and M. Lucas, “Credit card fraud detection using machine learning with integration of contextual knowledge To cite this version : HAL Id : tel-02951477 Computer Science Credit Card Fraud Detection using Machine Learning with Integration of Contextual Knowledge Before th,” 2020.