Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Improving Credit Card Fraud Detection Using Ensemble Machine Learning Techniques

Adel EL QASSOUARI

a.elqassouari.ced@uca.ac.ma

Cadi Ayyad University

Ayoub CHAREF

Cadi Ayyad University, UCA, National School of Applied Sciences

Zahi JARIR

Cadi Ayyad University

Research Article

Keywords: Credit Card Fraud Detection, Ensemble Learning, GANs, Machine Learning, Real-Time Transaction Analysis

Posted Date: August 14th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-10607534/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: No competing interests reported.

Improving Credit Card Fraud Detection Using Ensemble Machine Learning Techniques

Adel EL QASSOUARI[1*], Ayoub CHAREF[2†] and Zahi JARIR[1†]

[1*]Cadi Ayyad University, UCA, Faculty of Science, Laboratory of Computer Systems Engineering (LISI), Department of Computer Science, Marrakesh, Morocco.
[2]Cadi Ayyad University, UCA, National School of Applied Sciences, LMSC Laboratory, Marrakesh, Morocco.

*Corresponding author(s). E-mail(s): a.elqassouari.ced@uca.ac.ma;
Contributing authors: ay.charef@uca.ac.ma; jarir@uca.ac.ma;
†These authors contributed equally to this work.

Abstract

The rapid growth of online transactions has significantly increased the risk of credit card fraud, making accurate and real-time fraud detection systems essential for financial platforms. This paper proposes an enhanced fraud detection framework that combines Generative Adversarial Networks (GANs) for data balancing with ensemble machine learning techniques to improve detection performance. The study is conducted on a real-world dataset containing 76,900 transactions, where fraudulent operations represent only 2.5% of the total data, leading to a severe class imbalance problem.

To address this issue, a GAN-based approach is employed to generate synthetic samples and create a balanced training dataset. Several machine learning models, including XGBoost, Random Forest, Logistic Regression, SVM, KNN, and Naïve Bayes, are evaluated using performance metrics such as Accuracy, Precision, Recall, F1-score, and AUC-ROC, along with execution time to assess real-time feasibility.

Experimental results demonstrate that ensemble models, particularly XGBoost, achieve superior performance, significantly improving fraud detection recall while maintaining high precision and low latency. The findings confirm that integrating synthetic data generation with ensemble learning provides a robust, scalable, and practical solution for real-world credit card fraud detection systems.

Keywords: Credit Card Fraud Detection, Ensemble Learning,GANs, Machine Learning, Real-Time Transaction Analysis

1 Introduction

The rapid growth of online payment systems and digital financial services has led to a significant increase in both the volume and velocity of credit card transactions, making fraud detection a critical challenge for financial institutions and e-commerce platforms. With millions of transactions processed daily across heterogeneous platforms, even a small fraud rate can result in substantial financial losses. Beyond direct monetary impact, credit card fraud erodes customer confidence, damages institutional reputation, and increases operational costs due to chargebacks, dispute management, and regulatory compliance. These risks are particularly pronounced in real-time transaction environments, where fraud detection systems are required to deliver highly accurate decisions within milliseconds, without disrupting legitimate user activity [1].

Detecting fraudulent transactions is inherently complex due to several factors. Fraud patterns continuously evolve as attackers adapt their strategies to bypass existing security mechanisms, making static or rule-based systems increasingly ineffective. Additionally, transaction datasets are typically high-dimensional, heterogeneous, and noisy, incorporating behavioral, temporal, and contextual features that are difficult to model jointly [2]. A major challenge arises from the severe class imbalance between legitimate and fraudulent transactions, where fraudulent activities usually constitute only a very small fraction of the overall data. This imbalance causes traditional machine learning models to be biased toward the majority class, resulting in high overall accuracy but poor fraud detection capability, particularly in terms of recall and false-negative rates. Consequently, specialized learning strategies and robust evaluation frameworks are required to ensure reliable and efficient fraud detection in real-world deployment scenarios [3].

The primary objective of this study is to investigate the effectiveness of machine learning techniques for credit card fraud detection by conducting a comparative analysis of multiple classification algorithms. The goal is to identify the model that provides the best trade-off between high detection performance and rapid response time, which is essential for real-time transaction processing. In addition to evaluating predictive accuracy, the study emphasizes computational efficiency to ensure that the selected model can handle a high volume of transactions with minimal latency. To further validate real-world applicability, a lightweight transaction simulation framework is developed to emulate concurrent transaction requests generated by clients. This simulator places the proposed models under realistic operational conditions, enabling a practical assessment of their robustness, throughput, and speed of fraud detection in dynamic and time-sensitive environments.

The remainder of this paper is organized as follows. The next section reviews related work on credit card fraud detection, with particular emphasis on machine learning approaches and techniques for handling class imbalance. This is followed by the methodology section, which describes the proposed framework and details the use of Generative Adversarial Networks (GANs) to balance the dataset and enhance classification performance. Subsequently, the results and discussion section presents a comparative evaluation of the selected classification models, focusing on detection effectiveness and, in particular, on response time and computational efficiency under

simulated real-time transaction conditions. Finally, the paper concludes by summarizing the main findings, highlighting the practical implications of the proposed approach, and outlining potential directions for future research.

2 Related work

The rapid growth of electronic payment systems has led to a significant increase in credit card fraud, both in scale and sophistication, making automated detection a critical research challenge. Traditional rule-based systems lack flexibility and fail to adapt to evolving fraud strategies, which has motivated the widespread adoption of machine learning and artificial intelligence–based approaches [4, 5]. Despite notable progress, most existing methods are hindered by the extreme class imbalance inherent in fraud datasets, where fraudulent transactions constitute only a small fraction of the data.

Early research in credit card fraud detection focused on classical supervised machine learning models such as Logistic Regression, Decision Trees, Naïve Bayes, KNN, and Support Vector Machines due to their simplicity and low computational cost [6, 7]. To mitigate class imbalance, these models are frequently combined with resampling techniques such as random undersampling and SMOTE, leading to improvements in recall and F1-score. Among these approaches, Random Forest consistently demonstrates strong robustness and superior predictive performance, particularly when trained on balanced datasets [8, 9].

Building upon single-model approaches, ensemble learning techniques have been extensively investigated to improve generalization and reduce false negatives. Methods based on bagging, boosting, and soft or weighted voting combine multiple classifiers to achieve more stable and accurate predictions [10, 11]. Gradient boosting models such as XGBoost and LightGBM further enhance detection performance by capturing complex non-linear patterns in transaction data, especially under severe imbalance conditions [12, 13]. However, the increased complexity of ensemble models often limits their interpretability and practical deployment.

In parallel, feature engineering and feature selection have been identified as critical components for improving fraud detection systems. Several studies employ genetic algorithms, meta-heuristic optimization, and hybrid filter–wrapper strategies to eliminate redundant and irrelevant features while preserving discriminative information [14, 15]. By optimizing feature subsets using imbalance-aware metrics, these methods improve sensitivity to fraudulent transactions and reduce computational overhead [3]. Nevertheless, most feature selection techniques remain static and struggle to adapt to concept drift in real-world environments.

More recently, deep learning approaches have gained attention due to their ability to model complex temporal and spatial dependencies in transaction data. Autoencoders, CNNs, and LSTM-based architectures have demonstrated strong performance on benchmark datasets, often outperforming traditional machine learning models when appropriately tuned [16, 17]. Dilated CNNs further improve detection accuracy by capturing long-range temporal patterns in sequential transaction data [18]. Despite these

advantages, deep learning models typically require large labeled datasets and involve high computational costs.

Beyond conventional deep learning architectures, advanced representation learning techniques have been explored. Transformer-based models such as BERT leverage contextual embedding to handle high-dimensional transaction features, achieving state-of-the-art detection accuracy [19]. Similarly, graph neural networks exploit relational structures between transactions, users, and merchants, enabling improved fraud detection in complex financial networks [20]. However, scalability, explainability, and deployment complexity remain significant challenges for these approaches.

With increasing regulatory constraints on financial data sharing, privacy-preserving learning has become a major research focus. Federated learning frameworks enable collaborative training across institutions without sharing raw transaction data, improving detection accuracy while preserving privacy [21, 22]. Hybrid approaches integrating federated learning with ensemble techniques or data augmentation further enhance robustness against evolving fraud patterns [23]. Nevertheless, these methods introduce communication overhead and additional system complexity.

Given the scarcity of labeled fraud data, unsupervised and semi-supervised learning approaches have also been widely investigated. Autoencoder-based anomaly detection, clustering-based undersampling, and label synthesis techniques demonstrate promising results in identifying previously unseen fraud patterns [24, 25]. Additionally, GAN-based data augmentation has been extensively studied to address class imbalance by generating synthetic fraud samples [26], although concerns regarding training stability and data realism persist [27].

Several comprehensive surveys and systematic reviews consolidate existing research and identify persistent limitations in the field [28, 29]. Commonly reported issues include over-reliance on benchmark datasets, insufficient use of imbalance-sensitive evaluation metrics, limited integration of explainable AI, and inadequate handling of real-time concept drift. These findings highlight the need for more adaptive and realistic fraud detection frameworks.

Despite the strong performance reported by existing machine learning, ensemble, deep learning, and privacy-preserving approaches, several critical limitations remain. A large proportion of prior studies rely on publicly available benchmark datasets that are repeatedly reused, increasingly consumed, and often outdated. These datasets fail to accurately reflect the evolving nature of real-world fraud behaviors, leading to models that perform well in offline evaluation but exhibit limited generalization when exposed to new and unseen fraud patterns [30, 31]. Moreover, the extensive reuse of the same datasets raises concerns regarding data saturation and optimistic performance estimation.

In addition, many high-performing fraud detection models are evaluated solely in offline experimental settings, with limited consideration of real-world deployment constraints. Most studies focus primarily on predictive accuracy while neglecting practical aspects such as system scalability, response latency, and detection time. The absence of implementation-level evaluation makes it difficult to assess whether these approaches can meet the strict real-time requirements of operational financial systems, where rapid response is crucial to prevent financial losses and minimize fraud impact [32].

Furthermore, relatively little attention is given to strategies for generating or enriching data in a realistic and controlled manner to overcome the limitations of scarce or obsolete datasets. While synthetic data generation techniques are explored in some studies, their effectiveness is often evaluated only through standard classification metrics without assessing their impact on real-time detection performance or system responsiveness. These gaps highlight the need for fraud detection frameworks that combine adaptive learning, realistic data generation, and end-to-end system evaluation, including latency and detection speed.

Motivated by these limitations, the present work aims to address the challenges of dataset obsolescence, limited real-world validation, and lack of runtime performance analysis. By integrating realistic data handling strategies and evaluating the proposed approach in an operational context, this study seeks to develop a fraud detection framework that is not only accurate but also efficient, responsive, and suitable for real-world deployment.

3 Methodology

Machine learning enables fraud detection by exploiting historical records of both fraudulent and legitimate transactions. By learning underlying transaction patterns, these algorithms are capable of identifying anomalous behaviors at an early stage, before fraudulent activities escalate into significant financial losses. Figure 1 presents the flow diagram illustrating the process through which machine learning models detect credit card fraud.

3.1 DataSet

The dataset used in this study consists of credit card transactions carried out by WorldWays credit card holders over a six-year period on the AirportRental website in Morocco. During this period, the platform recorded a total of 76,900 payment transactions, of which approximately 2.5% were identified as fraudulent . Specifically, the dataset contains 1,922 fraudulent transactions out of the 76,900 recorded transactions.

The dataset exhibits a highly imbalanced class distribution, with fraudulent transactions representing only 2.5% of the total data . Each transaction is described by 15 additional features (see Table1) . Due to confidentiality and privacy constraints, these features were transformed using Linear Discriminant Analysis (LDA) [33, 34]. It should be noted that the (id_12 - id_38), Time and TransactionAMT attributes were excluded from this transformation and retained in their original form.

As previously highlighted, the pronounced class imbalance significantly affects the performance of machine learning models, particularly during training and evaluation phases. Models trained on such skewed datasets tend to be biased toward the majority class, resulting in poor fraud detection performance [35].

To address this issue, Generative Adversarial Networks (GANs) were employed to generate synthetic samples for the minority class [36]. This approach enabled the creation of a balanced dataset, achieving an equal class distribution of 50% legitimate transactions and 50% fraudulent transactions, thereby improving the robustness and effectiveness of the proposed fraud detection model [37].

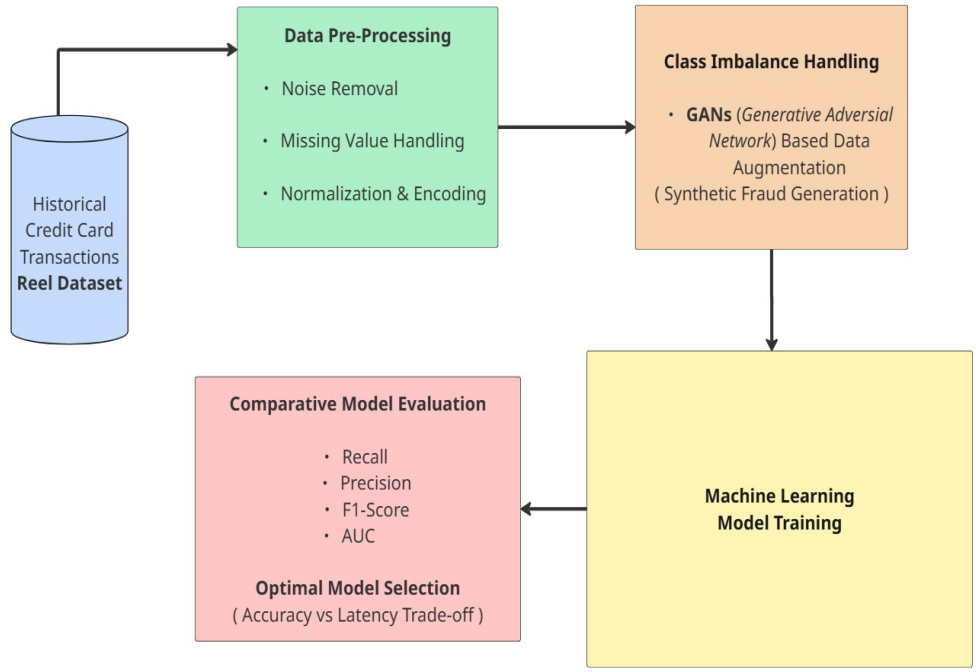

Fig. 1 Overview of the Proposed Credit Card Fraud Detection Pipeline.

Figure 2 illustrates the proposed GAN-based synthetic data generation framework, which is composed of three main phases.

- In Phase 1, the original dataset is first divided into positive (fraudulent) and negative (legitimate) samples based on the target label. The target column is then removed from both subsets to allow the generative models to learn the underlying feature distributions independently for each class.
- In Phase 2, two separate Generative Adversarial Networks (GANs) are trained in parallel: one GAN is trained using the positive samples, while the other is trained using the negative samples. Each GAN learns the statistical characteristics of its respective class and generates corresponding synthetic samples.
- In Phase 3, the generated synthetic positive and negative samples are assigned their respective target labels. These labeled samples are then concatenated to form a synthetic dataset, resulting in a balanced data distribution between legitimate and fraudulent transactions.

This multi-phase GAN-based approach effectively mitigates the class imbalance problem and enhances the robustness of the fraud detection model.

As shown in Table 1, the dataset is primarily divided into two categories: transaction-related features and identity-related features.

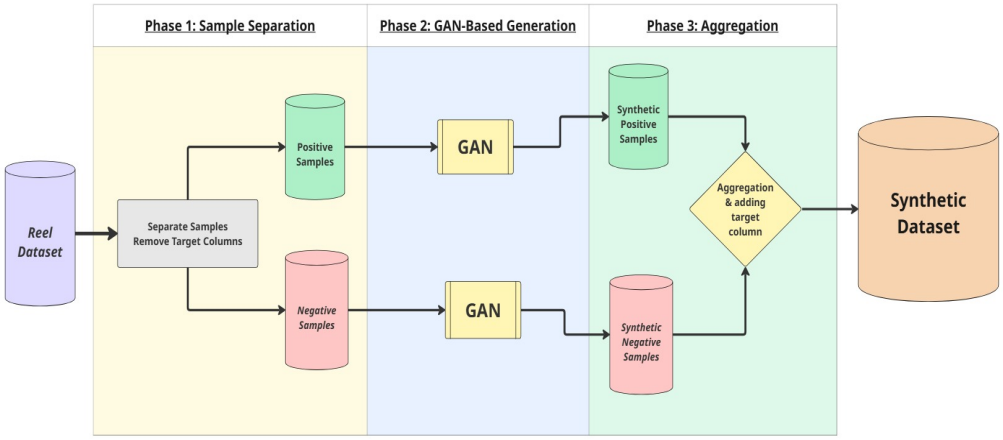

Fig. 2 GAN-Based Framework for Balancing Fraud and Legitimate Transactions.

Table 1 Transaction and Identity Features Description

Category	Feature	Description
Transaction data	TransactionDT	Transaction timestamp
	TransactionAMT	Transaction amount
	ProductCD	Product code for each transaction
	card1–card6	Card-related attributes
	addr	Address-related information
	dist	Distance-related feature
	P_emaildomain, R_emaildomain	Purchaser and recipient email domains
	C1–C14	Count-based features
	D1–D15	Time delta features
	M1–M9	Match-type features
	Vxxx	Anonymous numerical features
	Time	Transaction time feature
Identity data	DeviceType	Type of device used
	DeviceInfo	Device information string
	id_12–id_38	Identity-related anonymized features

3.2 Proposed Approach

The proposed approach focuses on identifying the most effective and efficient machine learning algorithm for detecting fraudulent credit card transactions within the AirportRental online platform. Following an extensive review of existing literature, several supervised and semi-supervised classification models are selected and evaluated through a comprehensive comparative analysis. The considered algorithms include One-Class Random Forest, K-Nearest Neighbors (KNN), Extreme Gradient Boosting (XGBoost), Naïve Bayes, and additional widely used classifiers such as Logistic Regression, Support Vector Machines (SVM), and Decision Trees. These models are

chosen due to their proven effectiveness in fraud detection tasks, their ability to handle imbalanced datasets, and their diverse learning paradigms [38, 39].

The evaluation strategy is based on two main criteria: detection performance and computational efficiency. Detection performance is assessed using standard metrics such as accuracy, precision, recall, F1-score, and AUC-ROC, with particular emphasis on recall to minimize false negatives in fraud detection [40]. Computational efficiency is measured by analyzing the execution time required for transaction classification, making it possible to identify the most suitable model for real-time deployment [41]. This dual evaluation allows the approach to balance high fraud detection capability with low latency, which is critical for processing online transactions in time-sensitive environments such as airport rental services.

The final model selection is determined by jointly considering classification accuracy and detection speed, ensuring that the chosen algorithm not only provides robust fraud detection but also meets the real-time operational constraints of online transaction processing systems.

Figure 3 illustrates the overall architecture of the proposed credit card fraud detection pipeline. Transaction requests are first generated by a client simulator to emulate real-world concurrent payment activity and are stored in the transaction database. Each incoming transaction undergoes an initial validation step to filter out incomplete or invalid records, which are immediately rejected. Valid transactions are then analyzed and forwarded to the machine learning layer, where trained classification models assess the likelihood of fraud based on learned transactional patterns. Depending on the prediction outcome, transactions are either approved or blocked in real time. This end-to-end pipeline reflects the operational flow of a real-world fraud detection system, enabling both accurate classification and low-latency decision-making under simulated deployment conditions.

3.3 Materials Used

The experimental environment is composed of dedicated hardware and software resources to support efficient machine learning training and evaluation. All experiments are conducted on a workstation equipped with an Intel Core i7 multi-core processor, 16 GB of RAM, and a high-performance NVIDIA GeForce RTX 3080 graphics card, enabling accelerated computation for training complex models and handling large transaction datasets. A solid-state drive (SSD) is used to ensure fast data access and reduced input/output latency. The system operates on a 64-bit operating system to guarantee compatibility with modern machine learning frameworks.

The software environment is based on the Python programming language, selected for its flexibility and extensive ecosystem in data science and machine learning. Data preprocessing and feature engineering are performed using NumPy and Pandas, while machine learning models such as K-Nearest Neighbors, Naïve Bayes, Logistic Regression, Support Vector Machines, Random Forests, and One-Class classifiers are implemented using Scikit-learn. Gradient boosting-based models are developed using XGBoost, with GPU acceleration leveraged where applicable to improve training efficiency. Model evaluation and execution-time analysis are carried out using

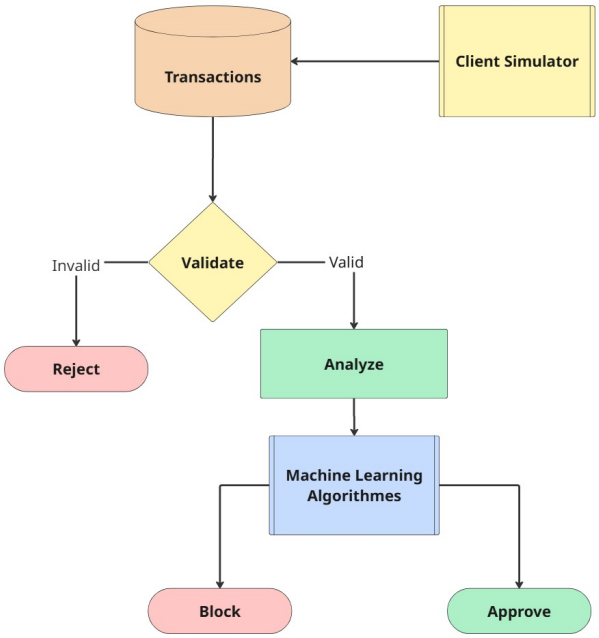

Fig. 3 Machine Learning–Driven Transaction Analysis and Decision Workflow.

Scikit-learn's evaluation metrics and timing utilities, allowing accurate comparison of detection performance and computational speed.

4 Results and Discussion

This section presents the experimental results obtained from the evaluation of the proposed fraud detection approach and provides a detailed discussion of the findings. The performance of the selected machine learning models is analyzed using standard evaluation metrics, with particular emphasis on their ability to detect fraudulent transactions under highly imbalanced and balanced data conditions. The results are compared to highlight the impact of the GAN-based data balancing strategy on detection accuracy, recall, and overall model robustness, as well as to identify the most effective model for real-time fraud detection within the AirportRental platform.

To compare the selected algorithms in terms of both detection performance and processing speed, a comprehensive experimental setup was adopted. In credit card transaction systems, rapid response time and the ability to handle a high volume of simultaneous transactions are critical requirements, particularly in real-time environments such as online rental platforms. Consequently, the evaluation focuses not only on classification accuracy but also on execution time and scalability.

Two different train–test split strategies were employed to assess model stability and generalization capability: a first configuration using 70% of the data for training and 30% for testing, and a second configuration with an equal split of 50% for training and

50% for testing. This dual setup makes it possible to analyze the impact of training data size on both detection effectiveness and computational efficiency.

Furthermore, to validate the practical applicability of the proposed approach, a lightweight transaction simulation environment was developed. This simulator generates and sends multiple transactions concurrently to the trained models within short time intervals (on the order of milliseconds), emulating real-world operational conditions. This experimental design enables the evaluation of each algorithm's response time, throughput, and robustness when processing a high number of transactions simultaneously, thereby identifying the most suitable model for real-time credit card fraud detection.

4.1 Comparative Performance Analysis under the 70%–30% Train–Test Split

This section presents the experimental findings of the proposed fraud detection framework, evaluated under the 70%-30% train-test split. The analysis is structured to first assess the impact of the GAN-based data balancing technique and then to provide a comparative evaluation of the selected machine learning models based on detection performance and computational efficiency.

The effectiveness of the proposed GAN framework for mitigating class imbalance is evaluated by comparing model performance on the original imbalanced dataset versus the GAN-synthesized balanced dataset. As illustrated in Table 2, the balanced dataset significantly enhances the models' ability to identify fraudulent transactions.

While overall accuracy sees a marginal and expected decrease on the balanced dataset as the model is no longer biased towards the overwhelming majority class all critical fraud-centric metrics show dramatic improvement. Most notably, Recall increased by an average of ˜25 percentage points across models. This confirms that the GAN-based synthesis successfully learned the underlying distribution of the minority class, enabling models to generalize more effectively and reduce false negatives, which are costlier in fraud detection.

Table 2 presents the comparative performance of seven machine learning models—XGBoost, Random Forest, Logistic Regression, and SVM (RBF Kernel)—evaluated on both imbalanced and balanced datasets using Accuracy, Precision, Recall, F1-Score, and AUC-ROC metrics.

Overall, XGBoost achieves the best performance across most evaluation metrics. On the balanced dataset, it records the highest Precision (0.95), Recall (0.93), F1-Score (0.94), and AUC-ROC (0.99), while maintaining very high Accuracy (0.9941). Random Forest shows competitive results, particularly on the balanced dataset, with strong F1-Score (0.93) and AUC-ROC (0.98).

Logistic Regression exhibits lower Recall and F1-Score on the imbalanced dataset (0.65 and 0.68, respectively), indicating sensitivity to class imbalance. However, its performance improves significantly after balancing. Similarly, SVM (RBF Kernel) demonstrates noticeable improvement on the balanced dataset, especially in Recall (0.90) and F1-Score (0.91).

Table 2 Comparative Model Performance on Imbalanced vs. Balanced Datasets (70-30 Split)

Model	Dataset	Accuracy	Precision	Recall	F1-Score	AUC-ROC
XGBoost	Imbalanced	0.998	0.89	0.78	0.83	0.96
	Balanced	**0.9941**	**0.95**	**0.93**	**0.94**	**0.99**
Random Forest	Imbalanced	0.9981	0.86	0.76	0.81	0.95
	Balanced	0.9938	0.94	0.92	0.93	0.98
Logistic Regression	Imbalanced	0.9975	0.72	0.65	0.68	0.91
	Balanced	0.9895	0.91	0.89	0.90	0.97
SVM (RBF Kernel)	Imbalanced	0.9979	0.81	0.70	0.75	0.93
	Balanced	0.9912	0.92	0.90	0.91	0.97
KNN	Imbalanced	0.9972	0.78	0.69	0.73	0.92
	Balanced	0.9905	0.90	0.88	0.89	0.96
Naïve Bayes	Imbalanced	0.9968	0.70	0.60	0.65	0.88
	Balanced	0.9888	0.88	0.86	0.87	0.95
One-Class Random Forest	Imbalanced	0.9960	0.68	0.58	0.62	0.89
	Balanced	0.9870	0.86	0.84	0.85	0.94

Figure 4 illustrates the comparative performance of the evaluated machine learning models under the 70%–30% train–test split. The models are assessed using four evaluation metrics: Accuracy, Precision, Recall, and F1-score.

Overall, all models achieve high Accuracy values (above 0.95), indicating strong general classification performance. However, more noticeable differences appear in Precision, Recall, and F1-score, which better reflect fraud detection capability. XGBoost demonstrates the best overall performance, achieving the highest Precision, Recall, and F1-score, followed closely by Random Forest. SVM and KNN show moderate performance, while Logistic Regression and Naïve Bayes exhibit lower Recall and F1-score values, suggesting weaker detection of minority (fraudulent) cases.

These results confirm that ensemble-based models, particularly XGBoost and Random Forest, provide more robust and balanced performance under the 70%–30% data split configuration.

4.2 Comparative Performance Analysis under the 50%–50% Train–Test Split

This section presents the experimental findings of the proposed fraud detection framework, evaluated under a 50%–50% train–test split. The analysis is first dedicated to examining the impact of the GAN-based data balancing strategy on class distribution and overall model robustness. Subsequently, a comprehensive comparative evaluation of the selected machine learning models is conducted, focusing on detection performance metrics, including Accuracy, Precision, Recall, F1-Score, and AUC-ROC, as well as computational efficiency. This balanced split setting enables a rigorous assessment of the generalization capability of the models while ensuring sufficient training and testing samples for reliable performance validation.

Table 3 presents a comparative analysis of the evaluated machine learning models on the original imbalanced dataset and the GAN-balanced dataset using a 50%–50%

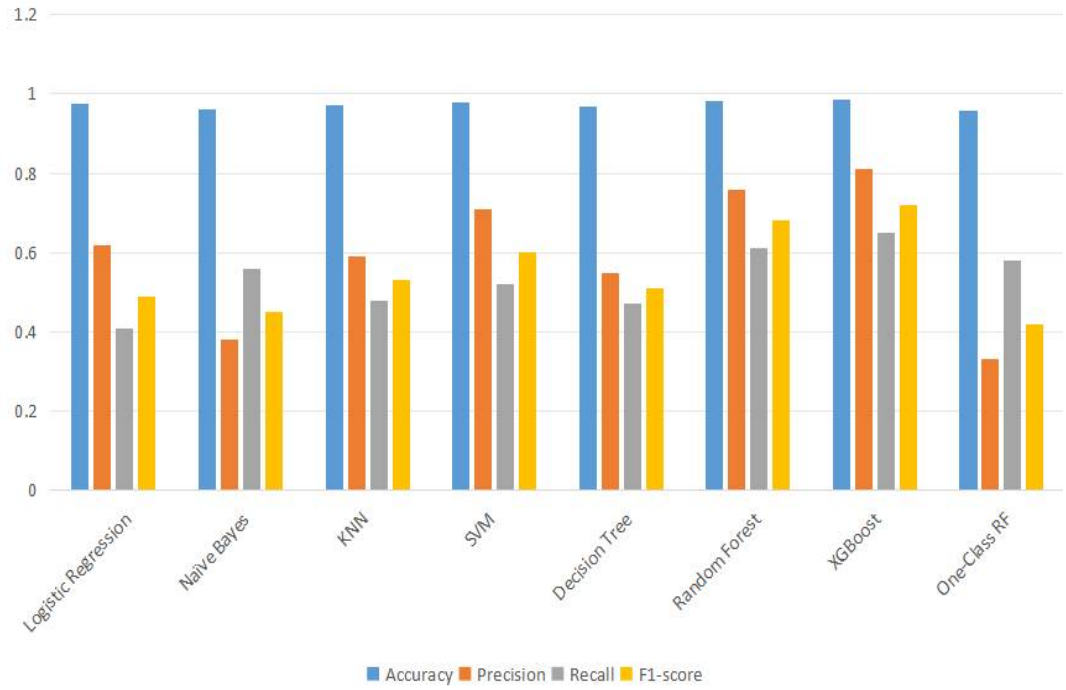

Fig. 4 Performance Comparison of Evaluated Models Using Accuracy, Precision, Recall, and F1-Score under the 70%–30% Train–Test Configuration.

train–test split. The results clearly demonstrate the significant impact of data balancing on fraud detection performance. While models trained on the imbalanced dataset achieve high Accuracy values, their Precision, Recall, and F1-Score remain relatively low, indicating limitations in detecting minority (fraudulent) instances.

After applying the GAN-based balancing technique, a substantial improvement is observed across all minority-class evaluation metrics. In particular, XGBoost and Random Forest achieve the highest overall performance, with notable gains in Recall, F1-Score, and AUC-ROC, confirming their strong capability to generalize under balanced conditions. Although a slight decrease in overall Accuracy is observed in some cases, the improvement in fraud detection effectiveness highlights the importance of addressing class imbalance. Overall, the balanced dataset leads to more reliable and robust detection performance across all evaluated models.

Figure 5 presents the comparative performance of the evaluated machine learning models using a 50%–50% train–test split. The models are assessed based on Accuracy, Precision, Recall, and F1-score.

Under this configuration, the performance differences between models become more balanced and consistent. XGBoost achieves the highest overall results, with Accuracy, Precision, Recall, and F1-score all approaching or exceeding 0.96–0.97, confirming its robustness. Random Forest follows closely, also demonstrating strong and stable performance across all metrics. SVM and KNN show competitive results, while Logistic

Table 3 Comparative Model Performance on Original Imbalanced vs. GAN-Balanced Datasets

Model	Dataset	Accuracy	Precision	Recall	F1-Score	AUC-ROC
XGBoost	Imbalanced	0.985	0.81	0.65	0.72	0.92
	GAN-Balanced	**0.968**	**0.97**	**0.97**	**0.97**	**0.99**
Random Forest	Imbalanced	0.982	0.76	0.61	0.68	0.89
	GAN-Balanced	0.957	0.95	0.96	0.96	0.98
SVM	Imbalanced	0.978	0.71	0.52	0.60	0.85
	GAN-Balanced	0.941	0.93	0.94	0.94	0.97
KNN	Imbalanced	0.973	0.59	0.48	0.53	0.81
	GAN-Balanced	0.928	0.91	0.93	0.92	0.96
Logistic Regression	Imbalanced	0.975	0.62	0.41	0.49	0.78
	GAN-Balanced	0.914	0.90	0.92	0.91	0.95
Decision Tree	Imbalanced	0.969	0.55	0.47	0.51	0.80
	GAN-Balanced	0.923	0.91	0.92	0.91	0.95
Naïve Bayes	Imbalanced	0.961	0.38	0.56	0.45	0.74
	GAN-Balanced	0.887	0.85	0.90	0.87	0.93
One-Class RF	Imbalanced	0.957	0.33	0.58	0.42	0.76
	GAN-Balanced	0.901	0.88	0.91	0.89	0.94

Regression and Decision Tree maintain acceptable but comparatively lower scores. Naïve Bayes and One-Class Random Forest exhibit relatively lower Precision and F1-score values.

Compared to the 70%–30% split, the 50%–50% configuration produces more stable and harmonized metric values, further highlighting the superiority of ensemble-based methods, particularly XGBoost and Random Forest, in fraud detection tasks.

4.3 Execution Time Analysis

Table 4 presents the comparative execution time (in milliseconds) of different machine learning models under three transaction-load scenarios: 1 transaction (1 Tx), 10 transactions (10 Tx), and 30 transactions (30 Tx). In addition, the percentage of performance degradation is reported to evaluate scalability as the transaction volume increases.

Table 4 Latency Degradation Under Increasing Concurrent Transactions

Model	1 Tx (ms)	10 Tx (ms)	30 Tx (ms)	Degradation (%) ↓
Logistic Regression	0.12	0.35	0.98	716
Naïve Bayes	0.09	0.28	0.81	800
KNN	1.84	6.92	18.45	902
SVM	2.31	7.85	21.10	813
Random Forest	0.65	2.10	5.90	808
XGBoost	0.50	1.98	2.90	**480**

When processing a single transaction, Naïve Bayes and Logistic Regression respond in just 0.09 ms and 0.12 ms respectively, which is remarkably fast and makes them

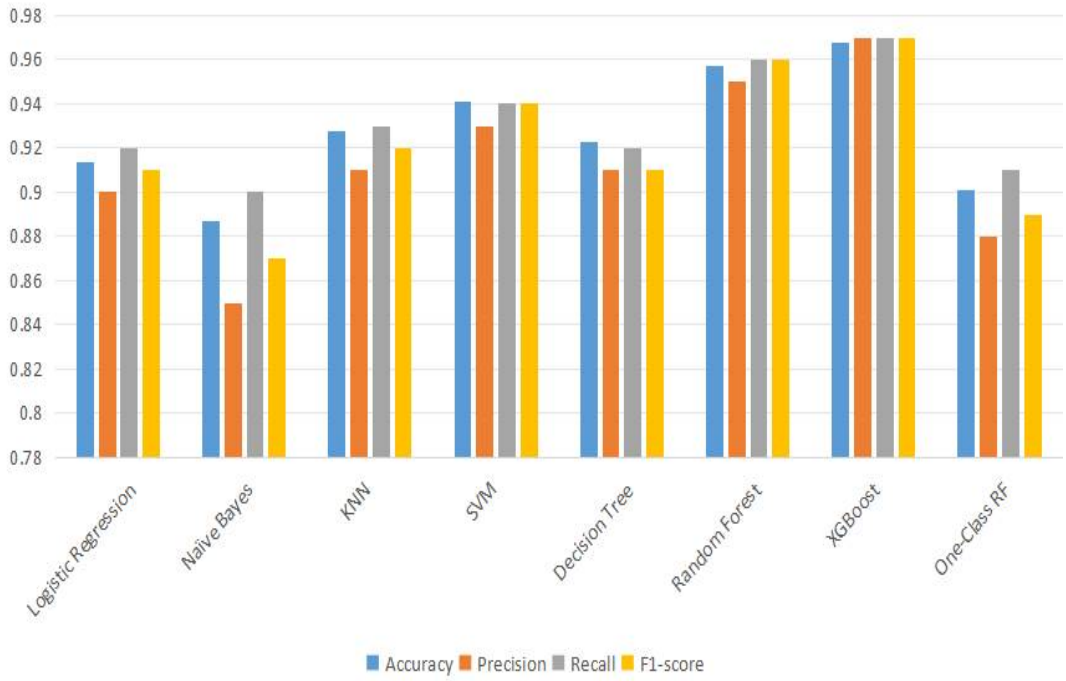

Fig. 5 Performance Comparison of Evaluated Models Using Accuracy, Precision, Recall, and F1-Score under the 50%–50% Train–Test Configuration.

a natural fit for systems where every millisecond counts. XGBoost is not far behind at 0.50 ms, and while that may seem slower in isolation, it is worth noting that it is the quickest among all ensemble-based models and already ahead of both KNN and SVM. Random Forest sits in the middle ground at 0.65 ms, whereas KNN and SVM lag noticeably at 1.84 ms and 2.31 ms, a reflection of the heavier computations they rely on internally.

Things get more interesting as the load grows. When scaling up to 10 and then 30 simultaneous transactions, every model slows down, which is expected. What is less expected is just how sharply KNN and SVM deteriorate, climbing to 18.45 ms and 21.10 ms respectively at 30 transactions. XGBoost, on the other hand, holds remarkably steady, topping out at only 2.90 ms under the same conditions. That is a striking result, especially when compared to Random Forest at 5.90 ms, and it suggests that XGBoost handles growing workloads with a kind of efficiency that its peers simply cannot match.

The degradation figures tell the same story in percentage terms. KNN degrades the most dramatically at 902%, while SVM, Random Forest, and Naïve Bayes all cluster in the 800% range, and Logistic Regression comes in at 716%. XGBoost, however, stands clearly apart with a degradation rate of just 480%, the lowest across the board. In practical terms, this means that as transaction volumes scale up, XGBoost's performance holds together far better than any other model tested.

Taken together, these findings paint a clear picture. XGBoost is not just another strong performer; it is the model that best combines speed with the ability to stay fast under pressure. Logistic Regression and Naïve Bayes do edge it out when processing a single transaction, but that advantage narrows quickly as load increases. For any system expected to handle fraud detection at scale and in real time, XGBoost emerges as the most dependable option. KNN and SVM, while computationally valid in other contexts, show too much strain under load to be practical choices here.

Stepping back to look at the full picture, ensemble models, XGBoost in particular alongside Random Forest, deliver the strongest predictive results regardless of how the data is split, whether at 70/30 or 50/50. Metrics like Recall, F1-score, and AUC-ROC, which matter most in fraud detection where missing a fraudulent transaction carries real consequences, consistently favour these models. Balancing the dataset also proves worthwhile, noticeably improving how well models detect the minority fraud class. When factoring in both detection quality and computational behaviour, XGBoost is the model that holds up best across every dimension tested. Logistic Regression remains a reasonable choice in scenarios where raw speed on a single transaction is the only concern, but for most real-world fraud detection deployments, XGBoost offers a combination of accuracy, efficiency, and resilience that none of the other models quite achieve.

5 Conclusion

This study presented an enhanced credit card fraud detection framework that combines GAN-based data balancing with ensemble machine learning techniques. By addressing the critical issue of class imbalance, the proposed approach significantly improved the detection of fraudulent transactions without compromising computational efficiency. The experimental results demonstrated that ensemble models, particularly XGBoost and Random Forest, consistently outperformed traditional classifiers in terms of precision, recall, F1-score, and AUC-ROC.

The integration of a transaction simulator further validated the robustness of the models under real-time and concurrent transaction scenarios, highlighting the practical applicability of the proposed solution. Notably, the GAN-based balancing strategy led to a substantial increase in recall, ensuring better identification of fraudulent activities while maintaining high overall accuracy.

Overall, the findings confirm that combining synthetic data generation with advanced ensemble learning methods provides a reliable, scalable, and efficient solution for real-world fraud detection systems. Future work may focus on incorporating adaptive learning mechanisms and explainable AI techniques to further enhance transparency, robustness, and resilience against evolving fraud patterns.

Acknowledgements. Not applicable

Data Availability Statement. The dataset used in this study contains sensitive and proprietary banking information. Due to privacy, confidentiality, and legal restrictions associated with personal financial data, the dataset cannot be publicly shared. Access to the data may be restricted and is not available for public distribution.

Authors Contributions. Adel EL QASSOUARI contributed to the conceptualization, methodology, implementation, data analysis, and manuscript writing. Ayoub CHAREF contributed to the methodology design, machine learning framework development, validation, experimental analysis, and manuscript revision. Zahi JARIR contributed to supervision, validation of the research methodology, critical review, and final manuscript editing. All authors read and approved the final manuscript.

Funding. The authors did not receive any specific funding for this work.

Declarations

Ethical Approval. Not applicable.

Consent to Participate. Not applicable.

Consent to Publish. Not applicable.

Competing Interests. The authors declare that they have no competing interests.

References

[1] Zhu, M., Zhang, Y., Gong, Y., Xu, C., Xiang, Y.: Enhancing credit card fraud detection: A neural network and smote integrated approach. Journal of Theory and Practice of Engineering Science **4**(02), 23–30 (2024) https://doi.org/10.53469/jtpes.2024.04(02).04

[2] Chen, Y.: A comparative study of machine learning models for credit card fraud detection. Academic Journal of Natural Science **2**(4), 11–18 (2025) https://doi.org/10.70393/616a6e73.333530

[3] Mosa, D.T., Sorour, S.E., Abohany, A.A., Maghraby, F.A.: Ccfd: Efficient credit card fraud detection using meta-heuristic techniques and machine learning algorithms. Mathematics **12**(14), 2250 (2024) https://doi.org/10.3390/MATH12142250

[4] Khalid, A.R., Owoh, N., Uthmani, O., Ashawa, M., Osamor, J., Adejoh, J.: Enhancing credit card fraud detection: An ensemble machine learning approach. Big Data and Cognitive Computing **8**(1), 6 (2024) https://doi.org/10.3390/BDCC8010006

[5] Cherif, A., Badhib, A., Ammar, H., Alshehri, S., Kalkatawi, M., Imine, A.: Credit card fraud detection in the era of disruptive technologies: A systematic review. Journal of King Saud University - Computer and Information Sciences **35**(1), 145–174 (2023) https://doi.org/10.1016/J.JKSUCI.2022.11.008

[6] Homepage, J., Unogwu, O.J., Filali, Y.: Fraud detection and identification in credit card based on machine learning techniques. Wasit Journal of Computer and Mathematics Science **2**(3), 16–22 (2023) https://doi.org/10.31185/WJCMS.185

[7] al., S.F.F.: Enhancing credit card fraud detection: A comprehensive study of machine learning algorithms and performance evaluation. Journal of Business and Management Studies **6**(3), 252–259 (2024) https://doi.org/10.32996/JBMS.2024.6.13.21

[8] Nuthalapati, A.: Smart fraud detection leveraging machine learning for credit card security. Educational Administration: Theory and Practice **2024**(2), 433–443 (2023) https://doi.org/10.53555/kuey.v29i2.6907

[9] al., F.O.A.: Enhancing the random forest model via synthetic minority oversampling technique for credit-card fraud detection. Journal of Computing Theories and Applications **1**(4), 407–420 (2024) https://doi.org/10.62411/JCTA.10323

[10] Hafez, I.Y., Hafez, A.Y., Saleh, A., El-Mageed, A.A.A., Abohany, A.A.: A systematic review of ai-enhanced techniques in credit card fraud detection. Journal of Big Data **12**(1), 6 (2025) https://doi.org/10.1186/S40537-024-01048-8

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

[12] Bakhtiari, S., Nasiri, Z., Vahidi, J.: Credit card fraud detection using ensemble data mining methods. Multimedia Tools and Applications **82**(19), 29057–29075 (2023) https://doi.org/10.1007/S11042-023-14698-2

[13] Noviandy, T.R., Idroes, G.M., Maulana, A., Hardi, I., Ringga, E.S., Idroes, R.: Credit card fraud detection for contemporary financial management using xgboost-driven machine learning and data augmentation techniques. Indatu Journal of Management and Accounting **1**(1), 29–35 (2023) https://doi.org/10.60084/IJMA.V1I1.78

[14] Mienye, I.D., Sun, Y.: A machine learning method with hybrid feature selection for improved credit card fraud detection. Applied Sciences **13**(12), 7254 (2023) https://doi.org/10.3390/APP13127254

[15] Seera, M., Lim, C.P., Kumar, A., Dhamotharan, L., Tan, K.H.: An intelligent payment card fraud detection system. Annals of Operations Research **334**(1), 445–467 (2021) https://doi.org/10.1007/S10479-021-04149-2

[16] Sulaiman, S.S., Nadher, I., Hameed, S.M.: Credit card fraud detection using improved deep learning models (2023) https://doi.org/10.32604/cmc.2023.046051

[17] Adejoh, J., Owoh, N., Ashawa, M., Hosseinzadeh, S., Shahrabi, A., Mohamed, S.: An adaptive unsupervised learning approach for credit card fraud detection. Big Data and Cognitive Computing **9**(9), 217 (2025) https://doi.org/10.3390/BDCC9090217

[18] Karthika, J., Senthilselvi, A.: Smart credit card fraud detection system based on dilated convolutional neural network with sampling technique. Multimedia Tools and Applications **82**(20), 31691–31708 (2023) https://doi.org/10.1007/S11042-023-15730-1

[19] Bao, Q., Wei, K., Xu, J., Jiang, W.: Application of deep learning in financial credit card fraud detection. Journal of Economic Theory and Business Management **1**(2), 51–57 (2024) https://doi.org/10.5281/zenodo.10960092

[20] Tang, Y., Liang, Y.: Credit card fraud detection based on federated graph learning. Expert Syst. Appl. **256**, 124979 (2024) https://doi.org/10.1016/J.ESWA.2024.124979

[21] Chatterjee, P., Das, D., Rawat, D.B.: Digital twin for credit card fraud detection: opportunities, challenges, and fraud detection advancements. Future Generation Computer Systems **158**, 410–426 (2024) https://doi.org/10.1016/J.FUTURE.2024.04.057

[22] Salam, M.A., Fouad, K.M., Elbably, D.L., Elsayed, S.M.: Federated learning model for credit card fraud detection with data balancing techniques. Neural Computing and Applications **36**(11), 6231–6256 (2024) https://doi.org/10.1007/S00521-023-09410-2

[23] Mathew, J.C., Nithya, B., Vishwanatha, C.R., Shetty, P., Priya, H., Kavya, G.: An analysis on fraud detection in credit card transactions using machine learning techniques. In: Proceedings of the 2nd International Conference on Artificial Intelligence and Smart Energy (ICAIS 2022), pp. 265–272 (2022). https://doi.org/10.1109/ICAIS53314.2022.9742830

[24] Ahmad, H., Kasasbeh, B., Aldabaybah, B., Rawashdeh, E.: Class balancing framework for credit card fraud detection based on clustering and similarity-based selection (sbs). International Journal of Information Technology **15**(1), 325–333 (2022) https://doi.org/10.1007/S41870-022-00987-W

[25] Kennedy, R.K.L., Villanustre, F., Khoshgoftaar, T.M., Salekshahrezaee, Z.: Synthesizing class labels for highly imbalanced credit card fraud detection data. Journal of Big Data **11**(1), 38 (2024) https://doi.org/10.1186/S40537-024-00897-7

[26] Strelcenia, E., Prakoonwit, S.: A survey on gan techniques for data augmentation to address the imbalanced data issues in credit card fraud detection. Machine Learning and Knowledge Extraction **5**(1), 304–329 (2023) https://doi.org/10.3390/MAKE5010019

[27] Devi, P.N., Mounika, K., Navanish, N., Reddy, S.B.: Credit card fraud detection using logistic regression. In: Applications of AI in Smart Technologies and Manufacturing, pp. 88–96 (2025). https://doi.org/10.1201/9781003682325-11

[28] Baisholan, N., Dietz, J.E., Gnatyuk, S., Turdalyuly, M., Matson, E.T., Baisholanova, K.: A systematic review of machine learning in credit card fraud detection under original class imbalance. Computers **14**(10), 437 (2025) https://doi.org/10.3390/COMPUTERS14100437

[29] Btoush, E.A.L.M., Zhou, X., Gururajan, R., Chan, K.C., Genrich, R., Sankaran, P.: A systematic review of literature on credit card cyber fraud detection using machine and deep learning. PeerJ Comput. Sci. **9**, 1278 (2023) https://doi.org/10.7717/PEERJ-CS.1278/TABLE-7

[30] Sulaiman, S.S., Nadher, I., Hameed, S.M.: Credit card fraud detection challenges and solutions: A review. Iraqi Journal of Science **65**(4), 2287–2303 (2024) https://doi.org/10.24996/IJS.2024.65.4.42

[31] Alraddadi, A.S.: A survey and a credit card fraud detection and prevention model using the decision tree algorithm. Engineering, Technology & Applied Science Research **13**(4), 11505–11510 (2023) https://doi.org/10.48084/ETASR.6128

[32] al., G.M.P.: The role of diversity and ensemble learning in credit card fraud detection. Advances in Data Analysis and Classification **18**(1), 193–217 (2022) https://doi.org/10.1007/S11634-022-00515-5

[33] Fisher, R.A.: The use of multiple measurements in taxonomic problems. Annals of Eugenics **7**(2), 179–188 (1936) https://doi.org/10.1111/j.1469-1809.1936.tb02137.x

[34] Hastie, T., Tibshirani, R., Friedman, J.: The Elements of Statistical Learning: Data Mining, Inference, and Prediction, 2nd edn. Springer, ??? (2009)

[35] He, H., Garcia, E.A.: Learning from imbalanced data. IEEE Transactions on Knowledge and Data Engineering **21**(9), 1263–1284 (2009) https://doi.org/10.1109/TKDE.2008.239

[36] Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., Courville, A., Bengio, Y.: Generative adversarial nets. Advances in Neural Information Processing Systems **27**, 2672–2680 (2014)

[37] Frid-Adar, M., Klang, E., Amitai, M., Goldberger, J., Greenspan, H.: Gan-based synthetic medical image augmentation for increased cnn performance in liver lesion classification. Neurocomputing **321**, 321–331 (2018) https://doi.org/10.1016/j.neucom.2018.09.013

[38] Leevy, J.L., Hancock, J., Khoshgoftaar, T.M.: Comparative analysis of binary and one-class classification techniques for credit card fraud data. Journal of Big Data **10**(1), 118 (2023)

[39] Sinap, V.: Comparative analysis of machine learning techniques for credit card

fraud detection: Dealing with imbalanced datasets. Turkish journal of engineering **8**(2), 196–208 (2024)

[40] Awoyemi, J.O., Adetunmbi, A.O., Oluwadare, S.A.: Credit card fraud detection using machine learning techniques: A comparative analysis. In: 2017 International Conference on Computing Networking and Informatics (ICCNI), pp. 1–9 (2017). IEEE

[41] Borketey, B.: Real-time fraud detection using machine learning. Borketey, B.(2024) Real-Time Fraud Detection Using Machine Learning. Journal of Data Analysis and Information Processing **12**, 189–209 (2024)