UMYU Scientifica, Vol. 4 NO. 4, December 2025, Pp 137 – 153
https://doi.org/10.56919/usci.2544.012
ISSN: 2955 – 1145 (print); 2955 – 1153 (online)
A periodical of the Faculty of Natural and Applied Sciences, UMYU, Katsina

ORIGINAL RESEARCH ARTICLE

Ensemble of Synthetic and Balancing Techniques for Credit Card Fraud Detection

Abdulmalik Umar[1]*, Armaya'u Umar Zango[2], Yakubu Ibrahim Galadima[1] and Mardiyya Lawal Bagiwa[1]

[1]Department of Computer Science and Information Technology, Al-Qalam University, Katsina, Nigeria
[2]Department of Software Engineering and Cyber Security, Al-Qalam University, Katsina, Nigeria

ABSTRACT

Globally, credit card fraud continues to afflict the financial sector, with losses projected to exceed $43 billion in the near future. The rise of online payments has introduced new challenges in identifying fraud due to a significant class imbalance in fraud data. Most traditional machine learning methods do not account for this imbalance, leading to biased classifiers with limited generalisability. The study proposes an innovative ensemble-based technique, EA-CT (Ensemble ADASYN-CTGAN), that combines ADASYN and Conditional Tabular Generative Adversarial Networks (CTGAN) to improve the accuracy and robustness of fraud detection. To execute the experiments, publicly available European credit card transactions obtained by Université Libre de Bruxelles (ULB) have been used; this dataset consists of 284,807 transactions with 0.172% fraud cases. To make our model computationally efficient and to avoid class imbalance issues during training and testing. The study begins by addressing the class imbalance problem using ADASYN, an oversampling technique that generates synthetic samples for the minority class, focusing on more difficult instances to learn. Subsequently, the dataset is enriched with realistic synthetic fraud samples via CTGAN to ensure diversity and representativeness. A 10% stratified sample of the dataset is then used after various preprocessing and augmentation techniques. The sampled dataset is then analyzed using various classifiers. Results indicate that fraud detection capabilities are improved with the EA-CT framework. XGBoost achieved the highest AUC-ROC of 95.20% and a recall of 87.99%. ANN attained the highest accuracy of 88.86%. KNN demonstrated an impressive F1-score of 87.05%. Despite these gains, the study is limited by its reliance on a single, anonymized dataset and a binary classification setting, which may affect its generalizability to real-world, evolving fraud scenarios. Overall, the proposed EA-CT framework demonstrates strong potential for enhancing fraud detection in highly imbalanced financial datasets.

ARTICLE HISTORY

Received June 23, 2025
Accepted December 03, 2025
Published December 30, 2025

KEYWORDS

Fraud Detection, Class Imbalance, Ensemble Learning, Synthetic Data, Data Augmentation

INTRODUCTION

Credit card fraud, defined as the deceptive and unauthorized use of credit card information for personal gain (Maina, 2023), has emerged as one of the most critical threats to global financial systems. Between 2020 and 2021, worldwide credit card fraud losses surged by 10%, marking the steepest rise since 2018, with merchants and card acquirers losing an estimated USD 30 billion; the United States alone accounted for USD 12 billion of these losses (Statista, 2024). As digital payments have become deeply embedded in modern commerce, fraud opportunities have expanded. Cash now represents only 20% of in-person transactions, while more than 2.8 billion credit cards are in circulation globally, greatly increasing exposure to fraudulent activity (Merchant Cost Consulting, 2024). Fraudsters continually exploit vulnerabilities in digital payment systems using increasingly sophisticated methods, thereby posing severe financial risks to merchants, cardholders, and financial institutions.

In response, a substantial body of research has focused on the application of machine learning (ML) to credit card fraud detection (Azarm *et al.*, 2024; Ghaleb *et al.*, 2023; Goran, 2023). Supervised learning algorithms have demonstrated strong potential for identifying fraudulent activities; however, their performance is heavily constrained by the significant class imbalance inherent in fraud datasets, where legitimate transactions vastly outnumber fraudulent ones (Thennakoon *et al.*, 2019; Majeed and Hwang, 2023; Ghaleb *et al.*, 2023). This imbalance leads to biased classifiers that fail to detect rare fraudulent cases, often resulting in high false-negative rates. The issue has drawn increased research attention, as fraudulent transactions and payment defaults continue to

Correspondence: Abdulmalik Umar. Department of Computer Science and Information Technology, Al-Qalam University, Katsina, Nigeria. ✉ abdulmalikumar1997@gmail.com.
How to cite: Umar, A., Umar, A. Z., Ibrahim, Y. G., & Lawal, M. B. (2025). Ensemble of Synthetic and Balancing Techniques for Credit Card Fraud Detection. *UMYU Scientifica*, 4(4), 137 – 153. https://doi.org/10.56919/usci.2544.012

rise annually (Seera *et al.*, 2024; Almazroi and Ayub, 2023). The combination of overlapping classes, skewed distributions, and scarcity of fraud samples presents critical challenges that impede accurate detection (Wu and Wang, 2021; Majeed and Hwang, 2023).

Class imbalance occurs when the majority and minority classes are disproportionately represented, and in real-world fraud detection, the imbalance can be extreme, ranging from 100:1 to 10,000:1 (Leevy *et al.*, 2018). To counter this, researchers have explored various data preprocessing strategies, oversampling, undersampling, and synthetic data generation, to improve classifier performance (Sadgali *et al.*, 2020; Nguyen *et al.*, 2020). Leevy *et al.* (2018) reviewed eight years of studies addressing this issue and categorized solutions into data-level methods (e.g., sampling), algorithm-level techniques (e.g., cost-sensitive learning), and hybrid approaches. Fernandez *et al.* (2017) and others (Lopez *et al.*, 2013; Krawczyk *et al.*, 2016) further identified three major categories of methods: pre-processing for data rebalancing (Batista *et al.*, 2004), algorithmic modifications for handling imbalance (Ramentol *et al.*, 2015), and cost-sensitive learning that penalizes misclassification of minority classes (Domingos, 1990; Lopez *et al.*, 2013).

SMOTE remains one of the most widely used oversampling techniques (Mienye and Sun, 2023; Ghaleb *et al.*, 2023), but it suffers from multiple limitations including oversampling noise, limited diversity in generated samples, and susceptibility to overfitting. ADASYN, proposed by He *et al.* (2008), improved on SMOTE by adaptively generating synthetic samples focused on complex minority regions, thereby improving decision boundaries and classifier performance. Other researchers have developed hybrid fraud detection strategies; for example, Azarm *et al.* (2024) combined personalized PageRank with SVMs to model social relationships among accounts, although computational demands and overfitting remain challenges.

Traditional ML models often struggle with imbalanced or small-sample data (Niaz *et al.*, 2022), leading to increased interest in deep learning (DL), which excels at modeling complex nonlinear patterns (Mienye and Jere, 2024). Several DL architectures have been proposed, including LSTM-based fraud detectors that leverage the sequential nature of transaction data (Benchaji *et al.*, 2021; Roseline *et al.*, 2022), though such sequence-dependent models may suffer in nonsequential fraud scenarios. BiLSTM-BiGRU architectures (Najadat *et al.*, 2020) demonstrated strong performance with oversampling but require long training times, limiting their real-time application. Ensemble LSTM models have also shown promise (Forough and Momtazi, 2021), though dependency on sequential data limits generalizability.

Deep learning models also face challenges related to interpretability and computational cost, especially in high-stakes domains such as fraud detection (Roy *et al.*, 2018). Recent work has emphasized model explainability, improved data augmentation, and synthetic data generation to overcome these limitations (Wu and Wang, 2021). The scarcity of publicly available fraud datasets further motivates the use of generative models such as GANs (Wang *et al.*, 2023), which expand training data by producing realistic synthetic samples. Ensemble techniques have similarly gained traction as a way to improve robustness on noisy and imbalanced datasets (Dong *et al.*, 2019). Hybrid incremental ensemble learning (HIEL), proposed by Yu *et al.* (2017), showed strong performance on noisy datasets through combined feature selection, bagging, and weighted voting.

Other studies have examined the effects of time inhomogeneity on fraud detection; Hsin *et al.* (2022) found that improper temporal sampling can distort evaluation metrics and that GAN-based augmentation outperforms SMOTE in such contexts. Further combining autoencoders and GANs, Goran (2023) demonstrated reduced false negatives and improved classifier behavior. Novel augmentation strategies such as CTGAN-MOS (Majeed and Hwang, 2023) significantly improved performance across major evaluation metrics by integrating synthetic data generation with noise mitigation. Techniques like PCA-GAN fusion have also been explored (Wang *et al.*, 2023), though their applicability to tabular fraud data remains limited.

Ensemble deep learning approaches continue to show strong potential. For example, Mienye and Sun (2023) combined LSTM-GRU ensembles with SMOTE-ENN resampling to achieve high sensitivity and specificity, though computational demands remain high. Ghaleb *et al.* (2023) achieved performance improvements with a weighted ensemble of random forest classifiers and reported a 0% false alarm rate. Collectively, existing research demonstrates substantial progress but also reveals persistent gaps. Current methods often struggle with class imbalance, data scarcity, interpretability, and real-time scalability. These limitations highlight the need for more robust, adaptive ensemble frameworks that integrate data balancing and augmentation to improve fraud detection accuracy and reliability in highly imbalanced environments; hence, the aim of this study and its contribution to knowledge.

An act of deception that an entity or a person commits is termed 'Fraud'. Credit card fraud occurs when an entity or person commits fraud using a credit card, knowing that it may result in benefits that are adverse to the individual or others (Maina, 2023). Between the years 2020 and the year 2021, card fraud losses exponentially rocketed worldwide by a staggering 10%, this marked the biggest increase since the year 2018, with an estimate of 30 billion dollars by merchants, including card acquirers also, with the United States accounting for about 12 billion of those losses (Statista, 2024).

Credit card transactions have become ubiquitous in the digital landscape, bringing with them the ever-present challenge of credit card fraud. Fraudsters have devised various methods to illegally obtain card information and

use it to make unauthorized purchases, posing a significant threat to credit card companies and merchants.

The widespread acceptance of credit and debit cards has changed how we process payments. Cash accounts for just 20% of all in-person transactions, indicating that plastic and digital wallets have penetrated far beyond the e-commerce space. But with roughly 2.8 billion credit cards in circulation worldwide, the opportunity for fraudsters to exploit this trend has never been greater (Merchant Cost Consulting, 2024).

A number of machine learning techniques have been put forth to detect credit card fraud (Azarm *et al.*, 2024; Ghaleb *et al.*, 2023; Goran, 2023). In particular, supervised learning algorithms have been shown to be highly effective at identifying credit card fraud. The mitigating factor would be to address the class imbalance (Thennakoon *et al.*, 2019; Majeed & Hwang, 2023; Ghaleb *et al.*, 2023).

The literature on credit card fraud detection and payment defaults has seen significant attention in recent years due to rising concerns over financial losses and security breaches associated with electronic payments (Seera *et al.*, 2024; Almazroi & Ayub, 2023). Credit card transactions, a prominent form of e-payment, are susceptible to fraud and defaults, which continue to escalate annually. Researchers have explored various machine learning techniques to address these challenges, focusing particularly on the inherent issues in credit card data, namely imbalanced class distributions and overlapping classes (Wu & Wang, 2021; Majeed & Hwang, 2023).

These issues pose substantial hurdles to accurately detecting fraudulent transactions and payment defaults, which are crucial for both card issuers and holders. Efforts have been made to develop effective methodologies to mitigate these challenges, leveraging techniques such as deep learning, ensemble learning, and specialized sampling methods to enhance detection capabilities (Ghaleb *et al.*, 2023; Kalid, 2024).

Any dataset, with unequal distribution between its classes, being the majority and minority classes, can be considered to have class imbalance; the severity of class imbalance in real-world applications, might vary from minor to severe (high or extreme). A dataset can be considered imbalanced if the classes, e.g., fraud and non-fraud cases, are not equally represented (Leevy, 2018).

Addressing the class imbalance inherent in credit card fraud data requires effective data preprocessing and balancing strategies. Techniques such as oversampling, undersampling, and synthetic data generation have been explored to improve the model's ability to accurately identify fraudulent transactions (Sadgali *et al.*, 2020; Nguyen et al., 2020).

A large survey of published studies within eight 8 years was reviewed by Leevy *et al.* (2018), that focused on high-class imbalance, where a majority-to-minority class ratio between 100:1 and 10,000:1, in big data, in order to assess adverse effects due to class imbalance. They covered two techniques, including Data-Level (e.g., data sampling) and Algorithm-Level (e.g., cost-sensitive and hybrid/ensemble) Methods. Data sampling methods are popular for addressing class imbalance, with random over-sampling methods generally yielding better overall results.

Fernandez *et al.* (2017) provided insight into imbalanced big data classification. To successfully address imbalanced classification, a number of solutions have been proposed, which mainly fall into three categories (Lopez *et al.*, 2013; Krawczyk *et al.*, 2016). The first category is pre-processing techniques that aim to rebalance the training data (Batista *et al.*, 2004). The second one concerns algorithmic approaches that alter the learning mechanism by accounting for different class distributions (Ramentol *et al.*, 2015). The third category comprises cost-sensitive learning approaches that assign different costs to the misclassification of each class (Domingos, 1990; Lopez *et al.*, 2013).

Numerous techniques have been employed to balance the number of fraudulent samples. Traditionally, the Synthetic Minority Oversampling Technique (SMOTE) has been widely used (Mienye & Sun, 2023; Ghaleb *et al.*, 2023). However, SMOTE has several limitations: it tends to oversample noisy data; the accuracy of nearest-neighbour selection depends on the data at hand; it may oversample uninformative samples; and it focuses on local information, resulting in a less diverse set of samples. Most models will overfit to the minority class because the same examples appear so often.

In machine learning, dealing with imbalanced datasets has been a significant challenge, as we have seen above. A novel approach to address this issue was presented by He *et al.* (2008). The Adaptive Synthetic Sampling (ADASYN) technique has gained attention for its ability to improve learning from such data. ADASYN uses a weighted sampling scheme to generate synthetic data for minority-class examples, placing greater weight on the harder-to-learn samples. This strategy ensures that the synthetic data generated for difficult examples helps the model better understand the challenging areas of the minority class. Consequently, ADASYN enhances classification by not only reducing the bias introduced by class imbalances but also by adaptively adjusting the decision boundary towards these difficult examples. Studies and simulations across various machine learning datasets have demonstrated ADASYN's effectiveness across multiple evaluation metrics, highlighting its potential to significantly improve classification performance in imbalanced learning tasks.

An instance of employing an approach to credit card fraud detection was done by Azarm *et al.* (2024), they explored a hybrid approach using the personalized PageRank (PPR) algorithm to capture the social dynamics of fraud by analyzing relationships between financial accounts, the proposed approach showed good percent of performance support vector machine (SVM) shows good accuracy in the proposed approach by classifying the test data to fraud and legal respectively. However, the algorithm used

requires more memory and storage than other classifiers and is prone to overfitting.

Traditional machine learning (ML) models have been widely used in credit card fraud detection due to their simplicity and interpretability. However, their performance tends to degrade significantly when applied to imbalanced datasets or when the number of fraud samples is insufficient to represent the underlying distribution (Niaz *et al.*, 2022). Nonetheless, the inherent assumptions and learning mechanisms of traditional models make them less suitable for highly imbalanced and small-sample problems than more flexible approaches such as deep learning methods.

Deep learning (DL), a branch of machine learning (ML), is the core technology in today's technological advancements and innovations. Deep learning-based approaches are state-of-the-art methods for analyzing and detecting complex patterns in large datasets, such as credit card transactions (Mienye & Jere, 2024).

Benchaji *et al.* (2021) developed a CCFD model by sequentially modelling credit card data using deep long short-term memory (LSTM) neural networks and attention mechanisms. The approach accounted for the sequential nature of the credit card data and enabled the classifier to identify which transactions in the input sequence were most relevant. To ensure sequential modelling of the data, the proposed approach used an LSTM, employed an attention mechanism to improve LSTM performance, and introduced uniform manifold approximation and projection (UMAP) to select the most significant attributes. The models yielded good performance with an accuracy of 96.7%. A credit card fraud detection model was developed by Roseline *et al.* (2022) to reduce losses from credit card fraud using an LSTM model. Since models with this structure have proven successful in sequence modelling, an attention mechanism was added to improve LSTM performance. Compared with other classifiers such as SVM, naïve Bayes, and ANN, the experimental results showed that the LSTM achieved 100% accuracy. However In real-world scenarios, fraudulent behaviour may not always follow consistent sequential patterns, thereby limiting the generalizability of sequence-dependent models

In another related study, Najadat et al. (2020) developed a model based on BiLSTM and BiGRU with MaxPooling layers. The dataset was preprocessed using three resampling techniques, random oversampling, random undersampling, and SMOTE. The performance of the deep learning-based classifier and other ML classifiers, including logistic regression, random forests, voting, naïve Bayes, AdaBoost, and decision trees, was compared. When random oversampling was applied, the BiLSTM-BiGRU obtained an impressive performance, with an AUC of 91.4%. Although their model was powerful for sequence modelling, it requires long training times, which may limit its scalability and practicality in real-time fraud detection environments.

Finally, a credit card fraud detection model developed by Forough and Momtazi (2021), were they considered the sequential structure of credit card transactions. When tested on two credit card datasets, the suggested LSTM ensemble outperformed other techniques, achieving AUCs of 0.879 and 0.88 on the European and Brazilian datasets, respectively. The approach used LSTM models as base classifiers in an ensemble implementation, with a feed-forward neural network (FFNN) serving as the voting mechanism. While they achieved notable improvements, their model's reliance on sequential data limits its adaptability in fraud-detection contexts where temporal patterns are either weak, nonlinear, or absent.

However, in deep learning-based credit card fraud detection systems, model explainability is particularly important. They also lack interpretability and are attributed with high computational complexity. In high-stakes applications such as fraud detection, there is a growing emphasis on understanding the brass tacks of these models and their decision-making processes, driven by advances in machine learning (Roy *et al.*, 2018). A variety of techniques have been explored, including synthetic data generation and balancing methods, to enhance the performance of credit card fraud detection systems and address these challenges (Wu & Wang, 2021).

The evolving nature of fraudulent activities and the limited availability of public data sets present ongoing challenges that require further research and innovation (Rawat & Tiwari, 2023; Nguyen *et al.*, 2020). This issue is addressed by employing the generation of synthetic data. Generative adversarial networks (GANs), a form of synthetic data generation technique, have been explored to address data scarcity (Wang *et al.*, 2023). Researchers have expanded the available training data to build more robust, accurate fraud detection models by generating realistic synthetic data that captures the underlying patterns and characteristics of fraudulent transactions (Wang *et al.*, 2023).

Despite significant successes in knowledge discovery, traditional machine learning and deep learning methods may fail to achieve satisfactory performance when dealing with complex data, such as imbalanced, high-dimensional, and noisy data. Ensemble learning aims to integrate data fusion, data modeling, and data mining into a unified framework (Dong et al,.2019).

Consequently, Yu *et al.* (2017) proposed the hybrid incremental ensemble learning (HIEL) approach, which considers both the feature and sample spaces simultaneously. The HIEL first adopted linear discriminant analysis and the bagging technique to remove noisy attributes, generating a set of bootstrap samples and the corresponding ensemble members. Then two criteria were used, the classifier-specific criterion function and an ensemble criterion function to incrementally select classifiers. Weights for the classifiers were assigned during the same process. Finally, the label was summarised using a weighted voting scheme, which served as the classification result. Results showed that HIEL performed well on noisy datasets and outperformed most compared

classifier ensemble methods on 14 of 24 noisy real-world datasets.

In 2022, a study found that generating training/testing sets via random sampling falsely eliminates time inhomogeneity, resulting in misleading assessments of the robustness of machine learning models. These time-inhomogeneous phenomena also entailed various patterns which influenced the performance of different resampling methods for addressing data imbalance in fraud detection. Improper linear interpolation of SMOTE-related approaches led to poor performance due to varying patterns of modi operandi. However, synthesizing fraudulent samples with simple oversampling and GANs mitigated that problem (Hsin *et al.*, 2022)

Goran (2023) combined two deep learning techniques, autoencoders and generative adversarial networks. A trivial autoencoder (TAE) was used to change the data representation, and modified generative adversarial networks (GANs) were used to create new instances from random noise. The results showed that datasets balanced by the new framework influenced the classifier to change the types of prediction errors, significantly reducing false negatives. All these studies show the significant potential of combining different techniques with generative adversarial networks (GANs).

In 2023, Majeed and Hwang used a novel data augmentation scheme called CTGAN-MOS, it offered a solution to addressing the imbalance problem via six primary procedures, these key steps were employing advanced preprocessing methods for data engineering, identifying specific data vulnerabilities, generating high-quality synthetic data through the CTGAN model, intelligently integrating real and synthetic data, employing a coin-throwing algorithm to mitigate noise in augmented data, and constructing classifiers using the enhanced augmented dataset. The findings demonstrated that CTGAN-MOS significantly outperformed state-of-the-art (SOTA) methods across accuracy, recall, precision, F1 score, and G-mean. This further justifies the need for ensemble methods.

Over time, for data compression, one of the most widely used algorithms is Principal Component Analysis (PCA). In PCA, the data from the original coordinate system is converted into a new coordinate system. This method was utilized when Wang *et al.* (2023) optimized the PCA with a Generative Adversarial Network, by compressing and reducing the original data to generate the input of the confrontation network, so that the input data retains the characteristics of the original data to some extent, thereby improving the data generation performance and reducing the training time cost. They applied it to image classification, and the experimental results showed that the model effectively improved classification accuracy and enhanced model stability. However, their approach is primarily designed for image data and may not effectively generalize to tabular, highly imbalanced datasets such as those encountered in credit card fraud detection.

To address the difficulty of machine learning classifiers achieving optimal performance, Mienye and Sun (2023) presented a deep learning approach that uses an ensemble framework with long short-term memory (LSTM) and gated recurrent unit (GRU) neural networks. A hybrid synthetic minority oversampling technique and the edited nearest neighbour (SMOTE-ENN) method are used to balance the dataset's class distribution. The findings demonstrate that combining the suggested deep learning ensemble with the SMOTE-ENN method yields sensitivity and specificity of 1.000 and 0.997, respectively. Although they achieved impressive sensitivity and specificity, their approach relies heavily on deep architectures that are computationally intensive and require substantial training resources.

A proposed model by Ghaleb *et al.* (2023) was found to improve detection, reduce the cost of manual analysis, and improve performance on a fraud dataset. A set of random forest classifiers was trained using a proposed ensemble technique. The probabilistic outputs of the trained classifiers were combined using a weighted voting scheme for decision-making. The results showed that the proposed model achieved 1.9% and 3.2% improvements in overall performance and the detection rate, respectively, with a 0% false alarm rate.

The inadequacy of existing solutions highlights the urgent need for robust models that can effectively address the challenges posed by class imbalance, data scarcity, and model interpretability. To overcome these limitations, this research aims to develop an innovative ensemble technique that combines data balancing with data augmentation. The study seeks to enhance the detection of fraudulent transactions while maintaining high validation accuracy and reliability in imbalanced settings.

MATERIALS AND METHODS

3.1 Proposed ensemble technique: EA-CT

This study proposes an ensemble technique, termed EA-CT, that combines ADASYN (Adaptive Synthetic Sampling) and CTGAN (Conditional Tabular Generative Adversarial Network). Given that CTGAN requires substantial training data, directly applying it to the fraudulent dataset is ineffective due to the dataset's paucity of samples. ADASYN, a data-level resampling technique, is introduced to tackle the issue of highly imbalanced datasets. The method employs techniques to create diverse, balanced, and minimally overlapping subsets of the training dataset, generating moderately imbalanced subsets to eliminate noise. These subsets are then used to train ensembles of CTGAN networks, which help remove noise and enhance the representativeness of the synthesized fraudulent samples.

3.2 Methodological Framework

The proposed methodological framework combines ensemble data augmentation techniques with synthetic data generation methods to address the prevalent class imbalance in credit card fraud detection. By generating

synthetic samples and employing balanced ensemble methods, this framework seeks to enhance detection performance for the minority class (fraud instances). The following steps detail the approach:

This structured framework enables a rigorous, balanced approach to credit card fraud detection, with emphasis on handling class imbalance via synthetic data generation and ensemble learning.

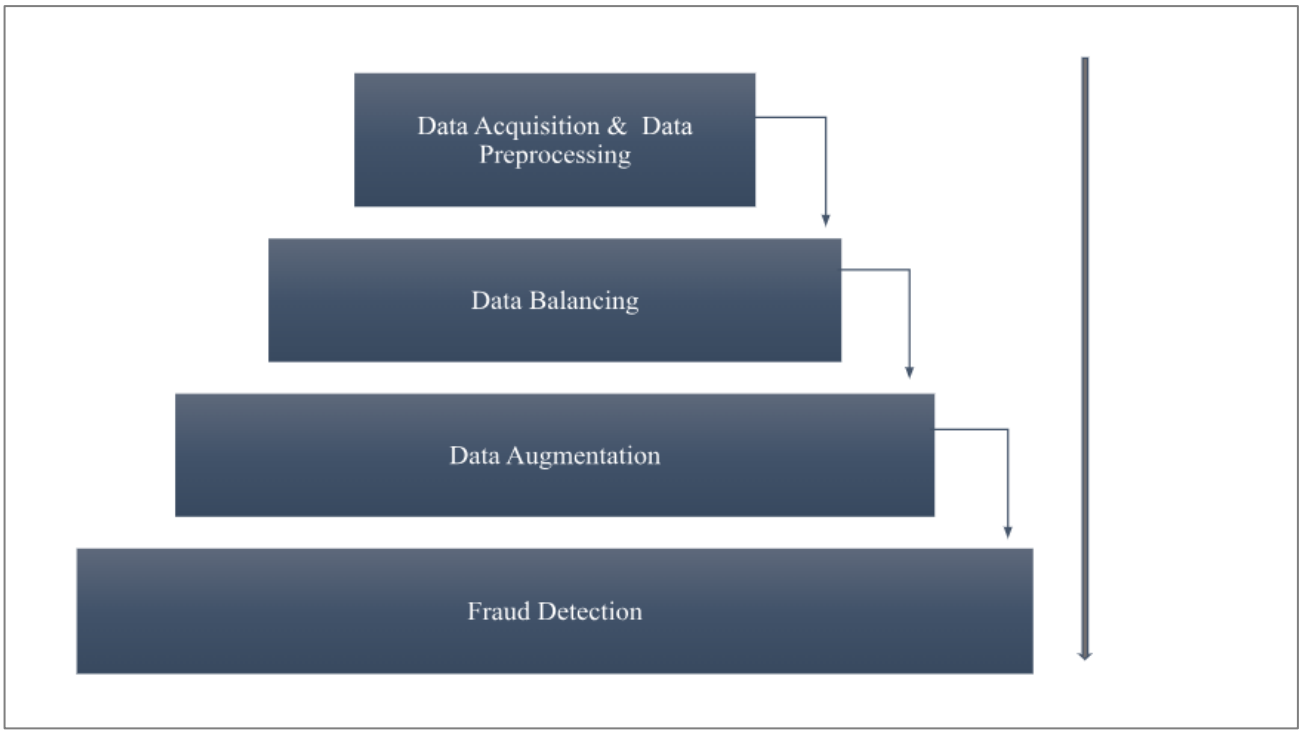


Figure 1: Methodical Structure depicting a holistic framework that captures every phase our research will traverse. See paragraphs below for breakdown of phase activities.

Table 1: Dataset Features

S/n	Feature	Description
1	Account number	Related to account numbers
2	Open to buy	Availability of balance
3	Credit limit	Maximum amount of credit of the associated account
4	Card number	Number of credit card
5	Transaction amount	Transaction amount submitted by merchant
6	Transaction time	Time of transaction
7	Transaction date	Date of transaction
8	Transaction type	Type of transaction e.g cash withdrawal and purchase
9	Currency code	The currency code
10	Merchant Category code	Merchant business type code
11	Merchant Number	Reference number of Merchant
12	Transaction Country	Country where the transaction took place
13	Transaction City	City where the transaction took place
14	Approval Code	Response to authorization request e.g approve or reject

3.2.1 Data Acquisition and Pre-Processing Phase

The dataset used in this study comprises credit card transactions made in September 2013 by European cardholders. This dataset includes transactions spanning two days, totalling 284,807 (Machine Learning Group - ULB Andrea, 2018). The dataset is highly unbalanced; the positive class (frauds) account for 0.172% of all transactions. It contains only numerical input variables, which are the result of a PCA transformation. PCA (Principal Component Analysis) is a well-established, widely used method that makes these datasets more interpretable while minimizing information loss. The process of PCA involves generating uncorrelated independent variables and progressively maximizing their variances. Pre-augmented features were calculated for most features to preserve the privacy and security of both customers and merchants. Furthermore, due to concerns about the confidentiality of consumer transaction details, the majority of the dataset's features were subjected to PCA to reduce dimensionality (Ghaleb *et al.*, 2023).

Consequently, original features and more background information about the data were not provided. Features V1, V2, … V28 are the principal components; the only features which have not been transformed with PCA are 'Time' and 'Amount'. Feature 'Time' contains the seconds elapsed between each transaction and the first transaction

in the dataset. The 'Amount' feature is the transaction amount; it can be used, for example, in cost-sensitive learning. The 'Class' feature is the response variable, taking the value 1 for fraud and 0 otherwise (Machine Learning Group - ULB Andrea, 2018). A more detailed description of this dataset can be found in Table 1 and Table 2.

Table 2: Summary of Dataset Characteristics

Aspect	Description
Dataset	Transactions made by credit cards in September 2013 by European cardholders.
Transaction Period	Two days
Total Transactions	284,807
Fraudulent Transactions	492 (0.172%)
Class Imbalance	Highly unbalanced; the positive class (frauds) accounts for 0.172% of all transactions.
Features	Numerical input variables resulting from PCA transformation (V1 to V28). 'Time' and 'Amount' are not PCA-transformed.
Time Feature	Seconds elapsed between each transaction and the first transaction in the dataset.
Amount Feature	Transaction amount, useful, for example, for cost-sensitive learning.
Response Variable	'Class': 1 indicates fraud, 0 indicates non-fraud.
Principal Component Analysis (PCA)	PCA is used to reduce the dimensionality of the dataset while preserving variance, mitigating multicollinearity, and enhancing computational efficiency.

3.2.1.1 Feature Scaling

3.2.1.1.1 Min-Max Normalization

Min-max normalization is one of the most common methods for data normalization. For every feature, the minimum value of that feature gets transformed into a 0, the maximum value gets transformed into a 1, and every other value gets transformed into a decimal between 0 and 1. The formula is as follows:

$$X_{scaled} = \frac{X - X_{min}}{X_{max} - X_{min}} \quad (1)$$

3.2.1.1.2 Robust Scaling

Robust Scaling transforms the data using the interquartile range (IQR) rather than the mean and standard deviation, making it less sensitive to outliers. Robust Scaler scales features that are robust to outliers. The method it follows is almost similar to the MinMax Scaler, but it uses the interquartile range (rather than the min-max used in the MinMax Scaler). This scaling algorithm removes the median and scales the data according to the quantile range. It, thus, follows the following formula:

$$\frac{x_i - Q_{1(x)}}{Q_{3(X)} - Q_{1(x)}} \quad (2)$$

Where Q1 is the 1st quartile, and Q3 is the third quartile

3.2.2 Data Balancing Phase

The second phase of this study focuses on preparing the training set by addressing the imbalanced class issue inherent in credit card fraud detection. The dataset is heavily skewed towards the majority class, with fraudulent transactions accounting for less than 1% of total transactions. This imbalance poses significant challenges in developing an effective and unbiased detection model. It is crucial to increase the number of fraudulent transactions in the training set to mitigate class imbalance.

In this research, we will use an oversampling technique called ADASYN (Adaptive Synthetic). Here, the minority class is copied *x* times until its size is similar to that of the majority class. ADASYN is an algorithm that generates synthetic data, and its greatest advantages are avoiding overfitting to the same minority data and generating more data for "harder to learn" examples.

Algorithm 1 ADASYN Algorithm

Input: Minority dataset X_s, Majority dataset X_l, k (nearest neighbors), β (balance ratio)

Output: Synthetic minority dataset X_syn

1. Initialize G = (n_l - n_s) * β
2. For each sample x_i in minority class X_s:
3. Compute the Euclidean distance from x_i to all other samples in X_s
4. Find the k-nearest neighbors S_ik of x_i
5. For each x_i in X_s:
6. Count A_i, the number of majority samples in the neighborhood S_ik
7. Compute τ_i = A_i / k
8. Normalize τ_i to get τ̂_i = τ_i / Σ(τ_i)
9. For each x_i in X_s:
10. Compute g_i = τ̂_i * G
11. Sample g_i neighbors from S_ik with replacement
12. For each neighbor x_ij sampled in step 5:
13. Generate a synthetic sample x_k = x_i + λ(x_i - x_ij), where λ is drawn from Uniform[0, 1]
14. Return the synthetic dataset X_syn

3.2.3 Data Augmentation Phase

CTGAN is a GAN-based method for modelling the distribution of tabular data and sampling rows from it (Xu and Veeramachaneni, 2018). A Conditional Tabular Generative Adversarial Network (CTGAN) can learn from the target distribution and generate artificial yet plausible samples from it. CTGAN is advantageous for data augmentation due to its capacity to simulate the distribution of real data. It comprises two deep learning networks: the generator and the discriminator.

Algorithm 2 CTGAN Algorithm

1. **Input:** Real dataset X (with both categorical and numerical columns), Latent space noise distribution Z, Learning rates for Generator (G_lr) and Discriminator (D_lr), Number of epochs, Batch size
2. **Output:** Synthetic tabular data X_syn
3. Preprocess dataset X:
4. Normalize numerical columns.
5. One-hot encode categorical columns.
6. Initialize Generator (G) and Discriminator (D).
7. for epoch in range(1, num_epochs):
8. for each batch of real data samples (X_real) from X:
9. Step 1: Sample noise and conditional vector
10. z = Sample noise from Gaussian distribution (Z)
11. c = Sample condition vector from real categorical data distribution
12. Step 2: Generate synthetic data
13. X_fake = G(z, c)
14. Step 3: Train Discriminator
15. D_real_loss = BinaryCrossEntropy(D(X_real, c), 1) # Real data
16. D_fake_loss = BinaryCrossEntropy(D(X_fake, c), 0) # Fake data
17. D_loss = D_real_loss + D_fake_loss
18. Update D using backpropagation with learning rate D_lr
19. Step 4: Train Generator
20. G_loss = BinaryCrossEntropy(D(X_fake, c), 1) # Fool the discriminator
21. Update G using backpropagation with learning rate G_lr
22. After training, generate new synthetic data X_syn using G(z, c).
23. Postprocess X_syn:
24. Reverse one-hot encoding for categorical columns.
25. Denormalize numerical columns.
26. Return synthetic tabular data X_syn

The generator is trained to produce samples (fake samples) from specific classes, while the discriminator determines if the generated samples resemble real samples. High similarity indicates successful generation by the generator, while low similarity suggests further training is needed. These networks are trained together recursively until the discriminator is fooled by at least half of the samples, indicating that the generative network is ready.

3.2.4 Detection Phase

The implementation of different algorithms will be carried out using Python 3.9.16. Classical machine learning models, Classification and Regression Trees (CART), Artificial Neural Networks (ANN), Logistic Regression (LR), Random Forest (RF), and XGBoost, were implemented using the Scikit-learn (Sklearn) 1.0.2 library. Deep learning models, on the other hand, were implemented using TensorFlow Keras 2.9.1

The choice of classifiers is driven by their ability to handle high-dimensional data and capture complex patterns, which are often present in fraud detection contexts. Logistic Regression (LR) and Classification and Regression Trees (CART) were selected as baseline models due to their simplicity and interpretability. Artificial Neural Networks (ANNs) and deep learning models were selected for their ability to learn complex patterns, especially when supported by synthetic data. Random Forest (RF) and XGBoost were chosen due to their stability, feature importance, and excellent performance on imbalanced classification problems.

3.2.4.1 Data Sampling and Stratification

Given the large size that will most likely result from the augmented dataset, this poses a significant challenge in terms of computational costs and training durations when attempting to analyze and process the entire dataset. To enhance efficiency while preserving the dataset's integrity, a stratified sampling method was used to derive a representative subset comprising 10% of the original dataset. This technique is designed to maintain the proportional distribution of classes within the sampled dataset, ensuring that subsequent analyses and model training are free from bias and accurately reflect the characteristics of the original dataset. By applying stratified sampling, our study maintains data integrity while optimizing computational efficiency for subsequent model training and evaluation.

$$n_s = [f \times N] \quad (3)$$

where:

n_s = Number of sampled instances

f = Sampling fraction (e.g., 0.1 for 10%)

N = Total number of instances in the dataset

For each class c (fraud and non-fraud in this case), stratified sampling ensures that the number of instances sampled maintains the same class proportions as in the original dataset. The number of sampled instances for each class c is given by:

$$n_s^c = [f \times N^c] \quad (4)$$

where

n_s^c = Number of sampled instances for class c

N^c = Number of instances belonging to class c in the original dataset

3.2.4.2 Model Training

To effectively detect fraud, the model is trained on an augmented dataset that combines original and synthesised data to improve robustness and generalizability.

3.2.4.3 Model Validation

To rigorously assess the model's performance and mitigate the risk of overfitting, stratified k-fold cross-validation will be utilized. Cross-validation divides the augmented data into k subsets, iteratively training the model on k−1 folds and testing on the remaining fold, yielding a comprehensive view of model performance across subsets.

Stratified K-Fold Cross-Validation is an advanced form of cross-validation that is particularly useful for datasets with unbalanced class distributions. This ensures that each fold of the dataset contains approximately the same percentage of samples from each class as the complete set, making the training and validation processes fairer and more reliable.

Table 3: StratifiedKFold Setting

S/n	Parameters	Settings
1	n_splits	5
2	Shuffle	True
3	random_state	42

where:

n_splits = Number of folds.

Shuffle = Whether to shuffle each class's samples before splitting into batches. Note that the samples within each split will not be shuffled.

random_state = ordering of the indices, which controls the randomness of each fold for each

3.2.4.4 Evaluation Metrics

The following evaluation metrics are applied to measure model accuracy, balance between precision and recall, and overall discriminative capability:

i. Accuracy (A):
Defined as the proportion of correct predictions among total predictions:

$$A = \frac{TP + TN}{TP + TN + FP + FN} \quad (5)$$

where TP, TN, FP, and FN represent true positives, true negatives, false positives, and false negatives, respectively.

ii. Recall (Sensitivity, R):
Recall measures the ability of the model to capture all actual positives:

$$R = \frac{TP}{TP + FN} \quad (6)$$

iii. F1 Score (F_1):
The harmonic mean of precision and recall, providing a balance between the two:

$$F_1 \frac{2 * P * R}{P + R} \quad (7)$$

Iv. Area under the curve (AUC)

The area under the ROC curve (AUC) represents the probability that the model, if given a randomly chosen positive and negative example, will rank the positive higher than the negative.

Algorithm 3 Pipeline

Input: Original dataset D

Output: Performance metrics (Mean ± SD)

1. Apply ADASYN on D_train to balance minority class
2. Train CTGAN using balanced D_train
3. Generate synthetic fraud samples using CTGAN
4. Augment D_train with synthetic samples
5. Split D into training set D_train and test set D_test (stratified)
6. Train classifiers using cross-validation on augmented D_train
7. Evaluate final models on untouched D_test
8. Repeat steps 1–7 for random seeds.

RESULTS AND DISCUSSION

4.1 Data Preprocessing

The first step in data processing is to understand the nature of our dataset and fully grasp the story we want to tell. We can interpret the histogram below as showing that the 'Time' values are between 0 and a larger number, while the rest have slightly wider distributions that are roughly normally distributed around 0, some wider, some skinnier, but mostly the same. Interestingly, we can see the number of transactions ranging from 0 to a couple of hundred euros, and the 'Class' column shows mostly non-fraudulent cases.

4.1.1 Scaling

The Amount column in our histogram appears to have a significant number of outliers due to its skew from one data point to the next. This will affect our scaling procedure. To tackle this, we used the Robust Scaler to normalise the Amount column and handle outliers. The Time column in our histogram also appears to have uneven features. Table 4 shows the results of both scaling procedures, with the Amount column now having a maximum transaction of €358 and the Time column ranging from 0 to 1. This confirms the accuracy of our scaling procedures.

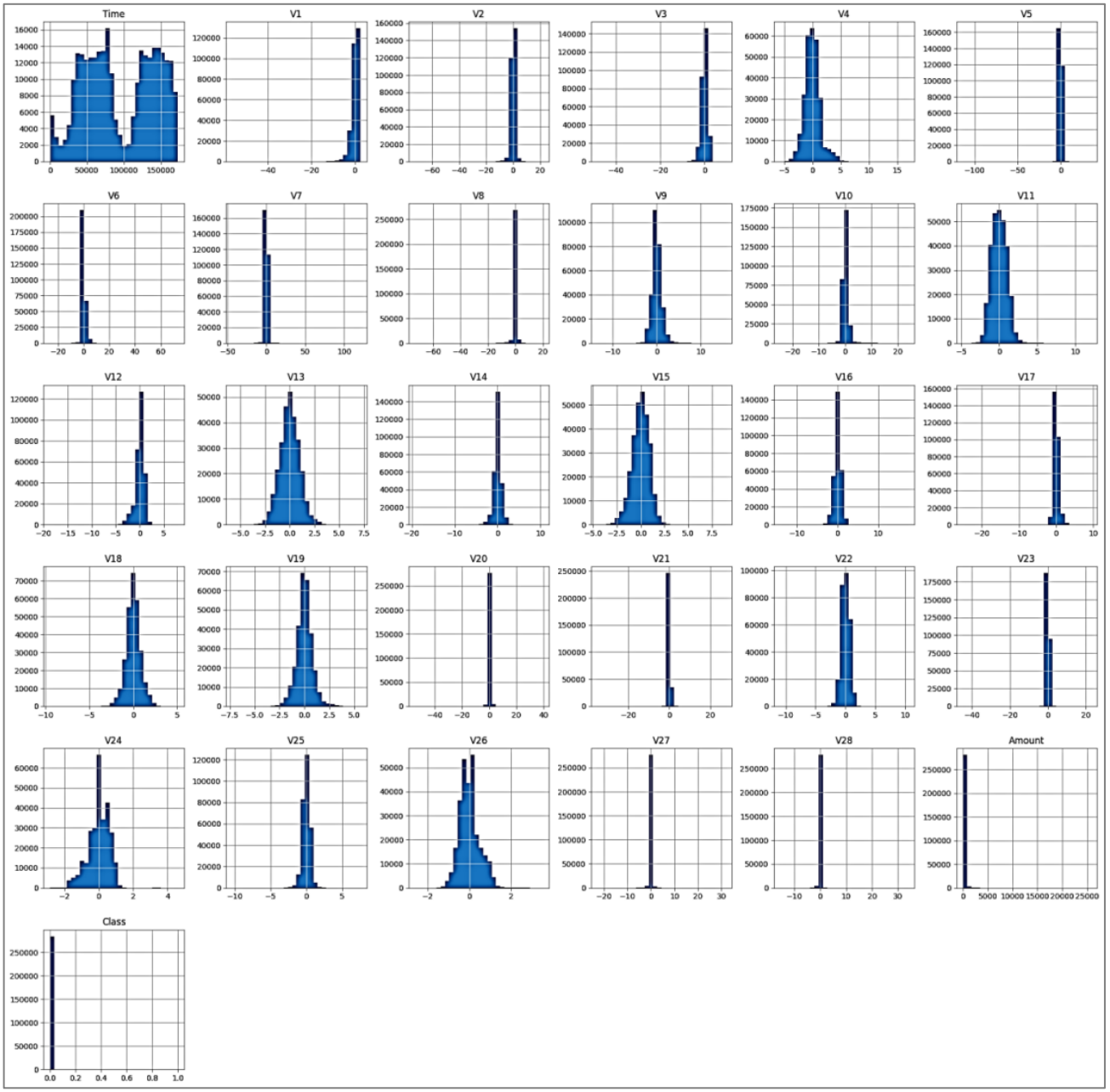


Figure 2: Nature of columns in the dataset during Exploratory Data Analysis

Table 4: Descriptive Statistics of Time and Amount Column after Scaling

Statistics	Time	Amount
Count	284807.000000	284807.000000
mean	0.548717	0.927124
std	0.274828	3.495006
min	0.000000	-0.307413
25%	0.313681	-0.229162
50%	0.490138	0.000000
75%	0.806290	0.770838
max	1.000000	358.683155

* We can now see that the Time column holds values from 0 and 1 respectively, which is good.

4.2 Results for Data Balancing

The data preprocessing we did above prepared our columns to have features on a more representative scale, so we can address class bias. While oversampling, ADASYN has a strategy (sampling_strategy='minority'), this parameter ensures that only the minority class (fraudulent transactions) is oversampled to match the distribution of the majority class. The algorithm adapts to the minority class density, ensuring better representation of complex fraudulent patterns. Both Class 0 and Class 1 now have 284,315 representative samples on both sides. See the results below. As seen vividly in ense 3, both fraudulent and non-fraudulent samples are now fully oversampled, and the class imbalance in the 'Class' column has been successfully eliminated.

4.3 Results for Data Augmentation

Even after applying ADASYN above to balance the dataset, our fraud detection models still faces a challenge because fraudulent cases often exhibit high variability,

making it difficult for models to generalize effectively. Our CTGAN model was initialized and trained for 5 epochs. The process learned the underlying distributions of the data, ensuring realistic synthetic sample generation. We then merged the newly formed synthetic data with the balanced dataset above to form an augmented dataset, a fraction of the augmented dataset is what we will save later for use in model training.

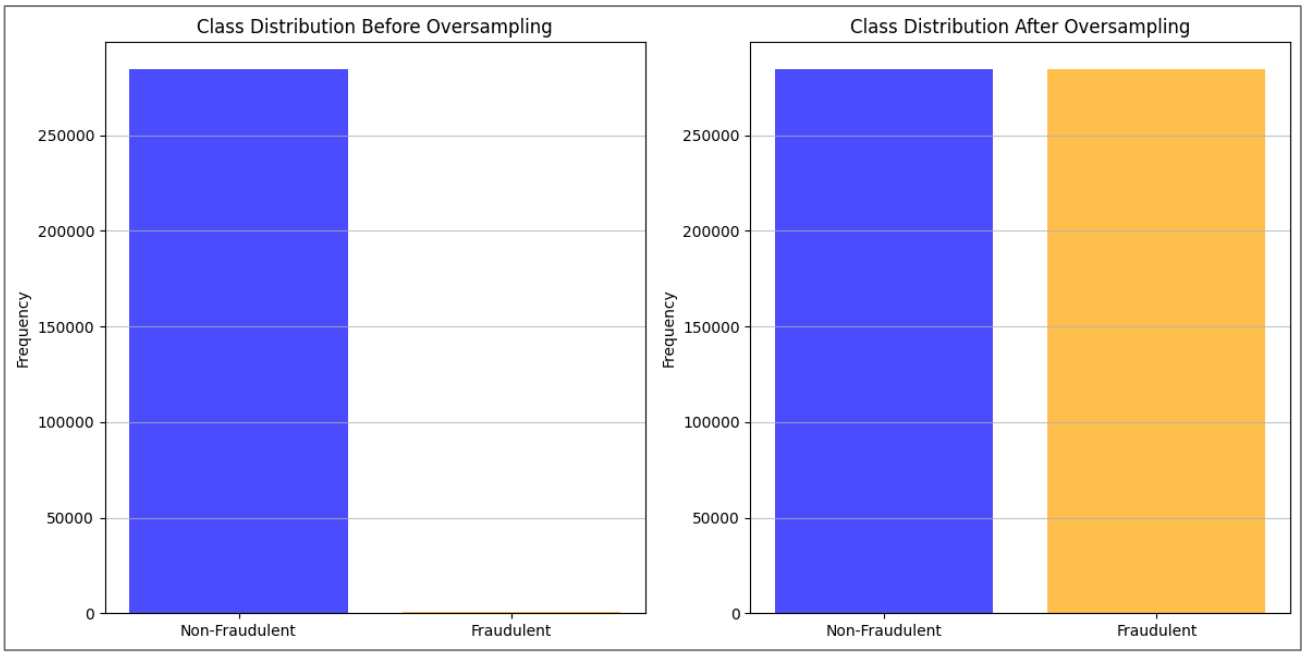


Figure 3: Class Distribution Before and After Oversampling

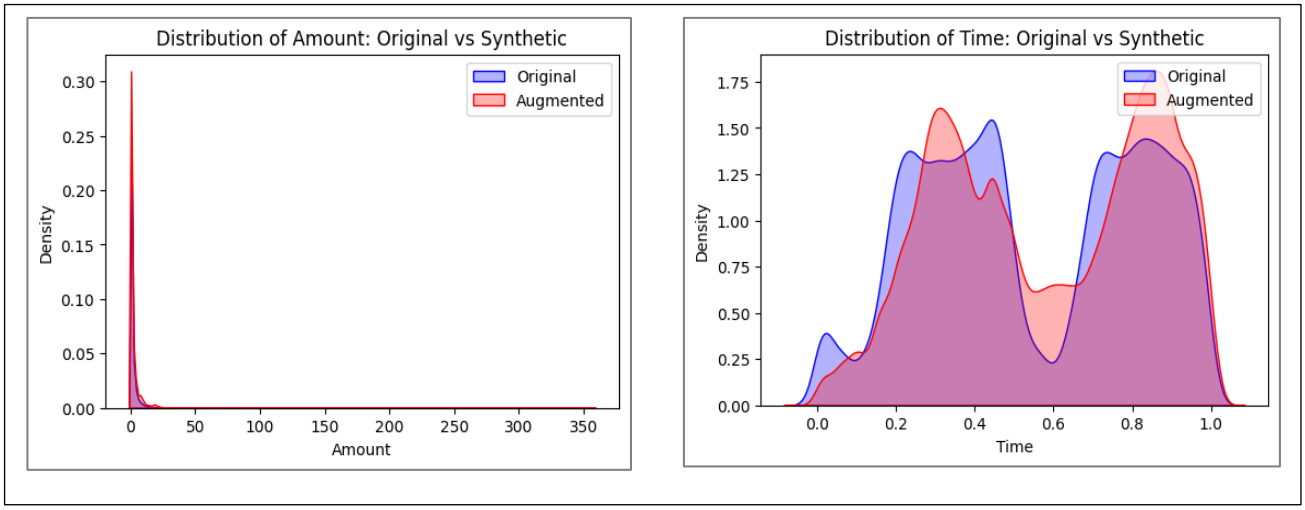


Figure 4: Distribution of Amount and Time: Original vs Augmented

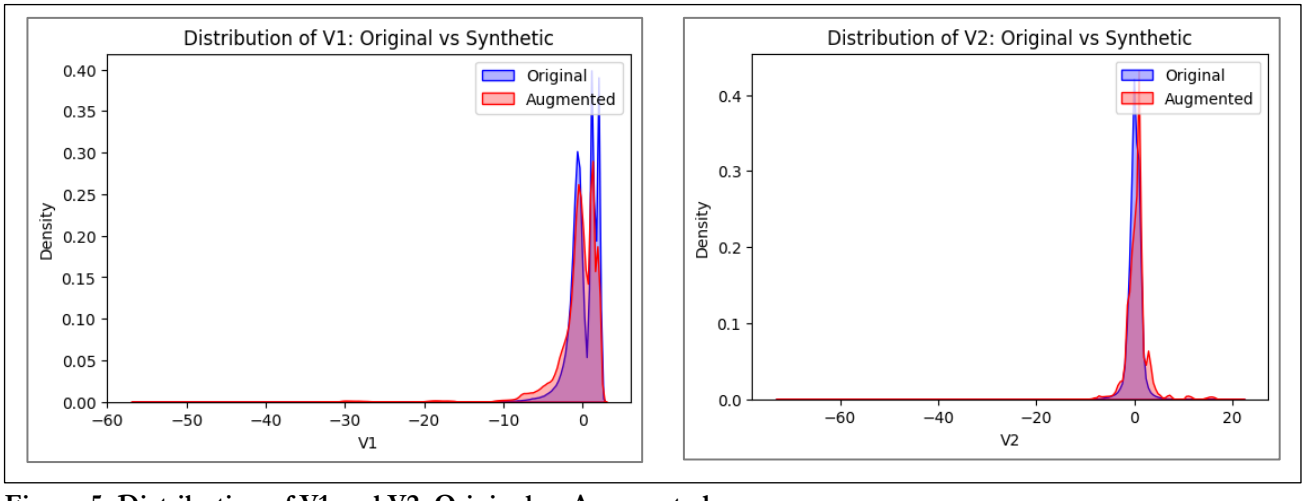


Figure 5: Distribution of V1 and V2: Original vs Augmented

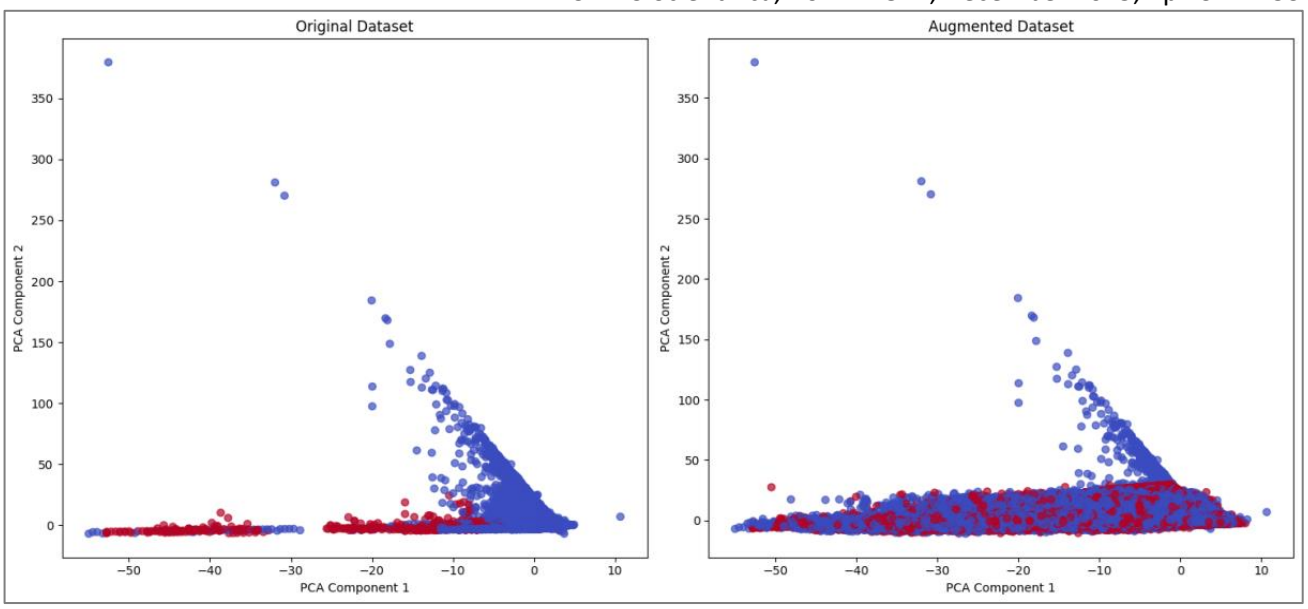


Figure 6: Original vs Augmented Components

Table 5: Learning Parameters

Algorithm	Parameters
KNN	*param_grid=param_grid, scoring='f1_weighted', cv=cv, verbose=1, n_jobs=-1*
CART	*estimator=cart, param_grid=param_grid, cv=5, scoring='f1', n_jobs=-1, verbose=1*
LR	*estimator=lr_base_model, param_distributions=param_grid, n_iter=5, scoring='f1', cv=2, random_state=42, n_jobs=1*
RF	*estimator=rf_base_model, param_distributions=param_grid, n_iter=5, scoring='f1', cv=2, random_state=42, n_jobs=-1*
XGBoost	*n_iter = 10, scoring = 'f1', cv=2, random_state=42, n_jobs=1, param_distributions=param_grid*
ANN	*activation='relu','sigmoid',optimizer=Adam(),loss='binary_crossentropy',random_state=42, epoch = 10*

Table 6: Learning Parameters

S/n	Classifiers	MODEL EACT Metrics			
		f-measure	accuracy	recall	auc roc
1	KNN	87.05	87.11	87.11	93.35
2	LR	76.78	80.65	74.49	87.62
3	RF	85.26	87.58	83.61	94.37
4	XGBOOST	85.53	87.21	87.99	95.2
5	ANN	86.62	88.86	83.96	88.25
6	CART	83.42	86.08	81.54	85.52

Table 7: Performance across five independent runs (mean ± SD)

	Model	AUC-ROC	F1-score	Recall	Accuracy
1	KNN	92.39 ± 0.98	84.91 ± 1.28	84.79 ± 1.40	84.99 ± 1.27
2	LR	84.34 ± 1.84	74.33 ± 1.52	73.53 ± 0.86	77.96 ± 1.58
3	RF	93.14 ± 0.70	82.74 ± 1.54	81.93 ± 0.96	85.09 ± 1.42
4	XGBoost	93.89 ± 0.80	82.94 ± 1.50	87.05 ± 0.70	84.81 ± 1.49
5	ANN	92.52 ± 3.09	84.40 ± 1.61	81.04 ± 1.90	87.64 ± 1.43
6	CART	82.29 ± 1.83	80.22 ± 2.32	78.14 ± 2.26	82.90 ± 1.79

*Results demonstrate stable performance across runs, with low variance for most classifiers, indicating robustness of the proposed EA-CT framework.

By leveraging CTGAN, this study not only addresses class imbalance but also ensures that the dataset captures realistic fraud characteristics, thereby strengthening fraud detection models' ability to generalize to unseen cases. Figure 4 and Figure 5 show the density plot of how closely similar our samples are between the original data and the augmented data. Find visuals for four columns, Amount, Time, V1 and V2 represented below.

Our research further employed PCA with 2 principal components (2D projection) to compare the distributions of the original and augmented datasets in a lower-dimensional space. The goal is to holistically determine and visualise whether the synthetic components generated by CTGAN align with real fraud cases across the entire scope. All components were successfully augmented,

ensuring they maintained meaningful characteristics as shown below.

4.4 Results for Fraud Detection

Sequel to the Figure 6 above, which shows the successful augmentation of representative samples, we further employed a stratified sampling technique to train our model with 10% of our recently augmented data, which now has over one million transactions. After implementation, the stratified data, now called sampled data, has 113726 transactions ready for detection.

For cost-sensitive machine learning and deep learning approaches, the classifiers were initially trained on the original imbalanced dataset. The probabilistic outputs of each classifier were fine-tuned to achieve optimal performance, as seen in Figure 7. Table 5 presents the parameters utilized in the experimental setup for the detection phase.

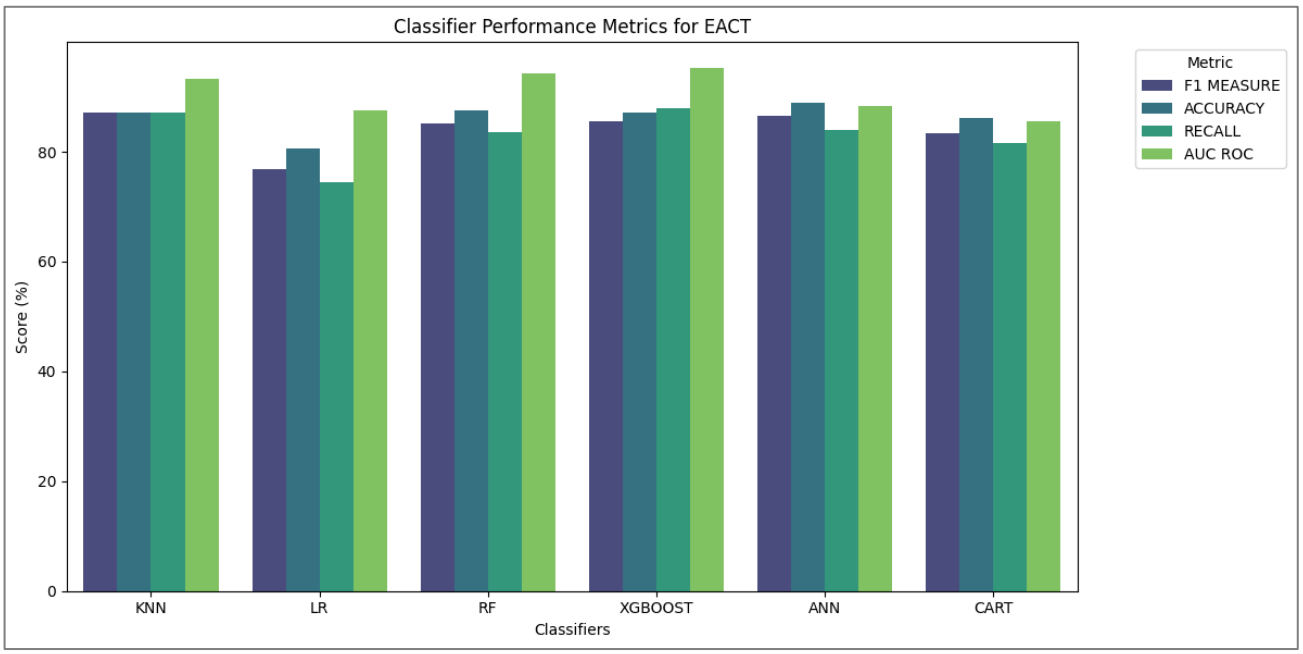


Figure 7: General Performance Measures for EA-CT

Table 8: Comparison between ESMOTE-GAN and Model EA-CT in terms of F-measure

S/n	Base Models	Unbalanced	ESMOTE-GAN	EA-CT
1	KNN	80.15	85.25	87.05
3	LR	83.33	49.66	76.78
4	RF	72.43	92.31	85.26
5	XGBOOST	83.15	92.44	85.53
6	ANN	78.49	85.71	86.62
7	CART	75.86	78.74	83.42

Table 9: Improvement in terms of F-measure

S/n	Base Models	EA-CT
1	KNN	1.8
3	LR	27.12
4	RF	-7.05
5	XGBOOST	-6.91
6	ANN	0.91
7	CART	4.68

4.4.1 Performance Evaluation

Table 8 shows the overall performance of the proposed model, EA-CT, in terms of F-Measure, accuracy, recall, and AUC. Results indicate that the ANN achieved the highest accuracy (88.86%), while XGBoost achieved the highest AUC-ROC score (95.20%), suggesting superior fraud-detection performance. KNN exhibited the highest F1-score (87.05%), indicating a balanced performance between precision and recall. Random Forest (RF) performed consistently well across all metrics, while Logistic Regression (LR) had the weakest recall (74.49%), making it less effective for detecting fraud.

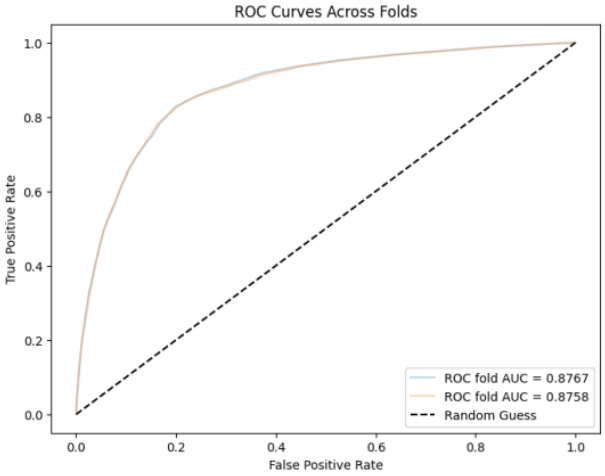
Every value mentioned has been taken straight from the held-out test set. Every value here has been collected across five independent exercises. Cross-validation was implemented solely for model selection within the training dataset.

As shown in Figure 7 above, the KNN and XGBoost algorithms appear to produce the best results on our EACT model. Consequently, model EACT exhibits a consistently very high AUC ROC as seen in Figure 8.

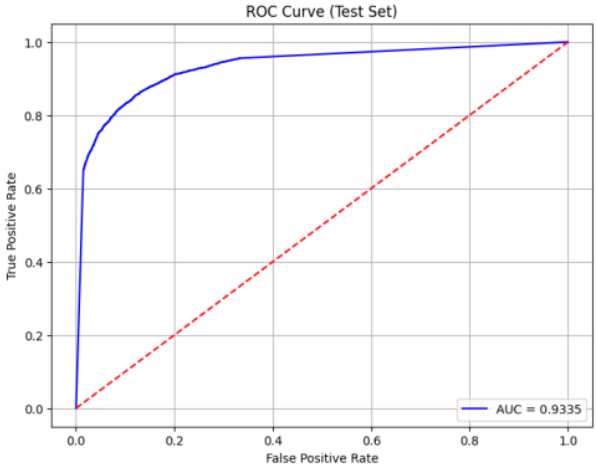
4.4.1 AUC ROC for EA-CT

A model with a high AUC-ROC consistently indicates that it generalises well to unseen data. This also suggests that the model is not overfitting (due to augmentation and cross-validation) and can perform reliably in real-world scenarios.

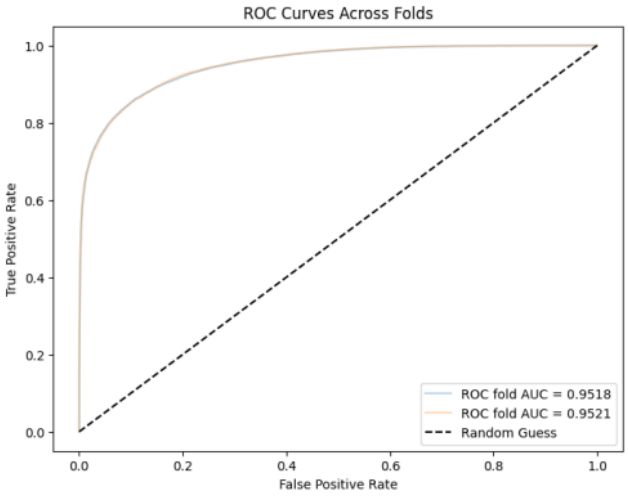
Furthermore, the model has a significantly high chance of correctly ranking a randomly chosen fraudulent transaction higher than a randomly chosen non-fraudulent transaction. The model is highly effective at distinguishing between the two classes and is likely to perform well in real-world applications.



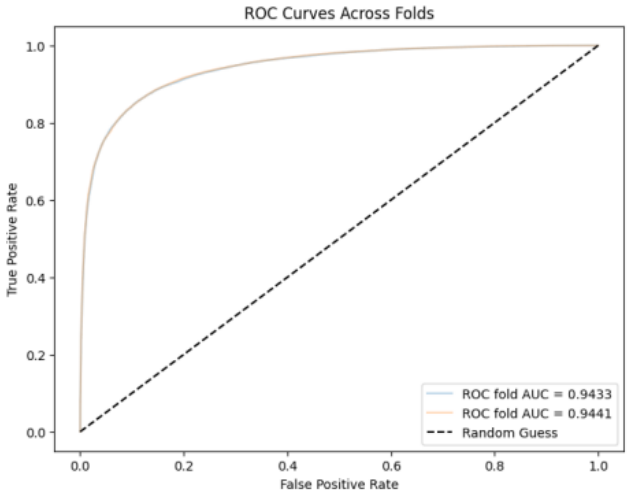
LR



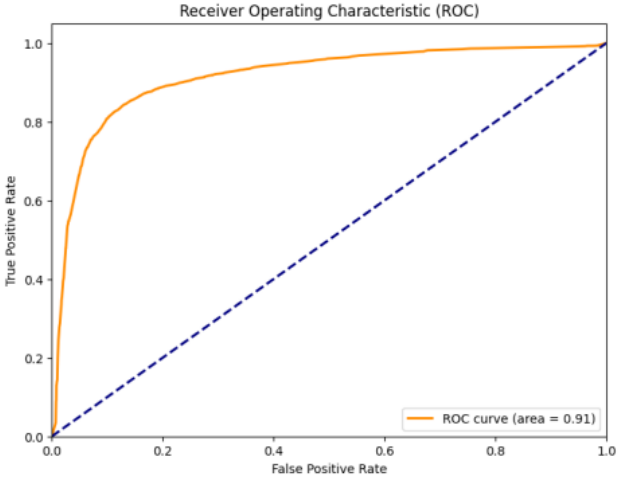
KNN



XGBOOST



RF



CART

Figure 8: **ROC Curves between models**

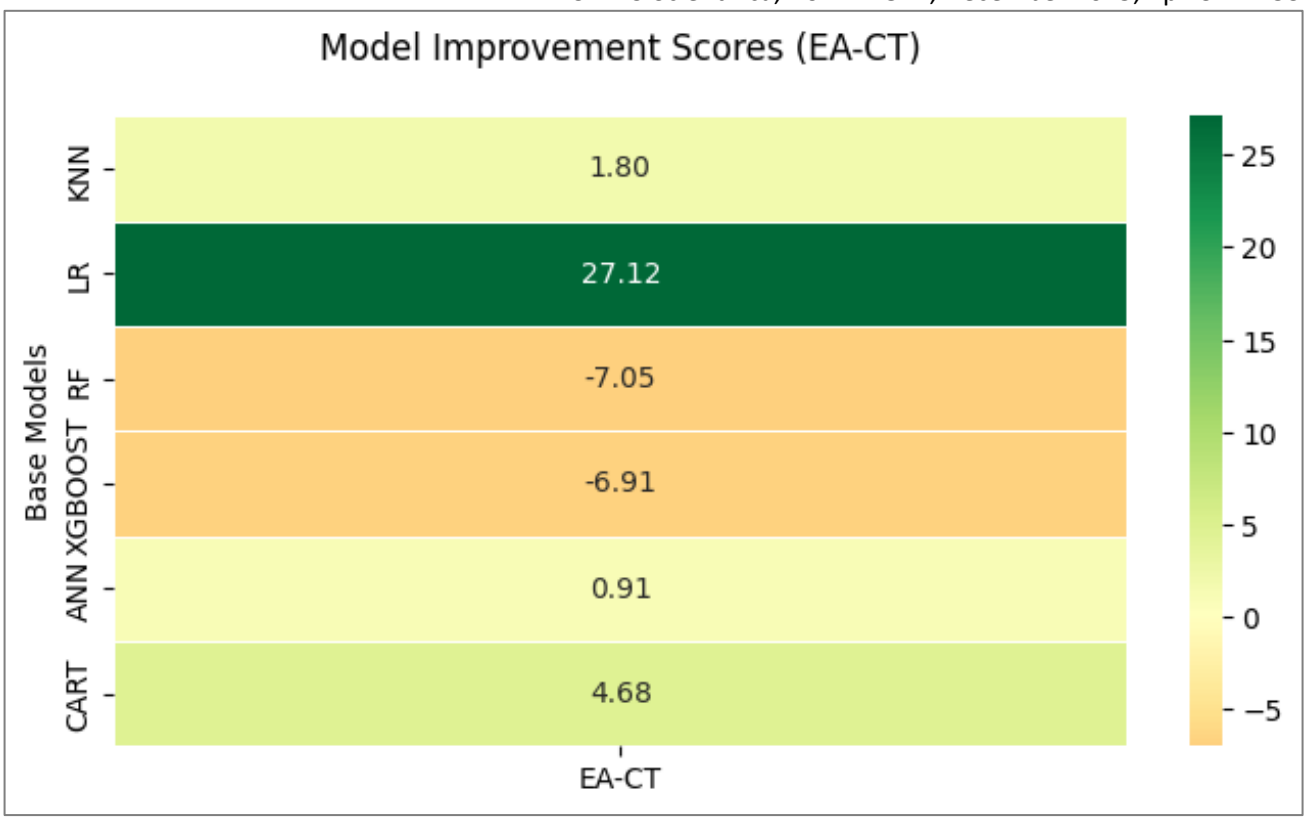


Figure 9: **Model EA-CT improvement in terms of F-measure**

4.4.2 Performance Comparison with the Related Work

The performance of models on unbalanced data varies significantly. This variability aligns with the existing literature by Ghaleb *et al.* (2023), which suggested that class imbalance often degrades classifier performance, particularly for minority classes.

The results of this study show how adopting the EA-CT approach can effectively address the problems of class imbalance and data scarcity in credit card fraud detection. The model achieved significant gains in detection performance across different classifiers, driven by the incorporation of ADASYN for data balancing and CTGAN for synthetic data augmentation, compared to the work of Ghaleb et al. (2023) and ESMOTE-GAN, as shown in Table 8.

Regarding the technique's discriminative power, it demonstrated strong performance across multiple runs, as evidenced by consistently high AUC-ROC scores in Figure 8, indicating good performance in distinguishing fraudulent transactions from true non-fraudulent transactions. A high ROC-AUC value indicates that the model achieves a good trade-off between sensitivity and specificity and can therefore be relied upon for binary classification problems in fraud detection. This demonstrates the model's ability to rank accurate predictions higher than less accurate ones, regardless of the thresholding criteria.

Performance improved when ensemble-based classification algorithms were applied. Ensemble approaches combine the predictive capabilities of multiple base learners to improve generalization and reduce the likelihood of overfitting. The combination of varied decision boundaries was probably responsible for the model's greater robustness and accuracy in identifying fraudulent patterns that could otherwise go undetected with individual classifiers.

CONCLUSION

The assessment incorporating XGBoost and KNN classifiers indicated that the previously stated technique, amongst others, achieved very good recall and AUC-ROC scores, thus confirming its usefulness in practical applications, where the occurrence of false negatives is highly undesirable. These findings indicate that the EA-CT technique effectively addresses class imbalance in fraud detection systems while providing a basis for further improvements in algorithmic performance and deployment in real-world settings.

LIMITATION

All classifiers exhibited AUC-ROC scores suggesting that the EA-CT technique successfully distinguished fraudulent from non-fraudulent transactions for the dataset in question. However, this evidence should be viewed as proof of internal generalisation. Since the evaluation was performed on a randomly stratified held-out test set drawn from the same data distribution, generalisation claims are constrained to the dataset. Internal claims of generalisation to other datasets are infeasible, especially in the absence of a deferred, external validation dataset, which precludes any claims about deploying this technique in novel, real-world fraud scenarios. This, however, can be a strong recommendation for future research.

Consequently, the augmented dataset is large. While this benefits training, the deployment environment requires sufficient memory to load the final model. Tree-based models are generally memory-efficient. If KNN is chosen, storing the entire reference sample for nearest-neighbour calculations would require more memory.

REFERENCES

Almazroi, A. A., & Ayub, N. (2023). Online payment fraud detection model using machine learning techniques. *IEEE Access, 11*, 137188–137203. [Crossref]

Azarm, C., Acar, E., & van Zeelt, M. (2024). *On the Potential of Network-Based Features for Fraud Detection*. [Crossref]

Batista, G. E., Prati, R. C., & Monard, M. C. (2004). A study of the behavior of several methods for balancing machine learning training data. *ACM SIGKDD Explorations Newsletter, 6*(1), 20–29. [Crossref]

Benchaji, I., Douzi, S., El Ouahidi, B., et al. (2021). Enhanced credit card fraud detection based on attention mechanism and LSTM deep model. *Journal of Big Data, 8*, 151. [Crossref]

Domingos, P. (1999). Metacost: A general method for making classifiers cost-sensitive. *Proceedings of the 5th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 155–164. [Crossref]

Dong, X., Yu, Z., Cao, W., et al. (2020). A survey on ensemble learning. *Frontiers of Computer Science, 14*, 241–258. [Crossref]

Fernández, A., del Río, S., Chawla, N. V., et al. (2017). An insight into imbalanced big data classification: Outcomes and challenges. *Complex & Intelligent Systems, 3*, 105–120. [Crossref]

Forough, J., & Momtazi, S. (2021). Ensemble of deep sequential models for credit card fraud detection. *Applied Soft Computing, 99*, 106883. [Crossref]

Ghaleb, F. A., Saeed, F., Al-Sarem, M., Qasem, S. N., & Al-Hadhrami, T. (2023). Ensemble synthesized minority oversampling-based generative adversarial networks and random forest algorithm for credit card fraud detection. *IEEE Access, 11*, 89694–89710. [Crossref]

Goran, O. (2023). *Synthesizing credit data using autoencoders and generative adversarial networks*. ResearchGate. [Link]

He, H., Bai, Y., Garcia, E. A., & Li, S. (2008). ADASYN: Adaptive synthetic sampling approach for imbalanced learning. *2008 IEEE International Joint Conference on Neural Networks (IEEE World Congress on Computational Intelligence)*, 1322–1328. [Crossref]

Hsin, Y. Y., Dai, T. S., Ti, Y. W., Huang, M. C., Chiang, T. H., & Liu, L. C. (2022). Feature engineering and resampling strategies for fund transfer fraud with limited transaction data and a time-inhomogeneous modi operandi. *IEEE Access, 10*, 86101–86116. [Crossref]

Kalid, S. N., Khor, K.-C., Ng, K.-H., & Tong, G.-K. (2024). Detecting frauds and payment defaults on credit card data inherited with imbalanced class distribution and overlapping class problems: A systematic review. *IEEE Access, 12*, 23636–23652. [Crossref]

Krawczyk, B. (2016). Learning from imbalanced data: Open challenges and future directions. *Progress in Artificial Intelligence, 5*, 221–232. [Crossref]

Leevy, J. L., Khoshgoftaar, T. M., Bauder, R. A., et al. (2018). A survey on addressing high-class imbalance in big data. *Journal of Big Data, 5*(1), 42. [Crossref]

Lopez, V., Fernández, A., García, S., Palade, V., & Herrera, F. (2013). An insight into classification with imbalanced data: Empirical results and current trends on using data intrinsic characteristics. *Information Sciences, 250*, 113–141. [Crossref]

Machine Learning Group - ULB · Andrea. (2018). Credit Card Fraud Detection. Kaggle. [Link]

Maina, D. G., Moso, J. C., & Gikunda, P. K. (2023). Detecting fraud in motor insurance claims using XGBoost algorithm with SMOTE. In *2023 International Conference on Information and Communication Technology for Development for Africa (ICT4DA)* (pp. 61–66). IEEE. [Crossref]

Majeed, A., & Hwang, S. O. (2023). CTGAN-MOS: Conditional generative adversarial network based minority-class-augmented oversampling scheme for imbalanced problems. *IEEE Access, 11*, 85878–85899. [Crossref]

Merchant Cost Consulting. (2024). *Credit card fraud statistics (2024)*. [Link]

Mienye, I. D., & Sun, Y. (2023). A deep learning ensemble with data resampling for credit card fraud detection. *IEEE Access, 11*, 30628–30638. [Crossref]

Mienye, I. D., & Jere, N. (2024). Deep learning for credit card fraud detection: A review of algorithms, challenges, and solutions. *IEEE Access*. Advance online publication. [Crossref]

Najadat, H., Altiti, O., Aqouleh, A. A., & Younes, M. (2020, April). Credit card fraud detection based on machine and deep learning. In *2020 11th International Conference on Information and Communication Systems (ICICS)* (pp. 204–208). [Crossref]

Nguyen, T. T., Tahir, H., Abdelrazek, M., & Babar, A. (2020, January 1). *Deep learning methods for credit card fraud detection*. arXiv. [Crossref]

Niaz, N. U., Shahariar, K. N., & Patwary, M. J. (2022, March). Class imbalance problems in machine learning: A review of methods and future challenges. In *Proceedings of the 2nd International Conference on Computing Advancements* (pp. 485–490). [Crossref]

Ramentol, E., Vluymans, S., Verbiest, N., Caballero, Y., Bello, R., Cornelis, C., & Herrera, F. (2015). IFROWANN: Imbalanced fuzzy-rough ordered weighted average nearest neighbor

classification. *IEEE Transactions on Fuzzy Systems, 23*(5), 1622–1637. [Crossref]

Rawat, A., & Tiwari, S. (2023). A comprehensive review on credit card fraud detection using machine learning techniques. *International Journal of Innovative Research and Growth, 12*(2), Article 12. [Crossref]

Roseline, J. F., Naidu, G. B. S. R., Pandi, V. S., alias Rajasree, S. A., & Mageswari, N. (2022). Autonomous credit card fraud detection using machine learning approach. *Computers and Electrical Engineering, 102*, 108132. [Crossref]

Roy, A., Sun, J., Mahoney, R., Alonzi, L. P., Adams, S., & Beling, P. A. (2018). Deep learning detecting fraud in credit card transactions. *2018 Systems and Information Engineering Design Symposium (SIEDS)*, 129–134. [Crossref]

Sadgali, I., Sael, N., & Benabbou, F. (2020). Adaptive model for credit card fraud detection. *International Journal of Interactive Mobile Technologies, 14*(03), 182–192. [Crossref]

Seera, M., Lim, C. P., Kumar, A., et al. (2024). An intelligent payment card fraud detection system. *Annals of Operations Research, 334*, 445–467. [Crossref]

Statista. (2024). *Total value of losses due to card fraud, either credit card fraud or debit card fraud, worldwide from 2014 to 2023*. [Link]

Thennakoon, A., Bhagyani, C., Premadasa, S., Mihiranga, S., & Kuruwitaarachchi, N. (2019). Real-time credit card fraud detection using machine learning. *2019 9th International Conference on Cloud Computing, Data Science & Engineering (Confluence)*, 488–493. [Crossref]

Wang, C., Wu, P., Yan, L., et al. (2021). Image classification based on principal component analysis optimized generative adversarial networks. *Multimedia Tools and Applications, 80*, 9687–9701. [Crossref]

Wang, S., Tricco, T. S., Jiang, X., Robertson, C. E., & Hawkin, J. (2023, January 1). *Synthetic demographic data generation for card fraud detection using GANs*. arXiv. [Crossref]

Wu, T., & Wang, Y. (2021). Locally interpretable one-class anomaly detection for credit card fraud detection. *2021 IEEE 33rd International Conference on Tools with Artificial Intelligence (ICTAI)*, 1–8. [Crossref]

Xu, L., & Veeramachaneni, K. (2018). *Synthesizing tabular data using generative adversarial networks*. arXiv. arXiv:1811.11264.

Yu, Z., Wang, D., Zhao, Z., Chen, C. P., You, J., Wong, H. S., & Zhang, J. (2017). Hybrid incremental ensemble learning for noisy real-world data classification. *IEEE Transactions on Cybernetics, 49*(2), 403–416. [Crossref]