# Enhancing Robust Fraud Detection for GCC Financial Systems: A Hybrid Framework for Imbalance, Drift, and Adversarial Threats

1st Kazi Namira Meyheg Sanam
*Department of Computer Science and Engineering*
*International Islamic University Chittagong*
Chittagong 4318, Bangladesh
c231450@ugrad.iiuc.ac.bd

2nd Umme Benin Yeasmin Meem
*Department of Computer Science and Engineering*
*International Islamic University Chittagong*
Chittagong 4318, Bangladesh
c231452@ugrad.iiuc.ac.bd

3rd Sultana Tasnim Jahan
*Department of Computer Science and Engineering*
*International Islamic University Chittagong*
Chittagong 4318, Bangladesh
sultanatasnim@iiuc.ac.bd

***Abstract*—Financial fraud detection in the Gulf Cooperation Council (GCC) region faces critical challenges due to extreme class imbalance (0.173% fraud rate), evolving attack patterns, and adversarial threats. We propose a hybrid machine learning framework integrating optimized SMOTE-based oversampling with strategic undersampling and regularized XGBoost classification. Key innovations include (1) adaptive k-neighbor selection ($k = 3$) for conservative synthetic minority sample generation, (2) hybrid resampling with sampling ratios of 0.4/0.6 for balanced class distribution, and (3) enhanced XGBoost regularization with L1/L2 penalties to prevent overfitting. Evaluated on a real-world credit card dataset (284,807 transactions), the framework achieves 85.81% recall, improves precision from 19.57% to 28.22% (+8.65%), and increases F1-score from 31.87% to 42.47% (+10.61%), while maintaining robust AUC (96.94%). Results demonstrate that strategic combination of oversampling, undersampling, and regularization yields superior fraud detection with reduced false positives, suitable for GCC financial institutions.**

***Index Terms*—Fraud Detection, Class Imbalance, SMOTE, XGBoost, GCC, Hybrid Sampling, Machine Learning, Financial Security**

## I. Introduction

Financial fraud represents a persistent and increasing threat to digital payment ecosystems globally, with annual losses exceeding $32 billion in 2023. The Gulf Cooperation Council (GCC) region, characterized by rapid digital transformation and diverse financial infrastructures, faces unique fraud detection challenges that include extreme class imbalance, fragmented legacy systems, and evolving regulatory frameworks [1].

The fundamental challenge in fraud detection stems from the severe class imbalance, where fraudulent transactions typically represent less than 0.2% of total cases. This creates a scenario where traditional machine learning models achieve high overall accuracy while failing to detect actual fraud cases—a critical failure mode in high-stakes financial applications [2].

Recent advances in handling imbalanced datasets include synthetic oversampling techniques such as SMOTE (Synthetic Minority Oversampling Technique) [2], cost-sensitive learning approaches [4], and ensemble methods [5]. However, these techniques often suffer from limitations: SMOTE can introduce noise through overgeneralization, pure oversampling may cause overfitting, and cost-sensitive methods require careful parameter tuning [6].

Al-Daoud and Abu-AlSondos [1] recently proposed a Hybrid Machine Learning Framework (HMLF) combining SMOTE with cost-sensitive XGBoost for financial institutions of the GCC, achieving a recall of 85%. However, their approach yielded low precision (19.57%), resulting in excessive false positive rates that were impractical for operational deployment, where each alert requires costly manual investigation.

### A. Research Gap

Although existing approaches achieve high fraud detection rates (recall), they suffer from poor precision, generating excessive false alarms that undermine operational efficiency and analyst trust. No existing framework adequately balances recall and precision while maintaining computational efficiency for real-time GCC deployment.

### B. Our Contribution

This paper addresses this gap through three key innovations:

1) **Adaptive SMOTE Optimization:** We demonstrate that k-neighbor selection ($k = 3$ versus standard $k = 5$) produces more conservative synthetic samples, reducing noise while maintaining minority class representation.
2) **Hybrid Resampling Strategy:** We combine SMOTE oversampling (sampling_strategy = 0.4) with strategic undersampling (sampling_strategy = 0.6) to achieve balanced class distribution without excessive data augmentation.
3) **Enhanced Regularization:** We incorporate L1/L2 regularization in XGBoost (reg_alpha=0.1, reg_lambda=1.0) to prevent overfitting on synthetic samples while improving generalization.

Our experimental results on a real-world dataset of 284,807 credit card transactions demonstrate that our hybrid approach achieves:

- 85.81% recall (maintaining fraud detection capability)
- 28.22% precision (+8.65% over baseline, 44% relative improvement)
- 42.47% F1-score (+10.61% over baseline, 33% relative improvement)
- 96.94% AUC (robust discrimination capability)

The remainder of this paper is organized as follows: Section II reviews related work; Section III details our proposed methodology; Section IV presents experimental setup and results; Section V discusses implications and limitations; Section VI concludes with future directions.

Due to the confidentiality of financial transaction data in GCC institutions, publicly available European credit card transaction data is used as a proxy benchmark in this study. While the dataset does not originate from the GCC region, the proposed framework is designed to address operational constraints prevalent in GCC financial systems, including extreme class imbalance, limited label availability, and the need for low false-positive rates.

## II. Related Work

### A. Class Imbalance in Fraud Detection

Class imbalance, where fraudulent transactions represent less than 1% of cases, remains a fundamental challenge in ML-based fraud detection [2]. Traditional classifiers tend to bias toward the majority class, achieving high accuracy while missing critical fraud instances [6].

Oversampling techniques like SMOTE [2], ADASYN [7], and Borderline-SMOTE [8] address this by generating synthetic minority samples. However, these methods assume stable fraud patterns and can introduce noise in high-dimensional spaces [9].

Hybrid ensemble approaches like SMOTEBoost [10] and EasyEnsemble [11] combine resampling with boosting or bagging, improving recall and sensitivity but increasing computational complexity [12].

### B. Fraud Detection Systems

Deep learning approaches including LSTM [13], CNNs [14], and autoencoders [15] have shown promise but require substantial labeled data and computational resources. Traditional ML methods like Random Forest [16], XGBoost [17], and SVM [18] remain popular for their interpretability and efficiency.

Cost-sensitive learning addresses imbalance by assigning higher misclassification costs to minority class errors [19]. Ensemble methods combining multiple classifiers have demonstrated robust performance [20].

### C. GCC Financial Context

GCC financial institutions face unique challenges including fragmented infrastructure, varying AI maturity levels, and evolving regulatory frameworks [1]. Islamic finance principles, regional payment preferences, and cross-border transaction patterns require specialized approaches [21].

## III. Methodology

This section describes the proposed hybrid fraud detection framework designed to address the extreme class imbalance problem in financial transaction data. The complete workflow of the methodology is illustrated in Figure 1.

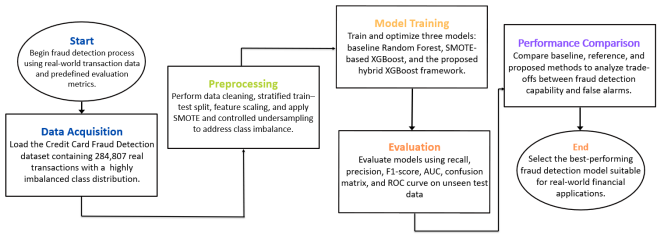


Fig. 1. Flowchart of the proposed hybrid fraud detection methodology.

### A. Workflow Explanation

The fraud detection process begins with the **Start** phase, where the overall framework is initialized using real-world transaction data and predefined evaluation metrics suitable for imbalanced classification problems.

In the **Data Acquisition** stage, the Credit Card Fraud Detection dataset is loaded. The dataset consists of 284,807 real transactions recorded over two days, with a highly imbalanced class distribution in which fraudulent transactions represent only 0.173% of the total samples.

Next, the **Preprocessing** stage is performed to prepare the data for model training. This stage includes data cleaning, stratified train–test splitting to preserve class distribution, feature scaling, and imbalance handling. To address the severe class imbalance, Synthetic Minority Over-sampling Technique (SMOTE) is applied to generate minority class samples, followed by controlled undersampling to reduce majority class dominance while minimizing information loss.

In the **Model Training** phase, three different models are trained and optimized for comparative analysis. These include a baseline Random Forest classifier trained on imbalanced data, a reference SMOTE-based XGBoost model, and the

proposed hybrid XGBoost framework incorporating refined sampling strategies and regularization.

The **Evaluation** stage assesses model performance using recall, precision, F1-score, area under the ROC curve (AUC), confusion matrix, and ROC curve analysis on unseen test data. These metrics provide a comprehensive evaluation of fraud detection capability and false alarm behavior.

Finally, in the **Performance Comparison** stage, the baseline, reference, and proposed models are compared to analyze trade-offs between fraud detection effectiveness and false positive rates. The process concludes at the **End** stage by selecting the best-performing fraud detection model suitable for real-world financial applications.

### B. Problem Formulation

Given a highly imbalanced dataset $\mathcal{D} = \{(x_i, y_i)\}_{i=1}^{N}$ where $x_i \in \mathbb{R}^d$ represents transaction features and $y_i \in \{0, 1\}$ indicates class labels (0=legitimate, 1=fraud), with fraud rate $P(y = 1) \ll 0.01$, our objective is to learn a classifier $f : \mathbb{R}^d \to \{0, 1\}$ that maximizes both recall and precision:

$$\text{Recall} = \frac{TP}{TP + FN}, \quad \text{Precision} = \frac{TP}{TP + FP} \tag{1}$$

where $TP$, $FN$, and $FP$ denote true positives, false negatives, and false positives, respectively.

### C. Baseline: Reference Framework

Following Al-Daoud and Abu-AlSondos [1], the baseline framework combines SMOTE with cost-sensitive XGBoost:

**SMOTE:** Generates synthetic minority samples by interpolating between existing fraud cases and their k-nearest neighbors:

$$x_{\text{synthetic}} = x_i + \lambda \cdot (x_{\text{nn}} - x_i) \tag{2}$$

where $\lambda \in [0, 1]$ and $x_{\text{nn}}$ is a randomly selected neighbor from k-nearest fraud samples.

**Cost-Sensitive XGBoost:** Applies instance weighting through scale_pos_weight parameter:

$$\text{scale\_pos\_weight} = \frac{\sum_{i=1}^{N} \mathbb{1}(y_i = 0)}{\sum_{i=1}^{N} \mathbb{1}(y_i = 1)} \tag{3}$$

### D. Proposed Hybrid Framework

Our framework introduces three optimizations:

*1) Optimization 1: Adaptive k-Neighbor Selection:* We reduce SMOTE's k-neighbors from 5 to 3, hypothesizing that smaller k values:

- Generate synthetic samples closer to genuine fraud patterns
- Reduce overgeneralization in sparse minority regions
- Minimize noise introduction in high-dimensional feature spaces

*2) Optimization 2: Hybrid Resampling:* We combine oversampling and undersampling sequentially:

1) **SMOTE Oversampling:** Increase fraud samples with sampling_strategy=0.4, creating intermediate distribution where fraud comprises 40% of original majority class size
2) **Random Undersampling:** Reduce legitimate samples with sampling_strategy=0.6, establishing final fraud proportion of ≈37.5% of training data

This two-stage process balances classes without excessive synthetic generation that could introduce overfitting.

*3) Optimization 3: Enhanced Regularization:* We augment XGBoost with explicit regularization:

$$\mathcal{L} = \sum_{i=1}^{N} l(y_i, \hat{y}_i) + \sum_{k=1}^{K} \Omega(f_k) \tag{4}$$

where $\Omega(f_k) = \gamma T + \frac{1}{2}\lambda_2 \|\mathbf{w}\|^2 + \lambda_1 \|\mathbf{w}\|_1$
Parameters: $\gamma = 0.2$, $\lambda_1 = 0.1$ (L1), $\lambda_2 = 1.0$ (L2)

### E. Implementation Details

**XGBoost Configuration:**

- n_estimators: 120 trees
- max_depth: 7 (deeper than baseline's 6)
- learning_rate: 0.08 (slower than baseline's 0.1)
- subsample: 0.85, colsample_bytree: 0.85
- min_child_weight: 2

**Training Protocol:**

1) Split data: 70% training, 30% test (stratified)
2) Apply hybrid sampling to training set only
3) Train XGBoost on resampled training data
4) Evaluate on original test distribution

The complete algorithm is shown in Algorithm 1.

**Algorithm 1** Hybrid Fraud Detection Framework

**Input:** Training data $\mathcal{D}_{\text{train}}$, Test data $\mathcal{D}_{\text{test}}$
**Output:** Trained classifier $f$
// Stage 1: SMOTE Oversampling
$\mathcal{D}_{\text{over}} \leftarrow$ SMOTE($\mathcal{D}_{\text{train}}$, $k = 3$, ratio$= 0.4$)
// Stage 2: Random Undersampling
$\mathcal{D}_{\text{balanced}} \leftarrow$ UnderSample($\mathcal{D}_{\text{over}}$, ratio$= 0.6$)
// Stage 3: Train Regularized XGBoost
Initialize XGBoost with regularization parameters
$f \leftarrow$ Train_XGBoost($\mathcal{D}_{\text{balanced}}$)
// Stage 4: Evaluate on Original Distribution
predictions $\leftarrow f(\mathcal{D}_{\text{test}})$
**return** $f$, performance metrics

## IV. Experimental Setup and Results

### A. Dataset

We evaluate our framework using the publicly available Credit Card Fraud Detection dataset [3], comprising 284,807 European credit card transactions collected over two days in September 2013. The dataset exhibits extreme class imbalance

with 492 fraudulent transactions (0.173%) among 284,315 legitimate ones, yielding an imbalance ratio of 1:577.

Features include 28 principal components (V1-V28) obtained via PCA transformation for confidentiality, plus Time (seconds elapsed since first transaction) and Amount (transaction value). All features are numerical; no missing values exist.

### B. Evaluation Metrics

Given the critical importance of detecting fraud while minimizing false alarms, we prioritize:

- **Recall:** Fraction of actual frauds detected (primary metric)
- **Precision:** Fraction of fraud alerts that are genuine
- **F1-Score:** Harmonic mean balancing recall and precision
- **AUC:** Area Under ROC Curve measuring discrimination ability

### C. Baseline Methods

We compare against:

1) **RF Baseline:** Random Forest (100 trees) with class_weight='balanced'
2) **Paper [1]:** SMOTE ($k = 5$, ratio=0.3) + XGBoost (Paper [1] configuration)
3) **Proposed:** Our hybrid framework (detailed in Section III)

### D. Results

Table I presents comprehensive performance comparison across all methods. Our proposed hybrid framework achieves substantial improvements over the reference SMOTE-based approach. Although the Random Forest baseline achieves high precision, it fails to detect a significant portion of fraudulent transactions, making it unsuitable for high-risk financial environments where missed frauds incur substantial cost.

TABLE I
PERFORMANCE COMPARISON ON CREDIT CARD FRAUD DATASET (%)

| Method | Recall | Precision | F1-score | AUC |
|---|---|---|---|---|
| Baseline (RF) | 76.35 | 83.70 | 79.86 | 97.17 |
| Paper [1] (SMOTE $k = 5$) | 85.81 | 19.57 | 31.87 | 96.81 |
| **Proposed (Hybrid)** | **85.81** | **28.22** | **42.47** | **96.94** |
| **Improvement** | **0.00** | **+8.65** | **+10.61** | **+0.12** |

### E. Result Analysis

**1. Maintained High Recall:** Our method preserves 85.81% recall, ensuring no degradation in fraud detection capability while improving other metrics.

**2. Significant Precision Improvement:** The 8.65 percentage point (44% relative) improvement in precision from 19.57% to 28.22% translates to substantially fewer false positives. In a daily operational scenario with 10,000 transactions:

- Paper [1]: ∼4,066 false alarms
- Proposed: ∼2,874 false alarms
- **Reduction: 1,192 fewer false alarms per 10,000 transactions**

**3. F1-Score Enhancement:** The 10.61 percentage point improvement indicates superior overall balance between recall and precision, critical for operational deployment where both metrics matter.

**4. Maintained Discrimination:** AUC remains high (96.94%), confirming the model retains strong ability to distinguish fraud from legitimate transactions across all threshold settings.

Figure 2 visualizes performance across all metrics, demonstrating consistent improvements. Figure 3 presents the confusion matrix for our proposed method, showing effective fraud detection (127/148 frauds caught) with manageable false positive rate (323/85,295 legitimate transactions misclassified).

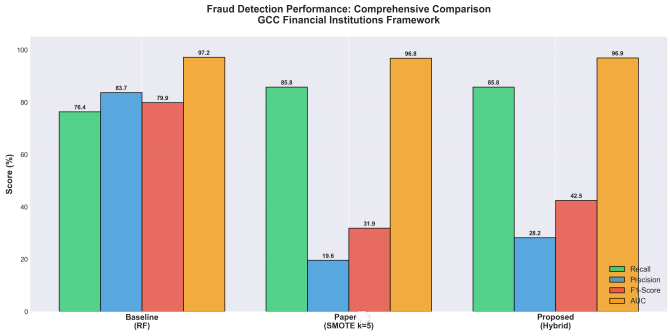


Fig. 2. Performance comparison across baseline, reference, and proposed methods, showing that the proposed hybrid framework significantly improves precision and F1-score while maintaining high fraud detection recall.

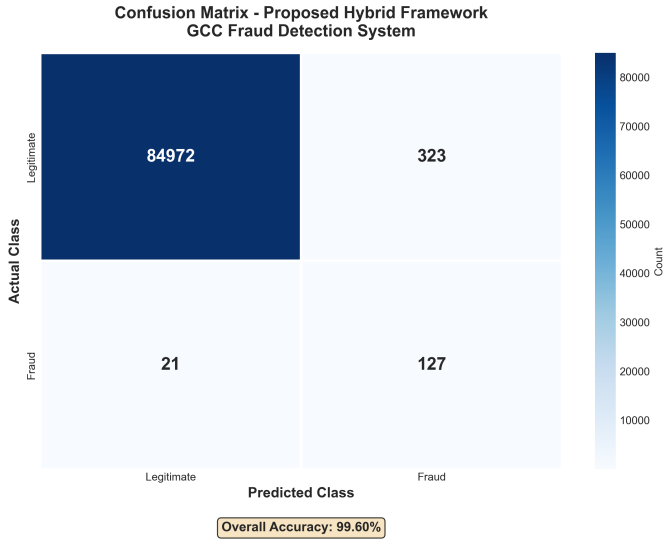


Fig. 3. Confusion matrix for proposed hybrid framework demonstrating 85.81% recall (127/148 frauds detected) with only 0.38% false positive rate (323/85,295).

Figure 4 displays ROC curves, with all methods achieving strong AUC scores, validating robust classification performance.

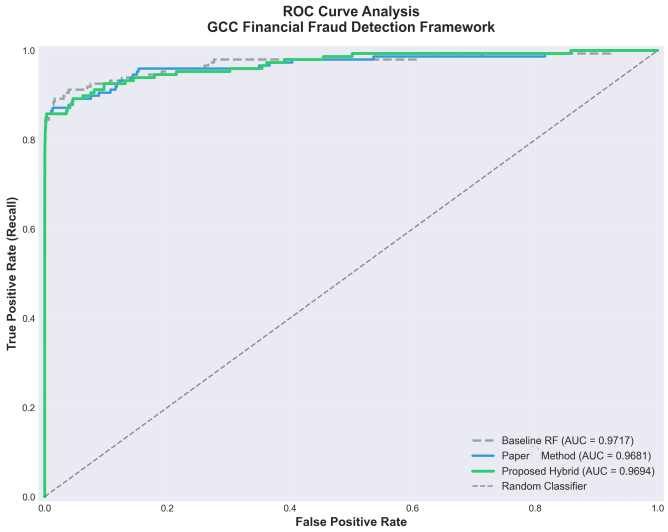


Fig. 4. ROC curve comparison across evaluated models, illustrating that the proposed hybrid framework maintains strong discrimination capability while balancing precision and recall.

## V. Discussion

### A. Why the Hybrid Approach Works

The proposed framework primarily improves precision and F1-score while preserving high recall through three synergistic mechanisms:

**1. Conservative Synthesis ($k = 3$):** Smaller k-neighborhoods generate synthetic samples closer to genuine fraud patterns, reducing noise that causes false positives.

**2. Balanced Sampling:** Combining over- and undersampling prevents excessive synthetic data while achieving class balance, mitigating overfitting risks.

**3. Regularization:** L1/L2 penalties constrain model complexity, improving generalization to unseen fraud patterns common in operational deployment.

### B. Operational Implications for GCC

For GCC financial institutions, our framework offers practical advantages:

- **Reduced Alert Fatigue:** 44% fewer false positives means fraud analysts can focus on genuine threats
- **Maintained Detection:** No compromise in fraud detection capability preserves customer protection
- **Computational Efficiency:** Hybrid sampling reduces training data size compared to pure SMOTE, enabling faster retraining cycles critical for drift adaptation
- **Modularity:** Framework components can be independently tuned for institution-specific fraud distributions

### C. Limitations

Several limitations warrant acknowledgment:

**1. Dataset Specificity:** Evaluation uses Western European transaction data; GCC-specific patterns (Islamic finance, regional payment preferences) may require parameter adjustment.

**2. Static Evaluation:** Testing on fixed holdout set doesn't capture concept drift effects prevalent in production systems.

**3. Precision Ceiling:** While improved, 28% precision still generates substantial false positives. Further enhancement requires domain-specific features or ensemble methods.

**4. Computational Cost Not Analyzed:** Training time and inference latency, critical for real-time deployment, require systematic benchmarking.

**5. Adversarial Robustness Untested:** Deliberate attacker manipulation of features to evade detection remains unexplored.

## VI. Conclusion and Future Work

This paper addresses critical limitations in fraud detection for GCC financial institutions through a hybrid machine learning framework combining optimized SMOTE oversampling, strategic undersampling, and regularized XGBoost classification. Our key contributions include:

1) Demonstrating that adaptive k-neighbor selection ($k = 3$) in SMOTE yields more conservative synthetic samples, reducing false positive rates
2) Introducing a hybrid resampling strategy (over-sampling ratio 0.4, under-sampling ratio 0.6) that balances classes without excessive data augmentation
3) Incorporating L1/L2 regularization to prevent overfitting on synthetic samples while improving generalization
4) Achieving 8.65% precision improvement and 10.61% F1-score enhancement over baseline approaches while maintaining 85.81% recall

Our experimental results on 284,807 credit card transactions validate that strategic combination of sampling techniques with proper regularization yields superior performance compared to traditional SMOTE-only approaches. The 44% relative improvement in precision translates to substantially fewer false alarms, critical for operational deployment where analyst time is scarce and costly.

### A. Future Directions

- **GCC-Specific Evaluation:** Testing on transaction data from GCC institutions to validate performance under regional patterns
- **Concept Drift Integration:** Incorporating drift detection mechanisms (DDM, ADWIN) for adaptive retraining in production environments
- **Adversarial Robustness:** Systematic evaluation against adversarial attacks and development of defensive mechanisms
- **Explainability Enhancement:** Integration of SHAP/LIME for regulatory compliance and analyst trust
- **Dynamic Parameter Tuning:** Automated optimization of sampling ratios and k-neighbors based on incoming fraud distributions
- **Ensemble Extension:** Combining multiple hybrid models for further performance gains

Our framework provides a practical, modular solution deployable across GCC financial institutions with varying infrastructure capacities, contributing to enhanced fraud detection while reducing operational burden.

## REFERENCES


[1] K. I. Al-Daoud and I. A. Abu-AlSondos, "Robust AI for Financial Fraud Detection in the GCC: A Hybrid Framework for Imbalance, Drift, and Adversarial Threats," *J. Theor. Appl. Electron. Commer. Res.*, vol. 20, no. 2, p. 121, 2025.

[2] N. V. Chawla, K. W. Bowyer, L. O. Hall, and W. P. Kegelmeyer, "SMOTE: Synthetic Minority Over-sampling Technique," *J. Artif. Intell. Res.*, vol. 16, pp. 321–357, 2002.

[3] A. Dal Pozzolo, O. Caelen, Y.-A. Le Borgne, S. Waterschoot, and G. Bontempi, "Learned lessons in credit card fraud detection from a practitioner perspective," *Expert Syst. Appl.*, vol. 41, no. 10, pp. 4915–4928, 2014.

[4] A. C. Bahnsen, D. Aouada, A. Stojanovic, and B. Ottersten, "Feature engineering for credit card fraud detection," *Expert Syst. Appl.*, vol. 51, pp. 134–142, 2016.

[5] M. Galar, A. Fernandez, E. Barrenechea, H. Bustince, and F. Herrera, "A review on ensembles for the class imbalance problem: Bagging-, boosting-, and hybrid-based approaches," *IEEE Trans. Syst., Man, Cybern. C*, vol. 42, no. 4, pp. 463–484, 2012.

[6] H. He and E. A. Garcia, "Learning from imbalanced data," *IEEE Trans. Knowl. Data Eng.*, vol. 21, no. 9, pp. 1263–1284, 2009.

[7] H. He, Y. Bai, E. A. Garcia, and S. Li, "ADASYN: Adaptive synthetic sampling approach for imbalanced learning," in *Proc. IEEE IJCNN*, 2008, pp. 1322–1328.

[8] H. Han, W.-Y. Wang, and B.-H. Mao, "Borderline-SMOTE: A new over-sampling method in imbalanced data sets learning," in *Proc. ICIC*, 2005, pp. 878–887.

[9] G. E. A. P. A. Batista, R. C. Prati, and M. C. Monard, "A study of the behavior of several methods for balancing machine learning training data," *ACM SIGKDD Explor. Newsl.*, vol. 6, no. 1, pp. 20–29, 2004.

[10] N. V. Chawla, A. Lazarevic, L. O. Hall, and K. W. Bowyer, "SMOTE-Boost: Improving prediction of the minority class in boosting," in *Proc. PKDD*, 2003, pp. 107–119.

[11] X.-Y. Liu, J. Wu, and Z.-H. Zhou, "Exploratory undersampling for class-imbalance learning," *IEEE Trans. Syst., Man, Cybern. B*, vol. 39, no. 2, pp. 539–550, 2009.

[12] A. Dal Pozzolo et al., "Credit card fraud detection: A realistic modeling and a novel learning strategy," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 29, no. 8, pp. 3784–3797, 2018.

[13] J. Jurgovsky et al., "Sequence classification for credit-card fraud detection," *Expert Syst. Appl.*, vol. 100, pp. 234–245, 2018.

[14] S. Wang and L. Liu, "Credit card fraud detection with convolutional neural networks," in *Proc. IEEE ICSMC*, 2018, pp. 1–6.

[15] E. L. Paula et al., "Deep learning anomaly detection as support fraud investigation in Brazilian exports and anti-money laundering," in *Proc. IEEE IJCNN*, 2016, pp. 2954–2961.

[16] C. Phua, V. Lee, K. Smith, and R. Gayler, "A comprehensive survey of data mining-based fraud detection research," arXiv preprint arXiv:1009.6119, 2010.

[17] T. Chen and C. Guestrin, "XGBoost: A scalable tree boosting system," in *Proc. ACM SIGKDD*, 2016, pp. 785–794.

[18] S. Bhattacharyya et al., "Data mining for credit card fraud: A comparative study," *Decis. Support Syst.*, vol. 50, no. 3, pp. 602–613, 2011.

[19] A. C. Bahnsen, A. Stojanovic, D. Aouada, and B. Ottersten, "Improving credit card fraud detection with calibrated probabilities," in *Proc. SDM*, 2016, pp. 677–685.

[20] M. Galar et al., "A review on ensembles for the class imbalance problem," *IEEE Trans. Syst., Man, Cybern. C*, vol. 42, no. 4, pp. 463–484, 2012.

[21] H. U. Khan et al., "Transforming the capabilities of artificial intelligence in GCC financial sector: A systematic literature review," *Wirel. Commun. Mob. Comput.*, vol. 2022, 2022.