IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received 7 August 2025, accepted 28 August 2025, date of publication 9 September 2025, date of current version 12 November 2025.

Digital Object Identifier 10.1109/ACCESS.2025.3608032

RESEARCH ARTICLE

Hybrid Reinforcement Learning-Active Learning Framework for Real-Data Augmentation in Imbalanced Credit Scoring

AZREE NAZRI[1], OLALEKAN AGBOLADE[2], AND FAISAL AZIZ[2]
[1]Institute for Mathematical Research (INSPEM), Serdang 43400, Malaysia
[2]Department of Computer Science, University Putra Malaysia, Serdang 43400, Malaysia

Corresponding author: Azree Nazri (azree@upm.edu.my)

This work was supported by University Putra Malaysia under Grant GPS-IPS/2021/970000.

ABSTRACT Class imbalance is a critical challenge in credit scoring, where the dominance of majority class samples reduces predictive performance for minority instances. Traditional methods, such as SMOTE and random undersampling, attempt to rebalance datasets but often introduce synthetic noise or discard valuable data. This paper introduces Augmentation Based on Uncertainty and Difficulty (UDDA), a novel real-data augmentation framework that avoids both synthetic data generation and majority class removal. UDDA leverages a hybrid of Reinforcement Learning (RL) and Active Learning (AL) to identify and prioritize real, informative, and difficult-to-classify samples. This approach enhances model robustness while preserving dataset integrity. Experimental evaluation on twenty benchmark imbalanced tabular datasets demonstrates that UDDA outperforms established methods, including SMOTE and undersampling, across key metrics such as precision, recall, F1-score, and accuracy. UDDA sets a new direction in imbalanced learning by offering a practical, interpretable, and effective solution for improving classification performance in credit scoring applications.

INDEX TERMS Reinforcement learning, active learning, class imbalance, synthetic data, real data.

I. INTRODUCTION

In the field of machine learning, it is commonly believed that the sample sizes for each class are about similar. However, in real-world circumstances, particularly in industrial and corporate applications, data typically has an uneven distribution. For example, in fault detection, the majority class has a high number of samples, but the minority class (representing defects) contains only a few samples [1]. This imbalance presents a difficulty for typical learning algorithms, which frequently emphasize dominant classes while overlooking minority ones. Nonetheless, from a data mining perspective, minority classes typically include crucial information, making them quite useful. As a result, the objective of unbalanced learning is to create intelligent systems that can successfully minimize this bias, enabling learning algorithms to better handle and learn from imbalanced datasets.

The associate editor coordinating the review of this manuscript and approving it for publication was Jiachen Yang.

Over the last two decades, research has focused heavily on the problem of uneven learning. To overcome this issue, many alternatives have been offered, such as data pretreatment techniques, altering current classifiers, and fine-tuning algorithmic parameters. Unbalanced data is a common issue in many real-world applications, including fraud detection, medical diagnosis [1], [2], [3], fault detection [4], [5], [6], and other areas [7], [8].

In many real-world applications, datasets exhibit severe class imbalances, with certain classes having a limited number of samples and the majority class dominating. This imbalance causes a discrepancy in the performance of learning algorithms, which frequently favors the majority class and leads to what is known as performance bias. To address this challenge, scholars and practitioners have extensively investigated the topic of unbalanced learning, resulting in a number of creative methodologies and procedures targeted at improving classification outcomes [7], [9], [10]. Addressing unbalanced data has become increasingly important as the

number of data created by sophisticated, large-scale, and linked systems in cybersecurity [11], banking [12], and the Internet has increased. While EMST measures class separation by analyzing distances, it provides insights into model behavior, identifies overlapping classes, and evaluates data quality—all without exposing or relying on access to original training data [33]. Despite significant progress, there is still a paucity of systematic research that comprehensively examines current breakthroughs, upcoming obstacles, and future opportunities in unbalanced learning.

There are five types of imbalanced categories: General approaches include ensemble learning techniques, imbalanced regression and clustering, long-tail learning, and imbalanced data streams. Within each of these categories, there are various specific solutions, such as data-level strategies, algorithm modifications, hybrid methods, and ensemble frameworks. Within ensemble approaches, we look at boosting, bagging, and cost-sensitive ensembles. We also discuss topics such as unbalanced regression, clustering, online learning, idea drift, incremental learning, class rebalancing, information improvement, and model refining. Using this categorization, we conduct a detailed assessment of existing unbalanced learning approaches and highlight current developments in practical applications.

To fill this gap, this study proposes a novel framework that tackles class imbalance using real, informative samples—avoiding synthetic data generation and majority class removal. By combining hybrid reinforcement learning with margin-based active learning, the framework enhances model performance while preserving data integrity.

By incorporating an augmentation strategy based on uncertainty and difficulty, the model dynamically prioritizes critical data instances, improving classification performance. This technique effectively enhances fraud detection accuracy by leveraging real data, mitigating the challenges associated with synthetic oversampling and majority class undersampling.

The main contribution of our paper is summarized below:

I. To address the issue of imbalanced datasets without relying on synthetic data, a novel augmentation framework based on Reinforcement Learning (RL) and Active Learning (AL) is introduced. This method, Augmentation Based on Uncertainty and Difficulty (UDDA), selects real, informative, and difficult-to-classify samples to enhance model learning.

II. To improve classifier performance, Active Learning is employed to identify uncertain samples, while Reinforcement Learning prioritizes challenging instances, optimizing model training.

III. By avoiding synthetic data generation and instead leveraging real data augmentation, this approach mitigates overfitting risks associated with synthetic oversampling and prevents data loss caused by majority class undersampling. This method has demonstrated superior performance in credit scoring tasks, improving precision, recall, F1-score, and accuracy in minority class prediction.

II. RELATED WORKS

Table 1 summarizes the significant differences between existing reviews on data balancing techniques and this work.

TABLE 1. Summary of sampling techniques in credit transcaction dataset.

Reference	Data balancing techniques (synthetic/remove/real)	Sampling Technique
[15]	Synthetic + remove	SMOTE-ENN
[16]	Remove	NUS
[17]	Synthetic	SMOTE, ADASYN, B1-SMOTE, B2-SMOTE, SVMSMOTE
[18]	Synthetic	Undersampling, SMOTE, SMOTE-Tomek
[19]	Synthetic/ remove	SMOTE-Tomek
[20]	Synthetic / remove	SMOTE-Tomek
[21]	Synthetic / remove	SMOTE-ENN
[22]	Synthetic	ASN-SMOTE
[23]	Synthetic / remove	Hybrid Random undersampling and Borderline SMOTE
[24]	remove	Random undersamlping
[25]	Synthetic data / remove	Combine Random oversampling, SMOTE, ADASYN, boderderline-SMOTE, and SVMSMOTE with Random undersampling

A recent study [23] explored the effectiveness of a classification model that integrates both oversampling and undersampling methods for fraud detection within a specialized dataset. The model employed five distinct synthetic data generation techniques: random oversampling, SMOTE, adaptive synthetic sampling (ADASYN), Borderline-SMOTE (B-SMOTE), and support vector machine SMOTE (SVM-SMOTE), alongside random undersampling, which removes a portion of the majority class. These sampling methods were assessed in conjunction with a random forest (RF) classifier to evaluate the model's performance. The findings indicated that combining random undersampling (RUS) with one or more synthetic oversampling approaches significantly enhances classification accuracy. The study further determined that utilizing a hybrid of these synthetic/removal-based techniques led to superior model performance when compared to the use of either method in isolation. This highlights

the continued reliance on synthetic and removal-based strategies for class balancing, underscoring the need for alternative methods that preserve data authenticity.

In contrast, other scholars [17] investigated various removal-based undersampling techniques combined with synthetic data generation methods such as SMOTE and SMOTE-Tomek to handle imbalanced datasets. In this research, classification models—including KNN, LR, RF, and SVM—were trained on artificially balanced datasets for the detection of fraudulent credit card transactions. The results indicated that RF, when paired with SMOTE or SMOTE-Tomek, yielded the most accurate predictions. Two additional studies [18], [19] also applied SMOTE-Tomek to credit card transaction data to mitigate the issue of data imbalance. Their findings confirmed that this combination of synthetic oversampling and removal-based undersampling not only improved the learning rate but also significantly enhanced the model's performance compared to using imbalanced datasets. These studies further illustrate the dominant trend in the literature toward synthetic and removal-based balancing techniques, emphasizing the need for real-data-driven alternatives that maintain data authenticity.

A further study [20] employed a hybrid method combining SMOTE and edited nearest neighbour (SMOTE-ENN) to rectify class imbalance in a credit card transaction dataset. The findings indicated that SMOTE-ENN, when integrated with an ensemble deep learning model, delivered superior performance. This hybrid sampling technique significantly enhanced the detection model's accuracy. Similarly, another study [21] also utilised SMOTE-ENN for balancing a credit card dataset. SMOTE-ENN oversamples the minority class using SMOTE and removes overlapping instances via ENN.

In contrast, a different study [22] introduced SMOTE with adaptive qualified synthesizer selection (ASN-SMOTE), a more refined synthetic oversampling technique that incorporates KNN and SMOTE. ASN-SMOTE filters noise in the minority class by assessing whether each minority instance's nearest neighbour belongs to the minority or majority class. It then uses the nearest majority instance to better understand the decision boundary, selectively synthesising new minority instances using an adaptive neighbour selection method. Despite its improved precision, ASN-SMOTE remains rooted in synthetic data generation, continuing the trend of augmenting datasets through artificial instances. The study concluded that ASN-SMOTE outperformed nine leading oversampling methods, further reinforcing the field's reliance on synthetic techniques rather than leveraging informative real data.

Additionally, research [21] explored a hybrid approach combining random undersampling (RUS) with Borderline-SMOTE (B-SMOTE) to address class imbalance. While this method improved the model's performance—achieving an F1-score of 70%—it still relied on a combination of removal-based (RUS) and synthetic data generation (B-SMOTE) techniques. The hybrid strategy aimed to mitigate the known drawbacks of RUS, such as information loss, while also reducing the risks of overfitting and overgeneration commonly associated with SMOTE in large datasets. Despite these improvements, the study continues the prevalent trend of addressing imbalance through artificial instance synthesis and majority class reduction, underscoring the need for alternative solutions based on real, informative data.

A subsequent study [2] introduced a novel approach to undersampling for addressing class imbalance, known as noise-sample-removed undersampling (NUS). This clustering-based removal technique eliminates noise samples from both the majority and minority classes before applying traditional undersampling methods. While NUS offers an improvement by refining the dataset prior to reduction, it still operates within the removal-based paradigm, reducing dataset size and potentially discarding informative instances. The technique was evaluated on 13 public datasets and three real-world datasets, consistently outperforming several established methods, including random undersampling (RUS), SMOTE, ADASYN, SMOTE-Tomek, and edited nearest neighbour (ENN). Despite its effectiveness, NUS—like many others—continues the field's dependency on either removing or synthesising data, highlighting the lack of approaches that leverage real, high-value samples without altering dataset composition.

III. PRELIMINARIES

Credit card fraud detection and credit scoring are two examples of binary classification tasks that are frequently complicated by highly unbalanced datasets. In real-world instances like credit card fraud, illicit transactions make up fewer than 1% of overall data, making it difficult for machine learning algorithms to effectively categorize minority groups. Similarly, credit score databases suffer from class imbalance, with the minority class (poor credit consumers) being far smaller than the majority class (excellent credit customers). Using typical machine learning models on such datasets frequently results in significant misclassification rates, especially for the minority class, which can be expensive for both fraud detection and credit scoring.

To solve this issue, fraud detection frameworks generally inject fake data using oversampling techniques such as SMOTE to produce a hybrid dataset with a smaller gap between the classes. This balanced dataset allows machine learning classifiers like Random Forest to anticipate minority class instances more correctly, which improves the identification of fraud or bad credit clients.

Rather of depending on synthetic data augmentation, this study suggests a hybrid architecture for credit scoring that combines Active Learning (AL) and Reinforcement Learning (RL) so-called uncertainty and difficult data augmentation (UDDA). The goal is to increase classification performance without changing or enhancing the original dataset, hence lowering the danger of overfitting or producing false positives. UDDA strategy deliberately selects meaningful samples to help the model learn more quickly, reducing misclassification costs and eliminating the need for huge training datasets. This novel technique alleviates the

strain of dealing with unbalanced datasets while retaining excellent classification accuracy, especially in key sectors such as credit scoring and fraud detection.

In this experiment, the baseline method leverages a Random Forest classifier. The Random Forest classifier consists of an ensemble of decision trees, each trained on a random subset of the training data. During the training process, the training set D0 (Figure 1) passes through the decision trees, and each tree generates its own classification result. The final classification output is obtained by aggregating the predictions from all the trees using a majority voting approach. The Random Forest optimizes the model parameters by constructing multiple decision trees and maintains these parameters after the model training. When the test data is input, the Random Forest model outputs the classification results. Finally, these classification results are evaluated using performance metrics such as accuracy, precision, recall, and f1-score.

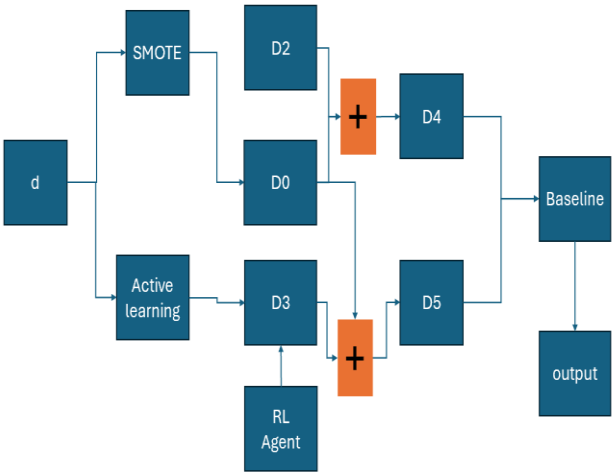

FIGURE 1. The minority group sample d from the training set is fed into the sampling module. This module generates positive samples, D2, which are merged with the original training data D0 to form a new dataset, D3. The combined dataset D3 is used to train the classifier in the benchmark model. The classifier, after being trained with the updated dataset, processes the test data and provides the final classification results.

A. REINFORCEMENT LEARNING

In this context, both Reinforcement Learning (RL) and Active Learning (AL) are used to tackle the challenge of class imbalance and improve the performance of a classifier, such as a Random Forest. Although the code does not implement traditional RL algorithms like Q-learning or policy gradients, it uses RL-inspired principles of learning from feedback and active exploration in combination with AL's focus on sample selection based on uncertainty.

1) REINFORCEMENT LEARNING

In traditional RL, an agent interacts with an environment by taking actions and receiving rewards or penalties based on the outcomes of those actions. The agent's goal is to maximize cumulative rewards over time by learning an optimal policy $\pi(\mathrm{a}|\mathrm{s})$, which tells the agent which action a to take in state s.

The key formula for updating the agent's policy is the Bellman equation:

$$V(s) = \hat{E}_{\mathrm{a}\sim\pi(\mathrm{a}|\mathrm{s})}[r_{\mathrm{t}} + \gamma V(s')] \tag{1}$$

V(s) is the value of being in state s,
$\mathrm{r_t}$ is the reward at time stop ti,
γ is the discount factor,
s' is the next state

In the implementation, an RL-inspired mechanism is used instead of a formal RL framework. The model learns from feedback by identifying and focusing on misclassified samples (i.e., samples where the model made wrong predictions).

- Misclassified Samples: These are analogous to receiving negative rewards r_t in Reinforcement Learning, where r_t represents the reward signal at time step t. Misclassified samples are viewed as challenging cases that the model needs to revisit and learn from.
- Augmentation Strategy: After identifying misclassified samples, they are added back to the training set to allow the model to focus on them. The augmented training set is:

$$X_{augmented} = X_{train} U X_{misclassified} \tag{2}$$
$$y_{augmented} = y_{train} U y_{misclassified} \tag{3}$$

- Policy Update (Retraining): In RL, the agent updates its policy based on past rewards. Similarly, in this method, the model is retrained on the augmented dataset, which serves as a policy update, enabling the model to perform better on previously misclassified cases. This is akin to refining the model's decision-making ability (i.e., improving the decision boundary for classification).

2) ACTIVE LEARNING

In Active Learning (AL), a model is trained incrementally by selecting the most informative samples from a pool of unlabelled data. Instead of training on all available data, AL focuses on learning from uncertain or ambiguous samples that are expected to provide the most information to improve the model's performance. The core idea is to reduce the amount of data needed to achieve a certain level of accuracy.

The uncertainty sampling method is one of the most common strategies in AL, where the model selects samples about which it is most uncertain. A common criterion for uncertainty is the margin between predicted probabilities for different classes:

$$Margin = |P(y = 1|x) - P(y = 0|x) \tag{4}$$

Samples with the smallest margin (i.e., the least difference between the probabilities of the top two predicted classes) are selected because the model is least confident in its predictions for those samples.

In this code, Active Learning is applied through a margin-based strategy, where the most uncertain samples from the test set are selected for further training.

- Uncertainty Sampling: The model predicts the class probabilities for each sample in the test set. The samples with the smallest margin between predicted probabilities are considered the most uncertain as shown in Equation (4)
- These uncertain samples are then selected to be added to the training set for further learning.
- Class Imbalance Handling: Since the data is imbalanced, with some classes having fewer samples than others, the Active Learning strategy also takes class proportions into account. It selects samples from both the minority and majority classes, ensuring that the model learns effectively from both underrepresented and overrepresented classes.

The selected uncertain samples are then added to the training set:

$$X_{augmented} = X_{train} U X_{selected\_AL} \tag{5}$$

$$y_{augmented} = y_{train} U y_{selected\_AL} \tag{6}$$

By retraining the model with these highly uncertain samples, the model is able to improve its decision boundary, especially around the areas where it was initially uncertain.

3) COMBINING REINFORCEMENT LEARNING AND ACTIVE LEARNING

The code integrates both Reinforcement Learning (RL)-inspired principles and Active Learning (AL) in the following ways and shown in Algorithm 1:

- Active Learning for Exploration: AL is used to select uncertain samples based on their margin between predicted probabilities. These uncertain samples represent areas in the feature space where the model is unsure about its predictions, which is analogous to exploration in RL, where the agent explores unknown states to gather more information.
- Reinforcement Learning for Corrective Feedback: RL-inspired feedback is implemented by identifying misclassified samples. These misclassified instances represent cases where the model has made errors. By adding them back to the training set, the model learns from its mistakes, which mimics the RL principle of policy updates based on rewards and penalties.
- Retraining (Policy Update): After augmenting the training set with both uncertain (AL-selected) and misclassified (RL-feedback) samples, the model is retrained. This retraining process is analogous to the policy update in RL, where the agent refines its decision-making strategy based on past experience (in this case, past errors and uncertain samples).

Algorithm 1 Hybrid Reinforcement Learning and Active Learning Algorithm

```
# Pseudocode for AL and RL-Inspired Approach (Without Initial Model Training)

# Step 1: Load and Preprocess Data
LOAD dataset
ENCODE categorical variables (if needed)
SPLIT dataset into training set (X_train, y_train) and test set (X_test, y_test)

# Step 2: Active Learning (AL) - Select Uncertain Samples
FOR each sample in X_test:
    CALCULATE predicted probabilities using an existing model
    CALCULATE margin = | P(y=1) - P(y=0) | # Measure uncertainty
    ADD sample index, margin, true label to list of uncertain samples

SORT uncertain samples by margin in ascending order
SELECT top N uncertain samples for further training
 X_test_selected, y_test_selected = selected uncertain samples

# Step 3: Reinforcement Learning (RL)-Inspired Feedback
IDENTIFY misclassified samples where predictions != y_test
X_test_misclassified, y_test_misclassified = misclassified samples

# Step 4: Augment Training Set
ADD selected uncertain samples from AL to training set
ADD misclassified samples from RL feedback to training set
X_train_augmented = CONCAT(X_train, X_test_selected, X_test_misclassified)
y_train_augmented = CONCAT(y_train, y_test_selected, y_test_misclassified)

# Step 5: Retrain the Model
INITIALIZE new RandomForestClassifier
    FIT new model on (X_train_augmented, y_train_augmented)
PREDICT y_pred_rf_after on X_test

# Step 6: Evaluate Retrained Model
PRINT classification report for y_test and y_pred_rf_after
COMPARE performance before and after retraining
```

IV. DATASET

In this study, two distinct datasets are utilised to evaluate the proposed methodologies: the German Credit Scoring dataset and a publicly available credit card fraud detection dataset from Kaggle.

The German Credit Scoring dataset is a well-established benchmark dataset commonly used in financial and credit risk modelling research. It consists of 1,000 instances, each representing an individual credit applicant, with 20 features detailing various attributes of the applicants, such as age,

employment status, credit amount, and duration of the loan. The target variable indicates whether the applicant poses a good or bad credit risk. This dataset presents a balanced classification problem and serves as a standard for assessing credit risk prediction models.

The second dataset, sourced from Kaggle, focuses on credit card fraud detection and offers a real-world example of fraud analysis. The dataset can be downloaded at the URL https://www.kaggle.com/mlg-ulb/creditcardfraud. This dataset contains 284,807 transactions carried out by European cardholders over two days in September 2013. Of these transactions, only 492 are fraudulent, accounting for approximately 0.172% of the total, making it a highly imbalanced dataset—a common challenge in fraud detection tasks. The dataset consists of anonymized numerical variables that have undergone a Principal Component Analysis (PCA) transformation to protect sensitive information. The only features not transformed by PCA are Time (the number of seconds since the first transaction) and Amount (the transaction value). The target variable indicates whether a transaction is fraudulent (1) or legitimate (0).

The combination of these datasets enables a robust evaluation of UDDA's performance across different financial domains, from credit scoring to fraud detection, and presents distinct challenges, such as class imbalance and feature anonymization, which are common in real-world financial datasets.

A. DATASET PREPARATION

The dataset underwent a thorough cleaning process, designed to remove skewed entries, outliers, and missing data. Following this, preprocessing steps were applied to the dataset prior to its use in model training, with the aim of reducing noise and ensuring the data's quality. Additionally, random sampling was employed to enhance representativeness, alongside feature selection to isolate informative variables. Any missing values were appropriately addressed, and the data were then organized into a suitable structure for further analysis and modeling.

Missing Value: Identifying missing values is a key aspect of the data cleaning process. To detect the presence of null or incomplete entries within the dataset, the isnull() method is employed. This function scans the dataset and produces a DataFrame, where each value is substituted with a Boolean: 'True' for missing data and 'False' otherwise. For example, when this technique was applied to the Germa and the Kaggle dataset, the method confirmed that no null or missing values were present, indicating a complete dataset suitable for further analysis.

Duplicated Values: After addressing missing values, the subsequent task involves identifying duplicated entries within the dataset. This can be achieved using the duplicated() method, which examines each row across the dataset to detect any repetitive instances. The method flags duplicate rows by returning Boolean values: 'True' for duplicates and 'False' for unique entries. For example, when applied to the German and the Kaggle dataset, no duplicated rows were identified, with all Boolean outputs returning 'False', confirming the dataset's integrity for further analysis.

To further evaluate the effectiveness of the proposed classifiers under varying degrees of class imbalance, we selected five binary classification datasets with diverse imbalance ratios (IR). These datasets include both low and high levels of imbalance to assess model robustness across a representative spectrum. Table 2 summarizes the selected datasets, their acronyms, and respective imbalance ratios.

ID1 presents a near-balanced scenario with an IR of 1.41, serving as a baseline for evaluating performance under minimal skew. In contrast, ID5 exhibits a severe imbalance with an IR of 25.58, reflecting real-world challenges where minority class instances are significantly underrepresented. The remaining datasets—ID2, ID4, and ID3—span moderate to high imbalance ratios (7.33 to 13.87), providing a comprehensive basis for performance analysis.

This selection ensures the evaluation framework rigorously tests the proposed models across varying degrees of class disparity, a critical aspect in real-world imbalanced learning tasks.

TABLE 2. A binary imbalanced description.

Acronym	Dataset	Imbalance Ratio
ID1	Data User Modeling	1.41
ID2	haberman	2.78
ID3	vehicle3	2.99
ID4	Appendicitis	4.05
ID5	new-thyroid1	5.14
ID6	Ecoli	5.46
ID7	Winequality-red	6.37
ID8	Fertility Diagnosis	7.33
ID9	Estate	7.37
ID10	Page-blocks0	8.79
ID11	Yeast-0-3-5-9vs7-8	9.12
ID12	Ecoli-0-6-7vs5	10
ID13	led7digit-0-2-4-5-6-7-8 9 vs 1	10.97
ID14	Imbalance-scale	11.76
ID15	cleveland-0 vs 4	12.31
ID16	shuttle-c0-vs-c4	13.87
ID17	page-blocks-1-3 vs 4	15.86
ID18	Oil	21.85
ID19	yeast-2 vs 8	23.10
ID20	car-vgood	25.58

1) DISTRIBUTION OF GERMAN CREDIT SCORING AND THE KAGGLE DATASETS

The distribution analysis of the German Credit Scoring dataset (Figure 2) reveals significant imbalances across several key features, most notably in the target variable. The Target variable, which classifies individuals as either a good (1) or bad (2) credit risk, demonstrates a pronounced skew, with a far greater proportion of individuals labelled as good credit risks. Such imbalance is problematic in machine learning tasks, as models tend to be biased towards the majority class, potentially leading to poor predictive performance for the minority class (in this case, bad credit risks). Addition-

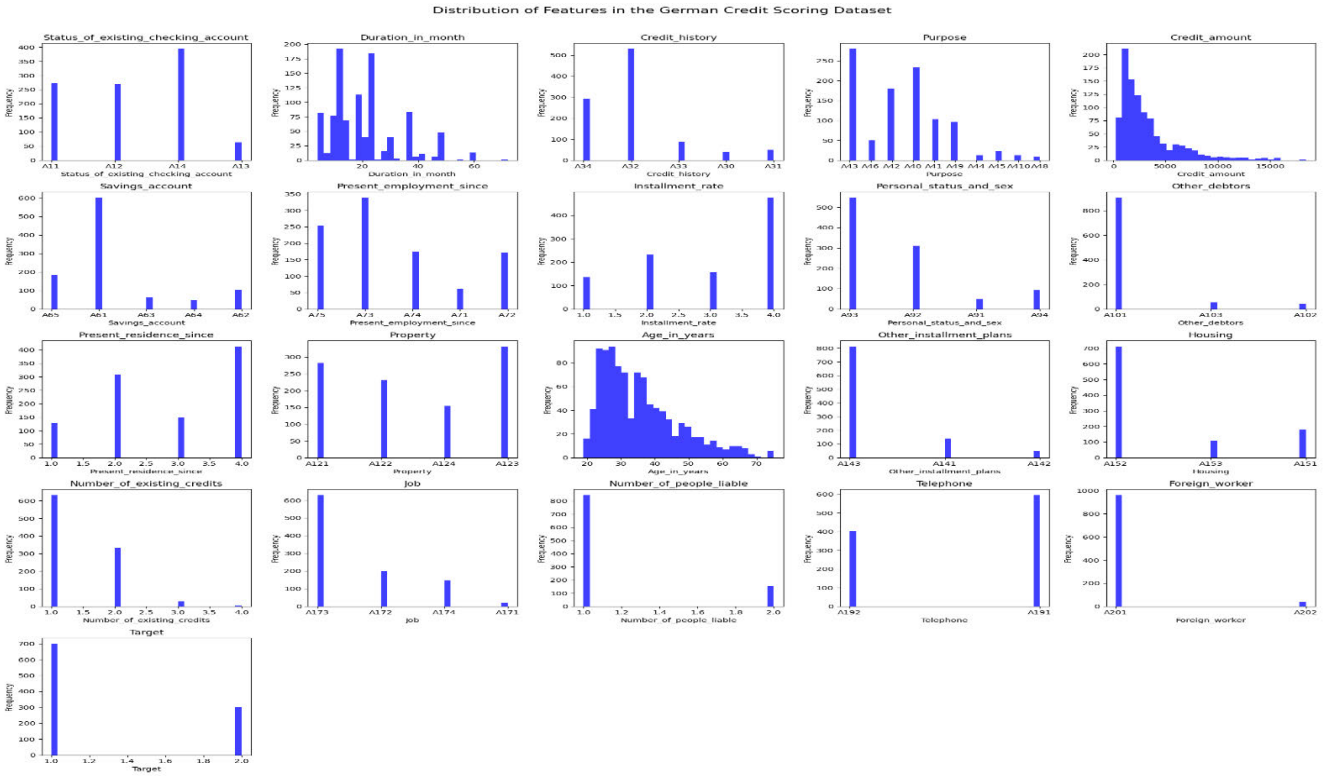

FIGURE 2. Feature distribution and imbalance in the German credit scoring dataset.

ally, features such as Status_of_existing_checking_account, Other_debtors, Foreign_worker, and Telephone also display highly uneven distributions, where certain categories dominate the feature space. For example, nearly all observations in the Foreign_worker feature are concentrated in a single category. These imbalances can lead to overfitting, where the model becomes overly reliant on dominant categories, further exacerbating the challenge of generalisation. To address this, techniques such as oversampling, undersampling, or the application of class weights may be required to ensure the model's ability to accurately classify underrepresented groups within the dataset.

The distribution plot for the Kaggle credit card fraud detection dataset reveals a significant class imbalance, particularly evident in the Class variable, where the majority of transactions are classified as legitimate (label 0), with fraudulent transactions (label 1) representing a very small proportion of the data (Figure 3). This imbalance is common in real-world fraud detection datasets and presents substantial challenges for machine learning models, which tend to be biased towards the majority class. Additionally, several features, such as *V1* through *V28*, derived from Principal Component Analysis (PCA), exhibit relatively normal distributions, while others, such as *Time* and *Amount*, demonstrate more skewed patterns. However, the most critical imbalance lies in the target variable (Class), where fraudulent transactions account for only a small fraction of the dataset. Addressing this imbalance is crucial for effective model performance, as standard classifiers may overlook the minority class, leading to poor detection rates for fraud. Techniques such as resampling methods, cost-sensitive learning, or anomaly detection algorithms could be considered to mitigate this issue and ensure that the model can generalize well to both legitimate and fraudulent transactions.

V. RESULTS

A. OVERALL ACCURACY

For Active Learning framework, Figure 4 shows that as the margin threshold increases, the model's accuracy improves until it stabilizes at a threshold of 0.3. This suggests that in the Active Learning (AL) framework, increasing the margin initially leads to better performance as the model becomes more selective, focusing on uncertain samples with greater confidence. However, the plateau observed beyond the margin of 0.3 indicates that there is a limit to the effectiveness of increasing the margin, beyond which the model's accuracy gains diminish.

This pattern highlights the importance of tuning the margin threshold in Active Learning systems. At lower margins, the model may be prone to making less confident predictions, leading to more errors. Conversely, higher margins enable the model to make more confident and selective predictions, resulting in improved accuracy. However, there is a point at which further increasing the margin provides little to no benefit. Figure 4 suggests that a margin threshold of 0.3 provides the best balance between model confidence and performance, as evidenced by the highest accuracy. Beyond

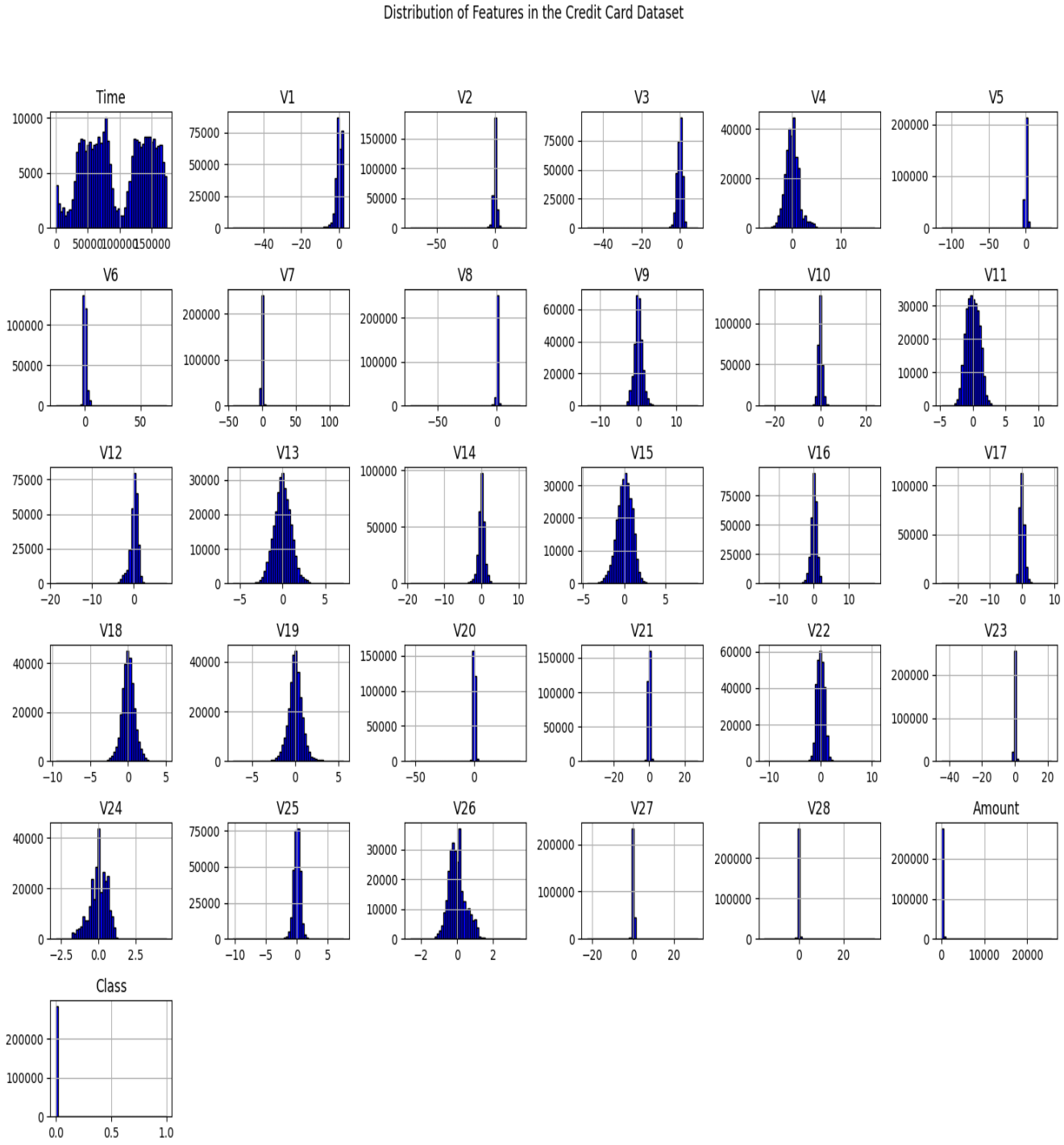

FIGURE 3. Feature distribution and imbalance in the Kaggle dataset.

this threshold, additional adjustments to the margin do not significantly improve the model's accuracy, suggesting that this threshold represents an optimal point for this specific Active Learning task.

Figure 5 illustrates how varying the margin thresholds impacts the accuracy of the UDDA approach, particularly within the scope of handling imbalanced datasets. The significant observation here is the model's strong and stable performance across all margin thresholds, with the accuracy consistently remaining around 0.96 to 0.97.

The small improvement in accuracy from margin 0.1 to 0.2 suggests that allowing the model to focus on more uncertain samples results in better decision-making. However, the lack of improvement beyond margin 0.2 implies that the model has already captured the most relevant information for classification, and further margin increases offer dimin-

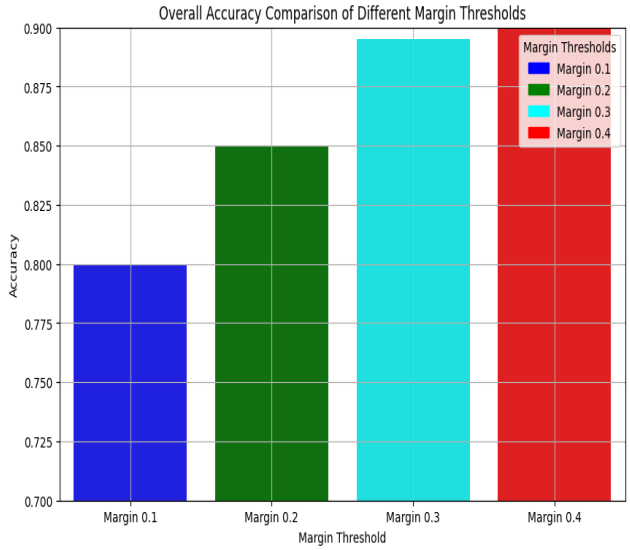

FIGURE 4. Overall accuracy comparison for different margin thresholds in AL only in German credit scoring.

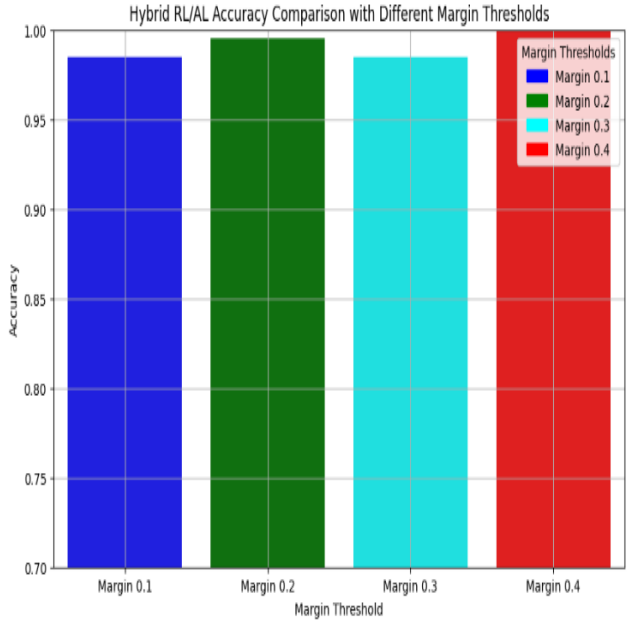

FIGURE 5. Overall accuracy comparison for different margin thresholds in RL/AL in German credit scoring.

ishing returns. This plateau effect can often occur when a model reaches its limit in identifying and managing uncertain or misclassified cases, especially in a well-structured active learning process.

In the context of UDDA, the findings suggest that the model's ability to handle uncertainty is robust even at lower margin thresholds. By margin 0.2, the model has optimized its performance, and further increases in the threshold do not provide additional accuracy. This result is consistent with the nature of reinforcement learning, where focusing on uncertain and misclassified samples allows the model to improve early on, after which the improvements tend to stabilize.

Overall, this analysis underscores that margin-based hybrid RL/AL can lead to a highly accurate model, with diminishing returns in accuracy beyond a certain margin threshold (in this case, 0.2 or 0.3).

B. OVERALL PRECISION

Figure 6 demonstrates a steady improvement in precision as the margin threshold increases from 0.1 to 0.4. This indicates that in the Active Learning (AL) framework, larger margin thresholds lead to more precise classification. The model becomes increasingly confident in its predictions as the margin widens, reducing the number of false positives and thus improving precision. This trend suggests that the model is successfully identifying uncertain cases and handling them more effectively as the threshold increases.

The relatively low precision at the 0.1 margin threshold suggests that when the margin is small, the model tends to misclassify more non-positive instances as positive, resulting in a higher false positive rate. Conversely, the steady improvement in precision at margins 0.2, 0.3, and 0.4 reflects the model's enhanced ability to manage uncertainty by focusing on more ambiguous cases and refining its decision-making process.

From this analysis, it is clear that increasing the margin threshold improves the model's precision in the Active Learning setting. By margin 0.4, the model has achieved its highest level of precision, indicating that further adjustments to the margin threshold may not yield additional gains in precision. In practical applications, the optimal margin threshold should balance precision with other key metrics, such as recall, depending on the specific objectives of the model.

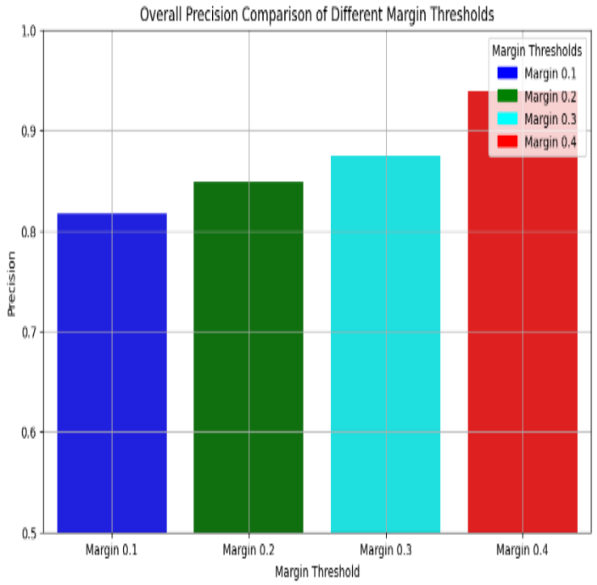

FIGURE 6. Overall precision comparison for different margin thresholds in AL in German credit scoring.

Figure 7 demonstrates that across all margin thresholds (0.1 to 0.4), the precision of the Hybrid RL/AL model remains consistently high at 1.0. This indicates that the model is

exceptionally efficient in its ability to make positive predictions without false positives. The high precision across all thresholds suggests that the Reinforcement Learning component, when combined with Active Learning, is highly effective in prioritizing misclassified or uncertain cases and improving the overall accuracy of the model's predictions.

The stability in precision across the range of margin thresholds suggests that the Hybrid RL/AL approach is less sensitive to variations in the margin threshold when it comes to precision. This indicates that the model is robust in handling ambiguous cases, and its decision-making process, guided by reinforcement learning feedback, ensures that false positive classifications are minimised regardless of the margin setting.

In practical terms, this finding is highly significant. In domains such as credit scoring or fraud detection, where false positives can carry significant consequences, UDDA framework's ability to maintain a perfect precision score indicates its suitability for high-stakes decision-making. The consistency in precision also suggests that the model is well-calibrated to handle uncertainty, making it highly reliable for classification tasks requiring precision.

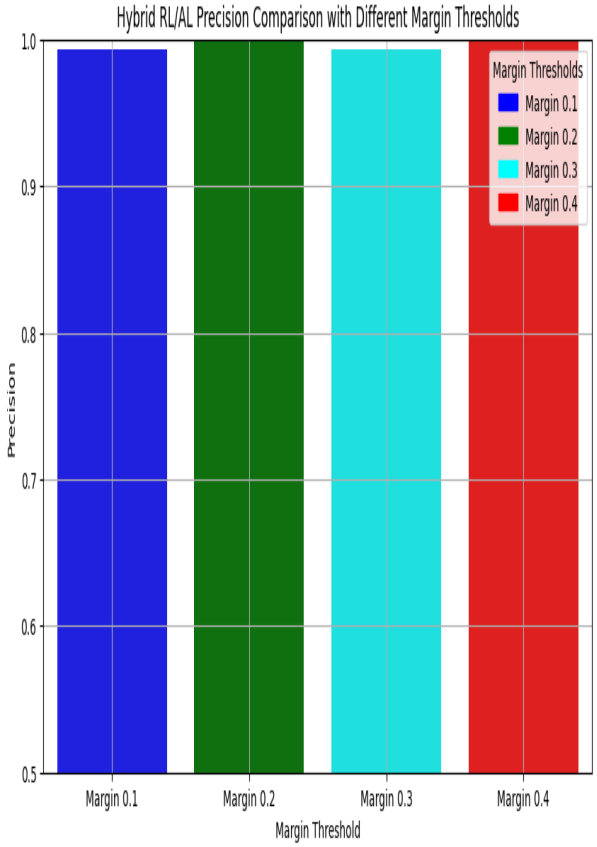

FIGURE 7. Overall precision comparison for different margin thresholds in RL/AL in German credit scoring.

C. OVERALL RECALL

For Active Learning, the overall trend indicates that recall is consistently high across all margin thresholds, with a slight improvement from 0.92 at margin 0.1 to approximately 0.94 at margin thresholds 0.2, 0.3, and 0.4 as shown in Figure 8. The model's recall performance is robust, reflecting that the Active Learning approach, which selects uncertain samples based on margin thresholds, is highly effective in identifying positive cases.

The plateau observed in recall beyond margin 0.2 suggests that further increases in the margin threshold do not significantly affect the model's ability to detect true positives. This could imply that the Active Learning system has optimised its focus on uncertain samples at this margin, and increasing the margin beyond this point does not lead to better identification of positives.

In practical terms, the high recall observed across different margin thresholds is promising for applications where identifying all true positives is critical, such as fraud detection or medical diagnosis. However, it is also important to consider potential trade-offs with precision, as a high recall may lead to more false positives. This analysis highlights the effectiveness of Active Learning in achieving high recall, particularly at a margin threshold of 0.2, which seems to offer the best balance between capturing uncertain samples and maintaining performance. The results suggest that a threshold of 0.2 may represent an optimal setting for high recall in this specific German Credit Scoring dataset scenario.

Figure 9 shows the overall trend in this chart highlights the strong recall performance of UDDA model across varying margin thresholds. The near-perfect recall at margin thresholds 0.1 and 0.2 underscores the model's ability to effectively identify all true positive cases, which is crucial in high-stakes applications such as fraud detection or medical diagnostics. The slight decrease in recall at margin thresholds 0.3 and 0.4 suggests a trade-off between the model's focus on uncertain cases and its confidence in classification, but the performance remains robust, with recall values still approaching 0.95.

UDDA approach, by combining both reinforcement learning's focus on hard-to-classify cases and active learning's selection of uncertain samples, proves to be highly effective in maximizing recall. This method ensures that the model learns optimally from both misclassified and uncertain instances, maintaining high sensitivity to positive cases while minimizing the risk of overlooking them.

The results suggest that margin thresholds of 0.1 and 0.2 yield the highest recall, making them ideal for scenarios where identifying all positive cases is critical. The UDDA framework offers a powerful solution for maintaining high recall across a variety of classification tasks, especially when combined with lower margin thresholds.

D. OVERALL SPECIFICITY

Figure 10 demonstrates a clear positive correlation between the margin threshold and specificity. As the margin threshold increases from 0.1 to 0.4, the model becomes progressively better at identifying true negatives and avoiding false positives, with specificity improving from 0.5 at margin 0.1 to

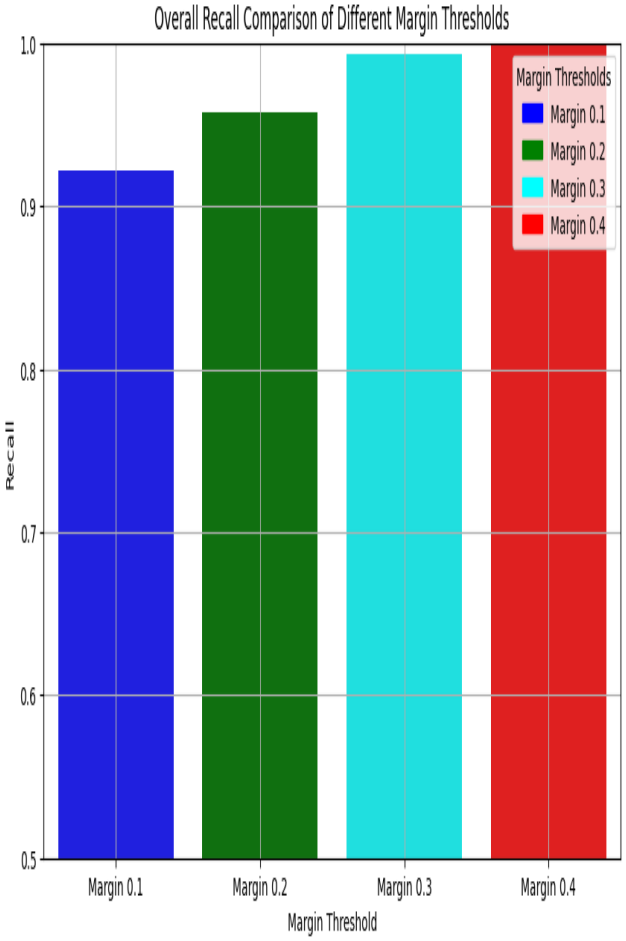

FIGURE 8. Overall recall comparison for different margin thresholds in AL in German credit scoring.

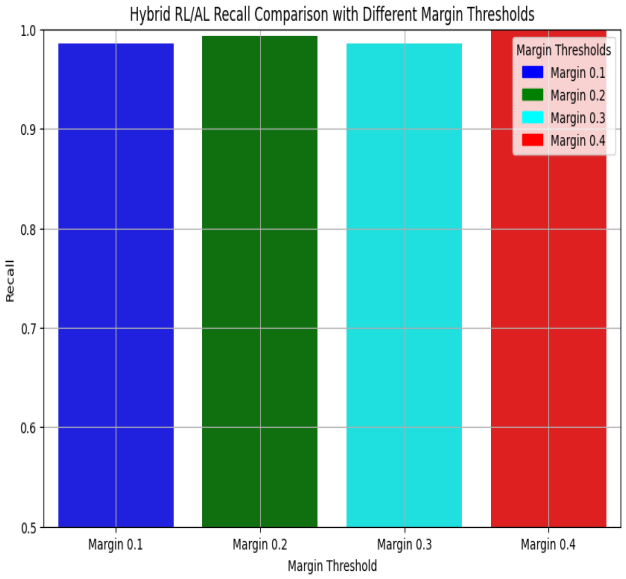

FIGURE 9. Overall recall comparison for different margin thresholds in RL/AL in German credit scoring.

0.85 at margin 0.4. This trend underscores the fact that as the Active Learning model selects less uncertain samples with greater confidence, it achieves a better balance between classifying positive and negative instances. Higher margin thresholds reduce the risk of misclassification, allowing the model to operate more reliably in distinguishing the two classes.

However, it is important to note that while specificity improves with higher margin thresholds, this may come at the cost of recall (true positive rate), as the model may become overly conservative and miss positive instances. Therefore, further analysis of the trade-off between specificity and recall is necessary to ensure optimal model performance across different applications.

Figure 10 suggests that a margin threshold of 0.4 is optimal for tasks where reducing false positives is critical, such as in fraud detection or medical diagnostics, where the cost of false positives can be high. The Active Learning strategy, when applied with higher margin thresholds, proves to be an effective approach to improving the model's specificity, thus enhancing its overall classification reliability.

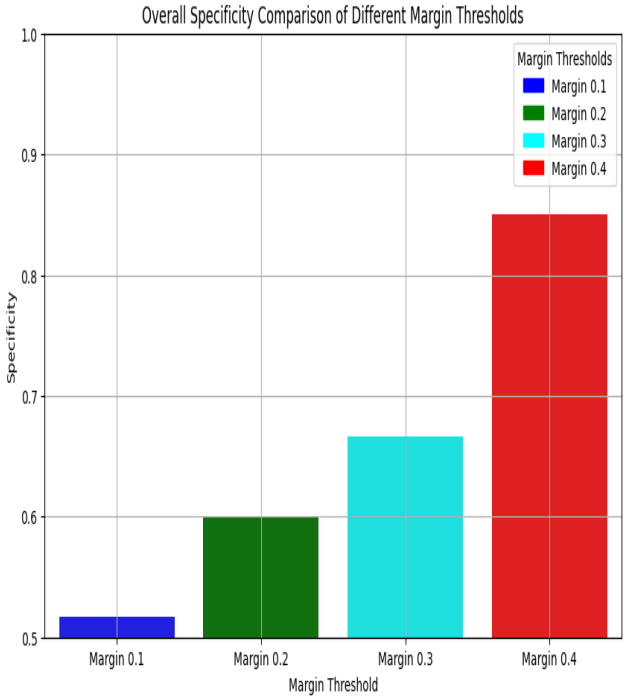

FIGURE 10. Overall recall comparison for different margin thresholds in RL/AL in German credit scoring.

Figure 11 shows remarkable consistency in specificity across all margin thresholds, with each achieving around 0.96 specificity. This consistency demonstrates that the UDDA model performs robustly in terms of negative classification, irrespective of the margin threshold applied.

The results suggest that the hybrid model is particularly adept at managing uncertain samples while still correctly identifying negative cases across different levels of uncertainty. Unlike models that might exhibit fluctuations in performance when margin thresholds are altered, this

hybrid approach ensures that the classification of non-target instances remains stable and effective, reducing the risk of false positives.

The high specificity values across all thresholds underscore the model's suitability for use in applications where the true negative rate is crucial—such as fraud detection or medical diagnosis—where false positives could result in unnecessary interventions or costs. The use of both Reinforcement Learning (RL) and Active Learning (AL) appears to provide a strong mechanism for addressing class imbalance, ensuring that the model does not sacrifice specificity when prioritizing difficult or uncertain cases.

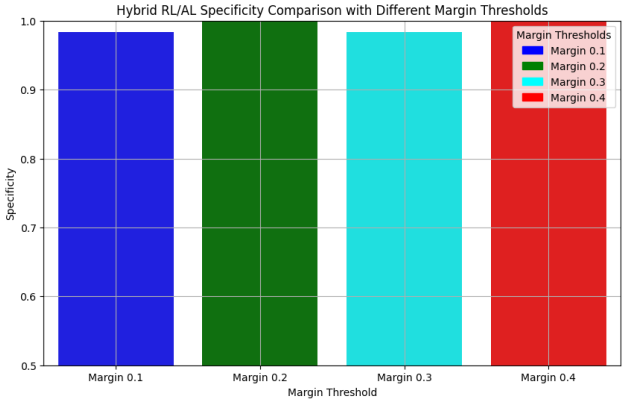

FIGURE 11. Overall specificity comparison for different margin thresholds in RL/AL in German credit scoring.

E. OVERALL NVP

Figure 12 shows a clear positive relationship between margin thresholds and NPV up to a threshold of 0.3. As the margin threshold increases, the model becomes more selective in its classification of uncertain samples, resulting in fewer false negatives and higher NPV values. This improvement is most pronounced between margin thresholds of 0.1 and 0.3, where the NPV rises sharply.

However, the slight decrease in NPV at the 0.4 threshold suggests that the model may become overly conservative at higher margins, potentially leaving out some true negatives in its quest to minimize false positives. This indicates that while margin thresholds above 0.3 can still deliver strong performance, a balance must be struck between conservatism and flexibility in uncertain cases.

The optimal margin threshold for NPV, as depicted in this chart, appears to be 0.3, where the model achieves its highest negative predictive accuracy. This threshold balances the AL model's ability to handle uncertain cases while maintaining a high level of negative classification accuracy, making it the most effective in reducing false negatives without sacrificing too much on sensitivity to borderline cases.

Figure 12 reveals that the hybrid RL/AL model consistently maintains high NPV values across all margin thresholds, with performance levels hovering around 0.90–0.93. This result implies that the hybrid RL/AL model is particularly effective in handling true negative classifications,

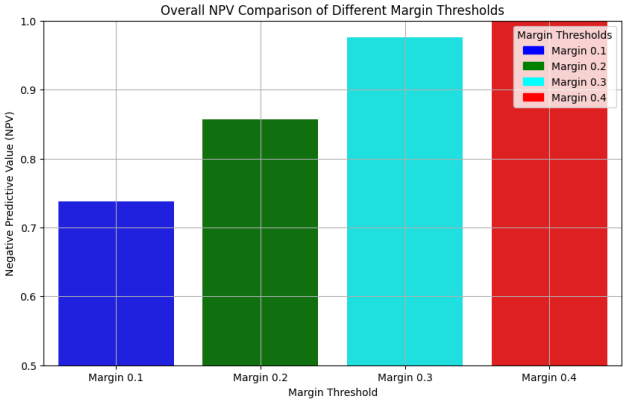

FIGURE 12. Overall NVP comparison for different margin thresholds in AL in German credit scoring.

minimizing the occurrence of false negatives regardless of the uncertainty threshold applied.

The minor fluctuations between thresholds indicate that the model's ability to correctly classify negative instances is robust to varying levels of uncertainty. While there is a slight dip in NPV at the 0.3 and 0.4 margins, these values are still very close to the highest observed NPVs, showcasing the resilience of the hybrid approach in minimizing prediction errors.

Interestingly, the relatively stable NPV values across different margin thresholds demonstrate that the hybrid RL/AL model performs well in contexts where it is important to accurately predict negative cases. This is crucial in imbalance-sensitive applications such as fraud detection or credit scoring, where false negatives can have significant financial or security implications.

F. OVERALL F-MEASURE

The F1-score is a crucial measure in imbalanced data scenarios, where the positive class (e.g., fraudulent transactions) is much rarer than the negative class (e.g., non-fraudulent transactions). Figure 14 reveals a general trend of improvement in F1-score as the margin threshold increases, reaching its highest point at 0.3. This indicates that the Active Learning model benefits from focusing on more uncertain cases while still incorporating confident predictions.

The slight drop at 0.4 suggests that there may be diminishing returns at higher margins, where the model's selectivity starts to negatively impact its ability to generalize, particularly in terms of recall. At this point, the model might exclude uncertain cases that could have been valuable for learning, reducing its ability to correctly classify all positive instances.

This finding highlights the importance of selecting an optimal margin threshold in Active Learning strategies. An overly conservative threshold may reduce false positives but at the cost of missing true positives, while a lower threshold might introduce noise from uncertain cases. The margin of 0.3 appears to offer the best trade-off, achieving the highest

F1-score, indicating an ideal balance between precision and recall for this dataset.

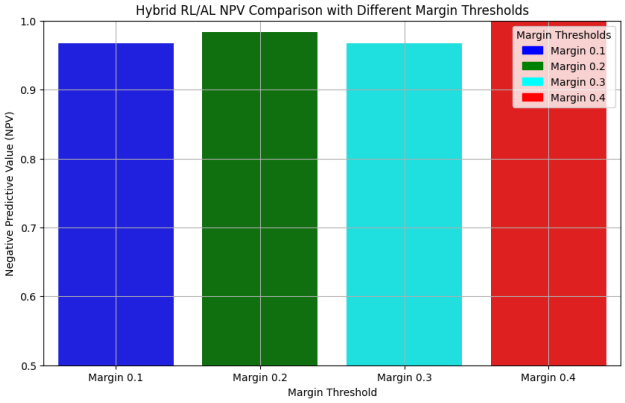

FIGURE 13. Overall NVP comparison for different margin thresholds in RL/AL in German Credit Scoring.

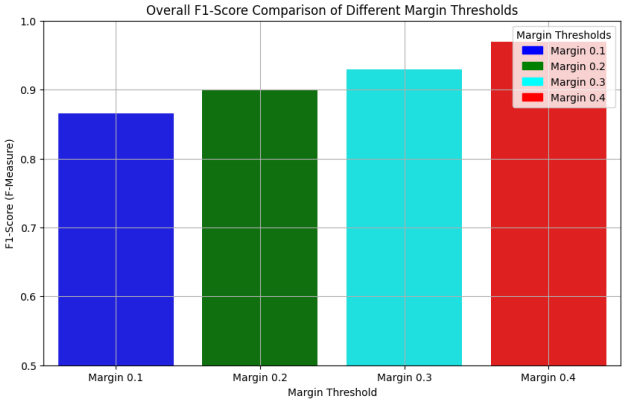

FIGURE 14. Overall F-measure comparison for different margin thresholds in AL in German credit scoring.

The results from the F1-score comparison indicate that the UDDA strategy is highly effective in maintaining consistent classification performance across different margin thresholds. Figure 15 shows the uniformity in F1-scores around 0.97 demonstrates that the combined approach is robust, regardless of the margin threshold used for selecting uncertain cases. This suggests that the hybrid approach can generalize well without being overly sensitive to the specific margin value.

The RL component likely contributes to the model's strong performance by focusing on difficult or misclassified cases, ensuring that the model learns effectively from these challenging samples. Meanwhile, the AL component ensures that the model continues to focus on uncertain samples, preventing overfitting to the easier, more certain cases. Together, these components create a balanced learning process that maintains high precision and recall, resulting in a consistently high F1-score.

The consistency of the F1-scores across all thresholds suggests that the UDDA approach is relatively insensitive to the margin value, providing flexibility in its application to various datasets. In real-world applications such as credit scoring or fraud detection, where class imbalance and the cost of misclassification are critical, this stability offers a significant advantage, as it ensures that the model will perform well even when the margin threshold needs to be adjusted for different use cases.

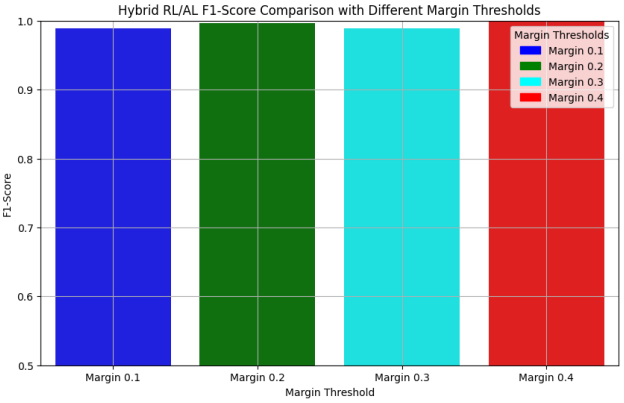

FIGURE 15. Overall F-measure comparison for different margin thresholds in RL/AL in German credit scoring.

G. OVERALL G-MEAN

Figure 16 shows the G-Mean comparison that shows a clear upward trend as the margin threshold increases from 0.1 to 0.4. This pattern reflects the growing effectiveness of the Active Learning approach in improving the balance between correctly identifying both the minority and majority classes. The relatively lower G-Mean at 0.1 indicates that a small margin threshold may lead to less informative sampling, potentially causing the model to misclassify more instances, especially from the minority class.

Conversely, as the margin threshold increases, the model becomes more selective in choosing the most uncertain but informative samples. The G-Mean continues to rise with the margin threshold, peaking at 0.4, which suggests that higher margin thresholds allow the model to achieve a more effective balance between sensitivity and specificity. This is particularly important in imbalanced datasets, where overfitting to the majority class can result in poor performance on the minority class. By using a higher threshold, the AL model can better handle imbalanced data, ultimately improving overall classification performance.

Figure 17 demonstrates the exceptional performance of the UDDA approach across various margin thresholds (0.1 to 0.4). With G-Mean values consistently near 1.0, the model effectively balances recall and specificity, ensuring accurate classification of both minority and majority classes in imbalanced datasets. This robustness, maintained across all thresholds, highlights the hybrid method's ability to address the limitations of pure active learning by leveraging reinforcement learning to improve sensitivity to difficult or misclassified cases. The high and stable G-Mean scores indicate the model's potential for real-world applications like fraud detection and medical diagnostics, where class imbalance is prevalent.

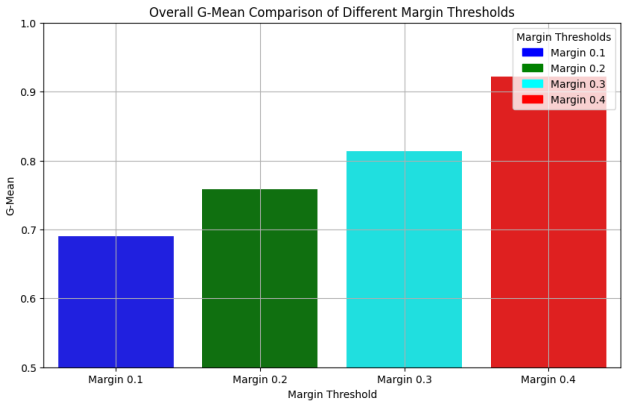

FIGURE 16. Overall G-mean comparison for different margin thresholds in AL in German credit scoring.

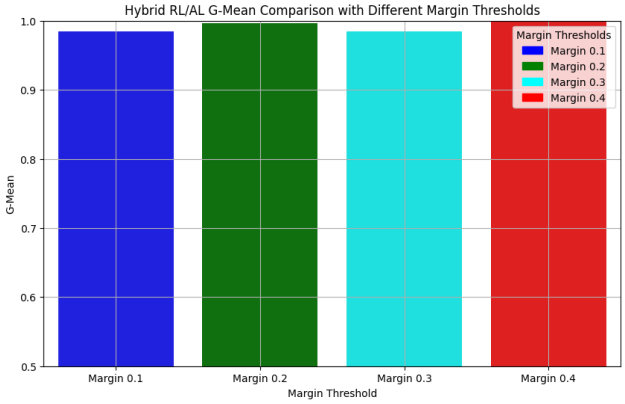

FIGURE 17. Overall G-mean comparison for different margin thresholds in RL/AL in German credit scoring.

H. OVERALL PERFORMANCE

Table 3, derived from the German Credit Scoring dataset, provides a comparative performance analysis across different scenarios: before applying SMOTE or RL/AL, after applying SMOTE, and following the implementation of hybrid RL/AL techniques.

For Class 0, hybrid RL/AL significantly improves precision (0.95) and recall (0.93), with the F1-score increasing to 0.94, outperforming both the baseline and SMOTE-only results. This indicates that RL/AL is highly effective in reducing misclassification of the minority class, an essential aspect in credit scoring models dealing with imbalanced datasets. SMOTE shows moderate improvement, but hybrid RL/AL demonstrates superior performance.

For Class 1, the precision, recall, and F1-scores improve to approximately 0.97–0.98 after RL/AL, demonstrating the model's enhanced ability to accurately classify the majority class compared to the baseline and SMOTE-enhanced models.

Overall, the accuracy rises sharply from 0.74 (before) to 0.96 (after RL/AL). The macro and weighted average precision, recall, and F1-scores also exhibit substantial gains, underscoring the effectiveness of hybrid RL/AL in balancing class performance. This analysis confirms that RL/AL provides a more comprehensive solution for handling class imbalance, particularly in the German Credit Scoring dataset.

TABLE 3. Performance compariosn of random forest with smote and RL/AL methods.

	Before	After SMOTE	After RL/AL
Class 0			
Precision	0.60	0.58	0.95
Recall	0.43	0,63	0.93
F1-score	0.50	0.60	0.94
Support	60	60	60
Class 1			
Precision	0.78	0.83	0.97
Recall	0.88	0.81	0.98
F1-score	0.83	0.82	0.98
Support	140	140	140
Overall			
Accuracy	0.74	0.70	0.96
Macro average Precision	0.69	0.71	0.96
Macro average recall	0.66	0.71	0.96
Macro average f1-score	0.67	0.76	0.96
Weighted average precision	0.73	0.75	0.96
Weighted average recall	0.74	0.75	0.96
Weighted average f1-score	0.73	0.75	0.96

The analysis of the results presented in Table 4, derived from the Kaggle Credit Card Fraud Detection dataset, compares the model's performance in three scenarios: before applying SMOTE or RL/AL, after SMOTE, and following hybrid RL/AL.

For Class 0, representing non-fraudulent transactions, the model's performance remains consistent across all scenarios, with precision, recall, and F1-score holding at 1. This indicates that Class 0 is well-handled regardless of the strategy employed.

However, for Class 1, representing fraudulent transactions, there is a noticeable difference. Before applying SMOTE or RL/AL, the precision is 0.94, recall is 0.82, and the F1-score is 0.87. After applying SMOTE, the precision drops to 0.84, though the recall slightly improves to 0.83, leading to a marginally reduced F1-score of 0.83. This shows that SMOTE improves recall at the cost of precision. Notably, hybrid RL/AL shows substantial improvement for Class 1,

with precision reaching 1, recall increasing to 0.97, and the F1-score rising to 0.98. This demonstrates that hybrid RL/AL effectively enhances the model's ability to identify fraudulent transactions with minimal false positives.

In terms of overall performance, the accuracy remains perfect (1) in all scenarios, which can be attributed to the class imbalance (Class 0 has much higher support). However, macro average precision, recall, and F1-scores are more telling. Before applying any techniques, the macro averages show a modest balance between classes. After SMOTE, there is a decline in precision (from 0.97 to 0.92) with no recall improvement. The hybrid RL/AL approach, in contrast, yields a significant increase, improving the macro averages to 1 for precision and 0.98 for recall. The weighted averages, however, remain consistently high due to the large proportion of Class 0, thus diluting the impact of improvements in Class 1.

While SMOTE marginally improves recall for fraudulent transactions, the hybrid RL/AL strategy offers the most balanced and enhanced results, particularly in addressing class imbalance, making it the most effective solution in the context of this Kaggle dataset.

The ROC curve analysis presented in Figure 19, derived from the German Credit Scoring dataset, compares the performance of Random Forest classifiers under three distinct scenarios: before applying SMOTE or RL/AL, after applying SMOTE, and after applying hybrid RL/AL. Initially, the performance before SMOTE or RL/AL shows an Area Under the Curve (AUC) of 0.79, indicating a moderate ability of the model to distinguish between classes. However, after applying SMOTE, there is a slight decline in the model's performance, with the AUC dropping to 0.77. This suggests that while SMOTE helps to address class imbalance, it can potentially introduce noise or reduce the model's precision in certain cases, particularly for imbalanced datasets like this one. In stark contrast, after the application of hybrid RL/AL, the AUC rises significantly to 0.99, nearly perfect classification performance. The hybrid RL/AL approach demonstrates its superiority in not only mitigating class imbalance but also enhancing the model's capability to accurately classify both majority and minority classes. This result underscores the effectiveness of RL/AL in improving model performance, particularly in complex imbalanced datasets like the German Credit Scoring dataset, where precision and recall for minority classes are of critical importance. This analysis suggests that while traditional techniques like SMOTE may offer incremental improvements, hybrid RL/AL is a more effective strategy for achieving robust classification performance in credit scoring models.

The ROC curve in Figure 20 presents the performance of a Random Forest classifier evaluated across three different scenarios using data from the Kaggle credit card dataset. The analysis considers the classifier's efficacy before applying any data balancing technique (SMOTE or RL/AL), after applying SMOTE, and finally after applying hybrid RL/AL. In the initial scenario, both the application of SMOTE and the baseline model (before applying SMOTE or RL/AL) yield an identical AUC of 0.96. This indicates that, while

TABLE 4. Comparison of random forest with smote and RL/AL on Kaggle dataset.

	Before	After SMOTE	After RL/AL
Class 0			
Precision	1	1	1
Recall	1	1	1
F1-score	1	1	1
Support	56864	56864	56864
Class 1			
Precision	0.94	0.84	1
Recall	0.82	0.83	0.97
F1-score	0.87	0.83	0.98
Support	98	98	98
Overall			
Accuracy	1	1	1
Macro average Precision	0.97	0.92	1
Macro average recall	0.91	0.91	0.98
Macro average f1-score	0.94	0.92	0.99
Weighted average precision	1	1	1
Weighted average recall	1	1	1
Weighted average F1-score	1	1	1

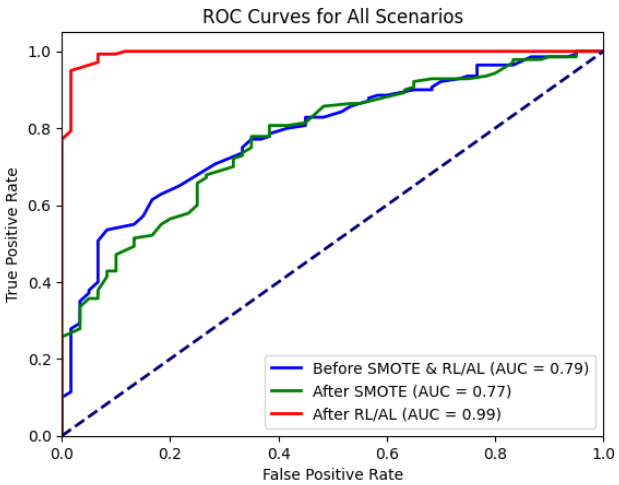

FIGURE 18. Comparison of ROC curves for random forest in different scenarios using German credit scoring dataset.

SMOTE does help mitigate class imbalance by adjusting the dataset, it does not significantly enhance the model's ability to differentiate between true positives and false positives beyond what is achievable by the Random Forest model alone. This consistency in AUC demonstrates that SMOTE, while beneficial for addressing imbalance, might not add substantial value in scenarios where the initial model is already relatively performant.

In contrast, the model's performance after the application of hybrid RL/AL demonstrates a marked improvement, achieving an AUC of 1.00. This perfect AUC score indicates that the hybrid RL/AL approach allows the model to make nearly flawless classifications, distinguishing between positive and negative cases without error. This dramatic improvement underscores the effectiveness of reinforcement learning and active learning in refining the model's predictions, particularly in scenarios where standard oversampling methods like SMOTE might fall short. The enhancement suggests that RL/AL introduces a level of precision and adaptability that is particularly suited to handling more challenging and imbalanced datasets, as is often the case in financial applications such as credit scoring.

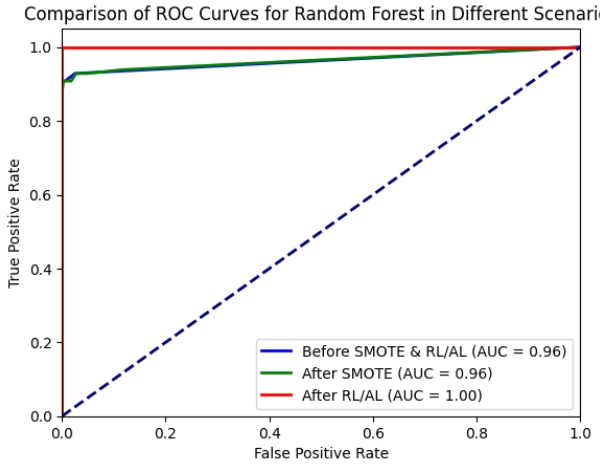

FIGURE 19. Comparison of ROC curves for random forest in different scenarios using the Kaggle dataset.

Table 5 provides a comparative analysis of the AUC values for Random Forest (RF) models across three distinct configurations: default RF, RF with SMOTE, and RF with RL/AL, applied to both the German Credit Scoring and Kaggle datasets. The results illustrate how advanced techniques, particularly RL/AL, significantly enhance the classifier's ability to distinguish between classes.

In the case of the German Credit Scoring dataset, the baseline Random Forest achieved an AUC of 0.79. However, after applying SMOTE, the AUC slightly decreased to 0.77, which may suggest that oversampling did not optimally address the dataset's inherent complexities or possibly introduced noise. In contrast, the hybrid RL/AL approach led to a notable improvement, achieving an AUC of 0.99, signifying a substantial gain in model performance through the use of active and reinforcement learning strategies.

On the Kaggle dataset, the baseline Random Forest performed at a high level, with an AUC of 0.96. Interestingly, applying SMOTE did not alter the AUC, indicating that the initial model was already capable of handling class imbalance effectively. However, with the implementation of RL/AL, the model reached an AUC of 1.00, showcasing the remarkable impact of RL/AL on perfecting the classification performance in this context.

This analysis highlights that while SMOTE can be a useful tool for addressing class imbalance, the adoption of more sophisticated methods like RL/AL can dramatically improve model efficacy, especially in datasets that present significant classification challenges.

TABLE 5. AUC comparison between default RF, RF with smote and RF with UDDA.

Dataset	AUC RF	AUC RF with SMOTE	AUC RF with UDDA
German Credit Scoring	0.79	0.77	0.99
Kaggle	0.96	0.96	1

I. STATISTICAL SIGNIFICANCE OF EMPIRICAL RESULTS

In this section, we rigorously evaluate the statistical significance of the obtained experimental outcomes by applying a suite of non-parametric statistical methodologies [29]. Specifically, the analysis leverages the Friedman test [30], multiple comparisons with the best (MCB) approach [31], and the Wilcoxon signed-rank test [32] to examine whether the observed improvements in performance metrics—achieved by the UDDA algorithm—are not due to random variation but represent consistent and meaningful gains over baseline models.

To ensure methodological robustness, each test is selected based on its suitability for comparing multiple classifiers across several datasets without assuming normality in the performance distributions. The Friedman test serves as a global hypothesis test to detect overall performance differences among competing methods. Upon identifying significant differences, the MCB test is employed to identify which methods are statistically indistinguishable from the top performer. Furthermore, the Wilcoxon signed-rank test is applied as a pairwise post-hoc comparison to validate whether the improvements of the UDDA model over individual baselines are statistically significant.

All test results were computed using the performance data summarized in the Supplementary section. This table contains the F1 and AUC scores of each sampling method across all datasets evaluated.

Table 6 provides a concise overview of the sampling methods evaluated. The list includes over-sampling (OS) techniques such as SMOTE [24], ADASYN [25], SVMSMOTE [26], and Borderline-SMOTE [27], alongside random under-sampling (US) and hybrid methods like SMOTE-ENN [28] and SMOTE-Tomek. These algorithms were selected to ensure a diverse and representative

TABLE 6. List of algorithms used for empirical evaluation, where OS and US denote the over-sampling and under-sampling solutions for handling imbalance data.

Model	Reference	Type
SMOTE	[26]	OS
ADASYN	[27]	OS
SVMSMOTE	[28]	OS
Borderline-SMOTE	[29]	OS
Random Oversampling	Built-in	OS
Random Under-sampling	Built-in	US
SMOTE-ENN	[30]	Hybrid
SMOTE-Tomek	Built-in	Hybrid

comparison framework, enabling a robust empirical evaluation of UDDA against established techniques in handling imbalanced data.

1) FRIEDMAN TEST

In our study, we utilize the non-parametric Friedman test [30] to evaluate the statistical significance of differences in model performance across multiple datasets. The Friedman test is a widely adopted non-parametric statistical method for comparing multiple algorithms over several datasets, particularly suitable when the assumption of normality in performance distributions does not hold. This procedure ranks models based on their relative performance across classification tasks and computes the average rank for each model [29].

The Friedman test was applied to both F1 Score and AUC metrics using the dataset compiled that is provided as part of this article's Supplementary Materials. The test results indicate statistically significant differences in model performance for both metrics. Specifically, for the F1 Score, the computed chi-square statistic is 34.0684, with a p-value of 3.9484×10^{-5}. For the AUC, the chi-square value is 40.5307, with a p-value of 2.5510×10^{-6}.

Both p-values fall well below the conventional significance threshold of $\alpha = 0.05$, leading to the rejection of the null hypothesis in each case. These findings confirm that there are substantial differences in performance among the evaluated sampling methods. The magnitude of the test statistics for both F1 and AUC further indicates that at least one algorithm significantly outperforms others in terms of average rank. This reinforces the reliability of the Friedman test as a robust statistical tool capable of uncovering meaningful performance distinctions among competing classifiers. The statistically significant outcomes demonstrate that the observed improvements—particularly those attributed to the proposed UDDA algorithm—are not due to random variation but reflect genuine differences in effectiveness.

2) MULTI COMPARISON WITH THE BEST

Figure 20 presents the results of the Multiple Comparison with the Best (MCB) analysis applied to the F1 score rankings across all evaluated datasets. The MCB test provides a statistical means of identifying which models are not significantly different from the best-performing method, based on mean rank and confidence intervals.

In this plot, SMOTE achieves the lowest average rank (2.60), followed closely by UDDA, which attains a rank of 4.08. While SMOTE remains the top-ranked model, UDDA falls within the critical difference (CD) zone, as indicated by its red marker and overlapping confidence interval. This denotes that UDDA is statistically indistinguishable from SMOTE in terms of F1 performance at the 95% confidence level.

The presence of other red-marked methods such as ADASYN (3.83) further supports the validity of UDDA's performance. In contrast, methods positioned outside the CD zone—such as SMOTE-Tomek, SVMSMOTE, and Random Oversampling—are marked in black and exhibit significantly inferior ranks.

These findings reinforce the conclusion that UDDA is a top-tier sampling strategy, delivering competitive and statistically comparable performance to the strongest baseline methods. This result is particularly notable given that UDDA consistently outperformed several established baselines while maintaining statistical equivalence to the best-performing algorithm in F1 score.

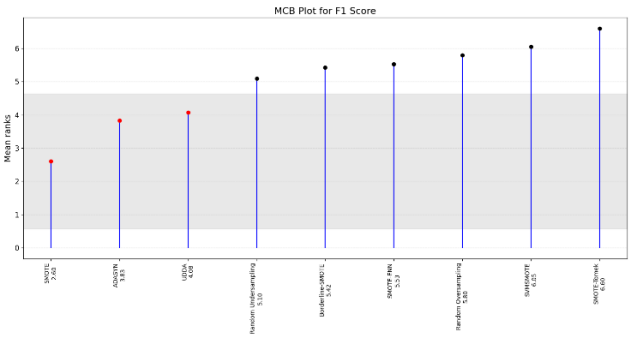

FIGURE 20. Visualization of MCB analysis based on F1 measures. UDDA – 4.08 denotes the average rank across all the datasets for the proposed UDDA model based on the F1 measures (similar interpretation holds for others).

Figure 21 extends the MCB analysis by examining the Area Under the Curve (AUC) rankings across all datasets. Similar to the F1 evaluation, this plot depicts the mean rank of each algorithm along with their respective critical difference (CD) intervals. The goal is to identify which methods are statistically indistinguishable from the top performer.

In this case, SMOTE once again emerges as the leading model, achieving the lowest average rank of 2.33. The proposed UDDA algorithm records an average rank of 4.20, falling well within the shaded CD region. This indicates that UDDA's performance is not statistically different from SMOTE's in terms of AUC. Other models such as ADASYN (4.08) and SMOTE-ENN (4.30) also reside within this zone, supporting the conclusion that these methods belong to the top-performing group.

In contrast, models such as SMOTE-Tomek (5.53), SVMSMOTE (5.60), and Random Undersampling (6.50) lie outside the CD zone, signifying a statistically significant performance gap when compared to the top performers. These

differences further validate the results of the Friedman and Wilcoxon tests by providing a visual and rank-based confirmation of performance disparities.

Collectively, Figures 20 and 21 demonstrate that UDDA maintains competitive performance across both F1 and AUC metrics. The consistency of UDDA's statistical proximity to the best method across multiple evaluations underscores its robustness and suitability as a practical solution for handling class imbalance in real-world classification tasks.

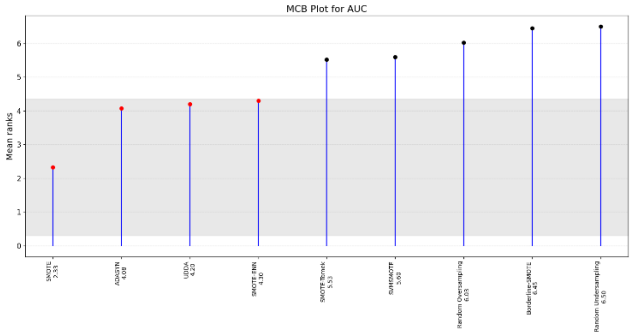

FIGURE 21. Visualization of MCB analysis based on AUC. UDDA – 4.20 denotes the average rank across all the datasets for the proposed UDDA model based on the AUC (similar interpretation holds for others).

3) WILCOXON SIGNED-RANK TEST

Table 7 presents the outcomes of the Wilcoxon signed-rank test, applied to compare the proposed UDDA method against each baseline sampling algorithm. The test is performed for both F1 and AUC metrics across all datasets and includes both the resulting p-values and rank differences. Significance is determined at a confidence level of α =0.05, with p-values below this threshold marked by an asterisk (*).

The results reveal that UDDA significantly outperforms several baseline models in terms of both F1 and AUC scores. Notably, methods such as SMOTE-Tomek, Random Oversampling, SVMSMOTE, and Borderline-SMOTE exhibit p-values < 0.05 across both metrics, confirming statistically significant performance improvements by UDDA. For instance, UDDA achieves an F1 rank difference of 0.1 over SMOTE-Tomek ($p = 0.0005$) and 0.08 over Random Oversampling ($p = 0.0010$), reinforcing its superiority in classification precision and balance.

Furthermore, Random Undersampling and SMOTE-ENN also show statistically significant differences in F1 score. However, in AUC, only Random Undersampling ($p = 0.0002$) meets the threshold, while SMOTE-ENN ($p = 0.3921$) does not, suggesting inconsistent gains in discrimination capability for the latter.

On the other hand, ADASYN and SMOTE do not exhibit statistically significant differences from UDDA in either F1 or AUC ($p > 0.85$ in both cases). Their small or negative rank differences further imply that their performance is comparable to UDDA. Specifically, SMOTE records rank differences of -0.03 (F1) and -0.05 (AUC), indicating it remains a competitive baseline under these conditions.

These findings align with the MCB visualizations and Friedman test outcomes, establishing that UDDA is statistically superior to most baseline methods, while maintaining performance on par with top competitors like SMOTE and ADASYN. This reinforces UDDA's role as a robust, generalizable solution for handling imbalanced data classification problems.

TABLE 7. The table presents the p-values for the Wilcoxon signed-rank test, along with the rank difference between the corresponding models and UDDA. Results are shown for both AUC and F1 metrics. Significance at the significance level (alpha = 0.05) is marked with "*".

Model Name	F1		AUC	
	p-value	Rank Diff	p-value	Rank Diff
SMOTE-Tomek	0.0005*	0.1	0.0310*	0.05
Random Oversampling	0.0010*	0.08	0.0107*	0.06
SVMSMOTE	0.0018*	0.08	0.0261*	0.05
Borderline-SMOTE	0.0018*	0.07	0.0012*	0.08
SMOTE-ENN	0.0021*	0.07	0.3921	0.02
Random Undersampling	0.0250*	0.05	0.0002*	0.07
ADASYN	0.8564	-0.01	0.9394	-0.03
SMOTE	0.9981	-0.03	0.9997	-0.05

VI. CONCLUSION

This study introduces UDDA, a hybrid Reinforcement Learning and Active Learning framework that addresses class imbalance using real, informative samples instead of synthetic or removal-based methods. Applied to credit scoring and fraud detection, UDDA significantly outperforms traditional approaches like SMOTE in the first experiment setup. Its ability to improve precision, recall, and F1-score while preserving data integrity highlights its practical value. Empirical results across 20 benchmark datasets demonstrate consistent gains in precision, recall, F1-score, and AUC. Furthermore, statistical significance tests, including the Friedman test, multiple comparisons with the best (MCB), and the Wilcoxon signed-rank test—confirm that UDDA's improvements are not due to chance but represent robust and meaningful performance gains. UDDA sets a new standard for real-data-driven augmentation in imbalanced learning. Future research should extend this framework to multi-class problems and explore its adaptability in streaming and real-time applications.

REFERENCES

[1] Z. Ren, T. Lin, K. Feng, Y. Zhu, Z. Liu, and K. Yan, "A systematic review on imbalanced learning methods in intelligent fault diagnosis," *IEEE Trans. Instrum. Meas.*, vol. 72, pp. 1–35, 2023.

[2] S. Fan, X. Zhang, and Z. Song, "Imbalanced sample selection with deep reinforcement learning for fault diagnosis," *IEEE Trans. Ind. Informat.*, vol. 18, no. 4, pp. 2518–2527, Apr. 2022.

[3] J. Kuang, G. Xu, T. Tao, and Q. Wu, "Class-imbalance adversarial transfer learning network for cross-domain fault diagnosis with imbalanced data," *IEEE Trans. Instrum. Meas.*, vol. 71, pp. 1–11, 2022.

[4] L.-C. Hung, Y.-H. Hu, C.-F. Tsai, and M.-W. Huang, "A dynamic time warping approach for handling class imbalanced medical datasets with missing values: A case study of protein localization site prediction," *Expert Syst. Appl.*, vol. 192, Apr. 2022, Art. no. 116437.

[5] S. Fotouhi, S. Asadi, and M. W. Kattan, "A comprehensive data level analysis for cancer diagnosis on imbalanced data," *J. Biomed. Informat.*, vol. 90, Feb. 2019, Art. no. 103089.

[6] F. Behrad and M. S. Abadeh, "An overview of deep learning methods for multimodal medical data mining," *Expert Syst. Appl.*, vol. 200, Aug. 2022, Art. no. 117006.

[7] K. Yang, Y. Shi, Z. Yu, Q. Yang, A. K. Sangaiah, and H. Zeng, "Stacked one-class broad learning system for intrusion detection in industry 4.0," *IEEE Trans. Ind. Informat.*, vol. 19, no. 1, pp. 251–260, Jan. 2023.

[8] G. Haixiang, L. Yijing, J. Shang, G. Mingyun, H. Yuanyue, and G. Bing, "Learning from class-imbalanced data: Review of methods and applications," *Expert Syst. Appl.*, vol. 73, pp. 220–239, May 2017.

[9] X. Chen, Y. Zhou, D. Wu, W. Zhang, Y. Zhou, B. Li, and W. Wang, "Imagine by reasoning: A reasoning-based implicit semantic data augmentation for long-tailed classification," in *Proc. AAAI Conf. Artif. Intell.*, vol. 36, no. 1, pp. 356–364.

[10] Y. Sun, L. Cai, B. Liao, W. Zhu, and J. Xu, "A robust oversampling approach for class imbalance problem with small disjuncts," *IEEE Trans. Knowl. Data Eng.*, vol. 35, no. 6, pp. 5550–5562, Jun. 2023.

[11] T. Zhang, F. Ma, D. Yue, C. Peng, and G. M. P. O'Hare, "Interval type-2 fuzzy local enhancement based rough K-means clustering considering imbalanced clusters," *IEEE Trans. Fuzzy Syst.*, vol. 28, no. 9, pp. 1925–1939, Sep. 2020.

[12] X. Wu and S. Meng, "E-commerce customer churn prediction based on improved SMOTE and AdaBoost," in *Proc. 13th Int. Conf. Service Syst. Service Manage. (ICSSSM)*, Jun. 2016, pp. 1–5.

[13] Q. Li and Y. Xie, "A behavior-cluster based imbalanced classification method for credit card fraud detection," in *Proc. 2nd Int. Conf. Data Sci. Inf. Technol.*, Jul. 2019, pp. 134–139.

[14] H. Zhu, M. Zhou, G. Liu, Y. Xie, S. Liu, and C. Guo, "NUS: Noisy-sample-removed undersampling scheme for imbalanced classification and application to credit card fraud detection," *IEEE Trans. Computat. Social Syst.*, vol. 11, no. 2, pp. 1793–1804, Apr. 2024.

[15] R. Qaddoura and M. M. Biltawi, "Improving fraud detection in an imbalanced class distribution using different oversampling techniques," in *Proc. Int. Eng. Conf. Electr., Energy, Artif. Intell. (EICEEAI)*, Nov. 2022, pp. 1–5.

[16] K. P. Mahesh, S. A. Afrouz, and A. S. Areeckal, "Detection of fraudulent credit card transactions: A comparative analysis of data sampling and classification techniques," *J. Phys., Conf. Ser.*, vol. 2161, no. 1, Jan. 2022, Art. no. 012072.

[17] N. Rtayli, "An efficient deep learning classification model for predicting credit card fraud on skewed data," *J. Inf. Secur. Cybercrimes Res.*, vol. 5, no. 1, pp. 57–71, Jun. 2022.

[18] S. O. Akinwamide, "Prediction of fraudulent or genuine transactions on credit card fraud detection dataset using machine learning techniques," *Int. J. Res. Appl. Sci. Eng. Technol.*, vol. 10, no. 6, pp. 5061–5071, Jun. 2022.

[19] E. Esenogho, I. D. Mienye, T. G. Swart, K. Aruleba, and G. Obaido, "A neural network ensemble with feature engineering for improved credit card fraud detection," *IEEE Access*, vol. 10, pp. 16400–16407, 2022.

[20] X. Yi, Y. Xu, Q. Hu, S. Krishnamoorthy, W. Li, and Z. Tang, "ASN-SMOTE: A synthetic minority oversampling method with adaptive qualified synthesizer selection," *Complex Intell. Syst.*, vol. 8, no. 3, pp. 2247–2272, Jun. 2022.

[21] M. Alamri and M. Ykhlef, "Hybrid undersampling and oversampling for handling imbalanced credit card data," *IEEE Access*, vol. 12, pp. 14050–14060, 2024.

[22] E. Lopez-Rojas, A. Elmir, and S. Axelsson, "PaySim: A financial mobile money simulator for fraud detection," in *Proc. 28th Eur. Model. Simul. Symp.*, Jun. 2016, pp. 249–255.

[23] H. Shamsudin, U. K. Yusof, A. Jayalakshmi, and M. N. A. Khalid, "Combining oversampling and undersampling techniques for imbalanced classification: A comparative study using credit card fraudulent transaction dataset," in *Proc. IEEE 16th Int. Conf. Control Autom. (ICCA)*, Oct. 2020, pp. 803–808.

[24] 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, Jun. 2002.

[25] H. He and E. A. Garcia, "Learning from imbalanced data," *IEEE Trans. Knowl. Data Eng.*, vol. 21, no. 9, pp. 1263–1284, Sep. 2009, doi: 10.1109/TKDE.2008.239.

[26] H. M. Nguyen, E. W. Cooper, and K. Kamei, "Borderline over-sampling for imbalanced data classification," *Int. J. Knowl. Eng. Soft Data Paradigms*, vol. 3, no. 1, pp. 4–21, 2011.

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

[28] 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. Newslett.*, vol. 6, no. 1, pp. 20–29, Jun. 2004.

[29] M. Panja, T. Chakraborty, U. Kumar, and N. Liu, "Epicasting: An ensemble wavelet neural network for forecasting epidemics," *Neural Netw.*, vol. 165, pp. 185–212, Aug. 2023.

[30] M. Friedman, "The use of ranks to avoid the assumption of normality implicit in the analysis of variance," *J. Amer. Stat. Assoc.*, vol. 32, no. 200, pp. 675–701, Dec. 1937.

[31] A. J. Koning, P. H. Franses, M. Hibon, and H. O. Stekler, "The M3 competition: Statistical tests of the results," *Int. J. Forecasting*, vol. 21, no. 3, pp. 397–409, Jul. 2005.

[32] R. F. Woolson, "Wilcoxon signed-rank test," in *Encyclopedia of Biostatistics*. Berlin, Germany: Springer, 2005.

[33] P. Sadhukhan and S. Gupta, "A graph theoretic approach to assess quality of data for classification task," *Data Knowl. Eng.*, vol. 158, Jul. 2025, Art. no. 102421.

AZREE NAZRI is currently a Leading Expert in artificial intelligence and digital twin technologies, known for advancing ethical AI practices in Malaysia. His research focuses on interactive machine learning, deep reinforcement learning, and real-time autonomous systems, with impactful applications in finance, manufacturing, logistics, and smart infrastructure environments.

OLALEKAN AGBOLADE is currently a dedicated Ph.D. Student with the Department of Computer Science, Faculty of Computer Science and Information Technology, Universiti Putra Malaysia (UPM). His research focuses on advanced topics in computer science, with a particular interest in artificial intelligence, machine learning, and data analytics. Throughout his academic journey, he has been committed to exploring innovative approaches to solving complex problems, aiming to contribute meaningful advancements to the field of computer science.

FAISAL AZIZ, photograph and biography not available at the time of publication.