

Supervised Machine Learning Algorithms for Predicting Academic Specialization: A Comparative Study

Samira Reguragui ^{1,*}, Ahmed Bendahmane ¹, and Abdellatif El Aidi ²

¹ Higher Normal School, Abdelmalek Essaadi University, Tetuan, Morocco

² Faculty of Letters and Human Sciences Sais-Fez, Sidi Mohamed Ben Abdellah University, Fez, Morocco

Email: samira.reguragui@etu.uae.ac.ma (S.R.); abendahman@uae.ac.ma (A.B.); abdellatif.elaidi@usmba.ac.ma (A.E.A.)

*Corresponding author

Manuscript received March 6, 2026; revised May 7, 2026; accepted July 3, 2026; published October 26, 2026

Abstract—The choice of a paramedical specialization suited to students' academic profiles and preferences is a challenge for Moroccan students. To respond to this challenge, this study evaluated seven tree-based algorithms on real data from 889 nursing students from two Moroccan training institutes to determine the best-performing model to predict the students' paramedical specialization. We implemented Bayesian hyperparameter optimization with Optuna, stratified 5-fold cross-validation, and final evaluation on an independent test set to ensure a robust and reliable evaluation of the models. The resampling techniques (SMOTE, SMOTEENN, ADASYN, SMOTETomek) were applied only in the training folds to avoid data leakage. Ensemble models (Random Forest, Extra Trees, XGBoost, Gradient Boosting, HistGradient Boosting) surpassed the baseline model (Decision Tree), while AdaBoost's performance remained weak and unstable in cross-validation and on the test set (test macro F1-score as low as 0.1384), confirming its strong sensitivity to resampling-induced noise. On the test set, Random Forest with SMOTEENN achieved the best overall performance (macro F1-score = 0.8836, macro Receiver Operating Characteristic curve and Area Under the Curve (ROC-AUC) = 0.9879, macro Precision-Recall Area Under the Curve (PR-AUC) = 0.9294), whereas XGBoost with SMOTE delivered near-identical predictive performance (macro F1-score = 0.8828) with a faster inference time (0.0060 s), offering the best deployment trade-off. By integrating these two optimal approaches, an orientation system can accurately predict the specialization best suited to a student's profile and preferences.

Keywords—tree-based algorithms, SMOTE, SMOTEENN, ADASYN, SMOTETomek, class imbalance, academic guidance systems, classification, resampling, machine learning algorithms, stratified K-fold cross-validation

I. INTRODUCTION

Academic guidance is a crucial stage that can strongly shape students' academic trajectories and career prospects [1]. With the growing availability of large-scale educational data across higher-education institutions, Machine Learning (ML) techniques offer concrete opportunities to support orientation by identifying patterns that link students' profiles to successful specialization choices [2]. Automated prediction of a student's study specialization can therefore help personalize orientation processes and enable more informed decision-making, especially in contexts where guidance resources are limited and student cohorts are large.

In this context, several orientation systems have been developed, and different algorithms have been tested to determine which predictive models are most useful for educational guidance tasks. Zahour *et al.* [3], for example, evaluated four algorithms for the automatic classification of

academic and career-guidance questions. Majjate *et al.* [4] designed a pedagogical advising system intended to support students in selecting the institution most aligned with their profiles and interests. Their comparison of 13 machine-learning algorithms showed that the Huber Regressor produced the most accurate results in that specific setting. Badrani *et al.* [5] also proposed a classification model for orienting middle-school students and reported that the Decision Tree performed better than Naïve Bayes and K-Nearest Neighbors (KNN). In a later comparative study [6], the same authors extended their analysis by including Random Forest (RF), showing that this ensemble method achieved stronger performance than the simple Decision Tree models.

Recent research has moved toward more advanced classification strategies for personalized academic guidance, particularly hybrid learning techniques and ensemble methods. Badrani *et al.* [7] reported that a hybrid model combining neural networks and Support Vector Machine (SVM) outperformed individual algorithms. Alsayed *et al.* [8] compared tree-based models with SVM and found that ensemble methods such as Random Forest (RF) and Gradient Boosting (GBC) reached high accuracy in predicting students' specialization. In general, ensemble models strengthen predictive robustness by combining several base algorithms, which helps reduce variance and improve stability in heterogeneous datasets [9–11]. However, although these methods are often presented as effective responses to modeling issues such as overfitting and underfitting [12], their performance in the orientation context still requires deeper investigation. This study addresses this gap by rigorously comparing several tree-based ensemble algorithms for predicting academic specialization.

Data imbalance remains a serious problem for prediction systems because ML algorithms often favor the majority class, which can weaken predictive performance for minority classes that may still be academically and institutionally important [13–15]. In such cases, overall accuracy may give a misleading impression of model quality, since strong results for dominant classes can hide poor results for minority classes [16, 17]. To address this issue, researchers increasingly recommend the use of strategies for handling class imbalance so that models can achieve stronger robustness, better generalization, and meaningful predictive performance across all classes [15, 18]. The usefulness of data-balancing methods has already been examined in several fields, including academic performance prediction [19–21], disease prediction [12, 22–24], and fraud detection [25–28]. Innuddin *et al.* [22], for instance, assessed SMOTE,

Borderline-SMOTE, and SVM-SMOTE for stroke classification on imbalanced data and found that these oversampling techniques improved the performance of RF and XGBoost classifiers. Another study [23] showed that applying K-Means SMOTE-ENN with Principal Component Analysis (PCA) improved RF model performance on imbalanced health data. However, data-balancing techniques remain insufficiently explored in academic and career guidance. The aforementioned studies on ML-based academic guidance systems, together with related works [29–33], do not directly address the class-imbalance problem.

This article responds to these identified gaps by presenting an in-depth comparative analysis of tree-based machine-learning algorithms while incorporating different resampling methods to handle class imbalance in the prediction of Moroccan students' paramedical specialization. More specifically, the study seeks to identify the most effective tree-based algorithm for imbalanced educational datasets in order to generate accurate and relevant recommendations for student guidance. For this purpose, seven ML models were selected because of their robustness and their established effectiveness in the literature, namely Random Forest (RF), Extra Trees (ETC), Decision Tree (DT), AdaBoost, Gradient Boosting (GBC), HistGradient Boosting (HTGB), and XGBoost. To address class imbalance, four resampling strategies are incorporated into the learning pipeline, namely SMOTE, ADASYN, SMOTEENN, and SMOTETomek, to improve the representativeness of minority specialties. The methodological contribution of this study lies in using an experimental protocol that limits data leakage and allows a reliable comparison of the models' predictive performance and generalization ability. The protocol combines Bayesian hyperparameter optimization, stratified cross-validation, and final evaluation on an independent held-out test set. Through the joint analysis of classifiers and resampling methods, the study identifies the most effective approaches for paramedical specialization prediction and clarifies how different rebalancing methods interact with tree-based algorithms in imbalanced multi-class classification problems.

The paper is structured as follows. Section II introduces the main resampling techniques and the stratified k-fold cross-validation method. Section III presents the adopted methodology. Section IV reports and discusses the main results. The final section summarizes the key findings, addresses the study's limitations, and proposes directions for future research.

II. BACKGROUND

A. Class Imbalance Handling Methods

Class imbalance is a frequent and important problem in machine learning because it may push classification models to favor the majority class and perform poorly on the minority class [34, 35]. To address this issue, several resampling techniques have been developed to rebalance the class distribution and improve model performance. Among the most widely used techniques are SMOTE, ADASYN, SMOTEENN, and SMOTETomek, whose effectiveness has been supported by several studies [36, 37].

1) Synthetic Minority Over-sampling Technique (SMOTE)

SMOTE [38] is an oversampling technique that generates synthetic instances of the minority class rather than duplicating existing instances [39–41]. For each sample in the minority class, SMOTE uses the KNN algorithm to find its k nearest neighbors. Synthetic instances are then created along the line segments connecting that sample to randomly chosen neighbors [38]. This synthetic-sample generation process is repeated until the desired oversampling rate is reached.

Unlike simple oversampling techniques, this method reduces the risk of overfitting by ensuring the diversity and distribution of synthetic samples in the feature space of the minority class, which can improve classification performance on minority classes and enhance the model's ability to generalize [38, 41].

2) Adaptive Synthetic Sampling (ADASYN)

ADASYN [42] is an extension of SMOTE that generates synthetic samples according to the learning difficulty of minority-class instances [37]. Unlike SMOTE, this technique focuses on the minority instances that are most difficult to classify, namely those surrounded by a larger number of majority class instances [43].

This adaptive oversampling technique shifts decision boundaries more aggressively in favor of the minority class in areas of high overlap by generating more synthetic samples for these difficult instances [37, 42]. This can potentially improve classification in these critical areas [43].

3) The SMOTE-ENN method

SMOTEENN, introduced by Batista *et al.* [44], is a hybrid method designed to handle class imbalance in datasets. It combines Synthetic Minority Over-sampling Technique (SMOTE) with Edited Nearest Neighbor (ENN).

SMOTEENN operates through two successive stages. First, SMOTE generates new minority-class samples, increasing their representation without merely duplicating existing cases. Second, ENN removes noisy and ambiguous instances that are misclassified by their k -nearest neighbors [37, 44], reducing both noise and class overlap. By combining oversampling and undersampling, this sequential process rebalances the class distribution and improves data quality, which can enhance the performance of the classification model [41].

4) SMOTETomek

SMOTETomek [44] is another hybrid method that combines the advantages of SMOTE and Tomek Links [37].

After SMOTE is used to generate new synthetic samples for the minority class, Tomek Links are used to clean the dataset. A Tomek Link is defined as a pair of instances belonging to different classes that are each other's nearest neighbors [45]. In SMOTETomek, majority class instances that are part of these links are removed, and sometimes minority-class instances are also removed if they are involved as well [37]. This hybrid method therefore aims both to balance the data and to eliminate class overlap by removing samples located near the decision boundary or that can be considered noise, which improves separation between classes.

B. The Stratified K-Fold Method

Stratified K-Fold cross-validation is widely used to assess

machine-learning model performance. It is particularly useful for classification problems with imbalanced data because it ensures that each fold retains approximately the same class distribution as the original dataset [46, 47]. This makes it possible to obtain a more reliable and representative evaluation of model performance [47].

More specifically, this technique divides the dataset into K stratified subsets. $K-1$ of them are used to train the model, and the remaining one is used to test it [47]. This process is repeated k times, changing the fold used for testing at each iteration so that each fold serves as the test subset only once [47].

This method not only reduces the risk of overfitting but also provides a more stable and generalizable estimate of the overall performance of the studied models.

III. METHODOLOGY

This study aims to analyze and compare different tree-based classification models combined with class-imbalance handling approaches, in order to identify the most effective strategy for predicting students' paramedical specialty. To this end, we adopted an experimental approach. Fig. 1 illustrates the overall workflow. In this section, we briefly describe the main stages of this process.

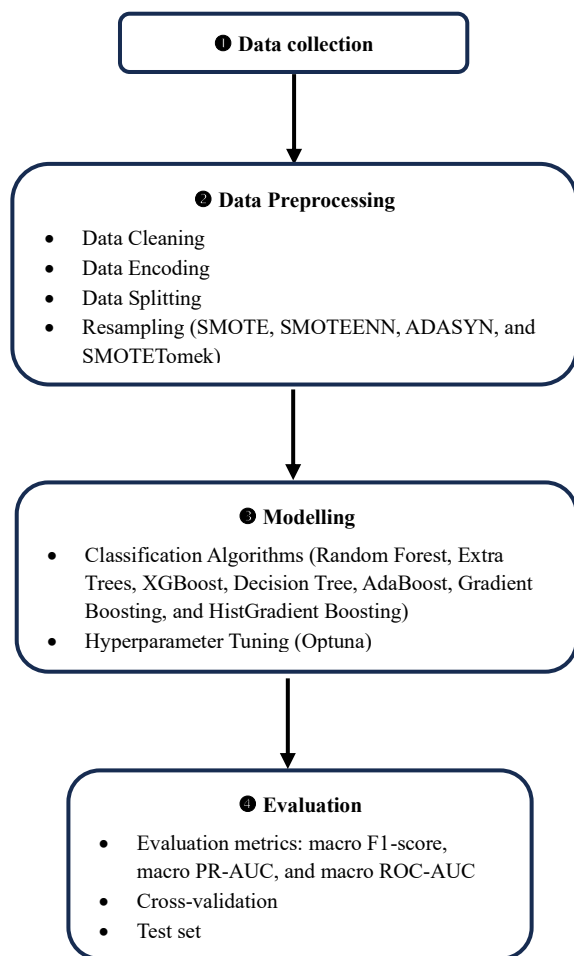


Fig. 1. Overall workflow.

A. Data Description

Our experimental approach is based on data from nursing students enrolled in two Moroccan training institutes. The

sample consists of 889 students who enrolled for the first time in the two training institutes during three academic years (2022–2023, 2023–2024, and 2024–2025), across different paramedical training programs.

The collected dataset contains 13 features related to demographic and academic information, as well as students' preferences. The study specialty is treated as the target variable, while the remaining features are considered predictors. Table 1 presents all variables, their descriptions, and their types.

Table 1. Dataset features used

Category	Variable	Description	Type
Demographic characteristics	Gender	Student gender	Categorical
	Area	Geographical origin (urban area, rural area)	Categorical
	O_Average	Overall baccalaureate average	Numeric
	N_Exa	Average of the national baccalaureate exam	Numeric
	R_Exa	Average of the regional baccalaureate exam	Numeric
Academic characteristics	Bac_Stream	Baccalaureate track/stream	Categorical
	Bac_Dist	Baccalaureate distinction (Excellent, Good, ...)	Categorical
	Prov_Del	Provincial delegation of the Ministry of National Education where the baccalaureate was obtained.	Categorical
	Bac_Session	Baccalaureate session (regular session, resit session)	Categorical
	Bac_Cand	Baccalaureate candidacy type (official candidate, independent candidate)	Categorical
Preferences	Domain	Chosen domain	Categorical
	Institute	Chosen institute	Categorical
Target variable	Specialty	Specialty of study: IP (general nurse), ISM (mental health nurse), AS (social worker), SF (midwife), SE (health and environment), Lab (laboratory).	Categorical

B. Data Preprocessing

Preprocessing is a crucial step for improving data quality and the performance of machine learning models. In this study, the collected dataset underwent rigorous preprocessing to convert it into a format suitable for the different classification models. This process involved the following steps and techniques:

- (1) Data cleaning: The dataset was checked to identify missing, redundant, or inconsistent values. Missing values were handled by imputation (mean or mode depending on the nature of the variable) or by deletion when the proportion of missing data was marginal.
- (2) Data encoding: Categorical variables were converted into numerical values using appropriate encoding techniques, such as one-hot encoding, label encoding, and ordinal encoding. One-hot encoding was used for nominal variables with no ordinal relationship in order to avoid any erroneous interpretation of a hierarchy between categories, while ordinal encoding was used for ordinal variables with a clear and meaningful hierarchical order. Label encoding was applied to the target variable.

C. Machine Learning Algorithms

This study examines seven machine learning classifiers, namely Decision Tree, Random Forest, Extra Trees,

AdaBoost, Gradient Boosting, HistGradient Boosting, and XGBoost, in order to predict students' paramedical specialization in an imbalanced multi-class data context. These models were selected to represent different ensemble learning paradigms recognized for their effectiveness on structured tabular data and predictive tasks. Random Forest and Extra Trees were retained for their generalization and variance reduction capacities, achieved through bagging and random feature selection. The four boosting models, AdaBoost, Gradient Boosting, HistGradient Boosting, and XGBoost, were chosen for their ability to optimize classification performance iteratively by focusing on complex or misclassified examples. The Decision Tree was included as an interpretable baseline model, allowing the study to compare simple tree-based learning with more advanced ensemble-learning approaches. This algorithmic diversity makes it possible to analyze the interaction between the internal mechanisms of each classifier and the synthetic data generated through resampling.

Table 2. Parameter settings of the machine learning models

Model	Parameter	Search range	Type
RF	n_estimators	[100, 200, 300, 400, 500]	Integer
	max_depth	[5, 10, 15, 20, 25, 30]	Integer
	min_samples_split	[2, 20]	Integer
	min_samples_leaf	[1, 10]	Integer
ETC	n_estimators	[100, 200, 300, 400, 500]	Integer
	max_depth	[5, 10, 15, 20, 25, 30]	Integer
	min_samples_split	[2, 20]	Integer
	min_samples_leaf	[1, 10]	Integer
XGBoost	n_estimators	[100, 200, 300, 400, 500]	Integer
	max_depth	[3, 10]	Integer
	learning_rate	[0.01, 0.30]	Floating
	Subsample	[0.60, 1.00]	Floating
	colsample_bytree	[0.60, 1.00]	Floating
DT	max_depth	[5, 10, 15, 20, 25, 30]	Integer
	min_samples_split	[2, 20]	Integer
	min_samples_leaf	[1, 10]	Integer
AdaBoost	n_estimators	[100, 200, 300, 400, 500]	Integer
	learning_rate	[0.01, 0.30]	Floating
GBC	n_estimators	[100, 200, 300, 400, 500]	Integer
	max_depth	[3, 10]	Integer
HTGB	learning_rate	[0.01, 0.30]	Floating
	max_iter	[100, 200, 300, 400, 500]	Integer
	max_depth	[5, 10, 15, 20, 25, 30]	Integer

The applicability of these tree-based algorithms to the guidance problem rests on several shared properties. They generate class probabilities, which makes it possible to quantify prediction certainty [48, 49]. In addition, they can effectively model complex and non-linear relationships

among students' demographic characteristics, academic profiles, and preferences without requiring extensive feature engineering. Finally, the compatibility of all these algorithms with Bayesian hyperparameter optimization through Optuna ensures that the comparisons reflect the real potential of each approach in its optimal configuration, thereby reducing biases linked to suboptimal tuning.

Table 2 outlines the optimization parameters and search ranges for each classifier used.

D. Experimental Scenario

To ensure a rigorous evaluation of machine learning algorithms for predicting paramedical specialization, the dataset was split into a training set and a test set using an 80/20 ratio. This split was performed before any preprocessing or data manipulation to ensure that final performance estimates are not biased by any exposure of the test data during model training, thereby ensuring that the evaluation accurately reflects the models' robustness and predictive capacity.

First, the seven algorithms selected for this study were trained on balanced training data using four class-imbalance techniques: SMOTE, SMOTEENN, ADASYN, and SMOTETomek. These resampling methods were applied only to the training folds within cross-validation, in order to avoid any leakage of synthetic data into the test set and to preserve the integrity of the evaluation. They were implemented using the imbalanced-learn library, with default parameters, in order to ensure methodological consistency and a fair comparison between the different resampling strategies. Table 3 summarizes the main parameter settings for each technique used in this study.

Table 3. Parameter Settings of the resampling methods

Methods	Component	Parameters
SMOTE		k_neighbors = 5
SMOTEENN	SMOTE	k_neighbors = 5
	ENN	n_neighbors = 3
SMOTETomek	SMOTE	k_neighbors = 5
	Tomek Links	
ADASYN		n_neighbors = 5

Hyperparameter optimization for each classifier was performed using Bayesian optimization implemented with Optuna [50]. This method enables more effective and efficient tuning than Grid Search or Random Search [51, 52]. Table 4 presents the results of this optimization for the seven models.

Table 4. Optimized models and their hyperparameters

Model	Hyperparameter	SMOTE	SMOTEENN	SMOTETomek	ADASYN
RF	n_estimators	500	100	300	200
	max_depth	30	10	10	5
	min_samples_split	18	6	19	15
	min_samples_leaf	4	3	8	7
ETC	n_estimators	200	400	300	100
	max_depth	10	15	15	30
	min_samples_split	17	13	18	16
	min_samples_leaf	5	8	10	5
XGBoost	n_estimators	100	300	100	100
	max_depth	4	5	4	4
	learning_rate	0.0423	0.0101	0.0145	0.0267
	subsample	0.8500	0.8019	0.6216	0.6231
	colsample_bytree	0.6453	0.6019	0.6002	0.8944
DT	max_depth	15	5	5	10
	min_samples_split	7	18	7	7
	min_samples_leaf	7	7	4	6

Model	Hyperparameter	SMOTE	SMOTEENN	SMOTETomek	ADASYN
AdaBoost	n_estimators	400	200	200	100
	learning_rate	0.0766	0.2950	0.0163	0.0187
GBC	n_estimators	400	400	200	100
	max_depth	4	8	4	8
HTGB	learning_rate	0.0127	0.1532	0.0146	0.2676
	learning_rate	0.0106	0.0146	0.0347	0.1402
	max_iter	300	400	100	300
	max_depth	20	20	25	25

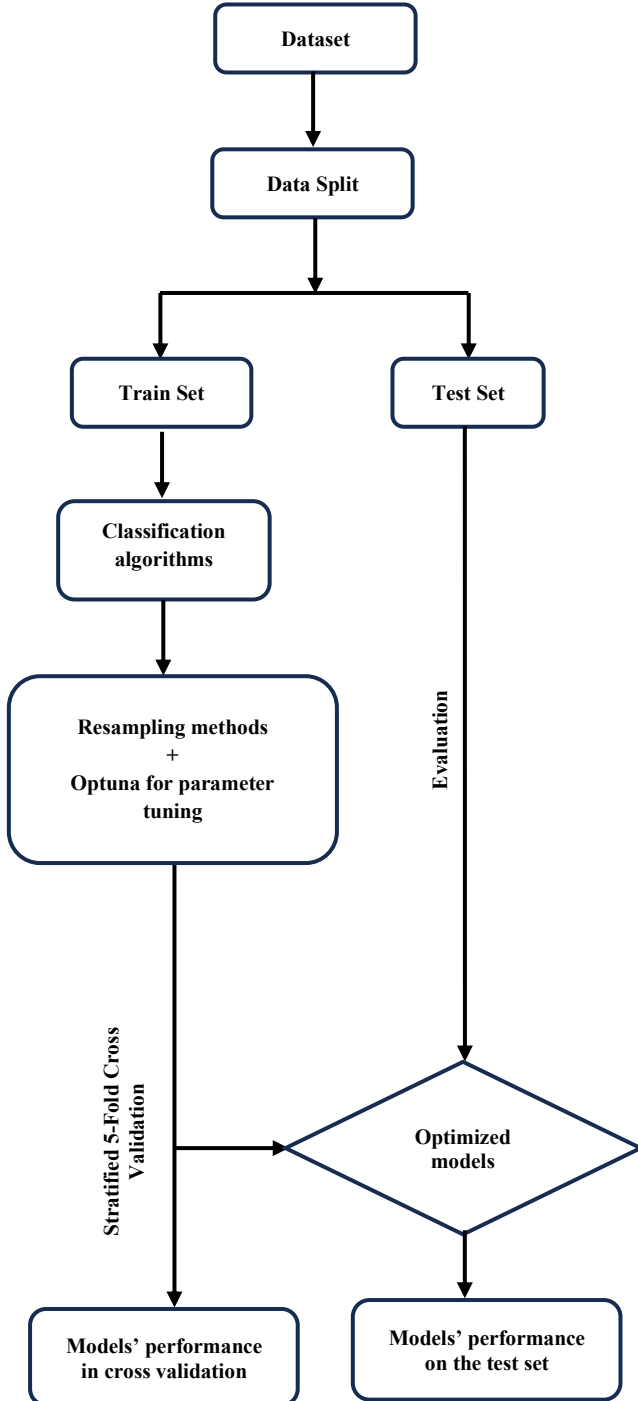


Fig. 2. Experimental design.

The models' performance was validated using stratified 5-fold cross-validation. This technique reduces the risk of overfitting and provides a more stable and generalizable estimate of the overall performance of the compared models [53, 54].

The optimized models were then tested on the independent test set to evaluate their generalization ability and check

whether overfitting was present. This final evaluation made it possible to assess the robustness of the models and their capacity to predict specializations accurately on unseen data (see Fig. 2).

E. Evaluation Metrics

The dataset used in this study exhibits substantial imbalance, which requires choosing appropriate evaluation measures that are not biased by unequal class representation. Accuracy can be misleading in this context because a model that mostly predicts the majority class can achieve high accuracy while failing to detect the minority class [16, 18, 55, 56]. For our imbalanced multi-class classification problem, we selected macro F1-score, macro Precision-Recall Area Under the Curve (PR-AUC), and macro Receiver Operating Characteristic curve and Area Under the Curve (ROC-AUC) as the main evaluation metrics.

Macro F1-score is the harmonic mean of precision and recall calculated for each class and then averaged in an unweighted manner:

$$\text{Macro F1-score} = \frac{1}{N} \sum_{i=1}^N F1_i \quad (1)$$

with

$$F1_i = 2 \times \frac{\text{Precision}_i \times \text{Recall}_i}{\text{Precision}_i + \text{Recall}_i} \quad (2)$$

$$\text{Precision}_i = \frac{TP_i}{TP_i + FP_i} \quad \text{and} \quad \text{Recall}_i = \frac{TP_i}{TP_i + FN_i} \quad (3)$$

where:

- N is the total number of classes.
- TP_i (True Positives): represents correct predictions for class i .
- FP_i (False Positives): refers to incorrect positive predictions for class i .
- FN_i (False Negatives): refers to incorrect negative predictions for class i .

Because each class contributes equally to the final score, this metric reduces the dominance of majority classes. A high macro F1-score therefore indicates that the model performs well across all classes rather than only on the most represented ones.

Macro PR-AUC is calculated by aggregating the precision-recall curves of each class. It is particularly useful in imbalanced contexts because it highlights the trade-off between precision and recall for minority classes. A PR-AUC value close to 1 indicates stronger performance.

ROC-AUC measures the model's ability to distinguish between classes by plotting the True Positive Rate (TPR) or recall against the False Positive Rate (FPR) across different

classification thresholds [57, 58].

For multi-class and imbalanced problems, the AUC is first computed for each class using a one-vs-rest approach, then these AUCs are averaged to obtain the macro ROC-AUC [57–59].

These three metrics are particularly useful for imbalanced data and provide a more nuanced and reliable evaluation of classifiers' performance by emphasizing their ability to handle minority classes.

IV. RESULTS AND DISCUSSION

This study evaluated seven supervised machine learning classifiers combined with four resampling techniques for predicting paramedical study specializations. The analysis followed a two-stage evaluation design: first, stratified 5-fold cross-validation was applied to the training data; second, the optimized models were assessed using independent test data to examine their final performance.

A. Model Performance in Cross-Validation

In the first evaluation stage, Random Forest, Extra Trees, XGBoost, Decision Tree, AdaBoost, Gradient Boosting, and HistGradient Boosting were trained and optimized using the training data. Their performance was then assessed through stratified 5-fold cross-validation using macro F1-score as the main metric. Table 5 presents the cross-validation results obtained by the seven classifiers under the four resampling techniques.

Table 5 indicates that the ensemble models generally performed well in cross-validation, with macro F1-scores ranging from 0.8470 to 0.8786. Random Forest obtained the strongest and most consistent results across the four resampling techniques, reaching its highest score with SMOTE (0.8786). This suggests that bootstrap aggregation

enabled the model to mitigate the variability introduced by synthetic samples while preserving balanced performance across classes.

Table 5. Cross-validation performance (Macro F1-score)

Model	SMOTE	SMOTEENN	ADASYN	SMOTETomek
RF	0.8786	0.8706	0.8727	0.8779
ETC	0.8621	0.8470	0.8632	0.8691
XGBoost	0.8700	0.8679	0.8659	0.8724
DT	0.8537	0.8555	0.8443	0.8632
AdaBoost	0.2794	0.5000	0.4230	0.3807
GBC	0.8673	0.8651	0.8515	0.8646
HTGB	0.8652	0.8632	0.8474	0.8645

The Decision Tree baseline also produced acceptable results, with macro F1-scores between 0.8443 and 0.8632, but consistently underperformed relative to the more robust ensemble models. This performance gap is expected because a single tree is more sensitive to local patterns and struggles to generalize when demographic, academic, and preference-related variables interact within a multi-class educational dataset. In contrast, ensemble methods mitigate this limitation by aggregating multiple base learners [60].

AdaBoost was clearly the weakest model in cross-validation, with its highest macro F1-score reaching only 0.5000 under SMOTEENN.

B. Model Performance on the Test Set

In the final experimental stage, the optimized models were evaluated on the independent test set to determine how well they generalized to unseen data and to identify the most suitable method for paramedical-specialization prediction. Test-set performance was assessed using metrics appropriate for class-imbalanced data, namely macro F1-score, macro-ROC-AUC, and macro-PR-AUC. Table 6 reports these results.

Table 6. Performance on the test set

Model	Macro F1-score (Test)	Macro PR-AUC	Macro ROC-AUC	Inference time (s)
SMOTE	RF	0.8765	0.9223	0.9891
	ETC	0.8557	0.9340	0.9903
	XGBoost	0.8828	0.9271	0.9900
	DT	0.8279	0.8620	0.9406
	AdaBoost	0.4844	0.7254	0.8872
	GBC	0.8312	0.9057	0.9839
	HTGB	0.8417	0.9245	0.9874
SMOTEENN	RF	0.8836	0.9294	0.9879
	ETC	0.8374	0.9286	0.9894
	XGBoost	0.8597	0.9157	0.9846
	DT	0.8570	0.8751	0.9791
	AdaBoost	0.1384	0.8333	0.9217
	GBC	0.8255	0.8789	0.9709
	HTGB	0.8425	0.8875	0.9795
SMOTETomek	RF	0.8645	0.9228	0.9892
	ETC	0.8494	0.9365	0.9903
	XGBoost	0.8659	0.9193	0.9886
	DT	0.8659	0.8553	0.9795
	AdaBoost	0.2110	0.5528	0.7890
	GBC	0.8508	0.9167	0.9858
	HTGB	0.8739	0.9203	0.9880
ADASYN	RF	0.8570	0.9121	0.9889
	ETC	0.8573	0.9297	0.9891
	XGBoost	0.8428	0.9188	0.9895
	DT	0.8424	0.8851	0.9624
	AdaBoost	0.2990	0.7003	0.8814
	GBC	0.8692	0.9024	0.9837
	HTGB	0.8504	0.9020	0.9808

As shown in Table 6, the experimental results confirm the robustness of the ensemble models. In particular, Random

Forest and XGBoost achieved the highest macro F1-scores on the test set, indicating a strong balance between precision and

recall across the six specialties. RF showed robust and relatively stable performance, with test F1-scores ranging from 0.8570 (ADASYN) to 0.8836 (SMOTEENN), while XGBoost demonstrated adaptability to different synthetic data distributions, with scores ranging from 0.8428 (ADASYN) to 0.8828 (SMOTE). This strong performance is further supported by high macro ROC-AUC values, which generally exceed 0.98, confirming the models' ability to distinguish effectively between specialties. The high macro PR-AUC values, above 0.91, reflect the models' effectiveness in classifying minority classes more directly than standard metrics.

AdaBoost consistently exhibited weak and unstable test-set performance, with macro F1-scores ranging from 0.1384 under SMOTEENN to 0.4844 under SMOTE. This poor performance highlights a critical limitation of the algorithm's adaptive weighting mechanism when applied to synthetic data. Because AdaBoost iteratively assigns higher weights to

misclassified instances, it tends to focus excessively on noisy observations or ambiguous samples generated by resampling techniques [61, 62]. In an imbalanced multi-class setting, this behavior destabilizes the decision boundaries, effectively biasing the model toward majority classes and significantly impairing its ability to generalize across all specialties. These findings indicate that AdaBoost is not well suited for predicting specializations in contexts that combine imbalanced multi-class classification with synthetic resampling, suggesting it is a suboptimal option for similar educational problems.

The comparison between cross-validation and independent test-set results suggests that most ensemble models generalized well to unseen data. The limited gap between these two evaluation stages, illustrated in Fig. 3, indicates the reliability of the adopted protocol and reduces the likelihood that the reported performance was driven by overfitting.

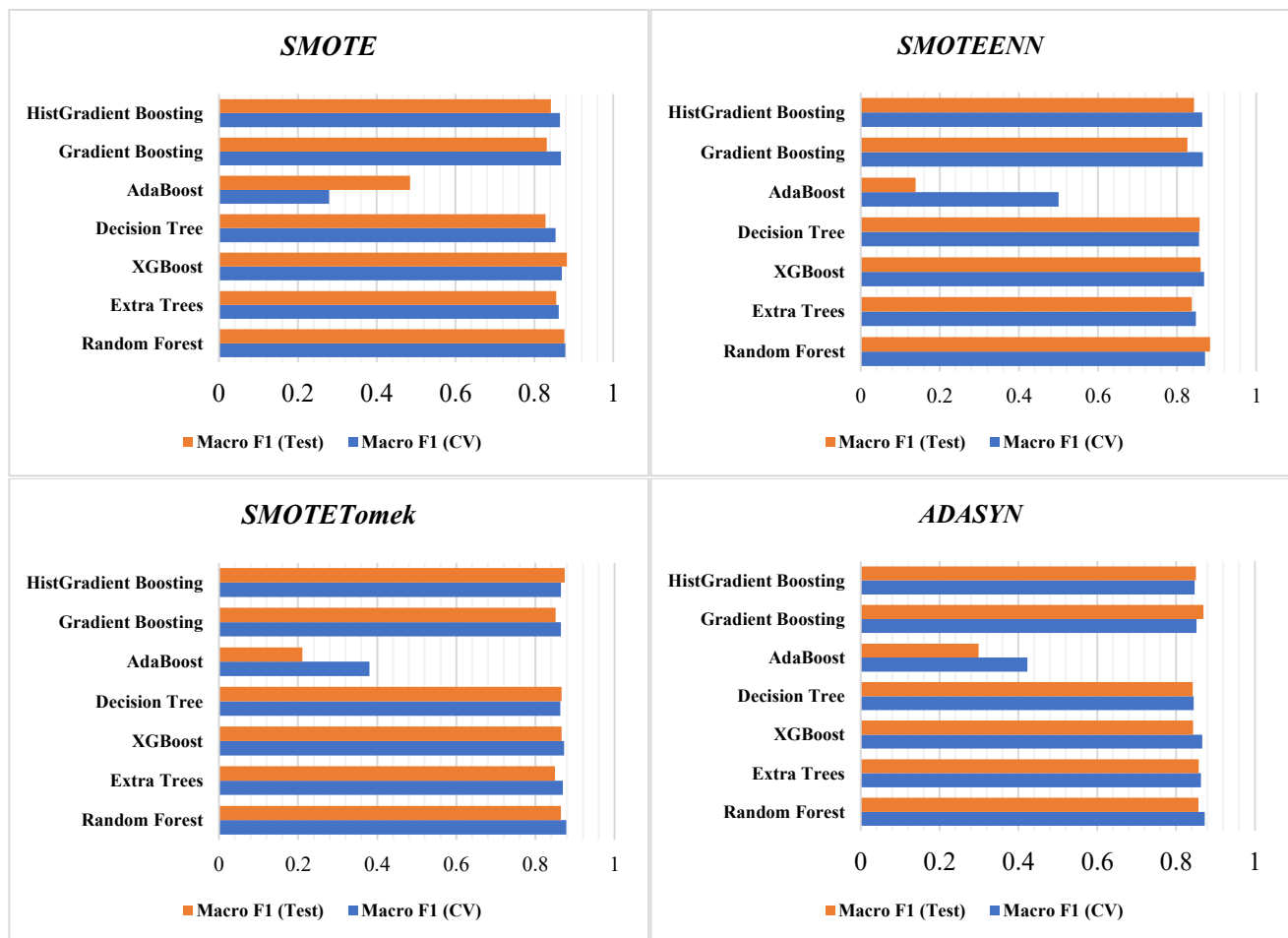


Fig. 3. Comparison of models' performance in cross-validation and on the test set under four resampling techniques: (a). SMOTE; (b). SMOTEENN; (c). SMOTETomek; (d). ADASYN.

Overall, Random Forest and XGBoost were the most effective models for predicting paramedical specialization. Random Forest combined with SMOTEENN achieved the strongest generalization performance, whereas XGBoost combined with SMOTE offered a nearly equivalent predictive result with a much lower inference time (0.0060 s). This makes XGBoost with SMOTE particularly attractive when fast prediction is required in an operational guidance system.

The best performance of Random Forest with

SMOTEENN can be explained by the complementarity between the classifier and the resampling method. SMOTEENN increases minority-class representation and then removes ambiguous observations, while RF reduces variance through bagging and random feature selection. This combination of oversampling and cleaning helps clarify decision regions while controlling the noise that may result from pure oversampling. This finding aligns with Priyanto *et al.* [23], who showed that Random Forest can benefit from a hybrid strategy that combines synthetic

balancing and cleaning in health-data classification.

XGBoost performs better with SMOTE than with the hybrid cleaning methods. This suggests that XGBoost benefits from a denser minority-class representation and does not require aggressive post-oversampling cleaning, since its regularized boosting mechanism can already manage complex interactions and reduce error through sequential learning. This result is also consistent with Innuddin *et al.* [22], who found that oversampling can markedly improve the performance of RF and XGBoost on imbalanced classification tasks. In the present educational dataset, SMOTE appears to provide XGBoost with sufficient minority-class support without removing borderline cases that may carry useful information.

These findings suggest that resampling methods should not be selected independently of the classifier. The same balancing strategy can improve one model while weakening another, depending on how the algorithm responds to noise, class overlap, and synthetic-example density.

V. CONCLUSION

This study aimed to identify the optimal modeling approach for predicting students' paramedical study specializations in an imbalanced multi-class context. A systematic comparative analysis of seven tree-based classifiers, combined with four resampling techniques (SMOTE, ADASYN, SMOTEENN, and SMOTETomek), was conducted using real-world data from nursing students at two Moroccan training institutes.

The findings show that the interaction between the model architecture and the resampling strategy is crucial. Random Forest with SMOTEENN provided the most reliable performance because the combination of oversampling and data cleaning complements the variance-reduction capacity of bagged trees. XGBoost with SMOTE provides the best operational trade-off, delivering nearly the same predictive performance while requiring far less inference time. These results provide a practical basis for developing academic guidance systems that are both accurate and sensitive to underrepresented specializations. The finding regarding AdaBoost is methodologically significant because it shows that compatibility between the algorithm's architecture and the resampling technique is as decisive as the raw performance reported in the literature. Consequently, AdaBoost should be used with caution in configurations that combine imbalanced multi-class data with synthetic oversampling methods.

The study has several limitations. First, the dataset was collected from two Moroccan training institutes across three academic years, which limits the generalizability of the findings to all Moroccan paramedical training institutes and to other national contexts. Second, the predictive models rely mainly on demographic, academic, and preference-related variables; future research should incorporate qualitative and psychometric characteristics, such as motivation, stress resistance, empathy, and spatial reasoning. Third, the study focused on tree-based algorithms; future work could compare these results with deep learning and hybrid architectures. Fourth, although the proposed framework achieved high predictive performance, model explainability was not implemented. This is a significant limitation because

educational guidance systems require transparent and interpretable recommendations.

Future research should therefore validate the proposed approach using larger, more diverse, and multi-institutional datasets. Future research should also prioritize the integration of Explainable Artificial intelligence (XAI) techniques, such as SHAP and LIME, to clarify how each recommendation is produced. Implementing such techniques would make the system more useful for academic advisors and more trustworthy for students.

CONFLICT OF INTEREST

The authors declare no conflict of interest.

AUTHOR CONTRIBUTIONS

Samira Reguragui conceived and designed the study, collected and prepared the dataset, conducted the machine learning experiments, analyzed and interpreted the results, and wrote the original manuscript. Ahmed Bendahmane supervised the research and provided academic and methodological guidance throughout the study, including guidance on the research design, interpretation of the findings, and revision of the manuscript. Abdellatif El Aidi wrote, revised and proofread the manuscript. All authors approved the final version.

REFERENCES

- [1] Y. Harraki, A. Moumen, A. Remaida, and D. Gretete, "Which actors influence student's university orientation in Morocco? An exploratory study with civil engineering students," *SHS Web Conf.*, vol. 119, 05011, 2021. doi: 10.1051/shsconf/202111905011
- [2] S. Lafdoul, S. Baadou, and A. Bendahmane, "Machine learning algorithms applied to post-bac academic guidance," in *Proc. 2024 Mediterranean Smart Cities Conference (MSCC)*, May 2024, pp. 1–4. doi: 10.1109/MSCC62288.2024.10697032
- [3] O. Zahour, E. H. Benlahmar, A. Eddauim, and O. Hourrane, "A Comparative Study of Machine Learning Methods for Automatic Classification of Academic and Vocational Guidance Questions," *International Journal of Interactive Mobile Technologies (IJIM)*, vol. 14, no. 8, pp. 43–60, May 2020. doi: 10.3991/ijim.v14i08.13005
- [4] H. Majjate, Y. Bellarhmouch, A. Jeghal, A. Yahyaouy, H. Tairi, and K. A. Zidani, "AI-Powered Academic Guidance and Counseling System Based on Student Profile and Interests," *Applied System Innovation*, vol. 7, no. 1, 6, 2023.
- [5] M. Badrani, A. Marouan, N. Kannouf, and A. Chetouani, "Smart Techniques for Moroccan Students' Orientation," in *Proc. 2nd International Conference on Emerging Technologies and Intelligent Systems*, M. A. Al-Sharafi, M. Al-Emran, M. N. Al-Kabi, and K. Shaalan, Eds. Cham: Springer International Publishing, 2023, pp. 361–366. doi: 10.1007/978-3-031-20429-6_33
- [6] M. Badrani, A. Marouan, N. Kannouf, and A. Chetouani, "Pathways to Success: A Study of Classification Algorithms for Guiding Middle School Students in Morocco," in *Proc. the 6th International Conference on Networking, Intelligent Systems & Security*, Larache Morocco: ACM, May 2023, pp. 1–5. doi: 10.1145/3607720.3607748
- [7] M. Badrani, A. Marouan, N. Kannouf, and A. Chetouani, "Personalized Guidance for Moroccan Students: An Approach Based on Machine Learning and Big Data," *International Journal of Engineering Pedagogy (IJEP)*, vol. 15, no. 1, 1, Jan. 2025. doi: 10.3991/ijep.v15i1.51985
- [8] A. O. Alsayed *et al.*, "Selection of the Right Undergraduate Major by Students Using Supervised Learning Techniques," *Applied Sciences*, vol. 11, no. 22, 10639, Jan. 2021. doi: 10.3390/app112210639
- [9] H. Karalar, C. Kapucu, and H. Gürüler, "Predicting students at risk of academic failure using ensemble model during pandemic in a distance learning system," *Int. J. Educ. Technol. High Educ.*, vol. 18, no. 1, 63, Dec. 2021. doi: 10.1186/s41239-021-00300-y
- [10] A. Surawatchayotin, W. Paireekreng, A. I. S. Shafie, and M. A. Shehab, "The Predictive Model of Higher Education Guidance for Information Overload of Learner Groups Using Hybrid Ensemble Techniques," *TEM Journal*, vol. 11, no. 4, 1792, 2022.

- [11] H. Şevgin, "A comparative study of ensemble methods in the field of education: Bagging and Boosting algorithms," *International Journal of Assessment Tools in Education*, vol. 10, no. 3, pp. 544–562, 2023.
- [12] E. Mbunge *et al.*, "Implementation of ensemble machine learning classifiers to predict diarrhoea with SMOTEENN, SMOTE, and SMOTETomek class imbalance approaches," in *Proc. 2023 Conference on Information Communications Technology and Society (ICTAS)*, Mar. 2023, pp. 1–6. doi: 10.1109/ICTAS56421.2023.10082744
- [13] D. Thammasiri, D. Delen, P. Meesad, and N. Kasap, "A critical assessment of imbalanced class distribution problem: The case of predicting freshmen student attrition," *Expert Systems with Applications*, vol. 41, no. 2, pp. 321–330, Feb. 2014. doi: 10.1016/j.eswa.2013.07.046
- [14] A. F. Gkontzis, S. Kotsiantis, C. T. Panagiotakopoulos, and V. S. Verykios, "A predictive analytics framework as a countermeasure for attrition of students," *Interactive Learning Environments*, vol. 30, no. 6, pp. 1028–1043, Jul. 2022. doi: 10.1080/10494820.2019.1709209
- [15] A. Villar and C. R. V. de Andrade, "Supervised machine learning algorithms for predicting student dropout and academic success: A comparative study," *Discov. Artif. Intell.*, vol. 4, no. 1, 2, Jan. 2024. doi: 10.1007/s44163-023-00079-z
- [16] P. Vuttipittayamongkol, E. Elyan, and A. Petrovski, "On the class overlap problem in imbalanced data classification," *Knowledge-Based Systems*, vol. 212, 106631, 2021.
- [17] S. Ossenov, "Developing a Dataset-Adaptive, Normalized Metric for Machine Learning Model Assessment: Integrating Size, Complexity, and Class Imbalance," arXiv: arXiv:2412.07244, Dec. 10, 2024. doi: 10.48550/arXiv.2412.07244.
- [18] M. Salmi, D. Atif, D. Oliva, A. Abraham, and S. Ventura, "Handling imbalanced medical datasets: Review of a decade of research," *Artif. Intell. Rev.*, vol. 57, no. 10, 273, Sep. 2024. doi: 10.1007/s10462-024-10884-2
- [19] R. Ghorbani and R. Ghousi, "Comparing Different Resampling Methods in Predicting Students' Performance Using Machine Learning Techniques," *IEEE Access*, vol. 8, pp. 67899–67911, 2020. doi: 10.1109/ACCESS.2020.2986809
- [20] A. Nabil, M. Seyam, and A. Abou-Elfetouh, "Prediction of Students' Academic Performance Based on Courses' Grades Using Deep Neural Networks," *IEEE Access*, vol. 9, pp. 140731–140746, 2021. doi: 10.1109/ACCESS.2021.3119596
- [21] M. F. Al-Hammouri, Z. A. A. Hammouri, I. T. Almalkawi, and A. Lafee, "Optimizing Multi-Class Classification in Educational Data with Ensemble Learning and Data Balancing Techniques," in *Proc. 2024 Fifth International Conference on Intelligent Data Science Technologies and Applications (IDSTA)*, Sep. 2024, pp. 12–17. doi: 10.1109/IDSTA62194.2024.10746987
- [22] M. Innuddin, Hairani, M. T. Jauhari, and L. Z. A. Mardedi, "Improving Classification Performance on Imbalanced Stroke Datasets Using Oversampling Techniques," *Jurnal RESTI (Rekayasa Sistem dan Teknologi Informasi)*, vol. 9, no. 5, pp. 1193–1199, Oct. 2025. doi: 10.29207/resti.v9i5.6859
- [23] D. Priyanto, H. Hairani, K. Marzuki, and M. Innuddin, "Optimization of Random Forest for Health Data Classification Using PCA and K-Means SMOTE-ENN," *Engineering, Technology & Applied Science Research*, vol. 15, no. 5, pp. 27646–27652, Oct. 2025. doi: 10.48084/etasr.12976
- [24] A. Ishaq *et al.*, "Improving the Prediction of Heart Failure Patients' Survival Using SMOTE and Effective Data Mining Techniques," *IEEE Access*, vol. 9, pp. 39707–39716, 2021. doi: 10.1109/ACCESS.2021.3064084
- [25] Z. Salekshahrezaee, J. L. Leevy, and T. M. Khoshgoftaar, "The effect of feature extraction and data sampling on credit card fraud detection," *J. Big Data*, vol. 10, no. 1, 6, Jan. 2023. doi: 10.1186/s40537-023-00684-w
- [26] I. Bouzgarne, M. Youssfi, O. Bouattane, and Q. Mohamed, "Empirical Oversampling Threshold Strategy for Machine Learning Performance Optimisation in Insurance Fraud Detection," *International Journal of Advanced Computer Science and Applications*, vol. 11, pp. 432–437, Oct. 2020. doi: 10.14569/IJACSA.2020.0111054
- [27] A. Muaz, M. Jayabalan, and V. Thiruchelvam, "A Comparison of Data Sampling Techniques for Credit Card Fraud Detection," *IJACSA*, vol. 11, no. 6, 2020. doi: 10.14569/IJACSA.2020.0110660
- [28] K. Praveen Mahesh, S. Ashar Afrouz, and A. Shaju Areeckal, "Detection of fraudulent credit card transactions: A comparative analysis of data sampling and classification techniques," *J. Phys.: Conf. Ser.*, vol. 2161, no. 1, 012072, Jan. 2022. doi: 10.1088/1742-6596/2161/1/012072
- [29] H. E. Mrabet and A. A. Moussa, "A framework for predicting academic orientation using supervised machine learning," *J. Ambient Intell. Human Comput.*, vol. 14, no. 12, pp. 16539–16549, Dec. 2023. doi: 10.1007/s12652-022-03909-7
- [30] M. Mimis, M. El Hajji, Y. Es-saady, A. Oued Guejdi, H. Douzi, and D. Mammass, "A framework for smart academic guidance using educational data mining," *Education and Information Technologies*, vol. 24, no. 2, pp. 1379–1393, Mar. 2019. doi: 10.1007/s10639-018-9838-8
- [31] N. Kamal, F. Sarker, and K. A. Mamun, "A comparative study of machine learning approaches for recommending university faculty," in *Proc. 2020 2nd International Conference on Sustainable Technologies for Industry 4.0 (STI)*, IEEE, 2020, pp. 1–6.
- [32] A. Tarik and Y. Farhaoui, "Recommender System for Orientation Student," in *Big Data and Networks Technologies*, vol. 81, Y. Farhaoui, Ed. in Lecture Notes in Networks and Systems, Cham: Springer International Publishing, 2020, pp. 367–370. doi: 10.1007/978-3-030-23672-4_27
- [33] F. Ouatik, M. Erritali, F. Ouatik, and M. Jourhmane, *Students' Orientation Using Machine Learning and Big Data*, 2021.
- [34] H. He and E. A. Garcia, "Learning from Imbalanced Data," *IEEE Transactions on Knowledge and Data Engineering*, vol. 21, no. 9, pp. 1263–1284, Sep. 2009. doi: 10.1109/TKDE.2008.239
- [35] C. Yang, E. A. Fridgeirsson, J. A. Kors, J. M. Reys, and P. R. Rijnbeek, "Impact of random oversampling and random undersampling on the performance of prediction models developed using observational health data," *J. Big Data*, vol. 11, no. 1, 7, Jan. 2024. doi: 10.1186/s40537-023-00857-7
- [36] M. H. Kotb and R. Ming, "Comparing SMOTE Family Techniques in Predicting Insurance Premium Defaulting using Machine Learning Models," *International Journal of Advanced Computer Science and Applications (IJACSA)*, vol. 12, no. 9, Sep. 2021. doi: 10.14569/IJACSA.2021.0120970
- [37] M. Rath and H. Date, "Adaptive Modelling Approach for Row-Type Dependent Predictive Analysis (RTDPA): A Framework for Designing Machine Learning Models for Credit Risk Analysis in Banking Sector," arXiv: arXiv:2311.10799, Nov. 17, 2023. doi: 10.48550/arXiv.2311.10799.
- [38] N. V. Chawla, K. W. Bowyer, L. O. Hall, and W. P. Kegelmeyer, "SMOTE: Synthetic Minority Over-sampling Technique," *Journal of Artificial Intelligence Research*, vol. 16, pp. 321–357, Jun. 2002. doi: 10.1613/jair.953
- [39] N. Budhathoki, R. Bhandari, S. Bashyal, and C. Lee, "Predicting asthma using imbalanced data modeling techniques: Evidence from 2019 Michigan BRFSS data," *PLOS ONE*, vol. 18, no. 12, e0295427, Dec. 2023. doi: 10.1371/journal.pone.0295427
- [40] W. Chen, K. Yang, Z. Yu, Y. Shi, and C. L. P. Chen, "A survey on imbalanced learning: latest research, applications and future directions," *Artif. Intell. Rev.*, vol. 57, no. 6, 137, May 2024. doi: 10.1007/s10462-024-10759-6
- [41] G. Husain *et al.*, "SMOTE vs. SMOTEENN: A Study on the Performance of Resampling Algorithms for Addressing Class Imbalance in Regression Models," *Algorithms*, vol. 18, no. 1, Jan. 2025. doi: 10.3390/a18010037
- [42] H. He, Y. Bai, E. A. Garcia, and S. Li, "ADASYN: Adaptive synthetic sampling approach for imbalanced learning," in *Proc. 2008 IEEE International Joint Conference on Neural Networks (IEEE World Congress on Computational Intelligence)*, Jun. 2008, pp. 1322–1328. doi: 10.1109/IJCNN.2008.4633969
- [43] A. Chatterjee, N. Pahari, A. Prinz, and M. Riegler, *ActieCoach: Personalized Recommendation Generation in Activity eCoaching with Meta-Heuristic Approach*, 2023.
- [44] 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," *SIGKDD Explor. Newsl.*, vol. 6, no. 1, pp. 20–29, Jun. 2004. doi: 10.1145/1007730.1007735
- [45] K. S. Selim and S. S. Rezk, "On predicting school dropouts in Egypt: A machine learning approach," *Educ. Inf. Technol.*, vol. 28, no. 7, pp. 9235–9266, Jul. 2023. doi: 10.1007/s10639-022-11571-x
- [46] S. Szeghalmy and A. Fazekas, "A Comparative Study of the Use of Stratified Cross-Validation and Distribution-Balanced Stratified Cross-Validation in Imbalanced Learning," *Sensors*, vol. 23, no. 4, 2333, Jan. 2023. doi: 10.3390/s23042333
- [47] S. Sutarman, P. K. Nasution, K. J. Sirait, and C. N. Yolanda Panjaitan, "Identifying the Impact of the Complexity of Datasets in Bayesian Optimized XGboost on the Performance of Classifications for Imbalanced Class Distribution Datasets," *Eastern-European Journal of Enterprise Technologies*, vol. 133, no. 4, 52, Jan. 2025. doi: 10.15587/1729-4061.2025.322626
- [48] A. Niculescu-Mizil and R. A. Caruana, "Obtaining Calibrated Probabilities from Boosting," arXiv: arXiv:1207.1403, Jul. 04, 2012. doi: 10.48550/arXiv.1207.1403

- [49] C. Strobl, J. Malley, and G. Tutz, "An introduction to recursive partitioning: Rationale, application, and characteristics of classification and regression trees, bagging, and random forests," *Psychological Methods*, vol. 14, no. 4, pp. 323–348, 2009. doi: 10.1037/a0016973
- [50] T. Akiba, S. Sano, T. Yanase, T. Ohta, and M. Koyama, "Optuna: A Next-generation Hyperparameter Optimization Framework," in *Proc. the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, New York, NY, USA: Association for Computing Machinery, Jul. 2019, pp. 2623–2631. doi: 10.1145/3292500.3330701
- [51] L. Yang and A. Shami, "On Hyperparameter Optimization of Machine Learning Algorithms: Theory and Practice," *Neurocomputing*, vol. 415, pp. 295–316, Nov. 2020. doi: 10.1016/j.neucom.2020.07.061
- [52] S. Putatunda and K. Rama, "A Comparative Analysis of Hyperopt as Against Other Approaches for Hyper-Parameter Optimization of XGBoost," in *Proc. the 2018 International Conference on Signal Processing and Machine Learning*, New York, NY, USA: Association for Computing Machinery, Nov. 2018, pp. 6–10. doi: 10.1145/3297067.3297080
- [53] Y. Zhang and L. Wu, "An MR Brain Images Classifier via Principal Component Analysis and Kernel Support Vector Machine," *PIER*, vol. 130, pp. 369–388, 2012. doi: 10.2528/PIER12061410
- [54] A. Gebauer, M. Ellinger, V. M. Brito Gomez, and M. Ließ, "Development of pedotransfer functions for water retention in tropical mountain soil landscapes: Spotlight on parameter tuning in machine learning," *SOIL*, vol. 6, no. 1, pp. 215–229, Jun. 2020. doi: 10.5194/soil-6-215-2020
- [55] I. Pratama, P. T. Prasetyaningrum, A. Y. Chandra, and O. Suria, "Measuring Resampling Methods on Imbalanced Educational Dataset's Classification Performance," *Register: Jurnal Ilmiah Teknologi Sistem Informasi*, vol. 10, no. 1, pp. 1–11, 2024.
- [56] J. Qin *et al.*, "IGL-Bench: Establishing the Comprehensive Benchmark for Imbalanced Graph Learning," arXiv: arXiv:2406.09870, Mar. 01, 2025. doi: 10.48550/arXiv.2406.09870
- [57] J. S. Aguilar-Ruiz and M. Michalak, "Classification performance assessment for imbalanced multiclass data," *Sci. Rep.*, vol. 14, no. 1, 10759, May 2024. doi: 10.1038/s41598-024-61365-z
- [58] J. Terven, D.-M. Cordova-Esparza, J.-A. Romero-González, A. Ramírez-Pedraza, and E. A. Chávez-Urbiola, "A comprehensive survey of loss functions and metrics in deep learning," *Artif. Intell. Rev.*, vol. 58, no. 7, 195, Apr. 2025. doi: 10.1007/s10462-025-11198-7
- [59] S. E. Gómez, L. Hernández-Callejo, B. C. Martínez, and A. J. Sánchez-Esguevillas, "Exploratory study on Class Imbalance and solutions for Network Traffic Classification," *Neurocomputing*, vol. 343, pp. 100–119, May 2019. doi: 10.1016/j.neucom.2018.07.091
- [60] M. Azad, T. H. Nehal, and M. Moshkov, "A novel ensemble learning method using majority based voting of multiple selective decision trees," *Computing*, vol. 107, no. 1, 42, Dec. 2024. doi: 10.1007/s00607-024-01394-8
- [61] A. A. Khan, O. Chaudhari, and R. Chandra, "A review of ensemble learning and data augmentation models for class imbalanced problems: Combination, implementation and evaluation," *Expert Systems with Applications*, vol. 244, 122778, Jun. 2024. doi: 10.1016/j.eswa.2023.122778
- [62] T. G. Dietterich, "An Experimental Comparison of Three Methods for Constructing Ensembles of Decision Trees: Bagging, Boosting, and Randomization," *Machine Learning*, vol. 40, no. 2, pp. 139–157, Aug. 2000. doi: 10.1023/A:1007607513941

Copyright © 2026 by the authors. This is an open access article distributed under the Creative Commons Attribution License which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited ([CC BY 4.0](https://creativecommons.org/licenses/by/4.0/)).