Annals of Healthcare Systems Engineering

www.ahse.reapress.com

Ann. Healthc. Syst. Eng . Vol. 3, No. 1 (2026) 13–31.

Paper Type: Original Article

Combining Resampling Methods, Multi Criteria Decision-Making and Clustering Analysis for Diabetes Detection in Imbalanced Data

Ali Ommi[1,*], Abbas Foroozanfar[2]

[1] Department of Industrial Engineering, Sharif University of Technology; ali.ommi@alum.sharif.edu; forouzanfar.abbas@ie.sharif.edu.

Citation:

Received: 10 December 2025 *Revised: 01 February 2026* *Accepted: 16 March 2026*	Ommi, A., & Foroozanfar, A. (2026). Combining resampling methods, multi criteria decision-making and clustering analysis for diabetes detection in imbalanced data. *Annals of healthcare systems engineering, 3*(1), PP.

Abstract

In this study, the challenges of classifying imbalanced datasets—particularly in medical applications such as diabetes detection—are investigated. The research evaluates the impact of various resampling techniques, including both oversampling and undersampling methods, on the performance of classification models. By comprehensively combining four oversampling and four undersampling approaches within a multicriteria decision-making framework for criterion weighting and ranking, the study proposes an integrated and practical framework for selecting optimal resampling strategies for diabetes detection using the large BRFSS dataset (Behavioral Risk Factor Surveillance System). Machine learning algorithms such as XGBoost and the Support Vector Machine (SVM) were employed and their performance assessed under different resampling regimes. Results show that, on average, sensitivity improved across all resampling methods, with a mean increase of 87.32%, an improvement that was most pronounced for XGBoost. The F1-score likewise exhibited substantial gains across all methods, with SVM contributing a relatively larger share to the F1-score improvements. Although AUC showed little change, the findings indicate a clear enhancement in the models' ability to detect the minority class (individuals with diabetes). To identify the best resampling approaches, a multicriteria decision-making (MCDM) procedure was applied, using the Analytic Hierarchy Process (AHP) for criterion weighting and MAIRCA for ranking and prioritizing the classification and resampling methods. In addition to the multicriteria ranking, an unsupervised clustering analysis based on the K means algorithm was conducted on the resampling–classifier combinations to further explore similarities and differences in their overall performance profiles. The optimal number of clusters was determined using the silhouette coefficient, leading to a partition that revealed distinct groups of methods characterized by different trade offs among accuracy, precision, sensitivity, F1 score, AUC, and computational cost. The clustering results were consistent with the MAIRCA ranking, with high ranked alternatives forming well separated, high quality clusters, while poorly performing methods were grouped into low performance clusters.

Keywords: Imbalanced data, Oversampling, Undersampling, Machine learning, Classification, Multicriteria decision-making (MCDM), Clustering, K-means

1 | Introduction

In today's world, medical data—particularly in the field of disease diagnosis—are increasingly recognized as a valuable resource for improving treatment processes and prevention efforts. One of the most significant

Corresponding Author: ali.ommi@alum.sharif.eduea

challenges in this field is dealing with imbalanced data [1]. In many chronic and complex diseases, such as diabetes, the number of positive samples (individuals who have the disease) is significantly lower than the number of negative samples (individuals who do not have the disease). This data imbalance can seriously affect the performance of machine learning algorithms and lead to reduced model accuracy, particularly in the accurate identification and diagnosis of patients. For this reason, the use of data enhancement methods and appropriate algorithms is of particular importance.

Another challenge in this area is the high dimensionality of the data, which—especially in medical cases with numerous and complex features—can complicate the modeling process.

In this study, we aim to optimize algorithms for processing high-dimensional and imbalanced medical data by using advanced machine learning techniques such as XGBoost and Support Vector Machines (SVM). These algorithms, due to their ability to handle complex data and identify nonlinear patterns, are particularly well-suited for addressing problems associated with medical data.

Moreover, in this study, to address the problem of imbalanced data, various oversampling and undersampling techniques, along with subsets of these methods, have been employed to improve model performance. Especially in areas such as diabetes diagnosis, where high sensitivity in detection and prevention is critical, the evaluation and optimal use of these techniques can represent an important step toward improving classification accuracy and clinical decision-making.

Finally, in order to select the most effective methods and models, Multi-Criteria Decision-Making (MCDM) techniques have been employed. The Analytic Hierarchy Process (AHP) was used to assign weights to the different criteria, and Multi Attributive Ideal-Real Comparative Analysis (MAIRCA) was employed to rank the models and select the best options. These approaches represent a significant innovation, especially in medical fields that require precise and rapid decision-making. This study seeks to present a comprehensive and optimized framework for improving the performance of classification models on medical data and can make a substantial contribution to the development of more accurate tools for disease screening and diagnosis in clinical settings.

To clarify the objectives of this study, three main Research Questions (RQs) are proposed:

RQ1. How do different resampling methods (oversampling and undersampling) affect the performance metrics of classification models (such as sensitivity, accuracy, F1-score, AUC, and run time) in diabetes diagnosis?

RQ2. Which combination of classification algorithms (e.g., XGBoost, SVM) and resampling methods provides the best balance between improving sensitivity and maintaining overall accuracy?

RQ3. Can the multi-criteria decision-making approach (using AHP for criteria weighting and MAIRCA for ranking) efficiently provide a practical and reliable solution for prioritizing resampling and classification methods, and how sensitive are the results to the parameters of synthetic data generation and their generalizability to other datasets?

The remainder of the paper is organized into five sections as follows: Section 2 reviews the related literature; Section 3 presents the materials and methods (dataset, preprocessing, resampling methods, algorithms, and evaluation metrics); Section 4 reports the experimental results and numerical analysis; and Section 5 includes the discussion, conclusions, and directions for future research.

2| Literature Review

2.1| Resampling Methods in Classification Models

In classification problems involving imbalanced data—particularly in the medical domain—the use of resampling methods plays a crucial role in improving the performance of machine learning models. By

adjusting class distributions, these methods enhance Precision, Recall, and the ability for early disease detection. In the following, several studies conducted in this area are reviewed.

Burnaev et al. [2] examined three classical resampling methods—oversampling, undersampling, and SMOTE—across more than 1,000 datasets. Their results showed that the success of each technique depends on the appropriate selection of parameters and the classification model, such that in some cases, not applying resampling yielded better performance than applying it improperly. The findings of this foundational study paved the way for more specialized analyses that investigated the more precise effects of specific resampling methods in medical applications. In this context, Celik [3], focusing on the diagnosis of pneumonia from X-ray images, compared the Adaptive Synthetic Sampling (ADASYN) and Random Under-Sampling (RUS) methods and demonstrated that ADASYN, by increasing minority-class samples, improved the Precision of the SVM algorithm to 98.388%, whereas the RUS method resulted in lower performance with an Area Under Curve (AUC) (area under the ROC curve) of 97.737%. This performance difference is also consistent with the findings of Afzal et al. [4], who, in the context of software defect prediction, confirmed that model performance depends on the correct choice of the type and level of resampling. The positive impact of oversampling techniques on imbalanced classification problems in medical data is also evident in the study by Gurcan and Soylu [5]. They found that the use of Synthetic Minority Oversampling Technique (SMOTE) and ADASYN led to significant improvements in the F1-score and AUC metrics across most of the applied algorithms, whereas data removal approaches—such as undersampling methods—resulted in a reduction in precision in some cases. These findings were also confirmed by Ghorbani and Ghousi [6]. In the context of predicting students' academic performance using a decision tree algorithm, they showed that the SMOTE method achieved higher precision of 88.4% compared with undersampling methods (81.2%), indicating the superiority of augmentation-based approaches for imbalanced data. Continuing this line of research, Saputra et al. [7] examined the performance differences between Random Over-Sampling (ROS) and RUS on the SVM algorithm in sentiment analysis. Similar to previous findings, they demonstrated the significant superiority of oversampling methods: the Precision of the ROS method reached 85% (at $k = 10$), whereas the RUS method achieved only 75% Precision. This performance gap was attributed to the potential removal of important samples from the majority class when using undersampling. However, the study by Welvaars et al. [8], by incorporating the BSMOTE method into their analyses, showed that more intelligent use of hybrid techniques—such as combining SMOTE with XGBoost—can further enhance performance, achieving precision of up to 97.3% and an AUC of 98.6%. These results are particularly noteworthy in terms of diagnostic quality for imbalanced medical datasets. Carvalho et al. [9], through a comprehensive review of resampling techniques such as SMOTE, ADASYN, ENN, and Tomek Links, emphasized that selecting an appropriate method should be based on an understanding of the type of imbalance and the algorithm in use. They reported that combining SMOTE with Edited Nearest Neighbors (ENN) improved the AUC metric in decision tree–based models to as high as 96%.

2.2| Performance Evaluation of Machine Learning Models Using Multi-Criteria Decision-Making

Over the past several decades, machine learning has become one of the key tools in many scientific and industrial fields. One of the main challenges in using these models is selecting the most suitable one from among a set of algorithms that are evaluated based on various performance criteria. Since different algorithms may compete with one another across various performance criteria—such as precision, training time, and computational cost—the use of MCDM methods has been recognized as an effective approach for comparing models and selecting the best one.

Kou et al. [10] investigated the selection of classification algorithms using five MCDM methods. In this study, the algorithms were ranked using methods such as TOPSIS, ELECTRE III, Grey Analysis, VIKOR, and PROMETHEE II. This study shows that each of these methods may yield different results in algorithm evaluation due to the use of different criteria. Therefore, to reduce the discrepancies among these rankings,

they employed Spearman's rank correlation coefficient and proposed a hybrid approach to improve agreement in algorithm selection. Alqaysi et al. [11] investigated hybrid models for autism diagnosis using medical and socio-demographic features and combining them with MCDM methods. In this study, three feature selection techniques (ReF, IG, and Chi2) and five machine learning algorithms (Decision Tree, SVM, KNN, Naive Bayes, and AdaBoost) were used to develop 15 hybrid models. The models were evaluated using the FDOSM method. The results indicate that the best models, including the ReF Decision Tree and Chi2 Decision Tree, demonstrate outstanding performance in diagnosing autism severity (mild, moderate, and severe), achieving high precision along with low training and testing times. Song and Peng [12] proposed an MCDM-based approach for evaluating imbalanced classification models in financial risk prediction. Using the TOPSIS technique, this method combined six evaluation criteria—G-mean, F-measure, AUC, FP rate, FN rate, and time—and showed that SMOTE-based methods, particularly the combination of SMOTE with the C4.5 and MLP algorithms, outperform other approaches. This study emphasizes the importance of using multiple criteria in the evaluation of imbalanced models. Akinsola et al. [13] investigated the selection of the best machine learning algorithms using MCDM. In this study, seven machine learning algorithms were evaluated based on criteria such as Accuracy, Kappa Statistic, True Positive Rate, and True Negative Rate, and the FAHP and TOPSIS methods were used to weight the criteria and rank the algorithms. The results indicate that the logistic regression algorithm achieved the best performance and had the highest Kappa statistic. Kumar and Kaur [14] proposed an MCDM-based framework for selecting the best machine learning techniques in diabetes prediction. In this study, three decision-making methods—WSM, TOPSIS, and VIKOR—were used to rank the models based on various performance criteria such as precision, recall, and the performance characteristic. Then, a hybrid approach called RPM was used to integrate the rankings and determine the final rank. Experimental results on the Pima dataset showed that the logistic regression model was identified as the best model for diabetes prediction. Moreover, the Bayesian test confirmed that the LR model performs significantly better than the other models. Das et al. [15] evaluated 27 screening formulas and 13 machine learning algorithms for identifying the beta-thalassemia trait in Indian pregnant women. In this study, in addition to using traditional criteria such as recall, precision, and AUC, two MCDM methods—SECA and TOPSIS—were employed to rank the models and select the best screening approach. The results showed that the ELM and GBC algorithms achieved the best performance according to the evaluation criteria.

2.3| Research Gap and Innovation

In this study, significant research gaps in improving clustering and classification algorithms for imbalanced data were identified and addressed. In high-dimensional data (which in this study include 21 features), the selection of XGBoost and SVM was evaluated as particularly appropriate due to these model's ability to handle complex data and identify nonlinear patterns, especially in high-dimensional settings. One of the innovative aspects of this study is its focus on improving classification performance in imbalanced data. To address this problem, two main strategies, oversampling and undersampling, were employed as key approaches. In addition, four subset methods from each of these strategies were implemented to improve classification results. Particularly in the context of diagnosing diseases such as diabetes, where recall is critical, the application of these techniques for improving the performance of classification models has not yet been comprehensively investigated. This aspect is considered one of the innovative contributions of the present research. To achieve a more accurate evaluation of model performance and to identify the most effective approach for improving classification, MCDM techniques were employed. To weight the different criteria and ensure greater precision in the evaluation, the AHP method was applied based on the judgments of domain experts. Finally, to optimally rank and select the best method, the novel MAIRCA approach was employed, which proved particularly effective in the process of criteria ranking and optimal alternative selection. This novel approach can have broad applications not only for diabetes data but also in other classification domains involving imbalanced datasets.

3| Problem Statement

In medical applications, particularly in the diagnosis of diseases such as diabetes, class imbalance in datasets is considered one of the fundamental challenges. In the data examined in this study, 383,402 samples correspond to No-Diabetes individuals (including healthy and Pre-diabetes subjects), while only 57,256 samples correspond to individuals with diabetes. In other words, the size of the majority class is approximately seven times that of the minority class, indicating the presence of severe class imbalance in the dataset. Such imbalance can bias classification models toward the majority class and, as a result, severely weaken the model's ability to accurately identify diabetic cases (the minority class). This issue is of even greater importance in the medical domain, where misdiagnosis or delayed diagnosis can have irreversible consequences for patient's health. Therefore, the main objective of this study is to apply and compare various resampling methods in order to address the class imbalance problem and improve the performance of classification models in diabetes diagnosis.

The overall workflow of the present study, from data preparation to the evaluation and final ranking of resampling methods, is illustrated in Figure 1.

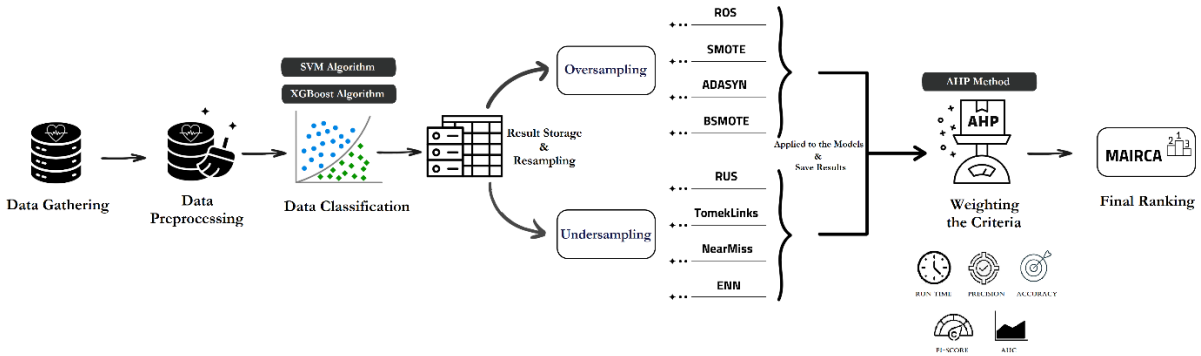

Fig. 1. The proposed approach of the present study for improving classification models on imbalanced data

Given the critical importance of selecting an appropriate resampling method to improve the performance of classification models, and considering the complexity of decision-making when multiple evaluation criteria are involved, this study employs MCDM techniques to enable a more comprehensive and precise analysis. Specifically, the AHP method is employed to determine the weights of the evaluation criteria, and subsequently, the MAIRCA method is used for the final ranking of the resampling methods. Additionally, K-means clustering is applied to group the resampling methods based on their performance similarities, providing further insight into the relationships among the evaluated approaches.

3.1| Resampling Methods

To overcome data imbalance, the development of specialized strategies and techniques for data analysis and modeling processes is essential. Resampling methods and balancing techniques play a vital role in this context. These techniques are implemented to ensure better representation of minority classes in datasets and to support classification algorithms in producing fairer and more accurate results. Resampling methods encompass various strategies, such as reducing majority-class samples (undersampling) and increasing minority-class samples (oversampling). These approaches facilitate dataset balancing and allow classification algorithms to produce more accurate and reliable results.

3.1.1 | Oversampling Methods

In oversampling approaches, the weight of the minority class is increased by duplicating existing samples or generating new instances from the minority class. Various oversampling methods exist; moreover, it is

noteworthy that oversampling approaches are generally applied more frequently than other strategies. Some common methods in this category include:

1. ROS method: This method balances class distributions by randomly duplicating samples from the minority class, without introducing new diversity into the data [5].

2. SMOTE method: This method is a statistical technique that increases the number of minority samples in datasets by generating new samples [5]. It is worth noting that the more precise discussions and mathematical formulation of the SMOTE algorithm have been comprehensively examined in the research by Elreedy et al. [16].

3. ADASYN method: This method adaptively and balancedly generates synthetic data by placing greater emphasis on samples that are harder to learn. The mathematical discussions and precise formulation of the ADASYN algorithm have been fully presented in the research by He et al. [17].

4. BSMOTE method: This method is an improved variant of the SMOTE technique that identifies borderline samples (i.e., samples close to the other class) and generates new synthetic instances specifically for them. The theoretical details and mathematical structure of the BSMOTE method, which aims to improve classification Precision by focusing on minority samples near the decision boundary, have been presented in detail in the study by Han et al. [18].

When selecting among these methods, it is essential to consider the degree of imbalance, the nature of the data, and the model's complexity.

3.1.2 | Undersampling Methods

This category of approaches is considered among the simplest strategies for handling imbalanced data. The performance of these methods is primarily based on the targeted removal of majority class samples, such that reducing the number of samples in this class leads to a relative balance in the data distribution.

1. RUS method: This method creates a relative balance in class distributions by randomly removing samples from the majority class.

2. Tomek Links method: In this approach, pairs of samples from different classes, known as Tomek links, are identified, and the sample belonging to the majority class is removed from the dataset. For a complete explanation of the theoretical foundations and the related mathematical formulations of this method, one may refer to Tomek's paper [19].

3. NearMiss method: This method establishes data balance with greater precision by selecting majority class samples that are located close to minority class samples. For a comprehensive explanation of the foundations of this method and its application in reducing data imbalance, one may refer to the study by Yen et al. [20], in which this method is examined as a majority-class reduction sampling technique, emphasizing the selection of nearest-neighbor samples to the minority class.

4. ENN method: In this technique, first introduced and algorithmically analyzed by Wilson [21], majority class samples that are misclassified and do not share the same class label as the majority of their neighbors (based on the k-NN criterion) are identified as noise and removed from the dataset.

Overall, undersampling methods help balance the data distribution through the targeted removal of majority class samples and are regarded as an effective approach for improving the precision of classification models in imbalanced data problems.

3.2| Model Performance Evaluation

In evaluating the performance of machine learning models, the selection of appropriate methods for weighting and ranking evaluation criteria plays a vital role in improving decision-making precision. The use of AHP as a valid method for weighting criteria makes it possible to systematically and structurally calculate the relative importance of each criterion. This is particularly important in situations where different criteria

have varying impacts on model performance. After determining the weights using AHP, the MAIRCA method is employed to rank the models. This method identifies existing discrepancies by comparing ideal and actual evaluations and ranks the models based on how closely their actual performance matches the ideal condition. The combination of these two methods not only increases evaluation precision, but also makes the decision-making process more transparent and reliable, as both approaches comprehensively and rigorously analyze the criteria and assess model performance.

3.2.1 | AHP Method

The AHP method is widely used for weighting criteria in complex decision-making problems. This method determines the relative importance of each criterion with respect to the others through pairwise comparative analysis of the criteria. In this process, the decision maker (DM) compares each pair of criteria in terms of importance and assigns specific weight values to them. AHP enables the calculation of precise and logical weights for each criterion based on the DM's priorities, without the need for complex mathematical formulas. The main steps of the AHP method are as follows:

I. Structuring the decision problem into a hierarchical model consisting of the overall goal, evaluation criteria, and alternatives.

II. Constructing the pairwise comparison matrix of the criteria using Saaty's nine-point scale to express the relative importance between each pair of criteria.

III. Calculating the normalized comparison matrix and deriving the priority weights (eigenvector) for each criterion.

IV. Evaluating the consistency of the pairwise comparisons by computing the Consistency Index (CI) and Consistency Ratio (CR). If the CR is less than 0.1, the judgments are considered consistent.

V. Determining the final weights of the criteria based on the normalized priority vector.

3.2.2 | MAIRCA Method

To rank the classification methods and improve their performance on imbalanced data, the MAIRCA method will be used; the steps of this method are presented below [22].

Step 1: The formation of the decision matrix (X)

At this stage, the criterion values are established. These values are defined such that i represents the methods $(1, \dots, m)$ and j represents the criteria $(1, \dots, n)$. The initial decision matrix is presented in Equation (1).

$$X = \begin{bmatrix} x_{11} & x_{12} & \cdots & x_{1n} \\ x_{21} & x_{22} & \cdots & x_{2n} \\ \cdots & \cdots & \cdots & \cdots \\ x_{m1} & x_{m2} & \cdots & x_{mn} \end{bmatrix} \quad (1)$$

Step 2: Determination of preferences according choice alternatives P_{A_i}

At this stage, the decision-maker has no specific preference for any of the proposed alternatives and adopts a neutral stance toward all of them, as expressed in Equation (2). However, this assumption is not mandatory, and priorities can be assigned to each method if desired.

$$P_{A_i} = \frac{1}{m} \quad i = 1,2, \dots, m; \sum_{i=1}^{m} P_{A_i} = 1 \quad (2)$$

Step 3: Matrix elements calculation of theoretical ponders (T_P)

At this stage, a theoretical evaluation is calculated for each alternative based on the criterion weights and the preferences of the alternatives, as shown in Equation (3), where W_n denotes the weight of the $n-th$ criterion.

$$T_p = \begin{bmatrix} P_{A_1}W_1 & P_{A_1}W_2 & \cdots & P_{A_1}W_n \\ P_{A_2}W_1 & P_{A_2}W_2 & \cdots & P_{A_2}W_n \\ \vdots & \vdots & \ddots & \vdots \\ P_{A_m}W_1 & P_{A_m}W_2 & \cdots & P_{A_m}W_n \end{bmatrix} = \begin{bmatrix} t_{p11} & t_{p12} & \cdots & t_{p1n} \\ t_{p21} & t_{p22} & \cdots & t_{p2n} \\ \vdots & \vdots & \ddots & \vdots \\ t_{pm1} & t_{pm2} & \cdots & t_{pmn} \end{bmatrix} \quad (3)$$

Step 4: Determination of the matrix elements of actual ponders (T_r)

At this stage, the actual evaluation for each alternative is calculated; for benefit (positive) criteria it is given in Equation (4), and for cost (negative) criteria it is presented in Equation (5).

$$t_{rij} = t_{pij}\left(\frac{x_{ij}-x_i^-}{x_i^+-x_i^-}\right) \quad (4)$$

$$t_{rij} = t_{pij}\left(\frac{x_{ij}-x_i^+}{x_i^--x_i^+}\right) \quad (5)$$

Step 5: Calculation of the matrix of the total gap (G)

The gap matrix is calculated as the difference between the ideal evaluations and the actual evaluations for each alternative, as presented in Equation (6).

$$G = T_p - T_r = \begin{bmatrix} t_{p11}-t_{r11} & t_{p12}-t_{r12} & \cdots & t_{p1n}-t_{r1n} \\ t_{p21}-t_{r21} & t_{p22}-t_{r22} & \cdots & t_{p2n}-t_{r2n} \\ \vdots & \vdots & \ddots & \vdots \\ t_{pm1}-t_{rm1} & t_{pm2}-t_{rm2} & \cdots & t_{pmn}-t_{rmn} \end{bmatrix} = \begin{bmatrix} g_{11} & g_{12} & \cdots & g_{1n} \\ g_{21} & g_{22} & \cdots & g_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ g_{m1} & g_{m2} & \cdots & g_{mn} \end{bmatrix} \quad (6)$$

Step 6: The calculation of the final values of criterion functions (Q_i) by alternatives

At this stage, the final values for each alternative are calculated using Equation (7), enabling comparison among them. An alternative with a smaller value is considered more desirable.

$$Q_i = \sum_{j=1}^{n} g_{ij} \quad ,\forall i = 1,2,..,m \quad (7)$$

4| Numerical Results

In this paper, the initial classification of the imbalanced data is performed using two widely used and well-established machine learning algorithms, namely SVM and XGBoost. The selection of these two algorithms is motivated by their strong performance, high generalization capability, and extensive use in previous studies, particularly in problems involving imbalanced data. It is worth noting that the primary focus of this study is on comparing and evaluating resampling methods for improving the performance of classification models, rather than on selecting the optimal classification algorithm.

All experiments were conducted on a system equipped with an Intel Core i7-1165G7 processor (quad-core, 2.8 GHz), 16 GB of RAM, and running Windows 10 Enterprise N LTSC (version 21H2). These algorithms were also implemented in the Python 3.11.5 environment, using the Scikit-learn library (version 1.6.1) and xgboost (version 3.0.3). In this study, the default configurations of these two libraries were used for parameter settings during model implementation. The key parameters are summarized in Table 1, which serves as the basis for the subsequent numerical analyses. Only two parameters—the objective function and the number of estimators (n_estimators)—were manually specified based on empirical experience and in accordance with the requirements and nature of the present study. These settings were applied to run the models in both scenarios: without resampling and with resampling (based on the two introduced approaches).

Table 1. Modeling parameters used in the XGBoost and SVM algorithms.

SVM		XGBoost	
Parameter	**Value**	**Parameter**	**Value**
Penalty coefficient	1	Objective parameter	binary:logistic
Kernel function	RBF	Number of estimators	100
Polynomial degree	ignore	Learning rate	0.1
Kernel coefficient (gamma)	Based on the number of features and variance of input data	Maximum depth	6
Independent term	0	Minimum sum of instance weight required in a child node	1
Shrinking parameter	active	Gamma	0
Probability parameter	inactive	Subsample ratio of the training instances	1 (Sampling the entire dataset)
Stop threshold	0.001	Column subsampling parameter	1 (Sampling from the entire data column)
Kernel cache size	200 MB	Lambda	1 (L2 Regularization)
Class weighting parameter	inactive (Equal weight for all classes)	Alpha	0 (L1 Regularization)
Maximum number of iterations	-1 (infinite)	Class weight balancing parameter	1 (Equal weight for all classes)
Decision function shape	ignore	Global bias	0.5 (inactive)
		Booster type	gbtree

4.1| Data Description

The Behavioral Risk Factor Surveillance System (BRFSS) is a health-related telephone survey conducted annually by the CDC, collecting data from more than 400,000 Americans on risk behaviors and chronic diseases. The dataset used in this study contains 441,455 records and 330 features, extracted from the 2015 version available on the Kaggle website. Based on prior research on diabetes and chronic diseases, 21 features and one target variable were selected to enable a more focused and accurate analysis of the influential factors. This selection was made due to their direct relevance to the factors influencing diabetes and chronic diseases, in order to provide a more precise analysis of the contributing factors. It should be noted that after performing the preprocessing operations, the number of records is reduced to 440,658.

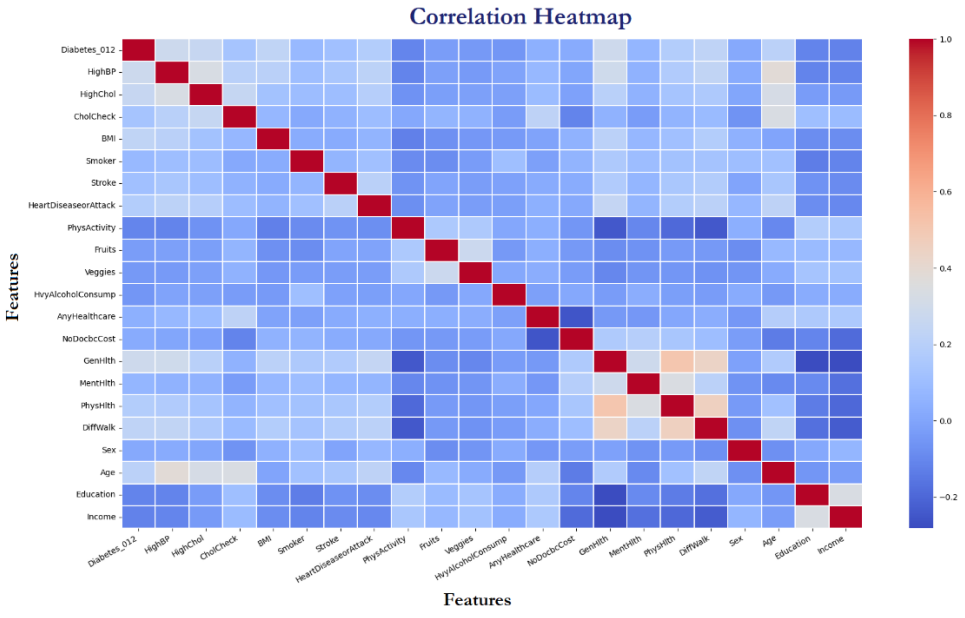

Fig. 2. Correlation analysis between variables

Based on the analysis of the correlation matrix and the heatmap shown in Figure 2, the features "General Health" (GenHlth) and "Physical Health" (PhysHlth) exhibited the strongest positive correlation with the target variable (diabetes diagnosis), whereas the feature "Sex" showed the weakest association. These findings confirm the decisive role of overall health status and mental health in diabetes diagnosis compared to demographic factors.

Finally, after performing the necessary preprocessing steps, the dataset was randomly divided into training (80%) and testing (20%) subsets while preserving the class distribution. These two subsets were subsequently used as inputs to the machine learning algorithms.

4.2| Results of Model Training Using the 8 Methods

In this section, the results obtained from applying the eight resampling methods to the classified data are presented in Table 2. These results were evaluated and compared using five performance criteria, including Accuracy, Precision, Recall, F1-score, AUC and run time (in seconds). In this regard, the Accuracy metric is used to evaluate the overall performance of the models, while the Precision and Recall metrics are employed to assess the model's ability to predict the minority class (diabetic people). As shown in Table 3, the baseline models for data classification perform well in predicting the majority class; however, they exhibit weaknesses in predicting the minority class, as evidenced by the low values of the Recall and F1-score metrics. Based on the values reported in Table 2, the use of resampling methods leads to improvements in the Recall and F1-score metrics, while the AUC criterion does not exhibit significant changes.

Table 2. Results obtained from implementing resampling methods on the classification models

Method	Accuracy	Precision	Recall	F1	AUC	Run Time
XGBoost-ROS	0.76	0.33	0.81	0.47	0.86	02.38
XGBoost-SMOTE	0.88	0.59	0.30	0.40	0.86	06.97
XGBoost-ADASYN	0.88	0.60	0.30	0.39	0.86	28.03
XGBoost-BSMOTE	0.88	0.59	0.30	0.40	0.86	23.13
XGBoost-RUS	0.75	0.32	0.81	0.46	0.86	00.73
XGBoost-ENN	0.85	0.44	0.59	0.50	0.86	119.59
XGBoost-TomekLinks	0.88	0.59	0.32	0.41	0.86	136.21
XGBoost-NearMiss	0.40	0.16	0.85	0.27	0.63	16.78
SVM-ROS	0.72	0.29	0.85	0.44	0.83	4730.28
SVM-SMOTE	0.73	0.30	0.82	0.44	0.84	3123.84
SVM-ADASYN	0.70	0.29	0.86	0.43	0.83	5084.55
SVM-BSMOTE	0.73	0.30	0.82	0.44	0.84	5010.92
SVM-RUS	0.72	0.29	0.84	0.44	0.84	237.78
SVM-ENN	0.86	0.45	0.45	0.45	0.82	370.64
SVM-TomekLinks	0.87	0.65	0.06	0.11	0.76	601.33
SVM-NearMiss	0.46	0.17	0.81	0.28	0.67	82.00

Table 3. Baseline results without implementing resampling methods

Method	Class	Precision	Recall	F1	Accuracy	AUC	Run Time
	0	0.87	1.00	0.93	0.87	0.84	2647.85
	1	0.66	0.00	0.01	0.87	0.84	2647.85
	0	0.90	0.98	0.94	0.88	0.86	2.82
	1	0.62	0.26	0.37	0.88	0.86	2.82

Based on the obtained results, for both the SVM and XGBoost algorithms, it is observed that applying resampling methods generally leads to a decrease in the Accuracy metric, except for the TomekLinks approach, for which the Accuracy remains comparable to the baseline results. The Precision metric decreases across all methods; however, the Recall metric, which is of particular importance in this study, shows a substantial increase for all methods, except for the XGBoost/SVM–TomekLinks approach, for which the changes are more limited. Furthermore, the F1-score metric shows a significant improvement across all methods, while the AUC metric remains approximately unchanged. Finally, the Run Time in the SVM models is significantly higher, indicating a greater computational time requirement for training these models. In an overall comparison, applying resampling methods leads to improved model performance in predicting the minority class (diabetic individuals) by increasing Recall and F1-score, which can contribute to enhancing the model's effectiveness in classifying imbalanced data.

4.3| Comparison and Evaluation of Method Performance

In this section, the final results of the multi-criteria analysis using the MAIRCA method are presented. Table 4 presents the pairwise comparison matrix used for weighting the criteria, which was completed by an expert in this field. In Table 5, following the calculations performed on the pairwise comparison matrix, the final weight for each criterion has been obtained.

Table 4. Pairwise Comparison Matrix

	Accuracy	Precision	Recall	F1	AUC	Run Time
Accuracy	1.00	2.00	1.00	1.00	2.00	4.00
Precision	0.50	1.00	1.00	1.00	1.00	3.00
Recall	1.00	1.00	1.00	2.00	2.00	5.00
F1-Score	1.00	1.00	0.50	1.00	2.00	4.00
AUC	0.50	1.00	0.50	0.50	1.00	3.00
Run Time	0.25	0.33	0.20	0.25	0.33	1.00

Table 5. Final weights of the criteria based on the AHP metho

Criteria	Accuracy	Precision	Recall	F1-Score	AUC	Run Time
Weight	0.22	0.16	0.25	0.19	0.13	0.05

Table 6 presents the final results of implementing the MAIRCA method; in this table, the final values of the criteria function have been calculated for each of the 16 alternatives, and based on these values, the final ranking of the alternatives is provided. This ranking enables a comprehensive comparison and precise prioritization of the alternatives and is used as a basis for decision-making to improve the performance of imbalanced data classification models.

Table 6. Results of algorithm ranking using the MAIRCA method

Method	Q_i	Rank	Method	Q_i	Rank
XGBoost-ENN	0.0105	1	SVM-ENN	0.0159	9
XGBoost-ROS	0.0118	2	SVM-SMOTE	0.0168	10
XGBoost-RUS	0.0127	3	SVM-BSMOTE	0.0179	11
XGBoost-TomekLinks	0.0145	4	SVM-ROS	0.0180	12
SVM-RUS	0.0151	5	SVM-ADASYN	0.0189	13
XGBoost-SMOTE	0.0153	6	SVM-TomekLinks	0.0318	14
XGBoost-BSMOTE	0.0153	7	SVM-NearMiss	0.0364	15
XGBoost-ADASYN	0.0154	8	XGBoost-NearMiss	0.0390	16

Based on the obtained results, the final ranking of resampling methods using the MAIRCA technique shows that the XGBoost-ENN method achieved the highest rank, indicating its superior performance in combining multiple criteria. In contrast, the XGBoost-NearMiss method holds the lowest rank among the 16 alternatives, indicating its limitations in improving the model's performance.

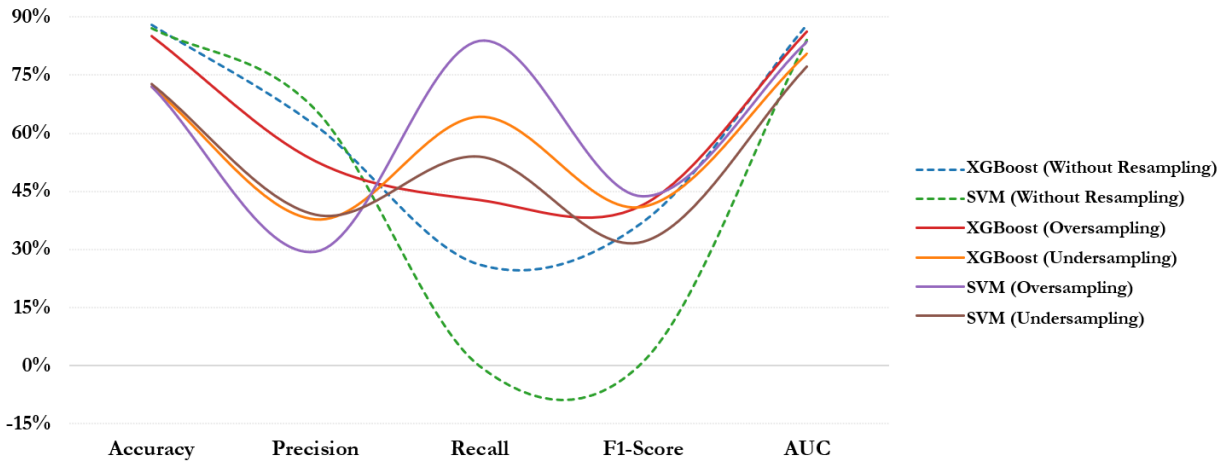

Fig. 3. Comparison of the performance of six implemented model types based on five evaluation criteria, with and without resampling

In Figure 3, the comparative performance chart of the six implemented model types is plotted based on the five selected criteria. Each curve in this chart represents the average (values reported in Table 1) corresponding to each category of models for its specific criterion. Among these, the two dashed curves, representing the states before applying the resampling methods, clearly illustrate the differences in results compared to the states after resampling was applied. As observed in the chart, the Accuracy criterion in the

state without resampling is significantly higher than the corresponding values in the states with resampling (undersampling and oversampling). This phenomenon occurs due to the high number of true positives (TP) under imbalanced data conditions, which allows the model to identify a large number of individuals as healthy. However, when resampling methods are applied and the data are balanced, the number of true negatives (TN) increases, which in turn leads to a more balanced Accuracy criterion. On the other hand, the Precision criterion has decreased in all resampling methods, which is due to the increase in the Recall criterion and the inherent trade-off that exists between these two metrics. The Recall criterion, which holds particular importance in this study, has increased in all resampling approaches compared to the states without resampling. This increase indicates a substantial improvement in the model's ability to identify diabetic samples (the minority class), and in other words, the number of false negatives (FN) has decreased. This is recognized as one of the key achievements of the present study, because reducing the number of diabetic individuals who are incorrectly classified as healthy is of great importance. On the other hand, the F1-score criterion has increased on average due to the significant increase in Recall and the relative decrease in Precision, indicating an improved balance between the Recall and Precision criteria. These changes indicate the success of the resampling techniques in improving the model's performance with respect to the F1-score criterion. Finally, the AUC criterion has remained stable, and no significant changes have been observed in it. On average, the AUC value has remained within a specific range, indicating no noticeable change in the discriminatory power of the models from this perspective, even after applying the resampling techniques.

Table 7. Changes in performance metrics resulting from the use of resampling methods

Criteria	XGB (Oversampling)	XGB (Undersampling)	SVM (Oversampling)	SVM (Undersampling)	Average
Accuracy	-3.41%	-18.18%	-17.24%	-16.38%	-13.80%
Precision	-14.92%	-39.11%	-55.30%	-40.91%	-37.56%
Recall	64.42%	147.12%	83.75%	54.00%	87.32%
F1	12.16%	10.81%	4275.00%	3100.00%	1849.49%
AUC	-2.10%	-8.52%	-0.60%	-8.04%	-4.81%

In Table 7, the percentage changes in model performance compared to the models without applying resampling methods are presented. Overall, the resampling methods have led to noticeable improvements in certain criteria compared to the models without resampling. The "Average" column indicates the overall improvement or lack of improvement across all resampling approaches based on the 5 evaluated criteria. As can be observed, the Accuracy criterion has decreased by an average of 10.33% in oversampling approaches and by an average of 17.28% in undersampling approaches; therefore, overall, it has decreased by an average of 13.80%. Similarly, for the Precision criterion, an average decrease of 37.56% is observed across all approaches. With the application of resampling approaches, the XGBoost algorithm shows an average increase of 105.77% in the Recall criterion compared to the non-resampled state; similarly, the SVM algorithm has also increased by an average of 68.88% in Recall. It should be noted that the F1-score criterion exhibits the largest changes among all evaluated criteria, with a substantial portion of this variation attributable to the SVM algorithm. Finally, based on the obtained results, the AUC criterion has not undergone any noticeable changes and has decreased by only 4.81% on average. It can be stated that applying resampling methods leads to a significant increase in the Recall and F1-score criteria, while, on the other hand, it reduces the Accuracy and Precision metrics.

4.4| Clustering Analysis of Resampling–Classifier Combinations

In addition to the multicriteria ranking obtained with the AHP–MAIRCA framework, an unsupervised clustering analysis was conducted on the sixteen resampling–classifier combinations. The goal of this analysis was twofold. First, we aimed to identify groups of methods with similar overall performance profiles across the evaluation criteria (accuracy, precision, sensitivity, F1-score, AUC, and Run time), rather than focusing solely on the global ranking from MCDM. Second, by grouping the alternatives, we sought to provide a more interpretable structure for decision makers: instead of choosing from sixteen isolated options, they can reason in terms of a small number of homogeneous clusters that reflect distinct trade-offs between performance metrics.

4.4.1 | Determination of the Optimal Number of Clusters

The decision matrix constructed for MAIRCA (Table 2) was used as the feature space for clustering. Each alternative (one resampling–classifier combination) is represented by its aggregated performance scores on the considered criteria. We applied the K-means algorithm to this matrix, varying the number of clusters k and computing the corresponding silhouette coefficient for each solution. The silhouette score quantifies the cohesion and separation of clusters, with higher values indicating more compact and better-separated groups.

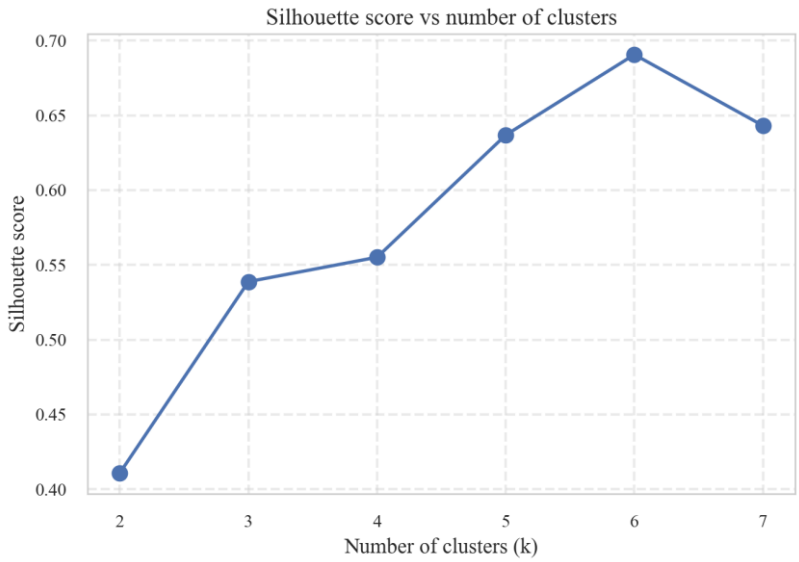

Fig. 4. Silhouette score to find the optimal number od clusters in K-means clustering

As depicted in Figure 4, the average silhouette score increases as k grows from 2 to 6, reaches its maximum at $k = 6$, and then decreases for $k = 7$. Therefore, according to the silhouette criterion, the most appropriate partition for this dataset is obtained with six clusters. This choice provides sufficiently fine granularity to distinguish qualitatively different behaviors, while still preserving good cluster separation.

4.4.2 | Cluster Composition and Interpretation

Using $k = 6$, the following assignments were obtained (Table 8):

Table 8. Results of K-means clustering

Model	Cluster
XGBoost-ENN	1
SVM-ENN	1

XGBoost-NearMiss	2
SVM-NearMiss	2
XGBoost-SMOTE	3
XGBoost-ADASYN	3
XGBoost-BSMOTE	3
XGBoost-TomekLinks	3
SVM-ROS	4
SVM-SMOTE	4
SVM-ADASYN	4
SVM-BSMOTE	4
SVM-TomekLinks	5
XGBoost-ROS	6
XGBoost-RUS	6
SVM-RUS	6

A two-dimensional visualization of these clusters based on principal component analysis (PCA) is presented in Figure 5. Each point corresponds to one resampling–classifier combination, and the plot confirms that the six clusters are well separated and internally coherent.

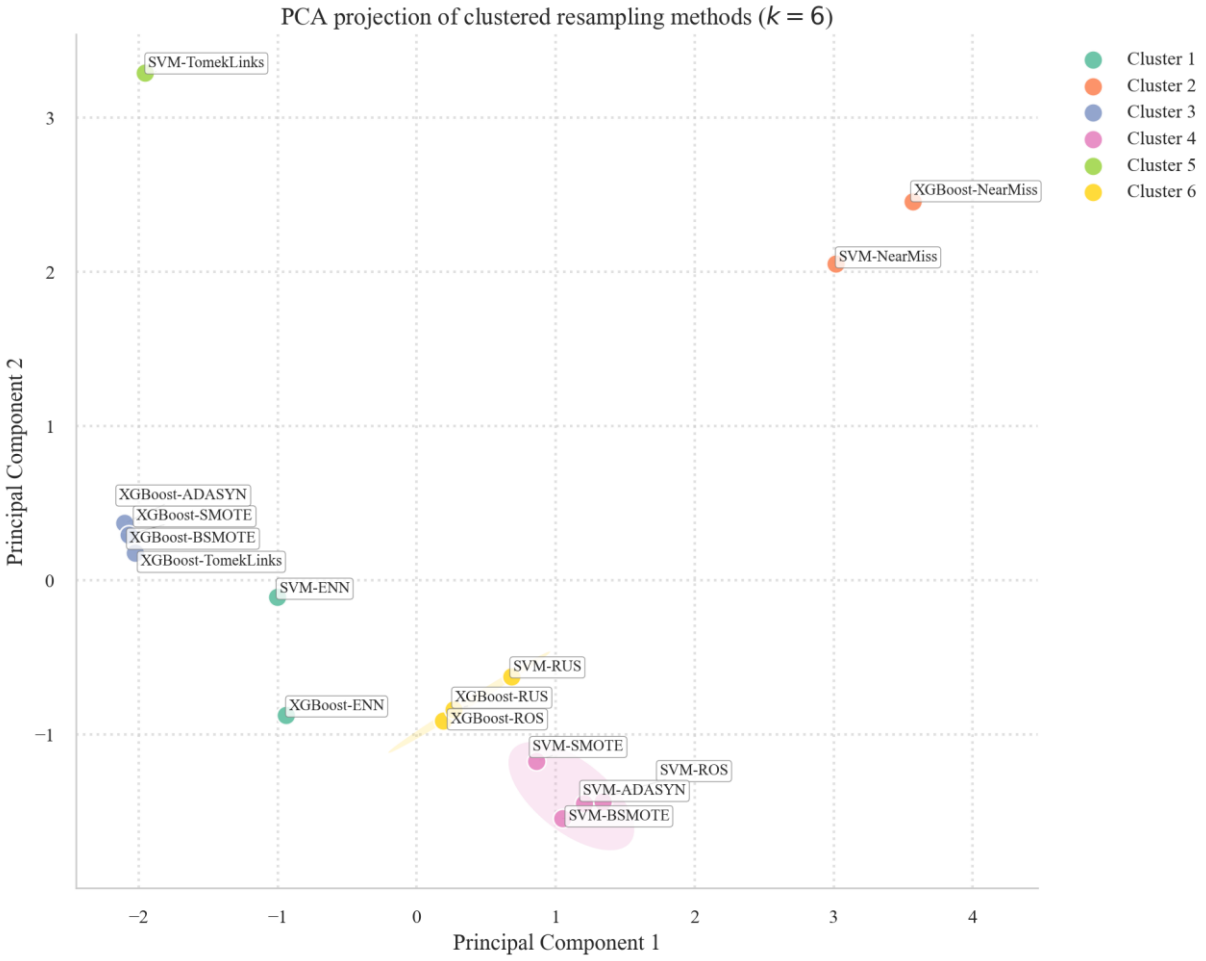

Fig. 5. PCA visualization of K-means clustering

Cluster 1

This cluster groups the two ENN-based methods, which, according to the MAIRCA scores occupy the top positions in the overall ranking, with XGBoost-ENN being the best alternative. Their joint clustering indicates that both variants share a similar performance profile: strong improvements in sensitivity and F1-score, relatively favorable accuracy and precision compared with other resampling strategies, and acceptable computational cost. This cluster can thus be interpreted as the high-performance, well-balanced group, offering the best compromise between minority-class detection and overall efficiency.

Cluster 2

Both NearMiss-based methods fall into the same cluster, which is located at the opposite side of Cluster 1 in the PCA projection. From the MAIRCA results, XGBoost-NearMiss is the lowest-ranked method, and SVM-NearMiss is also among the weakest options. Their clustering reflects a similar and unfavorable pattern: aggressive undersampling leads to large losses in accuracy and precision, while the gains in sensitivity and F1-score are insufficient to compensate for this degradation. Cluster 2 can therefore be characterized as the low-performance group, representing configurations that should be avoided in this application.

Cluster 3

This cluster gathers all oversampling and cleaning techniques combined with XGBoost, except for ENN and NearMiss. These methods typically exhibit considerable improvements in sensitivity (often exceeding 100% increase relative to the baseline) and F1-score, at the expense of moderate decreases in accuracy and precision. Their MAIRCA ranks are generally high, but slightly below XGBoost-ENN. Cluster 3 can thus be viewed as the XGBoost-centric oversampling cluster, which delivers strong minority-class performance with a balanced trade-off, but does not reach the same global desirability as Cluster 1.

Cluster 4

Analogously, Cluster 4 contains all oversampling variants associated with SVM, except ENN and NearMiss. These methods share a characteristic pattern: they substantially increase sensitivity and F1-score (in some cases by more than 80%), while experiencing more pronounced reductions in accuracy and precision than their XGBoost counterparts. In addition, SVM-based models incur higher computational time. Consequently, Cluster 4 can be interpreted as the SVM-based oversampling group, which achieves good minority-class detection but at a higher cost in both accuracy and runtime compared with Cluster 3.

Cluster 5

SVM-TomekLinks forms a singleton cluster, indicating that its performance profile is distinct from the other SVM-based methods. Previous results showed that TomekLinks yields only modest changes relative to the baseline: accuracy remains close to the original model, precision slightly decreases, and sensitivity increases less than in other resampling approaches. The isolated position of SVM-TomekLinks in Figure 5 suggests that this method behaves as a conservative SVM configuration, providing limited improvement in minority-class detection while preserving much of the original accuracy.

Cluster 6

The final cluster combines random oversampling and random undersampling strategies across both classifiers. These methods typically introduce more variance and instability, and their average performance—especially regarding precision and accuracy—is inferior to that of the more sophisticated oversampling techniques (SMOTE, ADASYN, BSMOTE). Nevertheless, they still benefit from the general resampling effect on sensitivity and F1-score. Cluster 6 therefore represents the simple random resampling group, with intermediate overall quality: better than NearMiss (Cluster 2), but clearly dominated by the ENN and advanced oversampling clusters.

4.4.3 | Relationship with the MCDM Ranking

The clustering results are consistent with, and complementary to, the MAIRCA-based ranking. Methods that received high MAIRCA scores (e.g., XGBoost-ENN, SVM-ENN, and the majority of XGBoost-based oversampling techniques) are grouped into clusters that correspond to high- or medium-quality regions of the decision space (Clusters 1 and 3). Conversely, the lowest-ranked method, XGBoost-NearMiss, appears in a cluster (Cluster 2) that is clearly separated and associated with poor performance across multiple criteria.

From a decision-support perspective, the combination of MAIRCA and K-means clustering provides a richer understanding of the alternatives. MAIRCA yields a strict total order that identifies XGBoost-ENN as the best single choice, while clustering reveals that there exists a broader group of methods (Cluster 1 and, to some extent, Cluster 3) with comparable behavior and attractive trade-offs between sensitivity, F1-score, accuracy, and computational cost. This structure is particularly valuable in practice, where secondary considerations (implementation constraints, available computational resources, or interpretability requirements) may lead practitioners to select any method within the same high-quality cluster, rather than solely the top-ranked option.

5| Conclusion

In this study, the effects of various resampling methods on improving the performance of classification models for imbalanced data in diabetes diagnosis were evaluated. In this research, it was demonstrated that applying various resampling methods to the BRFSS dataset led to a substantial improvement in the model's ability to identify the minority class (diabetic people). On average, the Recall of the models increased by 83% compared to the no-resampling scenario, with the largest contribution to this improvement attributed to XGBoost-based combinations. In addition, the F1-score showed a remarkable enhancement, and the analyses indicated that a greater share of the F1-score improvement was associated with SVM models. Meanwhile, overall accuracy decreased in most resampling methods, and the area under the ROC curve (AUC) did not exhibit any significant changes, reflecting a clear trade-off between increased Recall and reduced accuracy. Furthermore, the run time of SVM was noticeably higher. In the multi-criteria decision analysis—using AHP for criteria weighting and MAIRCA for ranking—the XGBoost-ENN combination was identified as the superior option.

In addition to the multicriteria ranking, the clustering analysis provided further insight into the structural relationships among the resampling–classifier combinations by grouping methods with similar performance profiles across multiple evaluation criteria. The clustering results were fully consistent with the MAIRCA rankings, as high-ranked methods were concentrated in well-separated, high-performance clusters, while low-ranked alternatives formed distinct low-quality groups. This complementary analysis enhances the interpretability of the decision-making process by revealing not only the single best method, but also sets of alternatives with comparable and practically acceptable trade-offs between recall, precision, accuracy, and computational cost. Consequently, the combined use of MCDM and clustering offers a more robust and flexible decision-support framework for selecting resampling strategies in imbalanced medical classification problems.

It is recommended that the results be validated through external validation on independent clinical databases and prospective studies. The use of cost-sensitive learning, decision threshold tuning, or ensemble frameworks should be investigated to mitigate the adverse effect of reduced experimental accuracy and to achieve a better balance between recall and precision. Computational optimization—particularly to reduce the run time of SVM—and the exploration of lighter or approximation-based alternative methods for practical applications are also advised. Finally, integrating the proposed approaches into clinical decision

support systems (CDSS) and evaluating clinical outcomes, including the costs of false positives and false negatives, represent valuable directions for future research.

Author Contributaion

Conceptualization, A.O. and A.F.; Methodology, A.O.; Software, A.O.; Validation, A.F.; formal analysis, A.F.; investigation, A.F.; resources, A.O.; data maintenance, A.O.; writing-creating the initial design, A.O.; writing-reviewing and editing, A.F.; visualization, A.O.; monitoring, A.F.

Funding

The authors declare that no funds, grants, or other support were received during the preparation of this manuscript.

Data Availability

Data will be available upon request.

Conflicts of Interest

The authors declare that they have no conflict of interest.

References

[1] Ramyachitra, D., & Manikandan, P. (2014). Imbalanced dataset classification and solutions: A review. International Journal of Computational Business Research, 5(4), 1–29.

[2] Burnaev, E., Erofeev, P., & Papanov, A. (2015). Influence of resampling on accuracy of imbalanced classification. In Eighth International Conference on Machine Vision (ICMV 2015) (pp. 423–427). SPIE.

[3] Celik, A. (2023). Diagnosis of the diseases using resampling methods with machine learning algorithms. Proceedings of the Bulgarian Academy of Sciences, 76, 1065–1076.

[4] Afzal, W., Torkar, R., & Feldt, R. (2012). Resampling methods in software quality classification. International Journal of Software Engineering and Knowledge Engineering, 22(2), 203–223.

[5] Gurcan, F., & Soylu, A. (2024). Learning from imbalanced data: Integration of advanced resampling techniques and machine learning models for enhanced cancer diagnosis and prognosis. Cancers, 16(19), 3417. https://doi.org/10.3390/cancers16193417

[6] Ghorbani, R., & Ghousi, R. (2020). Comparing different resampling methods in predicting students' performance using machine learning techniques. IEEE Access, 8, 67899–67911. https://doi.org/10.1109/ACCESS.2020.2986809

[7] Saputra, A. D., Arifianto, D., & Umilasari, R. (2025). Effect of random under sampling and random over sampling method on SVM performance. Computer and Information Systems Journal, 1(2), 78–86.

[8] Welvaars, K., et al. (2023). Implications of resampling data to address the class imbalance problem (IRCIP): An evaluation of impact on performance between classification algorithms in medical data. JAMIA Open, 6(2).

[9] Carvalho, M., Pinho, A. J., & Brás, S. (2025). Resampling approaches to handle class imbalance: A review from a data perspective. Journal of Big Data, 12(1), 71. https://doi.org/10.1186/s40537-025-00921-5

[10] Kou, G., Lu, Y., Peng, Y., & Shi, Y. (2012). Evaluation of classification algorithms using MCDM and rank correlation. International Journal of Information Technology & Decision Making, 11(1), 197–225. https://doi.org/10.1142/S0219622012500095

[11] Alqaysi, M. E., Albahri, A. S., & Hamid, R. A. (2022). Hybrid diagnosis models for autism patients based on medical and sociodemographic features using machine learning and multicriteria decision-making (MCDM) techniques: An evaluation and benchmarking framework. Computational and Mathematical Methods in Medicine, 2022(1), 9410222. https://doi.org/10.1155/2022/9410222

[12] Song, Y., & Peng, Y. (2019). A MCDM-based evaluation approach for imbalanced classification methods in financial risk prediction. IEEE Access, 7, 84897–84906. https://doi.org/10.1109/ACCESS.2019.2924923

[13] Akinsola, J. E. T., Awodele, O., Kuyoro, S. O., & Kasali, F. A. (2019). Performance evaluation of supervised machine learning algorithms using multi-criteria decision making techniques. In International Conference on Information Technology in Education and Development (ITED) (pp. 17–34). Retrieved from https://www.academiainformationtechnology.org/ited2019/uploads/8135_File_03ITED19041 IEEE Paper Format Performance Evaluation of Supervised Machine Learning Algorithms Using MCDM Techniques NEW (1).pdf

[14] Kumar, A., & Kaur, K. (2024). A novel MCDM-based framework to recommend machine learning techniques for diabetes prediction. International Journal of Engineering and Technology Innovation, 14(1), 29–43. https://doi.org/10.46604/ijeti.2023.11837

[15] Das, R., et al. (2022). Performance analysis of machine learning algorithms and screening formulae for β-thalassemia trait screening of Indian antenatal women. International Journal of Medical Informatics, 167, 104866. https://doi.org/10.1016/j.ijmedinf.2022.104866

[16] Elreedy, D., Atiya, A. F., & Kamalov, F. (2024). A theoretical distribution analysis of synthetic minority oversampling technique (SMOTE) for imbalanced learning. Machine Learning, 113(7), 4903–4923. https://doi.org/10.1007/s10994-023-06459-6

[17] He, H., Bai, Y., Garcia, E. A., & Li, S. (2008). ADASYN: Adaptive synthetic sampling approach for imbalanced learning. In 2008 IEEE International Joint Conference on Neural Networks (IEEE World Congress on Computational Intelligence) (pp. 1322–1328). IEEE. https://doi.org/10.1109/IJCNN.2008.4633969

[18] Han, H., Wang, W.-Y., & Mao, B.-H. (2005). Borderline-SMOTE: A new over-sampling method in imbalanced data sets learning. In International Conference on Intelligent Computing (pp. 878–887). Springer. https://doi.org/10.1007/11538059_91

[19] Tomek, I. (1976). Two modifications of CNN. IEEE Transactions on Systems, Man, and Cybernetics, SMC-6(11), 769–772. https://doi.org/10.1109/TSMC.1976.4309452

[20] Yen, S.-J., & Lee, Y.-S. (2006). Under-sampling approaches for improving prediction of the minority class in an imbalanced dataset. Lecture Notes in Control and Information Sciences, 344, 731–740. https://doi.org/10.1007/11760191_109

[21] Wilson, D. L. (1972). Asymptotic properties of nearest neighbor rules using edited data. IEEE Transactions on Systems, Man, and Cybernetics, SMC-2(3), 408–421. https://doi.org/10.1109/TSMC.1972.4309137

[22] Pamučar, D., Vasin, L., & Lukovac, L. (2014). Selection of railway level crossings for investing in security equipment using hybrid DEMATEL–MARICA model. In XVI International Scientific-Expert Conference on Railway (RAILCON) (pp. 89–92).