Pamukkale Üniversitesi Mühendislik Bilimleri Dergisi
Pamukkale University Journal of Engineering Sciences

Improvement of machine learning-based diabetes diagnosis via resampling techniques

Makine öğrenmesi tabanlı diyabet teşhisinin yeniden örnekleme teknikleri ile iyileştirilmesi

İrem Şenyer Yapıcı[1], Rukiye Arslan[2], Mustafa Alptekin Engin[3]*

[1]Department of Computer Engineering, Zonguldak Bülent Ecevit University, Zonguldak, Türkiye.
senyerirem@beun.edu.tr
[2]Department of Electrical and Electronics Engineering, Zonguldak Bülent Ecevit University, Zonguldak, Türkiye
rukiye.uzun@beun.edu.tr
[3]Department of Electrical and Electronics Engineering, Bayburt University, Bayburt, Türkiye
maengin@bayburt.edu.tr

Received/Geliş Tarihi: 26.11.2024 Revision/Düzeltme Tarihi: 06.08.2025 doi: 10.5505/pajes.2025.52882
Accepted/Kabul Tarihi: 20.08.2025 Research Article/Araştırma Makalesi

Abstract

The objective of this study is to enhance the accuracy of diabetes diagnosis through the utilisation of machine learning techniques and resampling methods. The imbalanced nature of diabetes datasets presents a significant challenge for traditional classification algorithms, which often struggle to accurately predict results. In order to enhance the efficacy of the model, a comparative analysis was conducted to assess the performance of a range of over-sampling and under-sampling techniques, including SMOTE, ADASYN, Borderline SMOTE, SVM SMOTE, Random Under Sampler, Near Miss, One Sided Selection, Neighbourhood Cleaning Rule, Edited Nearest Neighbours, Instance Hardness Threshold, AllKNN and Tomek Links. The aforementioned techniques were then applied to the Decision Tree, Random Forest, K-Nearest Neighbours, AdaBoost, Extra Tree Classifier, and machine learning classifiers, and their performance was evaluated using the accuracy, recall, precision, F-Score, and AUC-ROC performance metrics. The SVMSMOTE resampling technique was identified as the most successful method, achieving 99.06% accuracy when used in combination with the decision tree classifier. The findings demonstrate that the incorporation of resampling techniques markedly enhances diagnostic proficiency and yields more dependable forecasts. This research makes a significant contribution to the field of medical informatics, providing a robust framework for diabetes diagnosis and offering valuable insights into the application of machine learning in healthcare.

Key words: Diabetes diagnosis, Resampling techniques, Imbalanced dataset, Machine learning

Öz

Bu çalışmanın amacı, makine öğrenimi teknikleri ve yeniden örnekleme yöntemlerini kullanarak diyabet teşhisinin doğruluğunu artırmaktır. Diyabet veri setlerinin dengesiz yapısı, sonuçları doğru bir şekilde tahmin etmekte zorlanan geleneksel sınıflandırma algoritmaları için önemli bir zorluk teşkil etmektedir. Modelin etkinliğini artırmak amacıyla, SMOTE, ADASYN, Borderline SMOTE, SVM SMOTE, Random Under Sampler, Near Miss, One Sided Selection, Neighbourhood Cleaning Rule, Edited Nearest Neighbours, Instance Hardness Threshold, AllKNN ve Tomek Links dahil olmak üzere bir dizi aşırı örnekleme ve düşük örnekleme tekniklerinin performansını değerlendirmek için karşılaştırmalı bir analiz yapılmıştır. Yukarıda bahsedilen teknikler daha sonra Karar Ağacı, Rastgele Orman, K-En Yakın Komşular, AdaBoost, Ekstra Ağaç Sınıflandırıcı ve makine öğrenimi sınıflandırıcılarına uygulanmış ve performansları doğruluk, geri çağırma, kesinlik, F-Skoru ve AUC-ROC performans ölçütleri kullanılarak değerlendirilmiştir. SVMSMOTE yeniden örnekleme tekniği, karar ağacı sınıflandırıcısı ile birlikte kullanıldığında %99,06 doğruluk elde ederek en başarılı yöntem olarak belirlenmiştir. Bulgular, yeniden örnekleme tekniklerinin dahil edilmesinin teşhis yeterliliğini önemli ölçüde artırdığını ve daha güvenilir tahminler sağladığını göstermektedir. Bu araştırma, diyabet teşhisi için sağlam bir çerçeve sağlayarak ve makine öğreniminin sağlık hizmetlerinde uygulanmasına ilişkin değerli bilgiler sunarak tıbbi bilişim alanına önemli bir katkıda bulunmaktadır.

Anahtar kelimeler: Diyabet teşhisi, Yeniden örnekleme teknikleri, Dengesiz veri kümesi, Makine öğrenmesi

1 Introduction

Diabetes is a metabolic disorder resulting from insufficient production of insulin or the body's inability to effectively utilize the hormone. This chronic condition causes a sustained elevation in blood glucose levels. It affects millions of people worldwide and, if not properly managed, significantly increases the risk of serious complications. The management of diabetes primarily involves lifestyle modifications, regular monitoring of blood glucose levels, balanced nutrition, and consistent physical activity. Additionally, modern medical interventions, including insulin therapies and oral antidiabetic medications, assist patients in effectively controlling their blood glucose levels. Globally, 537 million adults aged 20-79 are living with diabetes, indicating that one in ten adults is affected by this condition. It is projected that the number of individuals living with diabetes will increase to 643 million by 2030 and further escalate to 783 million by 2045, reflecting a significant upward trend in global prevalence. Over 75% of individuals with diabetes reside in low- and middle-income countries, highlighting the significant burden of the disease in these regions. In 2021, diabetes accounted for 6.7 million deaths globally, which corresponds to a mortality rate of one individual every five seconds. These statistics underscore the substantial global health burden posed by diabetes and the critical need for effective management and prevention strategies to mitigate its impact [1, 2]. In recent years, the use

*Corresponding author/Yazışılan Yazar

of advanced technologies such as machine learning (ML) and data analytics in diabetes management has positively influenced the progression of the disease. The literature increasingly focuses on ML-based approaches for diabetes diagnosis, with numerous studies investigating various algorithms and models to enhance the accuracy and efficiency of early diagnosis and risk prediction [3]. These studies highlight the significant capabilities of ML techniques, including support vector machines (SVM), neural networks (NNs), and ensemble learning (EL), in processing complex datasets and uncovering patterns that may remain undetected using conventional methods. This approach has been instrumental in enhancing diagnostic accuracy and advancing diabetes management strategies. In their 2020 study, Pradhan et al. critically evaluate common data mining techniques for early diabetes prediction, such as Naïve Bayes (NB), Decision Tree (DT), and SVM, highlighting their limitations. They propose an artificial neural network (ANN) model for diabetes detection and classification, which also effectively reduces complications. Validation with the 'PIMA Indian Diabetes' dataset yielded a high accuracy (ACC) of 85.09%, demonstrating strong performance [4]. Maniruzzaman et al. (2020) developed a ML-based system for predicting diabetes. The study employed logistic regression (LR) to identify risk factors including age, education level, body mass index (BMI), blood pressure, and cholesterol levels. To predict diabetes, four classifiers (NB, DT, AdaBoost, and Random Forest (RF)) were utilized and evaluated across three partition protocols. The findings indicated that the combination of LR and RF provided the best performance, achieving an ACC of 94.25% [5]. Daghistani and Alshammari (2020) evaluated the performance of LR and Random Optimization (RO) algorithms in diabetes diagnosis using a dataset obtained from a healthcare institution in Saudi Arabia. Their analysis determined that RO achieved the highest classification ACC at 88% [6]. Shuja et al. (2020) compared the performance of five different ML algorithms in diabetes diagnosis using an imbalanced diabetes dataset obtained from a laboratory in Kashmir. They applied the Synthetic Minority Over-Sampling Technique (SMOTE) to address the dataset imbalance. The analysis revealed that the highest performance was achieved using DT in conjunction with SMOTE [7]. In the study conducted by Butt et al. (2021), a ML-based system is introduced for the early detection and classification of diabetes. The research also presents an Internet of Things (IoT) based system for monitoring blood glucose levels in individuals. The classification task employs RF, multilayer perceptron (MLP), and LR, with the MLP model achieving the highest ACC at 86.08%. Furthermore, in predictive analysis, the long short-term memory (LSTM) model demonstrated notable effectiveness, achieving an ACC of 87.26% [8]. In the study by Chaves and Marques (2021), various data mining techniques were comparatively analyzed for the early diagnosis of diabetes. The research utilized a publicly available dataset consisting of 520 instances, each with 17 attributes. The methods evaluated included NB, NNs, AdaBoost, k-Nearest Neighbors (KNN), RF and SVM. The findings of the study indicated that NNs achieved the highest performance in predicting diabetes, with the proposed model yielding an area under the curve (AUC) of 98.3%, demonstrating notable effectiveness [9]. Kumari et al. (2021) employ an ensemble of ML algorithms to predict diabetes with high ACC. Experiments were conducted using PIMA dataset, incorporating a soft voting classifier (SVC) that combines RF, LR and NB. The proposed methodology was compared against contemporary methods such as AdaBoost, SVM, Bagging, Gradient Boosting (GB), Extreme Gradient Boosting (XGBoost), and Categorical Boosting (CatBoost). The results demonstrate ACC, precision, recall, and F1-score values of 79.04%, 73.48%, 71.45%, and 80.6%, respectively, on the PIMA dataset [10]. In the study by Khanam et al. (2021), data mining, ML algorithms, and NN methods were utilized to predict diabetes using PIMA dataset. Seven ML algorithms were applied to the dataset for diabetes prediction. LR and SVM demonstrated effective performance. Additionally, a NN model with two hidden layers achieved an ACC of 88.6% [11]. Mesquita et al. (2021) examined the performance of ten different ML algorithms combined with six different over-sampling techniques for diabetes diagnosis using the PIMA dataset. Simulation studies revealed that the best result, with an ACC of 83.12%, was achieved using the AdaBoost algorithm in conjunction with SVM-SMOTE [12]. Özlüer Başer et al. (2021) analyzed the performance of six different ML algorithms for detecting diabetic conditions using K-fold cross-validation and SMOTE. The analysis revealed that RF achieved the highest performance with an ACC of 84.78% [13]. Harman (2021) investigated the performance of SVM and NB algorithms on an imbalanced diabetes dataset using SMOTE. The analysis revealed that the highest classification performance, at 90%, was achieved by SVM [14]. In the study by Saxena et al. (2022), various classifiers and feature selection methods were compared to enhance the accuracy of diabetes prediction. The classifiers evaluated included MLP, DT, KNN and RF. Using the PIMA dataset, the accuracies achieved were 77.60% for MLP, 76.07% for DT, 78.58% for KNN, and 79.8% for RF [15]. In the study by Mushtaq et al. (2022), various classifiers (LR, SVM, KNN, GB, NB and RF) were employed to evaluate prediction performance. After applying SMOTE, RF achieved the highest ACC at 80.7%. Additionally, three high-performing models were assessed using a voting algorithm, resulting in the model attaining an ACC of 82.0% on the default dataset and 81.7% on the balanced dataset [16]. Özkan et al. (2022) compared the performance of eight different ML algorithms using two different approaches to detect individuals' diabetic status. The study evaluated the performance of models that utilized statistically and clinically significant features for diabetes diagnosis using 10-fold cross-validation. The analysis revealed that, within both approaches, RF demonstrated superior classification performance compared to other algorithms [17]. Sevli (2022) analyzed the performance of six different ML algorithms on an imbalanced diabetes dataset using fourteen different resampling techniques. The study found that resampling techniques positively impacted classification performance, with the highest ACC of 96.296% achieved when applying the InstanceHardnessThreshold undersampling technique to RF [18]. In the study by Özoğur and Orman (2023), successful methods for addressing issues related to missing values in imbalanced data classification were compared using the PIMA dataset. The results show that the combination of the SMOTEENN algorithm and multiple imputation methods using chained equations achieved an F-score of 91%, outperforming other methods by approximately 9% [19]. Ali et al. (2023) introduced an optimized random forest algorithm (RFWBP) for early diabetes detection, utilizing RF algorithms and feature engineering. After applying data preprocessing and mining techniques, RFWBP achieved accuracies of 95.83% with 5-fold cross-validation and 90.68% without it. The results demonstrate that RFWBP surpasses traditional ML methods [20]. Febrian et al. (2023) conducted a comparative study of KNN and NB algorithms for diabetes prediction. Utilizing supervised ML techniques, the analysis on the PIMA dataset

revealed that NB outperformed KNN. Specifically, NB achieved an average ACC of 76.07%, precision of 73.37%, and recall of 71.37%, whereas KNN achieved an average ACC of 73.33%, precision of 70.25%, and recall of 69.37% [21]. In the study by Khaleel et al. (2023), a model for predicting diabetes onset is proposed. Evaluated using PIMA dataset, the model demonstrated precision rates of 94%, 79%, and 69% for LR, NB, and KNN, respectively [22]. In the study by Modak et al. (2024), various ML techniques and EL methods were employed to predict diabetes. The techniques included LR, SVM, NB and RF, alongside ensemble methods such as XGBoost, LightGBM, CatBoost, AdaBoost, and Bagging. The study found that CatBoost achieved the highest performance with an ACC of 95.4% and an approximate AUC-ROC score of 99%, while XGBoost achieved an ACC of 94.3% and an approximate AUC-ROC score of 98% [23]. In the study by NG et al. (2024), the En-RfRsK model is proposed for predicting diabetes risk. This ensemble approach combines RF, Radial SVM, and KNN. Testing with PIMA dataset demonstrated that the En-RfRsK model achieved an ACC of 88.89%, outperforming existing methods [24].

The PIMA dataset is a frequently referenced resource in existing literature pertaining to the prediction of diabetes. However, this study employs a publicly accessible dataset comprising health records from 130 hospitals across the United States, encompassing a substantial patient cohort (Diabetes 130-US Hospitals). The expanded and more diverse patient profile offered by this dataset enhances the robustness of our findings. However, as is the case with a significant proportion of datasets employed in the medical field, this dataset exhibits an imbalance in the number of samples drawn from different groups, which may impact the reliability of the results. In the literature, a variety of resampling techniques are employed with the objective of reducing the impact of data set imbalances on classification performance. Gaso et al. (2024) predicted early hospitalisations of diabetic patients by examining missing and imbalanced data problems in the preprocessing process with the SMOTE method using the Diabetes 130-US Hospitals database. Among the compared methods, the multilayer deep learning architecture (MDLA) showed the most successful performance with 98% accuracy and 99% recall value when used with SMOTE [25]. In a similar study, Zarghani (2024) compared classical machine learning models with the deep learning-based LSTM model. The LSTM model demonstrated an accuracy of 97.65%, exhibiting sensitivity to time series data. The SHAP analysis emphasised the effect of variables such as the number of laboratory procedures and discharge status on classification performance [26]. Kanu and Khanal (2023) aimed to develop machine learning models to predict hospital readmissions of diabetic patients by analysing the same dataset with big data analytics methods. Using big data tools Hadoop and PySpark, data preprocessing processes were applied to select 23 important features that directly affect patient readmission from an initial set of 50 features. Logistic Regression, Decision Trees and Random Forest algorithms were analysed comparatively; Random Forest algorithm showed superior performance compared to other algorithms with 100% accuracy in training and testing phases [27]. However, within the context of diabetes research, there is a notable absence of comprehensive comparative analyses examining the impact of different techniques on performance outcomes and the extent of this impact. This study examines a range of resampling techniques with the objective of minimising the impact of dataset imbalances on classification performance. The results of these analyses demonstrate the efficacy of classifiers in predicting diabetes risk. Furthermore, this study aims to present more effective methodologies for predicting diabetes by conducting analyses based on more recent and comprehensive datasets, while simultaneously providing a more profound understanding of the impact of resampling and feature selection on classifier performance.

2 Material and methods

The initial phase of the study entailed the implementation of editing operations within the database. Subsequently, a feature selection process was conducted, and all features proceeded through a normalisation stage. Following the execution of diverse resampling operations, the efficacy of various classifiers was evaluated, and the most optimal model was identified. The sequence of stages involved in the process is illustrated in Figure 1.

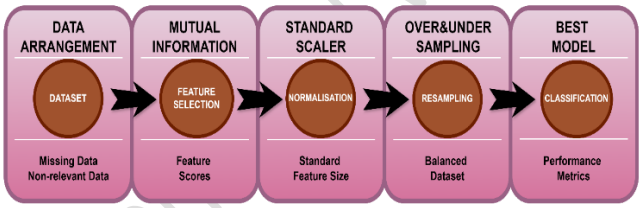


Figure 1. Process stages.

2.1 Dataset preprocessing

In this study, a publicly available dataset from the UCI Machine Learning Repository was utilized. This dataset comprises 10 years (1999–2008) of clinical care data from 130 hospitals across US, provided by the Cerner Corporation (Kansas City, MO) [28]. Initially, the dataset included 55 features. However, variables such as 'patient ID,' 'admission ID,' 'payment code,' and 'admission location' were excluded as they were deemed irrelevant for diabetes classification. Additionally, the 'medical specialty' variable was removed due to a large amount of missing data, and variables with highly imbalanced categories (e.g., repaglinide, nateglinide) were excluded as they contributed no significant information for classification. Expert consultations suggested that body weight could be associated with diabetes [29]. Consequently, the analysed dataset was limited to observations where body weight was recorded, yielding a total of 3,197 patients.

The data preprocessing involved removing duplicate entries based on patient IDs, retaining only the most recent entry for each patient. When multiple class labels were associated with the same patient ID, entries reflecting a diabetes diagnosis were prioritized, and non-relevant entries were excluded. Following this procedure, the dataset was reduced to 24 variables (23 independent variables and one class label) with a total of 2,866 observations. As depicted in Figure 2, the class distribution reveals that 1,961 observations correspond to diabetic individuals (labeled as 1), while 905 represent non-diabetic individuals (labeled as 0).

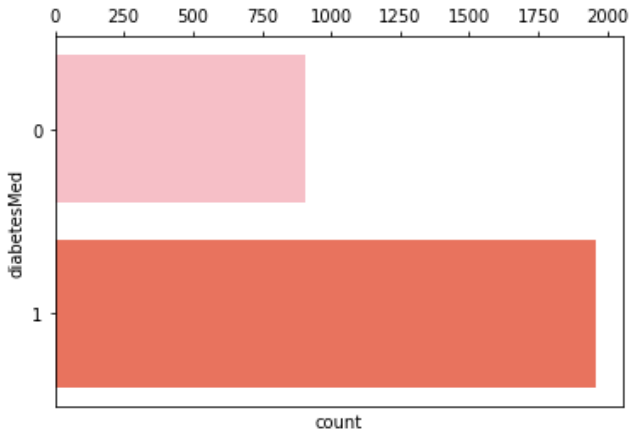


Figure 2. Distribution of diabetic and non-diabetic individuals in the dataset.

Following data preprocessing, the influence of each feature on diabetes classification was assessed through an analysis of dependencies and information gain. Mutual information (MI) was utilized to quantify the relationships between the independent variables and the class label. This method is a commonly employed approach in the field of medical data analysis [30]. Additionally, data scaling was applied using the Standard Scaler to normalize all features, which is essential for improving the consistency and performance of classification algorithms.

2.2 Feature importance assessment

Mutual Information (MI) is a statistical measure that quantifies the amount of information obtained about one variable through another. Mathematically, MI between two variables X (feature) and Y (class label) is defined as in Equation (1).

$$MI(X,Y) = \sum_{x}\sum_{y} p_{XY}(x,y)\log\left(\frac{p_{XY}(x,y)}{p_X(x)p_Y(y)}\right) \quad (1)$$

where $p_{XY}(x,y)$ represents the joint probability distribution of X and Y, $p_X(x)$ and $p_Y(y)$denote the marginal probability distributions of X and Y, respectively. This metric identifies which features contribute the most to the predictive power of the model. Higher MI values indicate a stronger relationship between a feature and the class label, making those features more relevant for classification purposes. In this study, MI was calculated to quantify how much each independent variable contributed to distinguishing between diabetic and non-diabetic individuals. The MI scores for all variables were computed and ranked to determine their significance in the model. As shown in Figure 3, features such as 'change', 'insulin', and 'metformin' exhibit the highest mutual information scores, indicating their significant contribution to the prediction of the target variable.

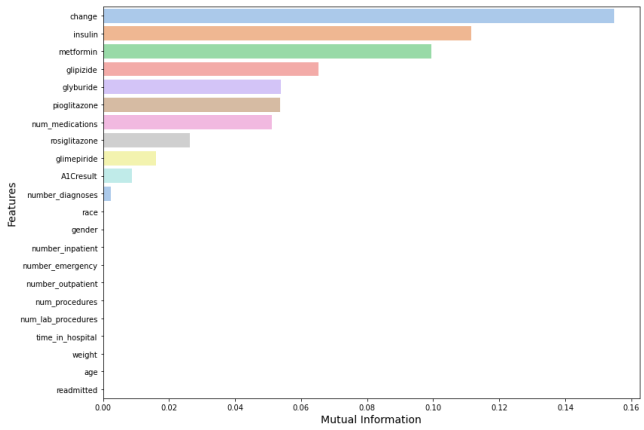


Figure 3. Feature importance ranking based on mutual information.

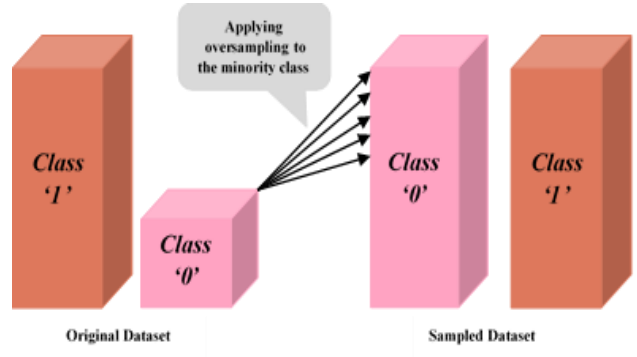
Once the feature importance had been determined using MI, data scaling was achieved using the Standard Scaler for the normalisation process. This is a necessary step to ensure that all features are on the same scale and thus improve the consistency and performance of the classification algorithms. The Standard Scaler method standardizes features by removing the mean and scaling them to unit variance. This ensures that each feature has a mean of 0 and a standard deviation of 1. Where x is the original feature value, μ is the feature mean and σ is the standard deviation, Mathematically, the transformation is given as in Equation (2).

$$z = \frac{x - \mu}{\sigma} \quad (2)$$

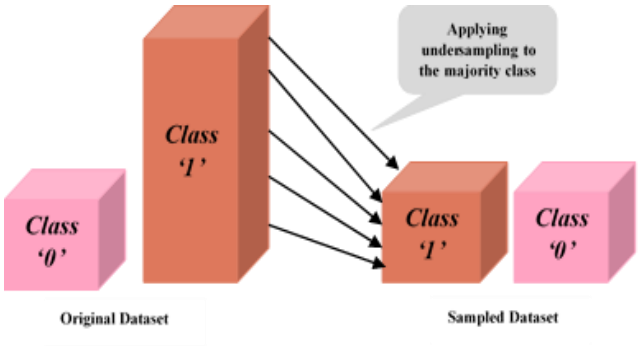
2.3 Resampling techniques

Resampling techniques are strategies developed to address class imbalance in datasets. When class imbalance is present, classification models often disproportionately focus on the majority class. To mitigate this issue and create a more balanced dataset, resampling techniques are employed. These techniques ensure that the model can effectively learn from both classes by either increasing the number of samples from the minority class or reducing the number of samples from the majority class. There are two main approaches to resampling: undersampling and oversampling. Undersampling seeks to achieve class balance by decreasing the number of samples from the majority class, whereas oversampling utilizes various techniques to increase the number of samples from the minority class [31, 32].

Figure 4 is the block diagram visually explaining undersampling and oversampling side by side. The diagram effectively highlights the before and after states of both techniques, demonstrating how the majority and minority classes are adjusted. In this study, the dataset used exhibits a distribution where 37.97% of the individuals are diabetic, while 62.03% are healthy, which could potentially lead the model to overfit on the healthy cases. To address this imbalance and assess its impact on model performance, twelve different resampling techniques were employed in addition to the default classification approach. The oversampling and undersampling methods implemented in this study were outlined in Table 1.



(a)



(b)

Figure 4. Block diagram of oversampling (a) and undersampling (b) techniques.

Table 1. The resampling techniques employed in the study.

Undersampling Techniques	Oversampling Techniques
RandomUnderSampler	SMOTE
NearMiss	ADASYN
OneSidedSelection	BorderlineSMOTE
NeighbourCleaningRule	SVMSMOTE
EditedNearestNeighbours	
InstanceHardnessThreshold	
AIIKNN	
TomekLinks	

2.4 Classifier algorithms

In this study, the ten most highly scoring features were input to the classifiers as a result of feature selection. The selected features were num_medications, A1Cresult, metformin, glimepiride, glipizide, glyburide, pioglitazone, rosiglitazone, insulin and change values. Validation methods are essential to prevent overfitting in machine learning classification models. Therefore, the k-fold cross-validation method, which is widely adopted in the literature, is also used in this study. In this method, the dataset is randomly divided into k parts. Each part is used once as the test set, while the remaining parts are used for training in each iteration [33]. The overall performance of the model is calculated as the average of the results in all iterations. Figure 5 illustrates a schematic representation of the 5-fold cross-validation procedure applied in this study.

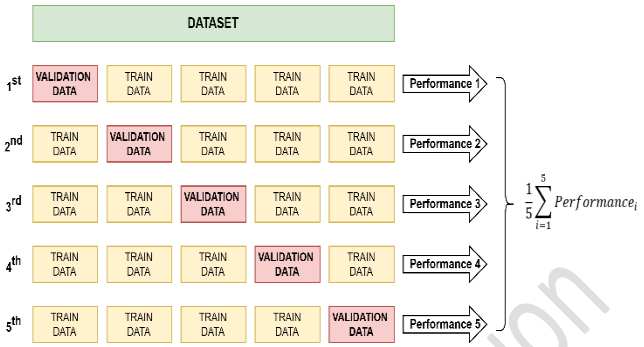


Figure 5. 5-fold cross validation procedure.

In this study, to achieve more stable and reliable results, the cross-validation process was repeated 10 times due to the randomness of the initial data partitioning, and the final performance metrics were calculated as the average of all repetitions. Furthermore, the classification model was constructed using a variety of algorithms, including Decision Tree, Random Forest, K-Nearest Neighbors, AdaBoost, and Extra Trees. The selection of these algorithms was made on the basis of their proven effectiveness in the handling of classification tasks involving complex, high-dimensional, and imbalanced medical datasets [34].

2.4.1 Decision tree (DT)

DT are a widely utilized ML algorithm, known for their ability to address classification and regression problems by iteratively splitting data into branches based on defined rules. The intuitive structure of DT makes them particularly valuable for interpretability, even in complex models. However, they are prone to overfitting, especially when dealing with intricate datasets, making it necessary to employ techniques like pruning to enhance the model's generalization performance [35].

2.4.2 Random forest (RF)

RF is an EL method that combines multiple decision trees to create a cohesive predictive model. In this technique, each DT is trained on a randomly selected subset of features from the dataset, fostering diversity among the trees. The final classification outcome is determined by aggregating the predictions of all individual trees, typically through a majority voting scheme. This ensemble strategy significantly reduces the risk of overfitting. Moreover, RF exhibits strong performance on high-dimensional datasets and provides valuable insights by generating feature importance rankings, which are crucial for uncovering underlying data patterns [36].

2.4.3 K-nearest neighbors (KNN)

KNN algorithm assigns a class to an unknown instance by considering the labels of its nearest neighbors, operating on the assumption that data points with similar feature values are likely to belong to the same class or yield comparable outcomes. The algorithm relies on spatial proximity within the feature space, typically employing distance metrics such as Euclidean distance to make predictions or classifications. A critical parameter, K, denotes the number of neighbors considered and significantly impacts the algorithm's performance. Although KNN is recognized for its simplicity and effectiveness, it is also prone to high computational costs, particularly when dealing with large datasets or high-dimensional spaces [37, 38].

2.4.4 AdaBoost

The primary goal of AdaBoost is to improve the model's generalization by modifying sample distribution, assigning greater weights to misclassified instances. Initially, all samples are weighted equally, and the classifier with the lowest error is selected. Misclassified instances are then assigned higher weights in subsequent iterations, allowing the algorithm to focus on correcting previous errors. This iterative process continues until the stopping criteria are met, forming a stronger classifier. In recent years, AdaBoost and its derivatives have attracted much attention due to their ability to handle complex and diverse datasets [39, 40].

2.4.5 Extra tree classifier

Extra Trees Classifier (ETC) is an advanced EL method designed to enhance predictive accuracy by constructing a large ensemble of decision trees and aggregating their outputs. Unlike traditional DT algorithms, ETC introduces a greater degree of randomness during the tree-building process. Specifically, it selects split points at each node from a randomly chosen subset of features rather than optimizing the split based on the entire feature set. This heightened randomness effectively reduces model variance and improves generalization performance on unseen data, making ETC a robust choice for various machine learning tasks [41].

2.5 Performance Metrics

The efficacy of a classifier is commonly assessed through metrics derived from the confusion matrix, which offers a comprehensive understanding of the model's capability to differentiate between various classes. In this study, the accurate classification of a diabetic was categorized as a True Positive (TP), whereas a misclassification as non-diabetic was categorized as a False Negative (FN). Conversely, the correct identification of a non-diabetic was classified as a True Negative (TN), and the misclassification of a non-diabetic as diabetic was labeled as a False Positive (FP). These categorizations served as the basis for calculating key performance metrics, which were detailed in Table 2. Furthermore, the calculation of accuracy is provided in Equation (3), while sensitivity is calculated in Equation (4). Precision is calculated in Equation (5), F-score in Equation (6) and AUC in Equation (7). AUC was utilized to provide a more comprehensive evaluation of the model's performance. AUC metric offers a broader perspective by illustrating the balance between sensitivity and specificity across varying classification thresholds, thereby facilitating a more nuanced assessment of the classifier's effectiveness under different conditions [42].

Table 2. Performance metrics.

Metric Name	Definition
Accuracy (ACC)	Accuracy represents the proportion of correctly predicted instances by the model.
Sensitivity (Recall)	Sensitivity measures the proportion of true positives within all positive instances, indicating how well the model identifies true positives.
Precision	Precision represents the proportion of true positives out of the total predicted positives, showing how many of the model's positive predictions are actually correct.
F1-Score	The F-Score represents the harmonic mean of precision and recall (sensitivity).
AUC-ROC	AUC-ROC denotes the area under the Receiver Operating Characteristic (ROC) curve, which plots the true positive rate (TPR) against the false positive rate (FPR) as a function.

$$ACC = \frac{TP + TN}{TP + TN + FP + FN} \quad (3)$$

$$Sensitivity = \frac{TP}{TP + FN} \quad (4)$$

$$Precision = \frac{TP}{TP + FP} \quad (5)$$

$$F1 - Score = \frac{2 \times precision \times sensitivity}{precision + sensitivity} \quad (6)$$

$$AUC - ROC = \int TPR\, d(FPR) \quad (7)$$

3 Results

In this study, a dataset related to diabetes from UCI was used to evaluate the performance of classifiers such as DT, RF, XGBoost, KNN, GB, and ETC. In order to address the adverse effects of data imbalance, a range of resampling techniques was individually applied to each classifier, followed by comprehensive performance evaluations. In order to improve the performance of the classifiers, the most relevant features in the dataset were selected using MI-based feature selection method. Additionally, the features were normalized using the Standard Scaler method. The performance of each classifier, with and without the application of resampling techniques based the extracted features, was evaluated based on performance metrics. All classification procedures were conducted using 5-fold cross-validation, and the average values of the results were reported. The effects of resampling techniques on the classifiers are presented in Table 3 to Table 7, respectively.

An examination of the results in Table 3 reveals that DT performed notably without resampling, achieving an ACC of 98.99% and an AUC of 99.40%. Among the oversampling techniques, SVMSMOTE demonstrated the highest performance, with an ACC of 99.06% and an AUC of 99.50%, reflecting a marked improvement in both metrics. SMOTE and ADASYN also resulted in accuracies of 99.02%, accompanied by AUC values of 99.45% and 99.41%, respectively, indicating a modest improvement over the non-resampled condition. In contrast, BorderlineSMOTE did not contribute to any performance improvement, as both ACC (98.99%) and AUC (99.40%) remained identical to the non-resampled condition. For undersampling techniques, the RandomUnderSampler yielded the best results, achieving the highest ACC of 99.06% and an AUC of 99.47%, as compared to the non-resamlping condition. NeighbourhoodCleaningRule, EditedNearestNeighbours, AIIKNN and InstanceHardnessThreshold achieved the same ACC (99.06%); however, their AUC values were comparatively lower, at 99.31%. Although TomekLinks exhibited a slight reduction in ACC (99.02%), its AUC value of 99.45% represented a little improvement. However, NearMiss yielded the lowest performance, with an ACC of 93.65% and an AUC of 95.79%.

As seen in Table 4, RF achieved 98.99% ACC and an AUC of 99.49% without resampling. Among the oversampling techniques, SVMSMOTE showed the highest performance, with 99.06% accuracy and 99.51% AUC. Both SMOTE and ADASYN achieved similar ACC (99.02%), with minor differences in AUC

values (99.50% and 99.48%, respectively). BorderlineSMOTE, with 98.99% ACC and 99.48% AUC, closely mirrored the results obtained without resampling. For undersampling techniques, RandomUnderSampler, AllKNN EditedNearestNeighbours and NeighbourhoodCleaningRule all surpassed the no-resampling condition, each reaching 99.06% ACC, where their AUC values ranged from 99.30% to 99.48%. TomekLinks, despite a slight drop in ACC (99.02%), maintained an AUC of 99.49%. OneSidedSelection and InstanceHardnessThreshold showed marginally lower ACCs (98.92%) with AUC values of 99.50% and 99.34%, respectively. NearMiss exhibited the lowest performance, with 93.65% ACC and a 95.84% AUC, making it the least effective resampling method.

As observed in Table 5, Adaboost classifier without any resampling scored 98.99% ACC and 99.47% AUC. Among the oversampling techniques, SVMSMOTE gave the highest performance with 99.06% ACC and 99.44% AUC. Both SMOTE and BorderlineSMOTE gave the same ACC (98.99%) results as in the non-resampled case, while SMOTE performed slightly better in terms of AUC (99.48%). In contrast, ADASYN produced a slightly lower AUC (99.43%), while improving accuracy (99.02%). Among undersampling techniques, RandomUnderSampler exhibited the best performance with 99.06% ACC and 99.52% AUC. NeighbourhoodCleaningRule, EditedNearestNeighbours, InstanceHardnessThreshold, and AllKNN all had the same ACC result, but with slightly lower AUC metric (99.31%). TomekLinks and OneSidedSelection maintained the same ACC and AUC as the non-resampled case, though other metrics varied. NearMiss, on the other hand, exhibited the lowest performance, with an ACC of 93.65% and an AUC of 96.08%.

Table 6 reveals that KNN classifier without resampling performed 98.99% ACC and 99.34% AUC. Among the oversampling techniques, SMOTE maintained the same ACC, though it resulted in a slight reduction in the AUC metric. In contrast, BorderlineSMOTE and SVMSMOTE provided the best results, achieving 99.02% ACC and 99.35% AUC. ADASYN, however, demonstrated a slightly lower ACC of 98.92%, while maintaining the same AUC of 99.35%. On the other hand, all undersampling techniques resulted in comparatively lower ACC than the case without resampling. The highest and lowest performance were obtained for OneSidedSelection and NearMiss. Based on the AUC metric, all methods, except NearMiss and TomekLinks, also exhibited lower performance. The lowest AUC value was obtained with NearMiss, while the highest AUC value was achieved with TomekLinks. Finally, as demonstrated in Table 7, the ETC classifier without resampling achieved an ACC of 99.02% and an AUC of 99.53%.

Table 2. Impact of various resampling techniques on DT classifier performance metrics.

Sampling Category	Resampling Technique	Mean ACC	Mean AUC	Mean Precision	Mean Recall	Mean F1 Score
No Resampling	-	98.99	99.40	99.90	98.62	99.26
Over Sampling	SMOTE	99.02	99.45	99.95	98.62	99.28
	ADASYN	99.02	99.41	99.95	98.62	99.28
	BorderlineSMOTE	98.99	99.40	99.90	98.62	99.26
	SVMSMOTE	99.06	99.50	100	98.62	99.31
Under Sampling	RandomUnderSampler	99.06	99.47	100	98.62	99.31
	NearMiss	93.65	95.79	99.89	90.82	95.12
	OneSidedSelection	98.99	99.40	99.90	98.62	99.26
	NeighbourhoodCleaningRule	99.06	99.31	100	98.62	99.31
	EditedNearestNeighbours	99.06	99.31	100	98.62	99.31
	InstanceHardnessThreshold	99.06	99.31	100	98.62	99.31
	AllKNN	99.06	99.31	100	98.62	99.31
	TomekLinks	99.02	99.45	99.95	98.62	99.28

Table 4. Impact of various resampling techniques on RF classifier performance metrics.

Sampling Category	Resampling Technique	Mean ACC	Mean AUC	Mean Precision	Mean Recall	Mean F1 Score
No Resampling	-	98.99	99.49	99.26	99.90	98.62
Over Sampling	SMOTE	99.02	99.50	99.28	99.95	98.62
	ADASYN	99.02	99.48	99.28	99.95	98.62
	BorderlineSMOTE	98.99	99.48	99.26	99.90	98.62
	SVMSMOTE	99.06	99.51	99.31	100	98.62
Under Sampling	RandomUnderSampler	99.06	99.48	99.31	100	98.62
	NearMiss	93.65	95.84	95.12	99.89	90.82
	OneSidedSelection	98.92	99.50	99.20	99.90	98.52
	NeighbourhoodCleaningRule	99.06	99.30	99.31	100	98.62
	EditedNearestNeighbours	99.06	99.32	99.31	100	98.62
	InstanceHardnessThreshold	98.92	99.34	99.20	100	98.42
	AllKNN	99.06	99.32	99.31	100	98.62
	TomekLinks	99.02	99.49	99.28	99.95	98.62

Table 5. Impact of various resampling techniques on Adaboost classifier performance metrics.

Sampling Category	Resampling Technique	Mean ACC	Mean AUC	Mean Precision	Mean Recall	Mean F1 Score
No Resampling	-	98.99	99.47	99.26	99.90	98.62
Over Sampling	SMOTE	98.99	99.48	99.26	99.90	98.62
	ADASYN	99.02	99.43	99.28	99.95	98.62
	BorderlineSMOTE	98.99	99.47	99.26	99.90	98.62
	SVMSMOTE	99.06	99.44	99.31	100	98.62
Under Sampling	RandomUnderSampler	99.06	99.52	99.31	100	98.62
	NearMiss	93.65	96.08	95.12	99.89	90.82
	OneSidedSelection	98.99	99.47	99.26	99.90	98.62
	NeighbourhoodCleaningRule	99.06	99.31	99.31	100	98.62
	EditedNearestNeighbours	99.06	99.31	99.31	100	98.62
	InstanceHardnessThreshold	99.06	99.31	99.31	100	98.62
	AllKNN	99.06	99.31	99.31	100	98.62
	TomekLinks	98.99	99.47	99.26	99.90	98.62

Table 6. Impact of various resampling techniques on KNN classifier performance metrics.

Sampling Category	Resampling Technique	Mean ACC	Mean AUC	Mean Precision	Mean Recall	Mean F1 Score
No Resampling	-	98.99	99.34	99.25	100	98.52
Over Sampling	SMOTE	98.99	99.35	99.25	100	98.52
	ADASYN	98.92	99.35	99.20	100	98.42
	BorderlineSMOTE	99.02	99.35	99.28	100	98.57
	SVMSMOTE	99.02	99.35	99.28	100	98.57
Under Sampling	RandomUnderSampler	98.74	99.28	99.07	100	98.16
	NearMiss	93.72	95.39	95.18	100	90.82
	OneSidedSelection	98.95	99.28	99.23	100	98.47
	NeighbourhoodCleaningRule	98.92	99.29	99.20	99.95	98.47
	EditedNearestNeighbours	98.92	99.29	99.20	99.95	98.47
	InstanceHardnessThreshold	97.91	99.21	98.44	99.95	96.99
	AllKNN	98.92	99.29	99.20	99.95	98.47
	TomekLinks	98.95	99.34	99.23	100	98.47

Table 7. Impact of various resampling techniques on ETC classifier performance metrics.

Sampling Category	Resampling Technique	Mean ACC	Mean AUC	Mean Precision	Mean Recall	Mean F1 Score
No Resampling	-	99.02	99.53	99.28	99.95	98.62
Over Sampling	SMOTE	99.02	99.54	99.28	99.95	98.62
	ADASYN	99.02	99.48	99.28	99.95	98.62
	BorderlineSMOTE	99.02	99.53	99.28	99.95	98.62
	SVMSMOTE	99.06	99.49	99.31	100	98.62
Under Sampling	RandomUnderSampler	99.06	99.48	99.31	100	98.62
	NearMiss	93.65	95.92	95.12	99.89	90.82
	OneSidedSelection	98.99	99.54	99.26	99.9	98.62
	NeighbourhoodCleaningRule	99.06	99.35	99.31	100	98.62
	EditedNearestNeighbours	99.06	99.35	99.31	100	98.62
	InstanceHardnessThreshold	99.06	99.33	99.31	100	98.62
	AllKNN	99.06	99.35	99.31	100	98.62
	TomekLinks	98.99	99.53	99.26	99.9	98.62

Among the oversampling methods, SMOTE, ADASYN, and BorderlineSMOTE all yielded identical ACC results compared to the non-resampled condition, although variations were observed in other performance metrics. Besides, the highest ACC (99.06%) was scored by SVMSMOTE with AUC value of 99.49%. For undersampling techniques, RandomUnderSampler, NeighbourhoodCleaningRule, EditedNearestNeighbours, InstanceHardnessThreshold, AllKNN all gave higher ACC (99.06%) than the non-resampling condition, with AUC values ranging from 99.33% to 99.48%. While OneSidedSelection and TomekLinks resulted in slightly lower ACC at 98.99%, their AUC remained strong at 99.54% and 99.53%, respectively. The lowest performance was observed with NearMiss, which recorded an ACC of 93.65% and an AUC of 95.92%. OneSidedSelection and TomekLinks demonstrated marginally lower ACC at 98.99%; however, their AUC values remained robust at 99.54% and 99.53%, respectively. The lowest performance was observed with NearMiss, which resulted in an ACC of 93.65% and an AUC of 95.92%.

Based on all findings, the SVMSMOTE oversampling technique achieved the highest performance in the DT classifier with 99.06% accuracy, 99.50 AUC, an F1 score of 99.31%, 100% precision, and 98.62% recall, indicating a balanced and robust classification. Similarly, SVMSMOTE demonstrated strong performance in the RF classifier, achieving 99.06% ACC, 99.51 AUC, 99.31% F1-score, 100% precision, and 98.62% recall. These metrics indicate that SVMSMOTE was an effective method not only in terms of ACC and AUC but also in providing balanced and robust classification performance overall. In the AdaBoost classifier, the highest ACC (99.06%) and AUC (99.52) values were achieved using the RandomUnderSampler undersampling technique. This superior performance was

further supported by additional key metrics, including an F1 score of 99.31%, 100% precision, and 98.62% recall, underscoring the model's strong predictive accuracy and high true positive detection rate. RandomUnderSampler thus emerged as a robust method for achieving balanced and effective classification. In the KNN classifier, BorderlineSMOTE and SVMSMOTE provided the best performance in terms of ACC and AUC. Both SVMSMOTE and BorderlineSMOTE demonstrated a solid and balanced classification performance, evidenced by an F1 score of 99.28%, 100% precision, and 98.57% recall. These results highlighted the model's high accuracy, near-error-free prediction capability, and substantial true positive detection rate. Lastly, in the ETC classifier, the SVMSMOTE technique delivered the highest performance, with an accuracy of 99.06% and an AUC of 99.49%. Metrics such as an F1 score of 99.31%, 100% precision, and 98.62% recall indicate that the model not only excels in ACC but also achieves a balanced classification with high sensitivity and precision. These findings demonstrate that SVMSMOTE efficiently reduces false positives while accurately identifying true positives, indicating a robust overall classification performance. In addition, the highest performance was observed in the AdaBoost classifier when the RandomUnderSampler technique was applied, achieving an accuracy of 99.06% and an AUC of 99.52. Notably, this approach led to a 7.07% improvement in accuracy and a 5.03% increase in AUC compared to the non-resampled scenario. Together, these enhancements underscore the substantial positive impact of the RandomUnderSampler technique on model performance. Figure 6 presents a comparison of the best-performing resampling techniques for different classifiers across five performance metrics: Accuracy, AUC, Precision, Sensitivity, and F1 Score. The resampling method yielding the best result is indicated in parentheses beneath each classifier. The figure clearly demonstrates that SVMSMOTE delivers superior performance for most classifiers, whereas BorderlineSMOTE achieves the best result for K-Nearest Neighbors.

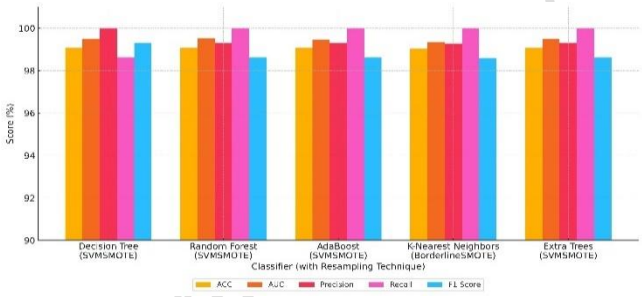


Figure 6. Performance metrics of classifiers using their top resampling technique.

4 Discussion

This study highlights the valuable role that resampling techniques play in enhancing the performance of ML-based decision support system for diabetes diagnosis, especially when dealing with imbalanced datasets. The challenge of class imbalance, commonly observed in medical datasets, often results in classifiers being biased toward the majority class, thereby reducing their ability to accurately predict outcomes for the minority class. By systematically applying various resampling methods, both oversampling and undersampling, the study demonstrates how these techniques can effectively balance the dataset and improve the overall predictive performance of machine learning algorithms.

Classifiers like DT, RF, KNN, AdaBoost and ETC showed significant improvements in their performance metrics when resampling was employed. Oversampling methods, particularly SMOTE and its variants (e.g., SVMSMOTE, ADASYN), were especially effective in boosting classification ACC, precision, recall, and AUC scores. SVMSMOTE stood out for its consistent enhancement of classifier performance, underscoring the importance of generating synthetic samples from the minority class to help the model learn more effectively from imbalanced data.

On the other hand, undersampling techniques such as RandomUnderSampler, NeighbourhoodCleaningRule, and EditedNearestNeighbours also delivered promising results, with RandomUnderSampler consistently improving classifier ACC and AUC. However, the effectiveness of undersampling techniques can be mixed, as seen with NearMiss, which caused a noticeable drop in performance for classifiers like KNN. This suggests that while undersampling can help address imbalance, its application must be carefully balanced to avoid discarding too much useful data from the majority class. Additionally, the use of feature selection through MI and data scaling via Standard Scaler proved essential in optimizing model performance. By ensuring that only the most relevant features were included in the classification models, and that these features were normalized, the study-maintained consistency across different classifiers, enabling more reliable predictions [43].

To further contextualize these findings, a comparative analysis with previous studies that employed resampling techniques for imbalanced medical datasets is provided in Table 8. This comparison aims to highlight both the similarities and differences in the impact of these methods across various ML classifiers. The results of this study reveal a marked improvement compared to the previous literature summarised in Table 8. Many studies in the literature have been conducted on the PIMA dataset, which usually has a small sample size and limited variable diversity, and mostly only SMOTE or its derivatives were used as resampling methods. However, in this study, the Diabetes 130-US Hospitals dataset, which has a larger and more representative patient profile, was preferred, thus enabling the classifiers to be tested under more realistic conditions. Moreover, while previous studies commonly used a single classifier and a limited number of sampling methods, here twelve different resampling techniques were systematically tested with five powerful machine learning algorithms. In particular, the high accuracy and AUC values achieved with SVMSMOTE and RandomUnderSampler techniques clearly demonstrated the impact of resampling strategies on classification performance. Thanks to this comprehensive approach, a more balanced, reliable and generalisable diabetes diagnosis model is presented compared to similar studies in the literature. Furthermore, the results are close and comparable to previous studies on the Diabetes 130-US Hospitals database used in this study.

5 Conclusions

This study emphasises the significant role of resampling techniques in optimising machine learning-based approaches for diabetes diagnosis, particularly in addressing the challenges associated with class imbalances that frequently impact model performance. The findings demonstrate that both

oversampling and undersampling techniques markedly enhance the predictive precision of classifiers, including Decision Tree, Random Forest, K-Nearest Neighbours, AdaBoost, and Extra Trees Classifier. Of these, SVMSMOTE and RandomUnderSampler were found to be the most effective, resulting in notable enhancements in key performance metrics such as accuracy, precision, recall, and AUC across a range of models. Furthermore, the incorporation of mutual information-based feature selection and data scaling with the standard scaler further optimised classifier performance, thereby ensuring both reliability and robustness. This methodological framework not only advances the study's objective of refining the accuracy of diabetes diagnoses but also highlights the potential for targeted resampling to enhance model robustness. In comparison to previous research, the findings confirm that the combined application of resampling, feature selection and scaling can result in significant performance improvements, thereby facilitating more accurate and reliable decision-making in the context of diabetes risk assessment. Furthermore, the proposed methodologies demonstrate adaptability to broader datasets, encompassing diverse patient populations, thereby expanding the scope of machine learning applications in diabetes recognition. Future studies could expand on this work by exploring additional feature engineering techniques and fine-tuning resampling methods to further enhance predictive accuracy and clinical applicability.

Table 8. Comparison of medical data classification studies.

Authors (Years)	Dataset (Size)	ML Algorithm	Resampling Status	Accuracy
Pradhan et al. (2020) [4]	PIMA (768)	ANN	-	85.09%
Maniruzzaman et al. (2020) [5]	National Health and Nutrition Examination (6561)	RF	-	94.25%
Daghistani and Alshammari (2020) [6]	Ministry of National Guard Hospital Affairs databases (66325)	RF		88.3%
Shuja et al. (2020) [7]	A diagnostic lab in Kashmir Valley (734)	DT	SMOTE	94.70%
Butt et al. (2021) [8]	PIMA (768)	LSTM	-	87.26%
Chaves ve Marques (2021) [9]	Sylhet Diabetes Hospital in Sylhet, Bangladesh (520)	Neural Networks	-	98.1%
Kumari et al. (2021) [10]	PIMA (768)	Soft Voting Classifier	-	79.08%
Khanam et al. (2021) [11]	PIMA (768)	NN	-	88.6%
Mesquita et al. (2021) [12]	PIMA (768)	AdaBoost	SVMSMOTE	83.12%
Özlüer Başer et al. (2021) [13]	Cerner Corporation, Kansas City, MO, US (70000)	RF	SMOTE	84.78%
Harman (2021) [14]	PIMA (768)	SVM	SMOTE	90%
Saxena et al. (2022) [15]	PIMA (768)	RF	-	79.8%
Mushtaq et al. (2022) [16]	PIMA (768)	RF	SMOTE	81.7%
Özkan et al. (2022) [17]	Endocrinology and Metabolic Diseases, Izmir Bozkaya Training and Research Hospital (232)	RF	-	84.48%
Sevli (2022) [18]	PIMA (768)	RF	InstanceHardnessThreshold	96.29%
Özoğur and Orman (2023) [19]	PIMA (768)	SVM	SMOTEENN	90%
Ali et al. (2023) [20]	PIMA (768)	Optimized RF	-	95.83%
Febrian et al. (2023) [21]	PIMA (768)	NB	-	76.07%
Khaleel et al. (2024) [22]	PIMA (768)	LR	-	-
Modak et al. (2024) [23]	Diabetic2 Dataset (5000)	CatBoost	-	95.4%
NG et al. (2024) [24]	PIMA (768)	En-RfRsK (RF, Radial SVM, KNN)	-	88.89%
Gaso et al. (2024) [25]	Diabetes 130-US Hospitals	MDLA	SMOTE	98%
Zarghani (2024) [26]	Diabetes 130-US Hospitals	LSTM	-	97.65%
Kanu and Khanal (2023) [27]	Diabetes 130-US Hospitals	RF, Hadoop, PySpark	-	100%
Proposed Model	Diabetes 130-US Hospitals	DT, MI	SVMSMOTE	99.06%

6 Author contribution statements

The conceptualization of this study was led by Authors 1 and 2. Author 3 conducted the literature review, identifying and analysing relevant sources and materials. The statistical analysis was performed by Author 1 with the input and collaboration of Author 2. All authors contributed to the drafting and writing of the manuscript. Additionally, Author 3 interpreted the results and revised the manuscript to ensure linguistic accuracy and content coherence.

7 Ethics committee approval and conflict of interest statement

The article does not require ethics committee approval and there is no conflict of interest with any person/institution.

8 References

[1] International Diabetes Federation. "IDF Diabetes Atlas". https://diabetesatlas.org. (11.11.2024)

[2] International Diabetes Federation. "Diabetes Now Affects One in 10 Adults Worldwide," https://idf.org/news/diabetes-now-affects-one-in-10-adults-worldwide/. (11.11.2024)

[3] Özmen T, Kuzu Ü, Koçyiğit Y, Sarnel H. "Early stage diabetes prediction by features selection with metaheuristic methods". *Pamukkale University Journal of Engineering Sciences.* 29(6), 596–606, 2023.

[4] Pradhan N, Rani G, Dhaka VS, Poonia RC. *Diabetes prediction using artificial neural network*. Editors: Basant A, Valentina EB, Lakhmi CJ, Ramesh CP. Deep Learning Techniques for Biomedical and Health Informatics. 327–339, Singapore, Springer Academic Press, 2020.

[5] Maniruzzaman M, Rahman MJ, Ahammed B, Abedin MM. "Classification and prediction of diabetes disease using machine learning paradigm". *Health Information Science and Systems*. 8(1), 7–14, 2020.

[6] Daghistani T, Alshammari R. "Comparison of statistical logistic regression and RandomForest machine learning techniques in predicting diabetes". *Journal of Advances in Information Technology*. 11(1), 78–83, 2020.

[7] Shuja M, Mittal S, Zaman M. "Effective prediction of type ii diabetes mellitus using data mining classifiers and SMOTE". *Advances in Computing and Intelligent Systems: Proceedings of ICACM 2019*. Singapore, 14-16 December 2020.

[8] Butt UM, Letchmunan S, Ali M, Hassan FH, Baqir A, Sherazi HHR. "Machine learning based diabetes classification and prediction for healthcare applications". *Journal of healthcare engineering*. 2021 (1), 933-985, 2021.

[9] Chaves L, Marques G. "Data mining techniques for early diagnosis of diabetes: a comparative study". *Applied Sciences*. 11(5), 2021.

[10] Kumari S, Kumar D, Mittal M. "An ensemble approach for classification and prediction of diabetes mellitus using soft voting classifier". *International Journal of Cognitive Computing in Engineering*. 2(1), 40-46, 2021.

[11] Khanam JJ, Foo SY. "A comparison of machine learning algorithms for diabetes prediction". *ICT Express.* 7(4), 432–439, 2021.

[12] Mesquita F, Aurício J, Marques G. "Oversampling techniques for diabetes classification: A comparative study". IEEE *2021 International Conference on e-Health and Bioengineering*. Virtual, 13 April 2021.

[13] Özlüer BB, Yangın M, Sarıdaş ES. "Makine Öğrenmesi Teknikleriyle Diyabet Hastalığının Sınıflandırılması". *Süleyman Demirel Üniversitesi Fen Bilim Enstitüsü Dergisi.* 25(1):112–120, 2021.

[14] Harman G. "Destek Vektör Makineleri ve Naive Bayes Sınıflandırma Algoritmalarını Kullanarak Diabetes Mellitus Tahmini". *European Journal of Science and Technology*, 32(1), 7-13, 2021.

[15] Saxena R, Sharma SK, Gupta M, Sampada GC. "A novel approach for feature selection and classification of diabetes mellitus: machine learning methods". *Computational Intelligence and Neuroscience*.;2022(1), 360-382, 2022.

[16] Mushtaq Z, Ramzan MF, Ali S, Baseer S, Samad A, Husnain M. "Voting Classification-Based Diabetes Mellitus Prediction Using Hypertuned Machine-Learning Techniques". *Mobile Information Systems*. 2022(1), 521-532, 2022.

[17] Özkan Y, Sarer YB, Suner A. "Diyabet tanısının tahminlenmesinde denetimli makine öğrenme algoritmalarının performans karşılaştırması". *Gümüşhane Üniversitesi Fen Bilim Enstitüsü Dergisi*. 12(1), 211-226, 2022.

[18] Sevli O. "Diyabet hastalığının farklı sınıflandırıcılar kullanılarak teşhisi". *Gazi Üniversitesi Mühendis-Mimar Fakültesi Dergisi*. 38(2), 989–1002, 2022.

[19] Özoğur HN, Orman Z. "Sağlık Verilerinin Analizinde Veri Ön işleme Adımlarının Makine Öğrenmesi Yöntemlerinin Performansına Etkisi". *Türkiye Bilişim Vakfı Bilgisayar Bilimleri ve Mühendisliği Dergisi*. 16(1), 23–33, 2023.

[20] Ali MS, Islam MK, Das AA, Duranta DUS, Haque MF, Rahman MH. "A novel approach for best parameters selection and feature engineering to analyze and detect diabetes: Machine learning insights". *BioMed Research International*. 2023(1), 1-15, 2023.

[21] Febrian ME, Ferdinan FX, Sendani GP, Suryanigrum KM, Yunanda R. "Diabetes prediction using supervised machine learning". *Procedia Computer Science*. 216(1), 21–30, 2023.

[22] Khaleel FA, Bakry AM. "Diagnosis of diabetes using machine learning algorithms". *Materials Today: Proceedings*. 8(1), 179-188, 2024.

[23] Modak SKS, Jha VK. "Diabetes prediction model using machine learning techniques". *Multimedia Tools and Applications.* 83(13), 38523–38549, 2024.

[24] Ng BA. "En-RfRsK: An ensemble machine learning technique for prognostication of diabetes mellitus". *Egyptian Informatics Journal*. 25(3), 1-8, 2024.

[25] Gaso M S, Mekuria R R, Khan A, Gulbarga M I, Tologonov I, Sadriddin Z. "Utilizing Machine and Deep Learning Techniques for Predicting Re-admission Cases in Diabetes Patients". *Proceedings of the International Conference on Computer Systems and Technologies,* 2024(1), 76–81, 2024.

[26] Zarghani A. "Comparative Analysis of LSTM Neural Networks and Traditional Machine Learning Models for Predicting Diabetes Patient Readmission". *arXiv preprint* arXiv:2406(19980), 1-21, 2024.

[27] Kanu D K S, Khanal M. "Implementation of Big Data Analytics on Diabetes 130-US Hospitals for year 1999-2008 for predicting patient readmission". *ResearchGate Preprint*, 2023.

[28] Strack B, Deshazo JP, Gennings C, Olmo JL, Ventura S, Cios KJ. "Impact of HbA1c Measurement on Hospital Readmission Rates: Analysis of 70,000 Clinical Database Patient Records". *BioMed Research International*. 2014(1), 1-11, 2014.

[29] Janez A, Muzurovic E, Bogdanski P, Czupryniak L, Fabryova L, Fras Z. "Modern management of cardiometabolic continuum: From overweight/obesity to prediabetes/type 2 diabetes mellitus. Recommendations from the Eastern and Southern Europe Diabetes and Obesity Expert Group". *Diabetes Ther*. 15(9), 1865–1892, 2024.

[30] Adanur Dedeturk B, Bakir Gungor B. "Evaluation of Sub-Network search programs in epilepsy-related GWAS

dataset". *Pamukkale University Journal of Engineering Sciences*. 28(2), 292–298, 2022.
[31] Sevli O. "Farklı sınıflandırıcılar ve yeniden örnekleme teknikleri kullanılarak kalp hastalığı teşhisine yönelik karşılaştırmalı bir çalışma". *Journal of Intelligent Systems: Theory and Applications*. 5(2), 92–105, 2022.
[32] Akgül G, Çelik AA, Ergül Aydin Z, Kamişli Öztürk Z. "Hipotiroidi Hastalığı Teşhisinde Sınıflandırma Algoritmalarının Kullanımı". *Bilişim Teknolojileri Dergisi*. 13(3), 255–68, 2020.
[33] Kohavi R. "A study of cross-validation and bootstrap for accuracy estimation and model selection". *Proceedings of the 14th International Joint Conference on Artificial Intelligence (IJCAI),* Palais des Congrès, Montreal, Quebec, Canada, 20–25 August 1995.
[34] Fernández A, García S, Galar M, Prati RC, Krawczyk B, Herrera F. *Learning from Imbalanced Data Sets*. Cham, Switzerland, Springer, 2018.
[35] Charbuty B, Abdulazeez A. "Classification based on decision tree algorithm for machine learning". *Journal of Applied Science and Technology Trends*. 2(1), 20–8, 2021.
[36] Cesur E, Efe C. "Kampüs İçi kapalı alanlarda Hava kalitesinin modellenmesi ve Karar destek sistemi geliştirilmesi". *Zeki Sistemler Teori ve Uygulamaları Dergisi*. 6(2), 181–90, 2023.
[37] Turan T. "Optimize edilmiş denetimli öğrenme algoritmaları ile obezite analizi ve tahmini". *Mehmet Akif Ersoy Üniversitesi Fen Bilimleri Enstitüsü Dergisi*. 14(2), 301–312, 2023.
[38] Yapıcı İŞ, Arslan RU, Erkaymaz O. "Kalp Yetmezliği Tanılı Hastaların Hayatta Kalma Tahmininde Topluluk Makine Öğrenme Yöntemlerinin Performans Analizi". *Karaelmas Fen ve Mühendislik Dergisi*. 14(1), 59–69, 2024.
[39] Shan W, Li D, Liu S, Song M, Xiao S, Zhang H. "A random feature mapping method based on the AdaBoost algorithm and results fusion for enhancing classification performance". *Expert Systems with Applications.* 256(1), 124902, 2024.
[40] Zhijun W, Yuqi L, Meng Y. "A hybrid intrusion detection method based on convolutional neural network and AdaBoost". *China Communications*. 1–10, 2024.
[41] Hama SMA. "Diabetes type 2 classification using machine learning algorithms with up-sampling technique". *Journal of Electrical Systems and Inf Technol.* 10(8), 1-10, 2023.
[42] Arslan RU, Yapıcı İŞ. "Farklı Örnekleme Tekniklerine ve Farklı Sınıflandırıcılara Dayanarak Kalp Yetmezliği Tanılı Hastaların Sağkalımlarının İncelenmesi". *EMO Bilimsel Dergi*. 14(2), 35–47, 2024.
[43] Akalın F, Yumusak N. "Classification of acute leukaemias with a hybrid use of feature selection algorithms and deep learning-based architectures". *Pamukkale University Journal of Engineering Sciences*. 29(3), 256–263, 2023.