Turkish Journal of Science & Technology
21(2), 413-423, 2026

Research Article
https://doi.org/10.55525/tjst.1888687

A New Approach in Diabetes Diagnosis: Performance Comparison using Machine Learning Algorithms

Cemal AKTÜRK[1], Mustafa Turan ARSLAN[2], Tarık TALAN[3*]
[1,3] Department of Computer Engineering, Faculty of Engineering and Natural Sciences, Gaziantep Islam Science and Technology University, Gaziantep, Türkiye
[2] Department of Computer Technologies, Kırıkhan Vocational School, Hatay Mustafa Kemal University, Hatay, Türkiye
[1] cemalakturk79@gmail.com, [2] mtarslan@mku.edu.tr, [3*] ttalan46@hotmail.com,

(Geliş/Received: 13/02/2026; Kabul/Accepted: 14/06/2026)

Abstract: According to the International Diabetes Federation (IDF), approximately 589 million adults are expected to be living with diabetes in 2025. This number continues to rise due to unhealthy eating habits and poor quality of life. Early and accurate diagnosis is crucial to prevent organ damage and improve patient outcomes. This study aims to develop a method for predicting diabetes without requiring individuals to visit healthcare facilities or undergo laboratory tests, using machine learning (ML) techniques to enable early intervention. A diabetes dataset containing 100,000 patient records was used. The dataset was reduced to 17,000 samples through undersampling. Ten different classification algorithms were applied to perform diabetes prediction. Initially, predictions were made using the full dataset, followed by a second phase where blood test data were excluded to assess the model's performance with fewer features. Among the ten algorithms tested, the BayesNet algorithm achieved the highest accuracy of 91% when using the full dataset. When blood-related features were removed, the BayesNet and Rotation Forest models still maintained a relatively high accuracy of 75.10%. The findings demonstrate that it is possible to predict diabetes with considerable accuracy even without blood test data. This approach can facilitate early diagnosis and preventive measures for diabetes, especially in settings where access to laboratory testing is limited.

Key words: Classification, diagnosis of diabetes, machine learning algorithms.

Diyabet Tanısında Yeni Bir Yaklaşım: Makine Öğrenmesi Algoritmaları Kullanılarak Performans Karşılaştırması

Öz: Uluslararası Diyabet Federasyonu'na (IDF) göre, 2025 yılı itibarıyla yaklaşık 589 milyon yetişkinin diyabetle yaşayacağı öngörülmektedir. Bu sayı, sağlıksız beslenme alışkanlıkları ve düşük yaşam kalitesi nedeniyle artmaya devam etmektedir. Organ hasarını önlemek ve hasta sonuçlarını iyileştirmek için erken ve doğru tanı kritik öneme sahiptir. Bu çalışma, bireylerin sağlık kuruluşlarına başvurmalarına veya laboratuvar testleri yaptırmalarına gerek kalmadan, makine öğrenmesi teknikleri kullanarak diyabeti tahmin etmeyi ve böylece erken müdahale imkânı sağlamayı amaçlamaktadır. Çalışmada 100.000 hasta kaydını içeren bir diyabet veri seti kullanılmıştır. Veri seti, alt örnekleme yöntemiyle 17.000 örneğe indirgenmiştir. Diyabet tahmini gerçekleştirmek için on farklı sınıflandırma algoritması uygulanmıştır. İlk olarak, tüm veri seti kullanılarak tahminler yapılmış, ardından kan test verilerinin çıkarıldığı ikinci bir aşamada modelin daha az özellik ile performansı değerlendirilmiştir. Test edilen on algoritma arasında, BayesNet algoritması tüm veri seti kullanıldığında %91 ile en yüksek doğruluk oranını elde etmiştir. Kanla ilgili özellikler çıkarıldığında ise BayesNet ve Rotation Forest modelleri hâlâ %75,10 gibi nispeten yüksek bir doğruluk oranı sağlamıştır. Elde edilen bulgular, kan testi verileri olmadan da diyabetin kayda değer bir doğrulukla tahmin edilebileceğini göstermektedir. Bu yaklaşım, özellikle laboratuvar testlerine erişimin sınırlı olduğu ortamlarda diyabetin erken teşhisi ve önleyici tedbirlerin alınmasını kolaylaştırabilir.

Anahtar kelimeler: Sınıflandırma, diyabet tanısı, makine öğrenmesi algoritmaları.

1. Introduction

Diabetes is a chronic disease associated with persistently high blood sugar levels. This disease occurs due to insulin production or impaired insulin function. It is a major global health problem that affects millions of people around the world and creates a serious burden on health systems [1]. Common symptoms of high blood sugar include frequent urination, constant thirst, and increased appetite. Without early diagnosis and appropriate treatment, diabetes can lead to serious complications, including diabetic ketoacidosis, a hyperosmolar hyperglycemic state, and even death. It can also cause various problems in the long term, such as cardiovascular

* Corresponding author: ttalan46@hotmail.com. ORCID Number of authors: [1] 0000-0003-3764-3862, [2] 0000-0001-5498-6571, [3] 0000-0002-5371-4520,

disease, stroke, kidney failure, and eye diseases [2]. Diabetes Mellitus is divided into three main types: Insulin-Dependent Diabetes Mellitus (IDDM), Non-Insulin-Dependent Diabetes Mellitus (NIDDM), and Gestational Diabetes. Type 1 diabetes is called IDDM, while type 2 diabetes is called NIDDM. Gestational diabetes, which occurs during pregnancy, is often called type 3 diabetes. The primary cause of type 1 diabetes is the body's inability to produce sufficient insulin. Therefore, patients must receive insulin injections. Type 1 diabetes usually occurs in young adults under the age of 30. Symptoms include excessive urination, excessive thirst, constant hunger, unintentional weight loss, vision changes, and fatigue. Type 2 diabetes commonly develops in adults over the age of 45 and is frequently linked to conditions such as obesity, high blood pressure, abnormal lipid levels, arteriosclerosis, and other related health issues. Type 3 diabetes is a form of diabetes that arises from an increase in blood sugar levels in pregnant women. Type 3 diabetes, or Gestational Diabetes, exclusively affects women during pregnancy [3, 4].

Although traditional diagnostic methods typically rely on clinical tests and physician assessments, this process can be both time-consuming and costly. In recent years, artificial intelligence (AI) and its subfield, machine learning (ML), have provided groundbreaking solutions for the diagnosis, prediction, and treatment of diseases within the medical healthcare [5-8]. ML algorithms, which are capable of learning from large medical datasets (including patient demographics, blood test results, and genetic data), enable the detection of diabetes in its early stages with high accuracy. In this context, supervised learning algorithms are among the most commonly used methods for the classification and prediction of diabetes. These algorithms build a model using data from patients with and without diabetes. When supplied with new data, the model can successfully predict whether a person has diabetes. Commonly used algorithms in this field include k-Nearest Neighbor (k-NN), Support Vector Machines (SVM), Random Forest (RF), Logistic Regression (LR), and Decision Trees (DT) [8-12].

In recent years, explainable artificial intelligence (XAI) approaches have gained increasing importance in studies on ML applications in diabetes prediction. Lundberg and Lee's work introducing the SHAP framework [14] has made significant contributions to classification models in the healthcare field by providing an integrated method for interpreting model decisions at the local and global levels. Furthermore, modern approaches to resampling and imbalanced data problems are also widely available in the literature. The SMOTE method proposed by Chawla et al. [15] has become the most widely used technique for addressing class imbalance by generating synthetic examples from the minority class. Recent studies on the prediction of diabetes and similar clinical outcomes also integrate both XAI and advanced resampling and modeling strategies. Farnoosh and Abnoosian's [16] pipeline-based ML framework for Covid-19 prediction demonstrates the effectiveness of a multi-stage modeling approach for achieving high accuracy in complex clinical data. Similarly, a pipeline-based framework [17] developed for early diabetes prediction and a study using ML ensemble models [18] demonstrate that multi-model architecture provides increased performance and better adapts to the structural characteristics of healthcare data. Taken together, these studies demonstrate that both XAI methods and advanced resampling and ensemble techniques are becoming increasingly important in healthcare classification studies.

Recent studies have further demonstrated the effectiveness of ML and XAI techniques in diabetes diagnosis. In particular, a 2025 study reported that models such as XGBoost and RF achieved high diagnostic performance, while SHAP and LIME-based explanations significantly improved model interpretability in clinical decision-making contexts [19]. Similarly, Navarro-Cerdán et al. highlighted that ML models combined with optimized feature selection techniques can achieve strong performance in both diabetes diagnosis and prognosis tasks, while also reducing feature dimensionality and improving computational efficiency through heterogeneous data analysis approaches [20]. In addition, emerging studies have explored hybrid approaches integrating ML with biomedical sensing technologies for diabetes diagnosis. For instance, Wei et al. proposed a ML-assisted photoelectrochemical biosensor for type 2 diabetes detection, achieving 95.8% diagnostic accuracy and demonstrating the potential of integrating ML with biomedical sensing systems for intelligent disease diagnosis [21]. Furthermore, comparative studies have confirmed the strong performance of ML models in diabetes diagnosis, particularly ensemble-based approaches. For example, Enríquez-Ortega et al. reported that RF and DT models achieved high diagnostic accuracy exceeding 97%, highlighting the effectiveness of properly preprocessed clinical datasets and ensemble learning methods for early diabetes detection [22].

This study examines the classification performance of various ML algorithms in diabetes diagnosis and proposes an approach to detect diabetes based on lifestyle characteristics such as age, gender, smoking status, and the presence of diagnoses such as hypertension and heart disease. The contributions of this study to the literature are listed below:

• Enables a comprehensive analysis of diabetes classification accuracy using ML methods. In this context, predicting diabetes diagnosis in patients can improve the quality of healthcare and positively impact patients' quality of life.

• We aim to contribute to the field by presenting the classification performance of diabetes prediction by selecting lifestyle and other disease-related features and excluding blood values from the dataset.

• It also provides new ideas for researchers and field experts to explore solutions to classification problems in other medical fields, potentially using new methodologies for diabetes classification.

• In conclusion, predicting diabetes diagnosis using ML algorithms without blood values represents a significant innovation from both clinical and academic perspectives.

• The findings obtained in this study can guide both medical practice and further research, and contribute significantly to the health management of the diabetic patient population.

Ultimately, this study distinguishes itself from similar early diabetes prediction studies in the literature in three key ways. First, it systematically compares 10 different ML algorithms on a comprehensive dataset of 100,000 individuals, both with the full feature set and with a reduced feature set containing only non-laboratory features. Second, the study comprehensively evaluates model performance using scenarios where blood-based measurements are completely excluded to increase applicability in low-resource settings. Third, while existing studies typically feature only a single method or a limited selection of algorithms, this study provides a broader methodological perspective with its multifaceted experimental design using 10 algorithms.

2. Material and Methods

In this study, the focus was on diabetes prediction by creating and preprocessing the dataset, followed by classifying it using ML algorithms. The main steps of the study are presented in Figure 1. For this purpose, firstly preprocessing processes were carried out on our dataset and then the first classification studies were carried out. We then extracted the blood data from the dataset, conducted further classification studies, and compared the performance metrics of the algorithms.

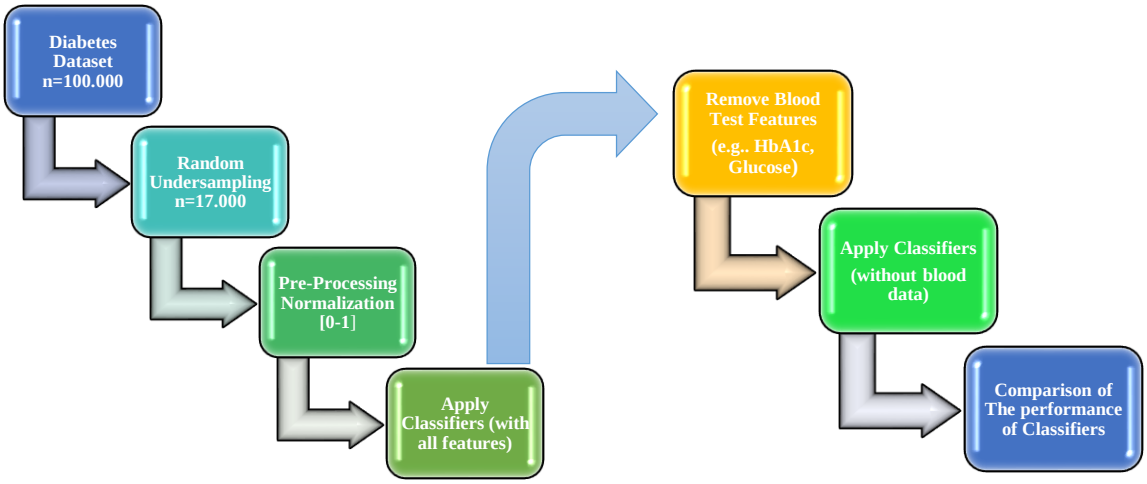


Figure 1. The workflow of the proposed model

2.1. Dataset description

In our study, DPD downloaded from the Kaggle repository was used [13]. The DPD consists of patients' medical and demographic information, along with their diabetes diagnosis (either positive or negative). It contains variables such as age, gender, body mass index (BMI), presence of hypertension, heart disease, smoking history, HbA1c values, and blood glucose measurements. The original dataset contains 100,000 records, of which 91,500 belong to the control group and 8,500 correspond to diabetes patients. To address the class imbalance present in the data, an equal number of samples were selected from each class as random undersampling method. Specifically, 8,500 records were randomly drawn from the control group to match the number of diabetes patient records. Consequently, the newly formed dataset consists of a total of 17,000 records, with an equal representation of both classes. The undersampling method was chosen to combat class imbalance by randomly selecting the same number of diabetes cases from among the non-diabetic cases, without causing any significant loss of representation in the

diabetes class. While alternative methods include SMOTE and hybrid resampling, the primary goal is to preserve the originality of the clinical data. The aim is to achieve a more homogeneous and true clinical dataset from a subset of the existing dataset, excluding the possibility that the newly generated random data will be squeezed around certain averages. In the newly created dataset, 9,549 patients are female, while 7,451 are male. Out of the individuals, 1,517 have heart disease, while 15,483 do not. There are 2,595 people with hypertension, whereas 14,405 do not have hypertension. Additionally, 6,308 people do not smoke, while 10,692 are smokers. The dataset contains a total of nine attributes: age, gender, BMI, presence of hypertension, heart disease, smoking history, HbA1c values, blood glucose measurements, and diabetes status. Among these attributes, the diabetes label is considered the class label. Table 1 shows the description of the attributes of the dataset.

Table 1. The attributes of the diabetes dataset

Attribute Names	Values
Age	0.08-80
Gender	Male, Female
Hypertension (Blood Pressure)	Yes, No
Heart disease	Yes, No
Smoking history	Never, No info, Current, Former
BMI - Body Mass Index (kg/m2)	10.01-95.69
HbA1c_level	3.5-9.0
Blood_glucose_level	80-300
Diabetes - Diabetes diagnose results (tested positive: 1, tested negative: 0)	Positive, Negative

The descriptive statistics of the dataset features are presented in Table 2.

Table 2. Descriptive statistics for diabetes dataset features

Attribute	Min	Max	Mean	Standard deviation
Age	0.08	80	50.62	21.46
BMI	10.3	88.72	29.53	7.45
HbA1c_level	3.5	9	6.16	1.28
Blood_glucose_level	80	300	163.47	56.90

Examining Table 2, the Age range varies between 0.08 and 80, while the BMI ranges from 10.3 to 88.72. The HbA1c level varies between 3.5 and 9. Finally, the Blood glucose level ranges from 80 to 300.

2.2. Data preprocessing

The DPD originally contains eight distinct features, excluding the class label. However, to conduct a classification analysis without relying on blood-based features, the HbA1c level and Blood glucose level features were removed from the dataset for this study. The resulting dataset thus includes the following features: age, gender, BMI, presence of hypertension, heart disease, and smoking history. Categorical variables, such as smoking history and gender, were transformed into numerical format using label encoding to make them suitable for the ML algorithms.

Preprocessing transforms data to facilitate the development of a more effective ML model, thereby yielding higher accuracy. It performs various functions aimed at improving data quality, such as data normalization and feature selection. Data normalization was performed by scaling the data to a range of 0 to 1, which served to increase the computational speed of the algorithms. Following normalization, the mean and standard deviation values for all attributes are presented in Table 3.

Table 3. Mean and standard deviation after normalization

Attributes	Mean	Standard Deviation
Age	0.632	0.269
BMI	0.245	0.095

As shown in Table 3, after the normalization process, the mean value of Age is 0.632, while the mean value of the BMI index is 0.245.

2.3. Implementation of classification models

In this study, an extensive analysis was conducted on the DPD using various ML classification algorithms, including Bayesian Network (BayesNet), Sequential Minimal Optimization (SMO), Multi-Layer Perceptron (MLP), Bagging, Dagging, Rotation Forest, ADTree, Functional Tree (FT), J48 Tree, and NBTree. The 10-fold cross-validation technique was used to evaluate the performance of different ML methods. The algorithms were implemented using the default hyperparameter settings.

2.4. Performance evaluations

Various statistical metrics can be employed to assess the performance of classifiers. In this study, we utilized Accuracy, Precision, and F-measure as evaluation measures. The accuracy of an ML algorithm can be determined using the confusion matrix. In general terms, the confusion matrix is presented as follows, Table 4.

Table 4. Confusion matrix of model's classification performance

	Predicted No (0)	Predicted Yes (1)
Actual No (0)	TN	FP
Actual Yes (1)	FN	TP

In this context, FP stands for False Positive, FN for False Negative, TN for True Negative, and TP for True Positive. Equations (1)– (4) are used to compute the performance metrics of the classification method.

$$Accuracy = \frac{(TP+TN)}{(TP+TN+FN+FP)} \quad (1)$$

$$Precision = \frac{TP}{(TP+FP)} \quad (2)$$

$$Recall = \frac{TP}{(TP+FN)} \quad (3)$$

$$F - Measure = \frac{2\ x\ (Precision\ x\ Recall)}{(Precision+Recall)} \quad (4)$$

ROC Area: It represents the area under the ROC curve and measures the classification performance of the model, especially the relationship between the true positive rate and the false positive rate. AUC ≥ 0.9 is considered high performance.

3. Results

The results of the correlation analysis between diabetes and other attributes are shown in Figure 2. When examining Figure 2, age, the most important determinant of diabetes, is seen as the attribute with the strongest correlation. BMI and hypertension are other attributes with strong correlations. Additionally, heart disease, smoking history, and gender are among the attributes identified in the correlation analysis of the dataset as related to diabetes. Thus, using correlation analysis, the most important attributes associated with the diabetes class in the dataset have been determined.

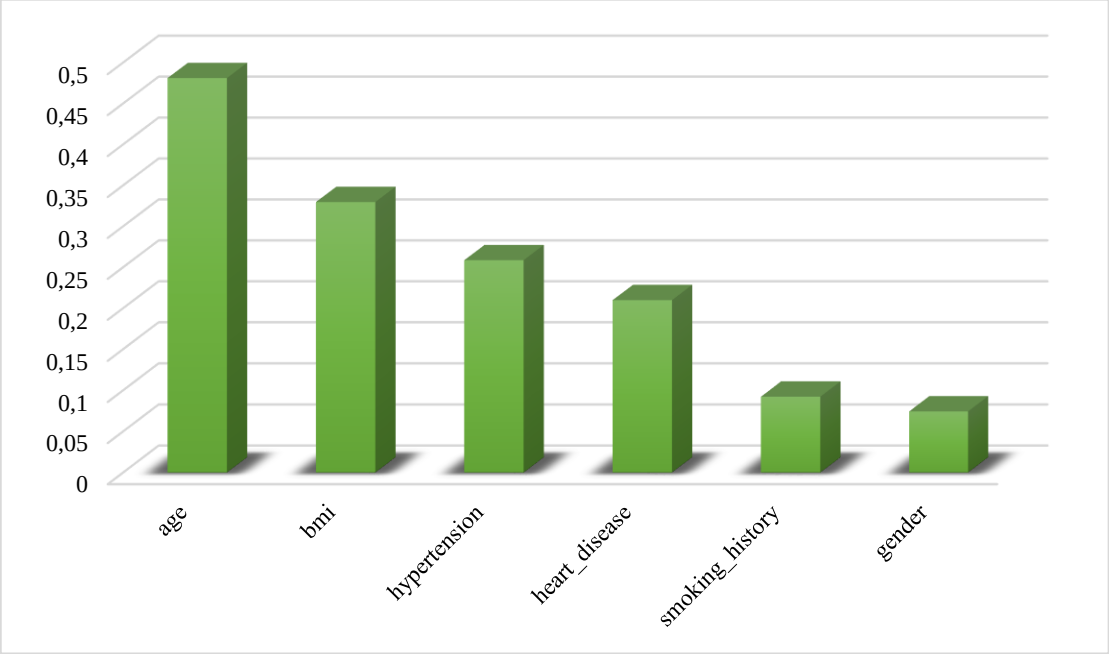


Figure 2. Correlation analysis for the most important characteristics related to the diabetes class.

To improve interpretability, feature importance was evaluated using Information Gain and Gain Ratio ranking techniques. These methods assess the significance of each input variable by measuring how much they reduce entropy related to class separation. While "age" ranks highest in Information Gain, it is only mid-ranked according to Gain Ratio. This suggests that although age contributes a substantial amount of information, its splitting effectiveness (after entropy normalization) is moderate. “BMI” holds moderate importance and is recognized as significant by both metrics, though it is not a top priority. The “smoking history” feature scores low in both measures, with Gain Ratio showing slightly better differentiation than Information Gain. Notably, “hypertension” has the highest Gain Ratio value, indicating it is the most efficient feature when considering normalized information gain. This implies that hypertension possesses strong discriminatory capability even in the presence of data imbalance. Despite “heart disease” scoring low in Information Gain, it ranks relatively high in Gain Ratio, meaning it produces more balanced data splits. “Gender” scores very low across both metrics, indicating minimal influence on the model. Overall, Information Gain tends to highlight features providing large information gains, whereas Gain Ratio adjusts for imbalanced splits, emphasizing features that yield more balanced partitions. Table 5 presents a complete ranking of attributes, shedding light on how the models weigh non-clinical factors during decision-making.

Table 5. Feature importance analysis of diabetes class

Attribute	Info Gain	Gain Ratio
Age	0.20151	0.007378
BMI	0.10158	0.04169
Hypertension	0.05125	0.08314
Smoking history	0.04487	0.02100
Heart disease	0.03446	0.07941
Gender	0.00399	0.00403

The feature contribution analysis shown in Figure 3 provides global interpretability for the model's decision structure and clearly demonstrates which clinical and lifestyle variables are more effective in predicting diabetes. Age appears to be the variable that provides the highest information contribution in predicting diabetes. Given that the risk of diabetes increases with age, it is expected that the model will use this variable as a strong discriminator. The fact that BMI is the second most predictive variable is consistent with obesity being a major risk factor for the development of diabetes. The moderate information contribution of cardiovascular and lifestyle variables such as

hypertension, smoking history, and heart disease reflects their indirect relationship with diabetes. Conversely, the low information gain of Gender suggests that gender does not provide significant discriminatory power for the model. These findings provide a practical framework for clinicians to better define risk groups, identify areas of focus for preventive healthcare, and prioritize which factors should be prioritized in population-based screening.

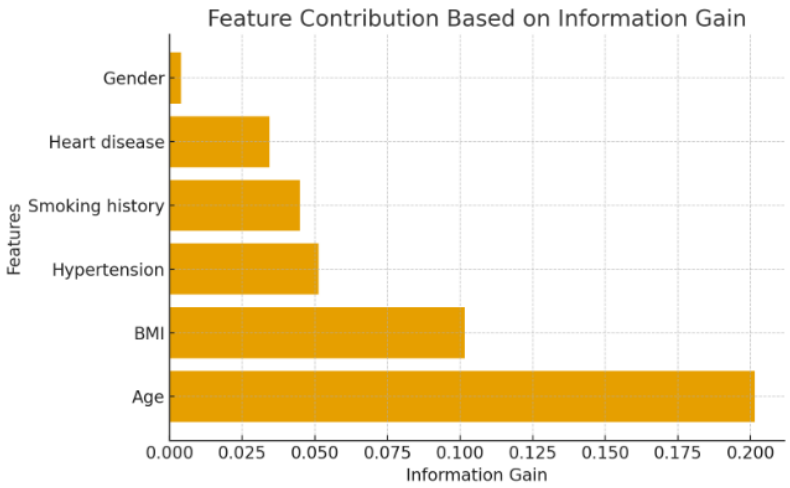


Figure 3. Feature Contribution Analysis of diabetes class.

This study focused on diabetes prediction without blood-based features and compared the performance of ten different ML algorithms on the DPD dataset. Table 6 presents the performance metrics for all classification algorithms applied to the DPD dataset. As shown in the Table 6, all classification methods achieved an accuracy rate exceeding 88%. Furthermore, both the BayesNet and NBTree methods achieved high performance compared to other algorithms, with a classification accuracy of 91%, a precision of 91.10%, and an F-measure of 91%, respectively. The lowest classification performance was demonstrated by SMO, with an accuracy, precision, and F-measure of 88.30%. The average accuracy rate, precision, Recall, F-Measure and ROC Area for all classifiers was found to be 90.18%, 0.902, 0.918, 0.902 and 0.958 respectively on this dataset. The ROC Area average of 95.8% is an important parameter in showing how high the classification performance is in all models.

Table 6. The performance measure of all classification methods.

	Accuracy(%)	**Precision**	**Recall**	**F-Measure**	**ROC Area**
BayesNet	91	0.911	0.910	0.910	0.979
SMO	88.30	0.883	0.883	0.883	0.839
MLP	90	0.900	0.900	0.900	0.975
Bagging	90.20	0.902	0.902	0.902	0.972
Dagging	90.60	0.907	0.906	0.906	0.976
Rotation Forest	90.20	0.903	0.902	0.902	0.975
ADTree	90.80	0.909	0.908	0.908	0.978
FT	89.80	0.898	0.898	0.898	0.94
J48 Tree	89.90	0.899	0.899	0.899	0.963
NBTree	91	0.911	0.910	0.910	0.978
Average	90.18	0.902	0.918	0.902	0.958

Table 7 shows the performance metrics for all classification algorithms applied to the DPD dataset without blood-based features. As seen in Table 7, the highest performance on the non-blood-based DPD dataset was demonstrated by BayesNet and Rotation Forest, both achieving a classification accuracy of 75.10%. For BayesNet, the precision and F-measure were 0.753 and 0.75, respectively, while for Rotation Forest, these values were 0.756 and 0.75, respectively. In contrast, the lowest classification performance was observed with J48 Tree, with an

accuracy rate, precision, and F-measure of 74.80%, 0.75, and 0.747, respectively. The average accuracy, precision, Recall, F-Measure, ROC Area and Specificity of all classification methods on the dataset was found to be 74.99%, 0.752, 0.749, 0.749, 0.809 and 69.88% respectively. The models' ROC Area parameters show more optimistic results than their accuracy rates. Rotation Forest and ADTree are the best models with an ROC Area of 0.828. SMO has the lowest ROC Area of 0.749. A higher ROC Area than the validation rate is a key indicator of the reliability of the predictions. In addition to standard metrics, the clinical discriminatory power of the models was further evaluated using the Specificity metric. As shown in Table 7, the average Specificity across all classification methods was found to be 69.88%. The Bagging algorithm achieved the highest Specificity at 71.45%, indicating its superior ability to correctly identify non-diabetic individuals compared to other models. It was followed by FT (70.45%) and ADTree (70.44%), while the SMO algorithm demonstrated the lowest Specificity at 67.92%. These results suggest that while the models are effective in predicting diabetes, they maintain a consistent performance in excluding healthy individuals even in the absence of blood-based features.

Table 7. The performance measure of all classification methods without blood-based features

	Accuracy (%)	**Precision**	**Recall**	**F-Measure**	**ROC Area**	**Specificity (%)**
BayesNet	75.10	0.753	0.751	0.750	0.824	70.28
SMO	74.90	0.754	0.749	0.748	0.749	67.92
MLP	75.00	0.753	0.750	0.749	0.821	68.94
Bagging	75.00	0.752	0.750	0.750	0.819	71.45
Dagging	75.00	0.752	0.750	0.749	0.824	69.60
Rotation Forest	75.10	0.756	0.751	0.750	0.828	68.59
ADTree	75.00	0.752	0.750	0.749	0.828	70.44
FT	75.00	0.752	0.750	0.749	0.766	70.45
J48 Tree	74.80	0.750	0.748	0.747	0.809	70.14
NBTree	75.00	0.752	0.75	0.749	0.823	69.99
Average	74.99	0.752	0.749	0.749	0.809	69.88

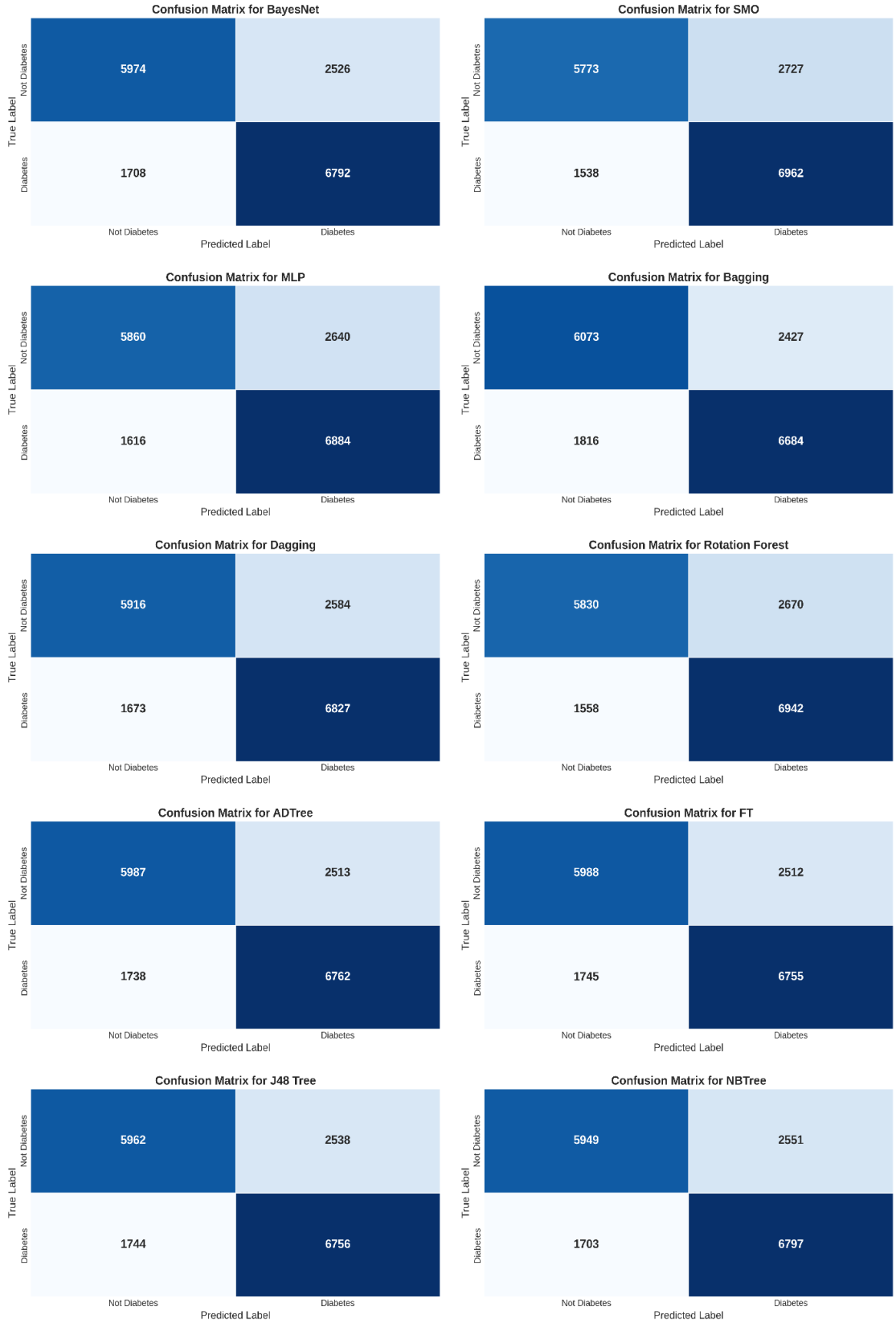


Figure 4. Confusion Matrices of ML Algorithms

Examining the Confusion Matrices in Figure 4, the model that best predicted the diabetes class was SMO with 6,962 correctly classified records. The model that performed worst in predicting the diabetes class was Bagging, with 6,684 correctly classified records. For the non-diabetes class, Bagging was the best performer with 6,073 correctly classified records, while SMO was the worst with 5,773 correctly classified records. Bagging made the most errors in predicting diabetes with 1,816 misclassified records, whereas SMO made the fewest errors with 1,538 misclassified records. Additionally, when predicting the non-diabetes class, SMO made the most errors with 2,727 misclassified records, while Bagging made the fewest errors with 2,427 misclassified records.

4. Conclusion

This study proposes a general perspective on diabetes diagnosis using routine laboratory tests and patient data. It primarily focuses on the critical importance of predicting diabetes without blood test data. The study first addressed class imbalance in a dataset containing blood data, and thus predicted diabetes reliably and generalizably. In this part of the study, accuracies ranging from 88.3% to 91% were achieved using 10 different classification algorithms. The most successful models were BayesNet and NBTree (91%). When the same algorithms were run on the DPD dataset, which did not contain any blood data, BayesNet and Rotation Forest achieved the highest accuracy rate of 75.10%, while J48 Tree achieved the lowest accuracy rate of 74.80%. Our results show that, while the performance of different ML algorithms is quite similar, BayesNet and Rotation Forest models consistently achieve higher accuracy rates than the others. This superiority stems from BayesNet's ability to effectively model conditional independence structures with weak to moderate dependencies between variables in the dataset, and Rotation Forest's transformation-based ensemble structure, which can capture multidimensional, nonlinear patterns. Therefore, our findings demonstrate that the structural properties of the relationships in the data structure directly affect algorithm performance and offer important methodological implications regarding the compatibility of appropriate model selection with data dependencies. When features based on blood values are removed, accuracy drops to approximately 75%. This performance alone may not be sufficient for clinical diagnosis. However, it can quickly provide a meaningful first-stage assessment in low-cost applications such as community screening or pre-screening. This level of accuracy is valuable for early referral and risk stratification, particularly in resource-limited settings where access to laboratory testing is limited. Because false negatives are clinically costly in medical decision support systems, classification performance should not be measured solely by accuracy. Therefore, in addition to accuracy, our analyses also included Precision, Recall, F-Measure, and ROC Area and Specificity metrics to enhance the clinical comparability of the models. The inclusion of Specificity alongside Recall (Sensitivity) provides a more comprehensive clinical perspective for diabetes screening. An average Specificity of approximately 70% implies that the proposed models can correctly identify seven out of ten healthy individuals without requiring invasive blood tests. Although this level is slightly lower than the overall Accuracy, it serves as a critical indicator for reducing unnecessary clinical referrals in low-resource settings. The high Accuracy combined with moderate Specificity suggests that the models are more 'optimistic' in detecting diabetes cases, which is often preferable in early-stage pre-screening to ensure fewer potential cases are missed.

While random undersampling effectively balanced the classes, it inherently involves discarding data from the majority class, which may limit the model's generalizability to the broader population. Future studies may consider oversampling-based methods such as SMOTE, ADASYN, or hybrid resampling approaches (e.g., SMOTE–Tomek Links) to preserve informative instances while mitigating imbalance. Furthermore, since the data set represents a specific population, re-validation of these models in different geographic or demographic groups will be important for future studies. This study highlights the potential of various ML models to improve diabetes detection and demonstrates that ML models can successfully diagnose diabetes without blood test data. In future studies, researchers could focus on improving the accuracy of diabetes prediction without using blood tests and explaining the models with explanatory AI techniques such as SHAP and LIME.

References

[1] Yameny AA. Diabetes mellitus overview 2024. J Biosci Appl Res 2024; 10(3): 641-645.
[2] Anjana RM, et al. Prevalence of diabetes and prediabetes (impaired fasting glucose and/or impaired glucose tolerance) in urban and rural India: Phase I results of the Indian Council of Medical Research–INDIAB study. Diabetologia 2011; 54(12): 3022-3027.
[3] Mujumdar A, Vaidehi V. Diabetes prediction using machine learning algorithms. Procedia Comput Sci 2019; 165: 292-299.
[4] Maniruzzaman M, et al. Classification and prediction of diabetes disease using machine learning paradigm. Health Inf Sci Syst 2020; 8(1): 7.

[5] Sarwar MA, et al. Prediction of diabetes using machine learning algorithms in healthcare. In: 24th Int Conf Autom Comput; 2018; pp. 1-6.
[6] Rani KJ. Diabetes prediction using machine learning. Int J Sci Res Comput Sci Eng Inf Technol 2020; 6(4): 294-305.
[7] Hasan MK, et al. Diabetes prediction using ensembling of different machine learning classifiers. IEEE Access 2020; 8: 76516-76531.
[8] Febrian ME, et al. Diabetes prediction using supervised machine learning. Procedia Comput Sci 2023; 216: 21-30.
[9] Soni M, Varma S. Diabetes prediction using machine learning techniques. Int J Eng Res Technol 2020; 9(9): 921-925.
[10] Wee BF, et al. Diabetes detection based on machine learning and deep learning approaches. Multimed Tools Appl 2024; 83(8): 24153-24185.
[11] Khanam JJ, Foo SY. A comparison of machine learning algorithms for diabetes prediction. ICT Express 2021; 7(4): 432-439.
[12] Khaleel FA, Al-Bakry AM. Diagnosis of diabetes using machine learning algorithms. Mater Today Proc 2023; 80: 3200-3203.
[13] Mustafa M. Diabetes prediction dataset. Kaggle; 2023.
[14] Lundberg SM, Lee SI. A unified approach to interpreting model predictions. In: Adv Neural Inf Process Syst; 2017; Long Beach, CA, USA. pp. 1-10.
[15] Chawla NV, Bowyer KW, Hall LO, Kegelmeyer WP. SMOTE: Synthetic minority over-sampling technique. J Artif Intell Res 2002; 16: 321-357.
[16] Farnoosh R, Abnoosian K. A robust innovative pipeline-based machine learning framework for predicting COVID-19 in Mexican patients. Int J Syst Assur Eng Manag 2024; 15(7): 3466-3484.
[17] Abnoosian K, Farnoosh R, Behzadi MH. A pipeline-based framework for early prediction of diabetes. J Health Biomed Inform 2023; 10(2): 125-140.
[18] Abnoosian K, Farnoosh R, Behzadi MH. Prediction of diabetes disease using an ensemble of machine learning multi-classifier models. BMC Bioinformatics 2023; 24: 337.
[19] Güler H, Avcı D, Ulaş M, Omma T. In-depth analysis of machine learning models and explainable artificial intelligence methods in diabetes diagnosis. J Fac Eng Arch Gazi Univ 2025; 40(3): 1995-2011.
[20] Navarro-Cerdán JR, Pons-Suñer P, Arnal L, Arlandis J, Llobet R, Pérez-Cortés JC, et al. A machine learning approach for type 2 diabetes diagnosis and prognosis using tailored heterogeneous feature subsets. Med Biol Eng Comput 2025;63:2733–2752.
[21] Wei J, Yan Z, Mei M, Chen S, Geng W, Yu Z, et al. Machine learning-assisted BMOFs-derived 1D NiCo2O4 nanozyme photoelectrochemical detection of RBP4 for type 2 diabetes diagnosis. Chem Eng J 2025;524:169192.
[22] Enríquez-Ortega D, Chulde-Fernández B, Pozo-Coral P, Vaca A, Zhinin-Vera L, Almeida-Galárraga D, Ramírez-Cando L, et al. Enhancing diabetes diagnosis through machine learning: a comparative study. Appl Sci 2025;15(18):10087.