



Research

A Machine Learning Model to Predict HbA1c Values of Patients

Hastaların HbA1c Değerlerinin Tahminlenmesi için Bir Makine Öğrenmesi Modeli

 Gökhan Silahtaroglu¹,  Özge Doğuş¹,  Yavuz Furuncuoğlu²

¹İstanbul Medipol University Faculty of Business and Management Sciences, Department of Management Information Systems, İstanbul, Türkiye

²Bahçeşehir University, Medical Park Göztepe Hospital, Department of Internal Medicine, İstanbul, Türkiye

ABSTRACT

Objective: To develop an automated method that accurately predicts patients' hemoglobin A1c (HbA1c) values.

Methods: Data from 118,265 unique patients were collected from 30 hospitals in İstanbul. Although the study began with a large dataset, extensive data cleaning reduced the final modeling dataset to 180 complete records. While this limits generalizability, it allows for a controlled evaluation of prediction models under realistic constraints of missing clinical data. For the analysis, the data was parted 70%-30% ratio as training and testing. A gradient boosting algorithm was used to train a machine learning model to predict HbA1c values.

Results: In this study, a machine learning model was developed to predict HbA1c using patients' readily obtainable vital data. For testing, the overall accuracy of the prediction is 80%, Cohen's kappa of 0.549, which is acceptable. Validation results are also promising, with an accuracy of 82%.

Conclusion: Although predicting HbA1c using different parameters can be advantageous, these predictions may be inaccurate and should be used and interpreted with caution in patients with anemia, polycythemia, or hemoglobinopathies. Our values, with specificity of 86.72%, sensitivity of 76.71%, and accuracy of 81.76%, can help eliminate this problem. In addition to satisfactory testing and validation results, the study not only explains the gradient-boosting machine learning model but also provides detailed information on cleaning noisy data and imputing missing health data.

Keywords: Data management, diabetes complications, internal medicine, machine learning

Öz

Amaç: Bu çalışmanın amacı hastaların hemoglobin A1c (HbA1c) değerlerini doğru bir şekilde tahmin edebilecek otomatik bir yöntem geliştirmektir.

Gereç ve Yöntem: İstanbul'daki 30 farklı hastaneden 118.265 benzersiz hasta verisi toplanmıştır. Başlangıçta büyük bir veri setiyle çalışılmış olsa da, kapsamlı veri temizleme işlemleri sonucunda modelleme için kullanılan nihai veri seti 180 tam doldurulmuş kayıttan oluşmuştur. Bu durum, sonuçların genellenebilirliğini sınırlasa da, eksik klinik verilerin oluşturduğu gerçekçi kısıtlamalar altında tahmin modellerinin kontrollü bir şekilde değerlendirilmesine olanak sağlamaktadır. Analiz için veriler %70-%30 oranında eğitim ve test olarak ayrılmıştır. Makineyi HbA1c değerlerini tahmin edecek şekilde eğitmek için gradient boost algoritması kullanıldı.

Bulgular: Bu çalışmada, hastaların kolayca elde edilebilen hayati verilerini kullanarak, HbA1c biyobelirteçlerini tahmin etmek için bir makine öğrenme modeli geliştirilmiştir. Test verileri için, tahminin genel doğruluğu, kabul edilebilir olan Cohen'in kappa katsayısı 0,549 değeriyle %80'dir. Doğrulama sonuçları ise %82 doğrulukla umut vericidir.

Sonuç: HbA1c'yi farklı parametreler kullanarak tahmin etmek bir avantaj olsa da anemi, polisitemi veya hemoglobinopatisi olan hastalarda kesin sonuçlar alınmadığından kullanılması ve yorumlanması yanlış olabilmektedir. %86,72-%76,71 özgüllük ve %81,76 duyarlılık ve doğruluk değerlerimiz, bu sorunun ortadan kaldırılmasına yardımcı olabileceği düşünülmektedir. Çalışma, gradient-boost makine öğrenmesi modelini açıklamanın yanı sıra yüksek doğruluk değerleri sağlamakta ve aynı zamanda sağlıktaki gürültülü verinin nasıl temizleneceği ve eksik sağlık verilerinin nasıl yeniden oluşturulacağı hakkında ayrıntılı bilgi vermektedir.

Anahtar Kelimeler: Veri yönetimi, diyabet komplikasyonları, iç hastalıkları, makine öğrenimi

Address for Correspondence: Özge Doğuş, Assoc. Prof., İstanbul Medipol University Faculty of Business and Management Sciences, Department of Management Information Systems, İstanbul, Türkiye
E-mail: odoguc@medipol.edu.tr **ORCID ID:** orcid.org/0000-0002-5971-9218

Cite as: Silahtaroglu G, Doğuş Ö, Furuncuoğlu Y. A machine learning model to predict HbA1c values of patients. Med J Bakirkoy. 2025;21(4):357-367

Received: 01.10.2024
Accepted: 13.06.2025
Publication Date: 18.12.2025



INTRODUCTION

Artificial intelligence (AI) and machine learning (ML) have rapidly become essential tools in healthcare for predicting disease outcomes, supporting diagnosis, and personalizing treatment plans. Among chronic conditions, diabetes mellitus stands out due to its prevalence and long-term complications. Accurate prediction and early intervention are critical to improving patient outcomes. One biomarker central to diabetes management is hemoglobin A1c (HbA1c), which reflects average blood glucose levels over a two- to three-month period and is a key metric for both diagnosis and monitoring of the disease (1).

While HbA1c is typically measured through laboratory tests, several real-world challenges hinder its timely availability (2), such as cost, limited access to laboratories in rural areas, and delays resulting from the required three-month measurement period. Moreover, in cases where patients have hemoglobin-related disorders (e.g., anemia, polycythemia, hemoglobinopathies), direct measurement may yield unreliable results. In such scenarios, a reliable prediction of HbA1c based on alternative, easily obtainable physiological parameters (e.g., age, glucose level, blood pressure) becomes clinically valuable, potentially enabling faster interventions and improved continuity of care.

Although various ML approaches have been applied to predict glycemic trends, most existing models fall short in predicting HbA1c. For instance, Plis et al. (3) used support vector regression and autoregressive integrated moving average models informed by glucose dynamics to predict short-term glucose levels, but these models rely heavily on patient-specific time-series glucose data, limiting their generalizability. Zou et al. (4) compared decision trees, random forests, and neural networks for diabetes prediction using physical exam data, but did not focus on HbA1c as a predictive target, nor did they address the implications of missing or noisy data—a common issue in real-world health datasets.

This study addresses these gaps by proposing a gradient boosting model specifically tailored to predict HbA1c levels using routinely collected patient data. Unlike prior approaches, the proposed model incorporates robust data-preprocessing strategies—including advanced outlier detection, synthetic data augmentation, and ML-based imputation of missing values—resulting in a cleaner, more usable dataset (5). While the study began with a large dataset, strict data quality protocols and completeness criteria narrowed the final sample used for modeling. Gradient boosting is chosen because of its strengths in handling imbalanced datasets, reducing bias, and offering

higher accuracy with less overfitting compared with traditional ensemble methods such as random forests (6).

In summary, this research contributes to the field by demonstrating that HbA1c can be predicted with high accuracy from non-invasive, widely available clinical parameters, using a refined gradient-boosting model, offering a practical solution in settings where HbA1c cannot be directly measured or trusted.

METHODS

A two-way analysis was performed in this study. The main purpose of the study is to estimate the HbA1c value, which is effective for predicting the development of diabetes. However, due to missing data, this estimation model could not directly adopted, and the data deficiencies were addressed beforehand, as explained above.

In the second stage, the HbA1c value, the main purpose of the study, was estimated. For this purpose, the HbA1c value was converted to discrete data. At this stage, HbA1c values were categorized into clinically relevant risk groups in alignment with American Diabetes Association guidelines and were further stratified as follows: values <5.7% were labeled as low risk (normal), 5.7-6.4% as moderate (prediabetes), 6.5-7.9% as high (controlled diabetes), and ≥8.0% as very high (uncontrolled diabetes). These thresholds were chosen to reflect increasing risk of diabetes-related complications and to align with clinical treatment targets. A total of 180 patients with complete HbA1c data were included in the study. Ethical approval was obtained from the Non-Invasive Clinical Research Ethics Committee of İstanbul Medipol University (approval no: 685, date: 03.09.2020). As this was a retrospective study using anonymized data, no informed consent was required from the patients. The data were divided into 70% training and 30% test sets using stratified sampling before ML model training. Synthetic minority over-sampling technique was applied to the 70% data slice used for learning, increasing the number of records to 375, and testing was performed with the remaining 30% slice (7,8). No synthetic data generation was carried out for the testing data.

ML algorithms are highly configurable by their hyperparameters. These parameters substantially influence the behavior, complexity, speed, and other aspects of the algorithm, and their values must be selected carefully to achieve optimal performance (9). In the estimation of the HbA1c variable, the best result was obtained when the gradient boosting algorithm was customized with the following hyperparameters: tree depth=8, number of models=200, and learning rate=0.05. A learning rate of

0.1 is usually a good starting point for gradient boosting algorithms, but the optimal learning rate depends on the number of models. For this reason, the model achieved its best performance when the number of models was high and the learning rate was low. Also, for the k-nearest neighbor algorithm, the Euclidean distance metric was used, and k was set to 5. The optimum values for the support vector machine (SVM) algorithm were kernel type radial basis function and sigma=1.6.

Selection and Identification of Cases

Data

In this study, data from 118,265 unique patients were collected at 30 hospitals in İstanbul. This patient dataset was anonymized before they were included in the study. By stripping off the identifying features. The dataset contains 12 patient attributes (variables). These variables are continuous or categorical and include variables such as age, gender, height, weight, patient diabetes status, family history of diabetes, and average glucose level. All variables are given in Table 1.

Preparation of Data

The main objective of the study is to estimate HbA1c levels, an important parameter in individuals with diabetes. However, before reaching this stage, some data preprocessing needed to be carried out to prepare the data for analysis. It is well known that one problem with health data is that they may be incomplete, inaccurate, missing, or noisy. The performance of the AI models may be poor, or the models' estimation accuracy may be low due to noisy data. Data preprocessing is probably the most important stage of data analysis and data mining, and therefore of ML.

Firstly, it was spotted that there were missing data in the raw dataset. Since the number of missing mean glucose values in the dataset is high (i.e., postprandial blood sugar; 107,012 missing values), these were completely excluded from the dataset. After removal of records with missing data, 11,253

unique patient records remained in the relevant dataset. After that, outlier cleaning was performed. To detect the outliers for each variable, the first and third quartiles (Q1 and Q3) were computed. An observation was flagged as an outlier if it lies outside the range $R=[Q1-k(IQR), Q3+k(IQR)]$ with $IQR=Q3-Q1$, where $k=1.5$, which corresponds to the smallest value in R (10,11). After this process, 6,754 records remained in the dataset.

Among these records, 6,686 mean fasting glucose values were missing and 68 were present. The aim is to fill records with missing mean fasting glucose values using existing values. As a first step, missing mean fasting glucose values were predicted and imputed into the dataset. In the original dataset, the fasting glucose value was a continuous variable. To handle missing values for this variable, the fasting glucose value was first converted into discrete categories. Accordingly, fasting glucose values between 0 and 70 were labeled as hypoglycemia; values between 70 and 100 were labeled as normal; those between 100 and 125 were labeled as prediabetes; and values higher than 125 were labeled as diabetes. After this transformation, the dataset included 16 patients with diabetes, 25 with hypoglycemia, and 27 with prediabetes.

Because the aim was to fill in the missing mean fasting glucose records using the non-missing values, 68 existing values were synthetically reproduced using the k-nearest neighbor technique, increasing the number of non-missing values to 204. This was done to increase the capability of the AI ML model.

An AI model was built, using these 204 records, to replace missing fasting glucose values with predicted values. The model was trained using a gradient boosting algorithm to impute the missing values for 6,686 patients. However, not all the predictions made by the AI model were added to the dataset. Only those predicted with a confidence of 95% or higher were added to the dataset. Subsequently, 4,331 records met this criterion. Together with the original non-missing dataset, the total number of records reached 4,399 (4,331+68).

Nevertheless, the dataset included missing values for the target variable HbA1c. All records with missing HbA1c were removed from the dataset, and, finally, 180 clean, complete records were ready for analysis. Figure 1 shows the workflow diagram of the data cleaning and preprocessing steps, including (1) exclusion of records with missing glucose values, (2) outlier detection using the IQR method, (3) imputation of missing fasting glucose values using a ML model, and (4) removal of incomplete HbA1c records.

Table 1. Variables used

Variable name
Gender
Age
Weight
Height
Pulse/heartbeat
Average blood pressure systolic
Average blood pressure diastolic
Average glucose
Has diabetes

The final dataset used for training consisted of 180 fully populated records.

A Close Look at the Cleaned Data

In the final case, the data set includes 80 female and 100 male patients; 78 patients with prediabetes, 72 with hypoglycemia, and 30 with diabetes. Table 2 summarizes descriptive statistics for all numeric variables in the dataset.

Technical Information

In this study, the KNIME platform (version 4.4.2; Konstanz Information Miner, Germany) (12) was used to create learning models, develop predictive models from data, and visualize data. The learning algorithm used throughout the study is gradient boosting trees (13). Boosting is an ensemble learning method that sequentially generates base models. Boosting algorithms aim to construct a strong learner by sequentially combining weak learners generated at each iteration according to predefined rules. Gradient boosting, a boosting algorithm, is used to reduce model bias (6).

Gradient-based decision-tree algorithms are also weak learners. A weak learner is a ML model that performs slightly better than chance. In the case of gradient-boosting trees,

the weak learners are shallow decision trees. Each new tree added to the ensemble model (i.e., the combination of all previous trees) minimizes the loss function associated with the ensemble (14). Decision trees rely on iteratively asking questions to split the data; however, they are prone to overfitting, and this is true for gradient-based trees as well. To reduce this risk, the gradient boosting algorithm employs a model that combines multiple decision trees.

The steps of the boosting method are listed below:

1. The weights of each sample are the same at the beginning,
2. It is used to learn basic learner 1 training examples,
3. When the learning is complete, the weight of the wrong samples is increased and the weight of the correct samples is decreased,
4. Basic learner 2 is used for learning,
5. Steps 2 and 4 are repeated to obtain M core-learners,
6. The results of M core learners are combined to produce the final learning outcome.

The weight of each base learner in the boosting method differs (15). Hyperparameters of the gradient boosting algorithm are set to tree depth=8, number of models=200, and learning rate=0.05. As a result, the average glucose hunger value was imputed from incomplete records using estimates with 95% confidence. Figure 2 depicts the overall model for training, testing, and validation.

Statistical Analysis

Testing

As mentioned above, 30% of the dataset, which consists of 55 records, was used for testing. As shown in Figure 3, the model achieved an accuracy of 80% and Cohen's kappa of 0.549 during testing. While accuracy alone can be misleading in imbalanced datasets, Cohen's kappa provides a more balanced measure of classification performance by accounting for agreement occurring by chance. A kappa value of 0.549 indicates moderate agreement, suggesting that the model produces reliable predictions beyond random chance but still has room for improvement.

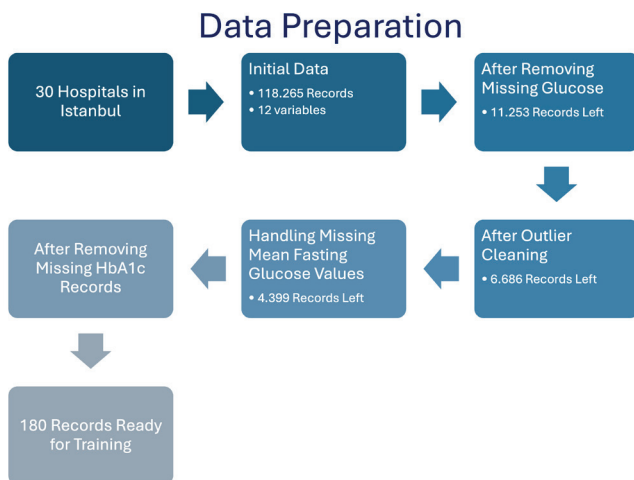


Figure 1. Data cleaning and preparation workflow
Hb1Ac: Hemoglobin A1c

Table 2. Sample data with minimum, maximum and mean point

Variables	Min	Max	Mean	Standard deviation
Age	16.00	94.00	56.64	16.53
Average blood pressure systolic	102.00	149.00	121.13	10.97
Average blood pressure diastolic	56.50	91.00	71.78	6.51
Weight	40.00	122.10	77.21	17.80
Height	144.00	190.00	165.85	9.21
Pulse/heartbeat	54.00	112.00	78.76	12.03
Average glucose	68.00	178.00	108.38	22.52

Implementation of ML

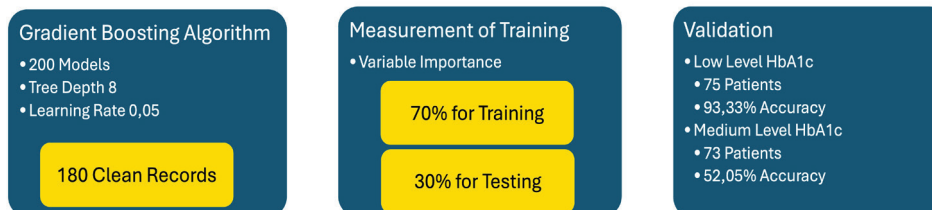


Figure 2. Implementation of gradient boosting algorithm
ML: Machine learning, HbA1c: Hemoglobin A1c

low (Actual)	0	34	2	0	94.44%
medium (Actual)	0	6	7	0	53.85%
veryhigh (Actual)	0	0	0	0	undefined
	100.00%	79.07%	77.78%	undefined	

Overall Statistics

Overall Accuracy	Overall Error	Cohen's kappa (κ)	Correctly Classified	Incorrectly Classified
80.00%	20.00%	0.549	44	11

	high (Predicted)	low (Predicted)	medium (Predicted)	veryhigh (Predicted)	
high (Actual)	4	0	0	0	100.00%
low (Actual)	1	24	3	0	85.71%
medium (Actual)	2	4	11	0	64.71%
veryhigh (Actual)	0	0	0	1	100.00%
	57.14%	85.71%	78.57%	100.00%	

Overall Statistics

Overall Accuracy	Overall Error	Cohen's kappa (κ)	Correctly Classified	Incorrectly Classified
80.00%	20.00%	0.655	40	10

	high (Predicted)	low (Predicted)	medium (Predicted)	veryhigh (Predicted)	
high (Actual)	2	2	0	0	50.00%
low (Actual)	0	28	0	0	100.00%
medium (Actual)	0	8	9	0	52.94%
veryhigh (Actual)	0	0	0	1	100.00%
	100.00%	73.68%	100.00%	100.00%	

Overall Statistics

Overall Accuracy	Overall Error	Cohen's kappa (κ)	Correctly Classified	Incorrectly Classified
80.00%	20.00%	0.608	40	10

Figure 3. Gradient boost, KNN and SVM evaluation of training with test data
KNN: K-nearest neighbors, SVM: Support vector machine

Examination of Figure 3 shows that the machine predicts a high level of HbA1c with 100% accuracy. On the other hand, there were no data to test for very high HbA1c values. Finally, the accuracies for the low and medium levels are 79.07% and 77.78%, respectively.

Table 3 shows a detailed evaluation of Gradient Boosting, k-nearest neighbors, and SVM algorithms. In the first table, all F-measures are acceptable and both sensitivity and specificity values are adequate for the gradient boost algorithm.

While the algorithm made the above predictions, it ranked the variables according to their importance. It should be

kept in mind that this grading is relative and is performed by the gradient boosting algorithm itself. It means that the algorithm has obtained information from the features shown in Table 4. As can be seen, the patient's average glucose, age, height, and blood pressure are more important to the algorithm than other features.

RESULTS

Since the testing data were drawn from the training data, were preprocessed, and had some missing values imputed using predictions from other variables, the results could have been biased. Due to this possibility, data collection

Table 3. Detailed evaluation of training

Gradient boost								
HbA1c	TP	FP	TN	FN	Precision	Sensitivity	Specificity	F-measure
High	3	0	86	1	1.0	0.75	1.0	0.86
Low	33	12	31	14	0.73	0.70	0.72	0.72
Medium	26	14	38	12	0.65	0.68	0.73	0.67
Very high	1	1	88	0	0.5	1.0	0.99	0.67
KNN								
HbA1c	TP	FP	TN	FN	Precision	Sensitivity	Specificity	F-measure
High	4	3	43	0	0.57	1.00	0.93	0.73
Low	24	4	18	4	0.86	0.86	0.82	0.86
Medium	11	3	30	6	0.79	0.65	0.91	0.71
Very high	1	0	49	0	1.00	1.00	1.00	1.00
SVM								
HbA1c	TP	FP	TN	FN	Precision	Sensitivity	Specificity	F-measure
High	2	0	46	2	1.00	0.50	1.00	0.67
Low	28	10	12	0	0.74	1.00	0.55	0.85
Medium	9	0	33	8	1.00	0.53	1.00	0.69
Very high	1	0	49	0	1.00	1.00	1.00	1.00

HbA1c: Hemoglobin A1c, TP: True positive, FP: False positive, TN: True negative, FN: False negative, KNN: K-nearest neighbors, SVM: Support vector machine

Table 4. Variable importance gradient boosting

Variables	Relative importance	Scaled importance	Percentage
Average glucose	297.9	1.0	0.3
Age	215.1	0.7	0.2
Average blood pressure diastolic	127.4	0.4	0.1
Height	125.1	0.4	0.1
Average blood pressure systolic	117.4	0.4	0.1
Pulse/heartbeat	87.1	0.3	0.1
Weight	86.9	0.3	0.1
Gender (male)	9.2	0.0	0.0
Gender (female)	0.0	0.0	0.0
Is diabetes	0.0	0.0	0.0
Diabetes in family	0.0	0.0	0.0

continued during the study, and these data were used to validate the learning. The test dataset used to evaluate the model's performance consisted of only 55 records, and the validation set included 73 medium and 75 low HbA1c values. No records representing the "high" or "very high" categories were present in the validation data. This limited sample size, particularly the absence of two of the four target classes during validation, reduces generalizability of the findings and undermines multi-class classification performance.

As it is seen in Table 5 and Figure 4 are examined it is seen that area under curve is 0.81 which close to 1 and sensitivity and specificity statistics are above 76% suggesting that machine predictions are accurate enough to be used. On the other hand, the area under the confidence curve of the machine appears to be approximately 50%. This value can be interpreted as: although the machine predicts with an accuracy of 81.76%, its confidence is low (Table 6, Figure 5). In this case, when the model is applied to real data, any prediction exceeding 55% may be sufficiently reliable.

DISCUSSION

HbA1c value is a criterion used in the diagnosis and treatment of diabetes. It is an important biomarker because, as the HbA1c level increases or decreases, the risk of diabetes-related complications changes accordingly. For this reason, it plays a critical role in assessing both the possibility of disease progression and the effectiveness of treatment. This value reflects the average of fasting and postprandial blood glucose levels over the preceding 2-3 months and is a robust indicator of long-term glycemic control. It is not only useful for monitoring diabetes but also predictive of cardiovascular complications.

While traditional glucose tests, such as fasting blood glucose, provide only a snapshot of a patient's current state, they do not reflect long-term trends. Continuous glucose monitoring addresses this gap partially, but remains costly and less accessible. In contrast, HbA1c provides a time-averaged measure; predicting it from easily collected variables may be useful in contexts where laboratory testing is unavailable or unreliable.

Study Limitations

One potential limitation of the study is the significant reduction in usable data due to missing values. While the initial dataset was large, our strict criteria for data quality and the need for complete records led to a smaller final sample size. This decision was intended to avoid bias and to ensure model integrity. However, it does mean that the findings should be interpreted as a proof-of-concept, and further research with larger, more complete datasets is warranted to validate the model's generalizability.

In our study, the gradient boosting model achieved a test accuracy of 80% and a Cohen's kappa of 0.549. This kappa score indicates moderate agreement between the predicted and actual HbA1c categories, suggesting the model performs better than random classification. In the validation phase, the model demonstrated 81.76% accuracy, with sensitivity and specificity values exceeding 76% for low and medium HbA1c categories. These metrics support the model's robustness and generalizability within the observed data range. However, due to the absence of high and very high HbA1c values in the validation dataset, further validation with a balanced sample is necessary to confirm performance across all classes.

Table 5. Evaluation of validation data for low and medium level HbA1c values

Gradient boost								
HbA1c	TP	FP	TN	FN	Precision	Sensitivity	Specificity	F-measure
Low	65	17	56	10	0.79	0.87	0.77	0.83
Medium	56	10	65	17	0.85	0.77	0.87	0.81
KNN								
HbA1c	TP	FP	TN	FN	Precision	Sensitivity	Specificity	F-measure
Low	10	1	4	1	0.91	0.91	0.80	0.91
Medium	3	1	11	1	0.75	0.75	0.92	0.75
SVM								
HbA1c	TP	FP	TN	FN	Precision	Sensitivity	Specificity	F-measure
Low	11	3	2	0	0.79	1	0.40	0.88
Medium	2	0	12	2	1	0.5	1	0.67

HbA1c: Hemoglobin A1c, TP: True positive, FP: False positive, TN: True negative, FN: False negative, KNN: K-nearest neighbors, SVM: Support vector machine

Confusion Matrix

	high (Predicted)	low (Predicted)	medium (Predicted)	veryhigh (Predict...	
high (Actual)	0	0	0	0	undefined
low (Actual)	0	65	10	0	86.67%
medium (Actual)	0	17	56	0	76.71%
veryhigh (Actual)	0	0	0	0	undefined
	undefined	79.27%	84.85%	undefined	

Overall Statistics

Overall Accuracy	Overall Error	Cohen's kappa (κ)	Correctly Classified	Incorrectly Classified
81.76%	18.24%	0.635	121	27

	high (Predicted)	low (Predicted)	medium (Predicted)	veryhigh (Predict...	
high (Actual)	1	0	0	0	100.00%
low (Actual)	0	10	1	0	90.91%
medium (Actual)	0	1	3	0	75.00%
veryhigh (Actual)	0	0	0	0	undefined
	100.00%	90.91%	75.00%	undefined	

Overall Statistics

Overall Accuracy	Overall Error	Cohen's kappa (κ)	Correctly Classified	Incorrectly Classified
87.50%	12.50%	0.729	14	2

	high (Predicted)	low (Predicted)	medium (Predicted)	veryhigh (Predict...	
high (Actual)	0	1	0	0	0.00%
low (Actual)	0	11	0	0	100.00%
medium (Actual)	0	2	2	0	50.00%
veryhigh (Actual)	0	0	0	0	undefined
	undefined	78.57%	100.00%	undefined	

Overall Statistics

Overall Accuracy	Overall Error	Cohen's kappa (κ)	Correctly Classified	Incorrectly Classified
81.25%	18.75%	0.489	13	3

Figure 4. Gradient boost, KNN and SVM validation results of medium level HbA1c patients
 KNN: K-nearest neighbors, SVM: Support vector machine, HbA1c: Hemoglobin A1c

Another critical consideration in clinical ML is model interpretability. Gradient boosting models are often perceived as black boxes, which can undermine trust in their outputs. To address this, we conducted a variable importance analysis. Features such as average glucose, age, height, and blood pressure were identified as the most influential predictors of HbA1c. This aligns with

clinical knowledge and increases the model's transparency. Future work should incorporate advanced interpretability techniques, such as SHapley Additive exPlanations or local interpretable model-agnostic explanations, to further explain individual predictions and enhance physicians' confidence in the system.

Table 6. Validation statistics

Gradient boost				
Results	Low HbA1c validation		Medium HbA1c validation	
Statistic	Value	95% CI	Value	95% CI
Sensitivity	86.67%	76.84% to 93.42%	76.71%	65.35% to 85.81%
Specificity	76.71%	65.35% to 85.81%	86.67%	76.84% to 93.42%
Positive likelihood ratio	3.72	2.43 to 5.70	5.75	3.19 to 10.39
Negative likelihood ratio	0.17	0.10 to 0.31	0.27	0.18 to 0.41
Disease prevalence	50.68%	42.34% to 58.98%	49.32%	41.02% to 57.66%
Positive predictive value	79.27%	71.41% to 85.41%	84.85%	75.62% to 91.00%
Negative predictive value	84.85%	75.62% to 91.00%	79.27%	71.41% to 85.41%
Area under curve (ROC)	0.817			
Accuracy	81.76%	74.58% to 87.62%	81.76%	74.58% to 87.62%
KNN				
Results	Low HbA1c validation		Medium HbA1c validation	
Statistic	Value	95% CI	Value	95% CI
Sensitivity	90.91%	58.72% to 99.77%	75.00%	19.41% to 99.37%
Specificity	80.00%	28.36% to 99.49%	91.67%	61.52% to 99.79%
Positive likelihood ratio	4.55	0.78 to 26.50	9	1.27 to 63.89
Negative likelihood ratio	0.11	0.02 to 0.77	0.27	0.05 to 1.50
Disease prevalence	68.75%	41.34% to 88.98%	25.00%	7.27% to 52.38%
Positive predictive value	90.91%	63.17% to 98.31%	75.00%	29.70% to 95.52%
Negative predictive value	80.00%	36.98% to 96.46%	91.67%	66.64% to 98.38%
Area under curve (ROC)	87.50%	61.65% to 98.45%	87.50%	61.65% to 98.45%
Accuracy	90.91%	58.72% to 99.77%	75.00%	19.41% to 99.37%
SVM				
Results	Low HbA1c validation		Medium HbA1c validation	
Statistic	Value	95% CI	Value	95% CI
Sensitivity	100.00%	71.51% to 100.00%	50.00%	6.76% to 93.24%
Specificity	40.00%	5.27% to 85.34%	100.00%	73.54% to 100.00%
Positive likelihood ratio	1.67	0.81 to 3.41		
Negative likelihood ratio	0		0.5	0.19 to 1.33
Disease prevalence	68.75%	41.34% to 88.98%	25.00%	7.27% to 52.38%
Positive predictive value	78.57%	64.19% to 88.24%	100.00%	
Negative predictive value	100.00%		85.71%	69.25% to 94.11%
Area under curve (ROC)	81.25%	54.35% to 95.95%	87.50%	61.65% to 98.45%
Accuracy	100.00%	71.51% to 100.00%	50.00%	6.76% to 93.24%

HbA1c: Hemoglobin A1c, KNN: K-nearest neighbors, SVM: Support vector machine, ROC: Receiver operating characteristic, CI: Confidence interval

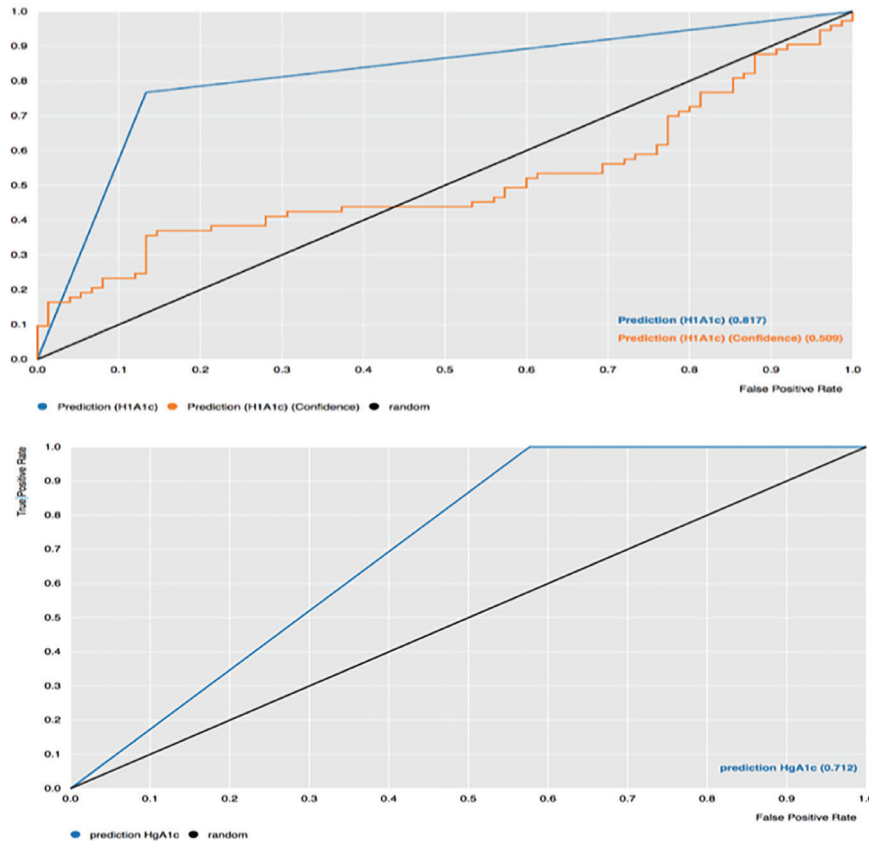


Figure 5. ROC of validation gradient boost and KNN
KNN: K-nearest neighbors, ROC: Receiver operating characteristic, HbA1c: Hemoglobin A1c

CONCLUSION

Although it is an advantage to predict HbA1c using different parameters, it can be wrong to use and interpret because accurate results cannot be obtained for the patients with anemia, polycythemia or hemoglobinopathy. Our values with 86.72-76.71% specificity and sensitivity and accuracy of 81.76% can help to eliminate this problem. Positive and negative predictivity values are also above acceptable levels. Based on these results, it is a useful estimation tool in situations where HbA1c cannot be measured, when testing is requested without requiring a 3-month waiting period, or when Hb-related diseases render HbA1c measurements unreliable.

ETHICS

Ethics Committee Approval: Ethical approval was obtained from the Non-Invasive Clinical Research Ethics Committee of İstanbul Medipol University (approval no: 685, date: 03.09.2020).

Informed Consent: As this was a retrospective study using anonymized data, no informed consent was required from the patients.

FOOTNOTES

Authorship Contributions

Surgical and Medical Practices: G.S., Y.F., Concept: G.S., Y.F., Design: Ö.D., Data Collection or Processing: Ö.D., Y.F., Analysis or Interpretation: G.S., Ö.D., Literature Search: Ö.D., Writing: G.S., Ö.D.

Conflict of Interest: No conflict of interest was declared by the authors.

Financial Disclosure: The authors declare that this study received no financial support.

REFERENCES

1. Canbolat ZN, Silahtaroglu G, Doğuç Ö, Yılmaztürk N. A machine learning approach to predict creatine kinase test results. *Emerg Sci J.* 2020;4:283-96.

2. Sarwar MA, Kamal N, Hamid W, Shah MA. Prediction of diabetes using machine learning algorithms in healthcare. 2018 24th International Conference on Automation and Computing (ICAC); Newcastle Upon Tyne, UK; 2018. p. 1-6.
3. Plis K, Bunesco R, Marling C, Shubrook J, Schwartz F. A machine learning approach to predicting blood glucose levels for diabetes management. Proceedings of the AAAI 14 Workshop on Modern Artificial Intelligence for Health Analytics; 2014; Toronto, Canada.
4. Zou L, Lam NSN, Cai H, Qiang Y. Mining twitter data for improved understanding of disaster resilience. *Ann Am Assoc Geogr*. 2018;108:1422-41.
5. Sherwani SI, Khan HA, Ekhzaimy A, Masood A, Sakharkar MK. Significance of HbA1c test in diagnosis and prognosis of diabetic patients. *Biomark Insights*. 2016;11:95-104.
6. Jeatrakul P, Wong KW, Fung CC. Classification of imbalanced data by combining the complementary neural network and SMOTE algorithm. In: Wong K W, Mendis B S U, Bouzerdoum A (eds). *Neural Information Processing. Models and Applications*. ICONIP 2010. Lecture Notes in Computer Science, vol 6444. Berlin, Heidelberg: Springer-Verlag; 2010:152-9.
7. Basgall MJ, Hasperué W, Naiouf M, Fernández A, Herrera F. An analysis of local and global solutions to address big data imbalanced classification: a case study with SMOTE preprocessing. In: Rucci E, Naiouf M, Chichizola F, De Giusti L, De Giusti A (eds). *Cloud Computing and Big Data: Revised Selected Papers from JCC&BD 2019*. Communications in Computer and Information Science, vol 1050. Cham: Springer; 2019. p. 75-85.
8. Bischl B, Binder M, Lang M, Pielok T, Richter J, Coors S, et al. Hyperparameter optimization: foundations, algorithms, best practices, and open challenges. *WIREs Data Mining Knowl Discov*. 2023;13:e1484.
9. Tukey JW. Some thoughts on clinical trials, especially problems of multiplicity. *Science*. 1977;198:679-84.
10. Zimek A, Filzmoser P. There and back again: outlier detection between statistical reasoning and data mining algorithms. *WIREs Data Mining Knowl Discov*. 2018;8:e1280.
11. Berthold MR, Cebron N, Dill F, Gabriel TR, Kötter T, Meinl T, et al. KNIME – the Konstanz information miner: version 2.0 and beyond. *SIGKDD Explorations Newsletter*. 2009;11:26-31.
12. Friedman JH. Greedy function approximation: a gradient boosting machine. *Ann Stat*. 2001;29:1189-232.
13. Cao DS, Xu Q, Liang YZ, Zhang LX. The boosting: a new idea of building models. *Chemometr Intell Lab Syst*. 2010;100:1-11.
14. Callens A, Morichon D, Abadie S, Delpey M, Lique B. Using random forest and gradient boosting trees to improve wave forecast at a specific location. *Applied Ocean Research*. 2020;104:102339.
15. Cheng J, Li G, Chen X. Research on travel time prediction model of freeway based on gradient boosting decision tree. *IEEE Access*. 2018;7:7466-80.