

Original Article



Development of a Predictive Model for Glycated Hemoglobin Values and Analysis of the Factors Affecting It

HyeongKyu Park , MD, PhD¹, Da Young Lee , MD², So young Park , MD², Jiyoung Min , MS³, Jiwon Shinn , BA³, Dae Ho Lee , MD, PhD⁴, Soon Hyo Kwon , MD, PhD⁵, Hun-Sung Kim , MD, PhD^{3,6,*}, Nan Hee Kim, MD, PhD ^{2,7,*}

¹Division of Endocrinology and Metabolism, Department of Internal Medicine, Soonchunhyang University College of Medicine, Soonchunhyang University Seoul Hospital, Seoul, Korea

²Division of Endocrinology and Metabolism, Department of Internal Medicine, Korea University College of Medicine, Seoul, Korea

³Department of Medical Informatics, College of Medicine, The Catholic University of Korea, Seoul, Korea

⁴Department of Internal Medicine, Gachon University Gil Medical Center, Gachon University College of Medicine, Incheon, Korea

⁵Division of Nephrology, Soonchunhyang University College of Medicine, Soonchunhyang University Seoul Hospital, Seoul, Korea

⁶Division of Endocrinology and Metabolism, Department of Internal Medicine, Seoul St. Mary's Hospital, College of Medicine, The Catholic University of Korea, Seoul, Korea

⁷BK21 FOUR R&E Center for Learning Health Systems, Korea University, Seoul, Korea

OPEN ACCESS

Received: Sep 24, 2021

Accepted: Oct 31, 2021

Correspondence to

Nan Hee Kim, MD, PhD

Division of Endocrinology and Metabolism,
Department of Internal Medicine, Korea
University College of Medicine, 73 Goryeodae-
ro, Seongbuk-gu, Seoul 02841, Korea.
E-mail: nhkendo@gmail.com

Hun-Sung Kim, MD, PhD

Department of Medical Informatics, College of
Medicine, The Catholic University of Korea, 222
Banpo-daero, Seocho-gu, Seoul 06591, Korea.
E-mail: 01cadiz@hanmail.net

*Hun-Sung Kim and Nan Hee Kim are co-
corresponding author and equally contributed
to this paper

Copyright © 2021. Korean Society of
Cardiovascular Disease Prevention; Korean
Society of Cardiovascular Pharmacotherapy.
This is an Open Access article distributed
under the terms of the Creative Commons
Attribution Non-Commercial License (<https://creativecommons.org/licenses/by-nc/4.0/>)
which permits unrestricted non-commercial
use, distribution, and reproduction in any
medium, provided the original work is properly
cited.

ABSTRACT

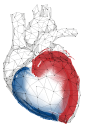
Background: Glycated hemoglobin (HbA1c), which reflects the patient's blood sugar level, can only be measured in a hospital setting. Therefore, we developed a model predicting HbA1c using personal information and self-monitoring of blood glucose (SMBG) data solely obtained by a patient.

Methods: Leave-one-out cross-validation (LOOCV) was performed at two university hospitals. After measuring the baseline HbA1c level before SMBG (Pre_HbA1c), the SMBG was recorded over a 3-month period. Based on these data, an HbA1c prediction model was developed, and the actual HbA1c value was measured after 3 months. The HbA1c values of the prediction model and actual HbA1c values were compared. Personal information was used in addition to SMBG data to develop the HbA1c predictive model.

Results: Thirty model training sessions and evaluations were conducted using LOOCV. The average mean absolute error of the 30 models was 0.659 (range, 0.005–2.654). Pre_HbA1c had the greatest influence on HbA1c prediction after 3 months, followed by post-breakfast blood glucose level, oral hypoglycemic agent use, fasting glucose level, height, and weight, while insulin use had a limited effect on HbA1c values.

Conclusions: The patient's SMBG data and personal information strongly influenced the HbA1c predictive model. In the future, it will be necessary to develop sophisticated predictive models using large samples for stable SMBG in patients.

Keywords: Blood glucose self-monitoring; Diabetes mellitus; Glycated hemoglobin A

**ORCID iDs**

Hyeongkyu Park 
<https://orcid.org/0000-0002-1832-7410>
Da Young Lee 
<https://orcid.org/0000-0002-8211-2393>
So young Park 
<https://orcid.org/0000-0002-7973-5944>
Jiyoung Min 
<https://orcid.org/0000-0002-5046-3316>
Jiwon Shinn 
<https://orcid.org/0000-0002-9131-6128>
Dae Ho Lee 
<https://orcid.org/0000-0002-8832-3052>
Soon Hyo Kwon 
<https://orcid.org/0000-0002-4114-4196>
Hun-Sung Kim 
<https://orcid.org/0000-0002-7002-7300>
Nan Hee Kim 
<https://orcid.org/0000-0003-4378-520X>

Funding

This research was supported by the Bio & Medical Technology Development Program of the National Research Foundation (NRF) funded by the Korean government (MSIT) (No. 2019M3E5D3073104).

Conflict of Interest

The authors have no financial conflicts of interest.

Author Contributions

Conceptualization: Park H, Kim HS, Kim NH;
Data curation: Lee DY, Park Sy, Min J, Shinn J, Kim HS, Kim NH; Formal analysis: Lee DY, Park Sy, Min J, Kim HS, Kim NH; Methodology: Kim HS, Kim NH; Writing - original draft: Park H, Kim HS, Kim NH; Writing - review & editing: Lee DH, Kwon SH.

INTRODUCTION

Several challenges are encountered when assessing the status of patients with diabetes from a long-term perspective based only on their self-monitoring of blood glucose (SMBG). Therefore, a method for checking the patient's blood sugar management status by measuring the glycated hemoglobin (HbA1c) is currently in use.^{1,2)} HbA1c reflects the average glycemic control over the previous 2–3 months,³⁾ which can be easily used for blood glucose measurement in patients, as well as for the screening, diagnosis, and treatment of diabetes.^{4,5)} Notably, HbA1c has shown a strong correlation with possible diabetic complications, such as cerebrocardiovascular diseases.⁶⁾

While SMBG can be performed at home, it may be inconvenient for patients to visit a hospital regularly and have their HbA1c levels measured. Therefore, a method for estimating the HbA1c value using SMBG at home would be convenient. Moreover, if the patient's HbA1c level is predictable, they will be able to set their own blood sugar goals and focus better on self-management.

Various HbA1c prediction models using SMBG exist⁷⁻⁹⁾; therefore, we aimed to create a model that predicts HbA1c in real-life situations using other variables in combination with the SMBG data. The actual HbA1c levels measured in the hospital and the predicted HbA1c values were compared. Furthermore, we evaluated the importance of factors affecting the prediction of HbA1c. The ability to identify the factors affecting HbA1c prediction in advance will be of great help to patients in self-glycemic management.

METHODS

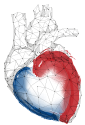
This study included patients with type 2 diabetes who visited Korea University Ansan Hospital and Soonchunhyang University Seoul Hospital between April 1 and August 31, 2020. We included patients aged 40–80 years who had visited the hospital for at least 2 years. Patients who agreed to participate in the study and fulfilled the selection criteria were included. In contrast, those with type 1 diabetes, an estimated glomerular filtration rate <30, kidney transplant, and ongoing dialysis were excluded from this study. A basic physical examination was performed for all included patients, and they were asked to record their blood sugar levels in a diary at home for a period of 3 months. This study was approved by the Institutional Review Boards of Korea University Ansan Hospital and Soonchunhyang University Seoul Hospital (IRB No. 2019AS0226).

Variables in the HbA1c predictive model

The variables of the HbA1c predictive model included demographic and SMBG data. Basic information, including sex, height, weight, and body mass index (BMI) of the patient at the first visit, HbA1c before SMBG (Pre_HbA1c), and insulin use were investigated. The SMBG data comprised blood glucose records from the diary maintained by the patient for 3 months, after which, the HbA1c (Post_HbA1c) was rechecked. The patient's SMBG values comprised the average values of fasting, post-breakfast, post-lunch, and post-dinner blood sugar.

Development and testing of the HbA1c predictive model

In this study, the eXtreme Gradient Boosting (XGBoost)¹⁰⁾ was used for the development of the HbA1c predictive model. LOOCV, a useful method for estimating the performance of a



small dataset, was used to train and evaluate the model. LOOCV used only one of the 30 data sets as the “test set” and the remaining 29 data sets as the “training set.” Thus, the process of training and testing the model was repeated 30 times. The mean absolute error (MAE) was used as the model test indicator. The Shapley value was used to measure the feature contribution to the model prediction.¹¹⁾¹²⁾

$$\text{MAE} = \frac{\sum_{i=1}^n |y_i - x_i|}{n}$$

y_i = prediction

x_i = true value

n = total number of data points

Statistical analysis

Descriptive statistics are presented as means (\pm standard deviation) or percentages. Python version 3.8.5 was used to perform the statistical analyses.

RESULTS

Study population

The mean age of the patients was 66.2 ± 8.0 years, and 63.3% (19/30) were male (Table 2). The mean BMI was 26 ± 2.8 kg/m², Pre_HbA1c was $7.3\% \pm 1.1\%$, and Post_HbA1c was $7.2\% \pm 1.0\%$ (Table 1). Moreover, 43.3% (13/30) of the patients were taking oral hypoglycemic agents (OHA) and 56.7% (17/30) were taking insulin. The mean fasting blood glucose level was 123 ± 17 mg/dL, whereas the mean blood glucose values post-breakfast, -lunch, and -dinner were 177 ± 44 mg/dL, 179 ± 60 mg/dL, and 170 ± 58 mg/dL, respectively. The average number of fasting blood glucose measurements conducted over the 3-month period was 59.4 ± 37.4 , while the average number of blood glucose measurements conducted post-breakfast, -lunch, and -dinner over the 3-month period was 25.7 ± 32.8 , 20.8 ± 35.5 , and 30.2 ± 37.8 , respectively.

Missing SMBG values

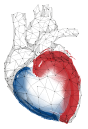
Post-lunch blood glucose levels were not recorded in 46.7% (14/30) of the patients. The rate of missing post-breakfast and post-dinner blood glucose values was 20% (6/30 patients). As the XGBoost model is able to handle missing values,¹³⁻¹⁵⁾ we did not proceed with separate imputation.

Model development and performance evaluation

In this study, 30 model training and evaluations were conducted using LOOCV (Table 2); the average MAE of the 30 models was 0.659 (range, 0.005–2.654). We visualized and compared the predicted HbA1c values of the XGBoost models and the true HbA1c values (Figure 1). If the model's predicted value and true value match, the data were located on the green dotted line. The solid blue line represents the “line of best fit” for the point, showing that the slope is close to the green dotted line.

Variables affecting the HbA1c predictive model

Among the 30 model predictions, the 10 variables that contributed the most to the predicted HbA1c values were selected (Figure 2). Pre_HbA1c had the greatest effect on HbA1c prediction

**Table 1.** Baseline characteristics

No.	True HbA1c	Predicted HbA1c	Differences	MAE
1	6.9	7.1	+0.2	0.188
2	7.2	7.0	-0.2	0.165
3	7.3	6.9	-0.4	0.402
4	7.0	7.7	+0.7	0.739
5	5.9	6.2	+0.3	0.277
6	6.4	6.6	+0.2	0.220
7	6.7	6.5	-0.2	0.201
8	8.0	6.0	-2.0	1.988
9	6.3	6.1	-0.2	0.226
10	8.6	7.4	-1.2	1.150
11	7.1	7.0	-0.1	0.066
12	7.2	6.5	-0.7	0.704
13	6.3	7.0	+0.7	0.702
14	5.5	6.8	+1.3	1.325
15	6.4	7.1	+0.7	0.703
16	8.4	7.6	-0.8	0.752
17	6.7	7.0	+0.3	0.268
18	8.2	8.1	-0.1	0.118
19	8.1	8.0	-0.1	0.059
20	9.8	7.6	-2.2	2.162
21	6.4	8.7	+2.3	2.328
22	6.2	6.6	+0.4	0.371
23	8.4	8.6	+0.2	0.189
24	7.1	6.9	-0.2	0.165
25	6.6	6.9	+0.3	0.344
26	7.2	7.2	0.0	0.005
27	9.1	6.4	-2.7	2.654
28	6.3	6.3	0.0	0.008
29	6.0	6.5	+0.5	0.479
30	7.8	7.0	-0.8	0.826

Values are presented as number (%) or mean±standard deviation.

BMI = body mass index; DM = diabetes mellitus; HbA1c = glycated hemoglobin; OHA = oral hypoglycemic agents.

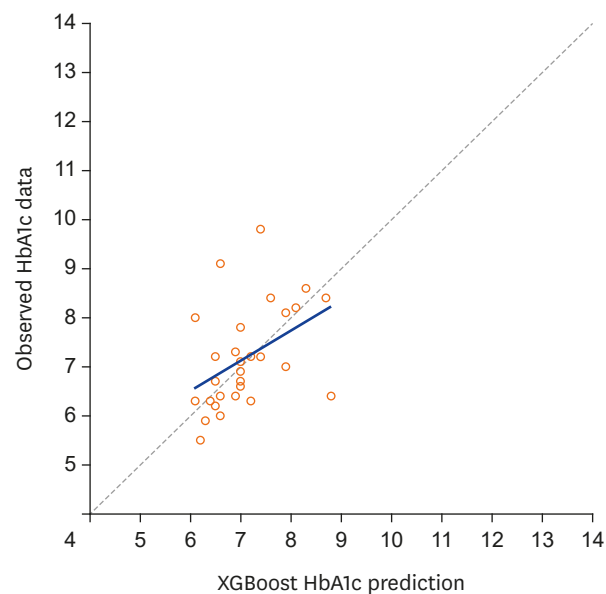
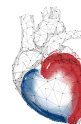
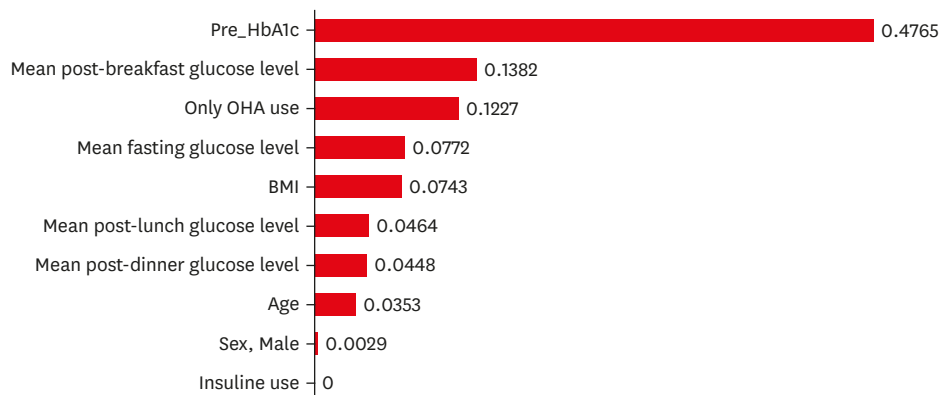


Figure 1. Plot of the predicted and observational glycated hemoglobin data. Solid blue line represents the “line of best fit” for the point.
HbA1c = glycated hemoglobin.

**Table 2.** Performance evaluation of each model using the leave-one-out cross-validation method

Variables	Values
Age (yr)	66.2±8.0
Sex	
Male	19 (63.3)
Female	11 (36.7)
Height (cm)	162.3±7.7
Weight (kg)	68.4±9.1
BMI (kg/m ²)	26.0±2.8
Pre_HbA1c (%)	7.3±1.1
<7.0	12 (40.0)
7.0–7.9	11 (36.7)
8.0–8.9	3 (10.0)
≥9.0	4 (13.3)
Post_HbA1c (%)	7.2±1.0
DM medication	
Only OHA use	13 (43.3)
Insulin use	17 (56.7)
Mean glucose level	
Fasting (mg/dL)	123±17
Post-breakfast (mg/dL)	177±44
Post-lunch (mg/dL)	179±60
Post-dinner (mg/dL)	170±58
Time of glucose level check	
Fasting	59.4±37.3
Post-breakfast	25.7±32.8
Post-lunch	20.8±35.5
Post-dinner	30.2±37.8

HbA1c = glycated hemoglobin; MAE = mean absolute error.

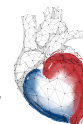
**Figure 2.** The contributors to model predictions.

BMI = body mass index; HbA1c = glycated hemoglobin; OHA = oral hypoglycemic agents.

after 3 months, followed by the post-breakfast blood glucose level, OHA use, fasting glucose level, height, and weight, while insulin use had a limited effect on the HbA1c prediction.

DISCUSSION

In this study, a model was developed to predict the HbA1c value using the patient's SMBG data. Moreover, the factors affecting it were identified. By providing this information to the patient, it is possible to predict the HbA1c value without the need for frequent hospital visits



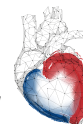
to undergo a separate test. Thus, this study provides a convenient method for the SMBG in patients by enabling easy prediction of HbA1c values.

A model that enables the prediction of HbA1c using SMBG data recorded in the patients' homes was developed. The SMBG data were divided into four types: fasting glucose, post-breakfast, post-lunch, and post-dinner glucose levels. Predictive models depend on how much SMBG data must be obtained to produce relevant results⁽⁶⁾⁽¹⁷⁾ and the limit of missing data allowed; this is because it is inconvenient for patients to measure all of the required blood glucose values daily.⁽¹⁸⁾ Moreover, in this study, the proportion of missing data was 20.0–46.7%, even though the average value of each postprandial glucose value was used. Therefore, we used XGBoost for the HbA1c prediction model.⁽¹⁰⁾⁽¹⁹⁾ XGBoost not only shows excellent performance in standardized data classification and prediction problems, but also permits cross-validation, and the missing values can be handled by themselves. Although a significant amount of postprandial glucose values were missing in this study, it was possible to proceed with the study using XGBoost without separate imputation.

Our study included 30 patients, which is a small number for creating a predictive model. Moreover, the LOOCV method used in this study requires a long time to develop the model because it depended on the amount of data.⁽¹³⁾⁽¹⁵⁾⁽²⁰⁾ Hence, LOOCV is often used to measure the performance of a relatively small data sample.⁽¹³⁾⁽¹⁵⁾ LOOCV uses one of the N datasets as the “test set” and the remaining $N-1$ data sets as the “training set”⁽²¹⁾; this process is repeated N times. The advantage of LOOCV is that since all samples are tested once, randomization is inexistent, and unlike the validation set approach, it is possible to obtain very reliable results. Furthermore, because only one sample was used as a “test set,” it was possible to create a model using a large amount of training data. However, it is difficult to include model diversity in LOOCV. Although, considering that our study is a pilot study concept, the use of LOOCV seemed appropriate.

Various models have been developed for predicting HbA1c.⁽⁷⁻⁹⁾ One such model predicted HbA1c based on the lifestyle, clinical, and biochemical information obtained at a health checkup center,⁽⁷⁾ while another predicted HbA1c after 6 years using various laboratory findings.⁽⁸⁾ Recently, given the diversity in the HbA1c prediction models, various laboratory findings are commonly used rather than simple SMBG data.⁽⁹⁾ Therefore, in this study, SMBG and simple personal information were included in the prediction model. Additionally, the pre-HbA1c value was added to reflect the patient's past self-glycemic control status.

The results of this study showed that the Pre_HbA1c value had the greatest influence on the HbA1c prediction model (**Figure 2**). Theoretically, Pre_HbA1c and Post_HbA1c are independent variables that do not affect HbA1c prediction. Although careful interpretation can be conducted in various ways, the most important reason for using SMBG data obtained from a patient's diary is that it is recorded in real time.⁽¹⁷⁾⁽²²⁾ Pre_HbA1c reflects the patient's past blood sugar management pattern and status. Although only the patient's average blood glucose value for 3 months can affect the prediction of HbA1c, management pattern and habits cannot be changed easily, which is thought to influence HbA1c prediction. This is considered a relevant finding. The prediction of the HbA1c value was only affected by fasting blood glucose or postprandial blood glucose levels; however, Pre_HbA1c demonstrated a rather significant effect, which may be related to the missing blood glucose values of the patients. This finding suggests that the Pre_HbA1c value compensates for the missing values; therefore, Pre_HbA1c was presumed to be the most powerful predictor of HbA1c in this study.



However, there are certain limitations in applying the results of this study. When patients make strong decisions about their blood sugar management and change their diet/exercise management rapidly, there is a high possibility that the predicted HbA1c value may be inaccurate.

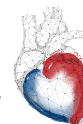
Various HbA1c prediction models are continuously being developed,⁷⁻⁹⁾ but researchers who wish to develop a prediction model in the future will have to consider several factors. For instance, it is necessary to consider whether to include personal information, such as Pre_HbA1c, as a variable rather than SMBG data only. Ultimately, the answer depends on the purpose of the study,¹⁶⁾ and it seems that a broad definition of predicted HbA1c is required.

If the main purpose of developing a prediction model for HbA1c is to improve the patient's blood glucose level or if the amount of SMBG data is sufficient, it would be more appropriate to use simple SMBG data. However, if there are few SMBG data or no significant difference occurred in the patients' willingness to manage blood sugar, it would be better to include data on personal information in addition to simple SMBG data in the predictive model. The fact that Pre_HbA1c had a significant effect in this predictive model suggests that there was almost no change in the pattern of the patient's blood glucose management. Given the retrospective nature of the cohort study, which can only estimate correlation and not causation,¹⁶⁾¹⁷⁾ we can assume that the study was conducted with patients exhibiting limited changes in blood glucose patterns.

Among the other factors influencing HbA1c prediction, the influence of post-breakfast blood glucose and fasting blood glucose was high in this study. This is theoretically consistent with the results of the original HbA1c prediction model, wherein SMBG correlated with the predicted HbA1c to some extent.²³⁾²⁴⁾ The influence of each postprandial glucose level on the prediction of HbA1c should be studied using large samples in the future. Furthermore, OHA use had a much greater influence than insulin use. For the latter, it is presumed that the predicted HbA1c value was affected by blood glucose changes as the insulin dose was gradually adjusted.

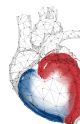
This pilot study was conducted with a small sample size; therefore, several limitations may have occurred. Considering that Pre_HbA1c is included as a variable, it is necessary to thoroughly evaluate the interpretation, which is different from the actual result of the patient. The results of this study may contribute to the development of various predictive models in the future, although it is difficult to generalize the research results. Hence, future studies should include larger samples and more variables.

Despite its limitations, the results of this study showed that self-management could be facilitated by allowing the patient to check their HbA1c level without visiting a hospital. By enabling easy prediction of HbA1c, early recognition of the degree of blood sugar control and blood sugar management status can be achieved, which would help patients in managing their blood sugar levels voluntarily and actively. Thus, we look forward to the creation of more diverse and sophisticated predictive models, and that more studies will be conducted to help patients manage their blood glucose levels. The existing results are insufficient, and it will be necessary to develop a model with high potential for practical use in the future by securing a large sample and more sophisticated methods of analysis.



REFERENCES

1. Sacks DB, Bruns DE, Goldstein DE, Maclaren NK, McDonald JM, Parrott M. Guidelines and recommendations for laboratory analysis in the diagnosis and management of diabetes mellitus. *Clin Chem* 2002;48:436-72.
[PUBMED](#) | [CROSSREF](#)
2. American Diabetes Association. Standards of medical care in diabetes. *Diabetes Care* 2004;27 Suppl 1:S15-35.
[PUBMED](#) | [CROSSREF](#)
3. Nathan DM, Kuenen J, Borg R, Zheng H, Schoenfeld D, Heine RJ; A1c-Derived Average Glucose Study Group. Translating the A1c assay into estimated average glucose values. *Diabetes Care* 2008;31:1473-8.
[PUBMED](#) | [CROSSREF](#)
4. Nathan DM, Genuth S, Lachin J, Cleary P, Crofford O, Davis M, Rand L, Siebert C; Diabetes Control and Complications Trial Research Group. The effect of intensive treatment of diabetes on the development and progression of long-term complications in insulin-dependent diabetes mellitus. *N Engl J Med* 1993;329:977-86.
[PUBMED](#) | [CROSSREF](#)
5. UK Prospective Diabetes Study (UKPDS) Group. Intensive blood-glucose control with sulphonylureas or insulin compared with conventional treatment and risk of complications in patients with type 2 diabetes (UKPDS 33). *Lancet* 1998;352:837-53.
[PUBMED](#) | [CROSSREF](#)
6. Singer DE, Nathan DM, Anderson KM, Wilson PW, Evans JC. Association of HbA1c with prevalent cardiovascular disease in the original cohort of the Framingham Heart Study. *Diabetes* 1992;41:202-8.
[PUBMED](#) | [CROSSREF](#)
7. Chien KL, Lin HJ, Lee BC, Hsu HC, Chen MF. Prediction model for high glycated hemoglobin concentration among ethnic Chinese in Taiwan. *Cardiovasc Diabetol* 2010;9:59.
[PUBMED](#) | [CROSSREF](#)
8. Huang CL, Iqbal U, Nguyen PA, Chen ZF, Clinciu DL, Hsu YH, Hsu CH, Jian WS. Using hemoglobin A1c as a predicting model for time interval from pre-diabetes progressing to diabetes. *PLoS One* 2014;9:e104263.
[PUBMED](#) | [CROSSREF](#)
9. Rauh SP, Heymans MW, Koopman AD, Nijpels G, Stehouwer CD, Thorand B, Rathmann W, Meisinger C, Peters A, de Las Heras Gala T, Glümer C, Pedersen O, Cederberg H, Kuusisto J, Laakso M, Pearson ER, Franks PW, Rutters F, Dekker JM. Predicting glycated hemoglobin levels in the non-diabetic general population: Development and validation of the DIRECT-DETECT prediction model - a DIRECT study. *PLoS One* 2017;12:e0171816.
[PUBMED](#) | [CROSSREF](#)
10. Tianqi C, Guestrin C. Xgboost: a scalable tree boosting system. *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*; 2016 Aug 13-17; San Francisco, CA. New York, NY: Association for Computing Machinery; 2016. pp.785-94.
11. Lundberg SM, Lee SI. A unified approach to interpreting model predictions. *Proceedings of the 31st International Conference on Neural Information Processing Systems*; 2017 Dec 4-9; Long Beach, CA. Red Hook, NY: Curran Associates Inc.; 2017. pp.4768-77.
12. Lundberg SM, Erion G, Chen H, DeGrave A, Prutkin JM, Nair B, Katz R, Himmelfarb J, Bansal N, Lee SI. From local explanations to global understanding with explainable AI for trees. *Nat Mach Intell* 2020;2:56-67.
[PUBMED](#) | [CROSSREF](#)
13. Shavitt I, Segal E. Regularization learning networks: deep learning for tabular datasets. *Proceedings of the 32nd International Conference on Neural Information Processing Systems* 2018; 2018 Dec 3-8; Montréal, Canada. Red Hook, NY: Curran Associates Inc.; 2018. pp.1386-96.
14. Abou Omar KB. XGBoost and LGBM for Porto Seguro's Kaggle challenge: a comparison semester project [Internet]. Zürich: ETH Zurich; 2018 [cited 2021 Sep 10]. Available from: <https://pub.tik.ee.ethz.ch/students/2017-HS/SA-2017-98.pdf>.
15. Cai H, Zhong R, Wang C, Zhou R, Zhou K, Lee H, Xu K, Gao Z, Zhong R, Luo J, Zhou Y, Ding M, Li L, Li Q, Li D, Jiang N, Cheng X, Cui S, Ye H, Shen J. KDD CUP 2017 travel time prediction [Internet]. KDD; 2017 [cited 2021 Sep 10]. Available from: https://www.kdd.org/kdd2017/files/Task1_3rdPlace.pdf.
16. Kim HS, Kim DJ, Yoon KH. Medical big data is not yet available: why we need realism rather than exaggeration. *Endocrinol Metab (Seoul)* 2019;34:349-54.
[PUBMED](#) | [CROSSREF](#)



17. Kim HS, Kim JH. Proceed with caution when using real world data and real world evidence. *J Korean Med Sci* 2019;34:e28.
[PUBMED](#) | [CROSSREF](#)
18. Hu ZD, Zhang KP, Huang Y, Zhu S. Compliance to self-monitoring of blood glucose among patients with type 2 diabetes mellitus and its influential factors: a real-world cross-sectional study based on the Tencent TDF-I blood glucose monitoring platform. *mHealth* 2017;3:25.
[PUBMED](#) | [CROSSREF](#)
19. Friedman JH. Greedy function approximation: a gradient boosting machine. *Ann Stat* 2001;29:1189-232.
[CROSSREF](#)
20. Yelin I, Snitser O, Novich G, Katz R, Tal O, Parizade M, Chodick G, Koren G, Shalev V, Kishony R. Personal clinical history predicts antibiotic resistance of urinary tract infections. *Nat Med* 2019;25:1143-52.
[PUBMED](#) | [CROSSREF](#)
21. DeCoste D, Wagstaff K. Alpha seeding for support vector machines. *Proceedings of the Sixth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*; 2000 Aug 20-23; Boston, MA. New York, NY: Association for Computing Machinery; 2000. pp. 345-9
22. Kim HS, Lee S, Kim JH. Real-world evidence versus randomized controlled trial: clinical research based on electronic medical records. *J Korean Med Sci* 2018;33:e213.
[PUBMED](#) | [CROSSREF](#)
23. Landgraf R. The relationship of postprandial glucose to HbA1c. *Diabetes Metab Res Rev* 2004;20 Suppl 2:S9-12.
[PUBMED](#) | [CROSSREF](#)
24. Ketema EB, Kibret KT. Correlation of fasting and postprandial plasma glucose with HbA1c in assessing glycemic control; systematic review and meta-analysis. *Arch Public Health* 2015;73:43.
[PUBMED](#) | [CROSSREF](#)