



Predictive models for posttransplant diabetes mellitus in kidney transplant recipients using machine learning and deep learning approach: a nationwide cohort study from South Korea

Seoyoung Choi^{1,*}, Mi Ryung Pyo^{2,*}, Sangwoong Kim³, Jong Cheol Jeong^{3,4}, Yu Ho Lee⁵, Hyejin Mo⁶, Jeong-Hoon Lee⁷, Jaeseok Yang⁸, Myoung Soo Kim⁹, Hye Eun Yoon^{2,†}, Sejoong Kim^{1,3,4,†}; the Korean Organ Transplantation Registry Study Group

For further information on the authors' affiliations, see **Additional information**.

Background: Posttransplant diabetes mellitus (PTDM) complicates kidney transplant recipients (KTRs) in morbidity and mortality. This study aimed to predict PTDM risk in KTRs using machine learning and deep learning models.

Methods: Data were obtained from the Korea Organ Transplantation Registry, a nationwide cohort study of KTRs. Four machine learning algorithms, including eXtreme Gradient Boosting (XGBoost), CatBoost, light gradient boosting machine and logistic regression, and deep learning were implemented on 41 pretransplant and 31 posttransplant variables to predict PTDM. Model performance was assessed using the area under the curve (AUC) of the receiver operating characteristic curve, accuracy, precision, recall, and F1 score.

Results: Among 3,213 KTRs, 497 patients (15.5%) developed PTDM within 1 year. The PTDM group had higher age, body mass index (BMI), triglyceride level, and prevalence of hypertension and cardiovascular disease, and lower total cholesterol level at baseline than the No-PTDM group. The XGBoost model showed the highest AUC (0.738) and F1 score (0.42), and modest accuracy (0.86), while the CatBoost model exhibited the highest accuracy (0.87) and precision (0.79). Feature importance in XGBoost was highest for recipient age, followed by baseline BMI, triglyceride level at posttransplant 6 months, baseline glycated hemoglobin and high-density lipoprotein cholesterol level, white blood cell (WBC) count and serum uric acid level at 6 months, baseline WBC count, and tacrolimus trough level at discharge.

Conclusion: The XGBoost model demonstrated the best performance for predicting PTDM within 1 year, offering an accurate tool for early identification and personalized care of high-risk KTRs for PTDM.

Keywords: Diabetes mellitus, Deep learning, Kidney transplantation, Machine learning

Received: April 22, 2024; **Revised:** September 29, 2024; **Accepted:** October 13, 2024

Correspondence: Sejoong Kim

Department of Internal Medicine, Seoul National University Bundang Hospital, 82 Gumi-ro 173beon-gil, Bundang-gu, Seongnam 13620, Republic of Korea. E-mail: sejoong2@snu.ac.kr
ORCID: <https://orcid.org/0000-0002-7238-9962>

Hye Eun Yoon

Division of Nephrology, Department of Internal Medicine, Incheon St. Mary's Hospital, College of Medicine, The Catholic University of Korea, 56 Dongsu-ro, Bupyeong-gu, Incheon 21431, Republic of Korea. E-mail: berryana@catholic.ac.kr
ORCID: <https://orcid.org/0000-0002-6347-7282>

*Seoyoung Choi and Mi Ryung Pyo contributed equally to this study as co-first authors.

†Sejoong Kim and Hye Eun Yoon contributed equally to this study as co-corresponding authors.

© 2025 by The Korean Society of Nephrology

© This is an Open Access article distributed under the terms of the Creative Commons Attribution Non-Commercial and No Derivatives License (<https://creativecommons.org/licenses/by-nc-nd/4.0/>) which permits unrestricted non-commercial use, distribution of the material without any modifications, and reproduction in any medium, provided the original works properly cited.

Introduction

Posttransplant diabetes mellitus (PTDM) is a common metabolic complication after kidney transplantation (KT), with an incidence ranging from 7% to 40% [1,2]. PTDM affects outcomes of KT recipients (KTRs), in terms of graft failure, cardiovascular disease, and mortality [3]. In addition, PTDM is associated with poor quality of life and increased healthcare costs [3]. Risk factors for PTDM are known as age, obesity, acute rejection, virus infection including cytomegalovirus and hepatitis B and C, and immunosuppressants [4–7]. Identifying KTRs with high risk for developing PTDM is crucial, since early diagnosis and management will help to reduce PTDM-related morbidity and mortality. Therefore, a risk scoring system will be helpful to allow robust pre- and posttransplant assessment of PTDM. For a generalizable scoring system, a larger cohort using a multicenter database is needed. In addition, a relatively recent database will be more useful as it will reflect the current immunosuppressive strategies.

Machine learning is used to create predictive models from data. Owing to high-performance computing, data availability, and algorithmic innovations, it effectively analyzes a large dataset [8]. Deep learning is an advanced subset of machine learning that uses artificial neural networks and big data [9]. Machine learning has the potential to detect possible interactions and new relationships between variables from a large dataset, which may provide more accurate prognostic models in medicine. Therefore, the aim of this study was to build a predictive model for PTDM using machine learning and deep learning algorithms from a nationwide multicenter cohort of KTRs.

Methods

Study population and data collection

Data were obtained from the Korea Organ Transplantation Registry (KOTRY), a prospective multicenter nationwide cohort study of KTRs in Korea. Forty-one transplantation centers participated in the KOTRY. Following the establishment of KOTRY in 2014, data for this study were requested in 2020 and subsequent multifaceted analyses have been conducted since. Consequently, the dataset included all 6,455 KTRs aged 18 years or more who underwent KT be-

tween May 2014 and August 2020 were included in this study. The KOTRY provided patient demographics and clinical and laboratory data at the time of transplantation. Follow-up data 6 months and 1 year after KT were collected including laboratory data, antirejection treatment, and complications of KT including PTDM. Data on medications taken posttransplantation including vitamin D analogs, tacrolimus, cyclosporine, mycophenolate acid, mammalian target of rapamycin inhibitor, and corticosteroids were also collected. Estimated glomerular filtration rate was calculated using the CKD-EPI (Chronic Kidney Disease Epidemiology Collaboration) equation [10]. Body mass index (BMI) was calculated as the patient's weight in kilograms divided by height in meters squared (kg/m^2). The original dataset consisted of a total of 236 variables related to the recipient and 125 variables related to the donor.

All patients provided written informed consent before KOTRY enrollment. The study was performed in line with the principles of the Declaration of Helsinki and approved by the Institutional Review Board of The Catholic University of Korea, Incheon St. Mary's Hospital (OC19ONDI0034).

Study flow chart and endpoint

Based on the criteria of American Diabetes Association/World Health Organization [11], patients were diagnosed as diabetes mellitus (DM) when they are subject to at least one of the following features: 1) fasting plasma glucose level ≥ 126 mg/dL, 2) random plasma glucose level ≥ 200 mg/dL, 3) 2-hour glucose after a 75-g oral glucose tolerance test of more than 200 mg/dL, and 4) hemoglobin A1c (HbA1c) $\geq 6.5\%$. Since the KOTRY data included patients whose HbA1c was $\geq 6.5\%$ or random plasma glucose was ≥ 200 mg/dL but not diagnosed as DM, DM diagnosis errors had to be revised first. Missing values and outliers of HbA1c and fasting plasma glucose at baseline were substituted with the median value. Missing values and outliers of follow-up serum plasma glucose 6 months and 1 year after KT were filled in using carry forward method.

KTRs who had been diagnosed as pretransplant DM ($n = 2,006$) and those who had been additionally determined as pretransplant DM following the criteria of the American Diabetes Association/World Health Organization [11] were excluded ($n = 771$), leaving 3,678 patients in the cohort. After excluding subjects without post-KT data at 6 months

and 1 year ($n = 465$), a total of 3,213 patients were included in the final analysis (Fig. 1).

The endpoint of this study was newly diagnosed PTDM within 1 year after KT. The data on the development of PTDM was recorded 6 months and 1 year after KT in the database. Additionally, PTDM was diagnosed from fasting plasma glucose level [12]. Four-hundred ninety-seven KTRs (15.5%) developed PTDM within 1 year.

Missing data and outlier preprocessing

For every feature, outliers were replaced with blank for the correct analysis of value distribution and missing percentages. Missing value percentages (proportion of number of patients missing value for the feature to the number of total patients) were calculated for every feature. For binary features, the proportion of the number of patients with the value 0, referred to as zero ratio from now on, was calculated. Missing values in all the selected features were substituted with representative values. Mean was used for con-

tinuous features; mode was used for binary and categorical features; interpolation of forward method was used for follow-up features. Derivative binary features were merged as categorical features.

Feature selection

Among the 236 variables, features to be used for training were selected according to a few criteria; baseline characteristics, laboratory values, medications, and complications relevant to allograft function and DM. Features with missing percentage higher than 50%, binary features with zero ratio higher than 90% or lower than 10% were eliminated.

As exceptions, features of patients' human leukocyte antigen alleles were eliminated, for the reason of excessive category numbers. Features for delayed graft function and anti-thymoglobulin dosage were included despite the missing percentage standard considering their medical importance in contribution to DM.

The 41 baseline variables and 31 follow-up variables included in the final analysis are shown in Table 1.

Training-machine learning

Fig. 2 shows the schema of machine learning. KTRs were classified into two classes based on the occurrence of PTDM within 1 year. Data set was split into training, validation, and test set as 7.5 to 1.5 to 1.5 ratio. Baseline and follow-up features were spread out and input into eXtreme Gradient Boosting (XGBoost), CatBoost, light gradient boosting machine (GBM) and logistic regression model. The model was optimized by finding the optimum hyperparameters through 300 trials of using Optuna library. StratifiedKFold cross validation was implemented to increase the validity, splitting the data set into five validation sets and assessing the model on each. Performances of algorithm were calculated by area under the curve (AUC) of the receiver operating characteristic (ROC) curve with 95% confidence intervals (CIs). The importance of each feature was determined using SHAP (SHapley Additive exPlanations) method.

Confusion matrix was generated from the prediction results, which had four categories for correct and incorrect prediction. The categories for correct predictions were true positive (TP) and true negative (TN), whereas those for

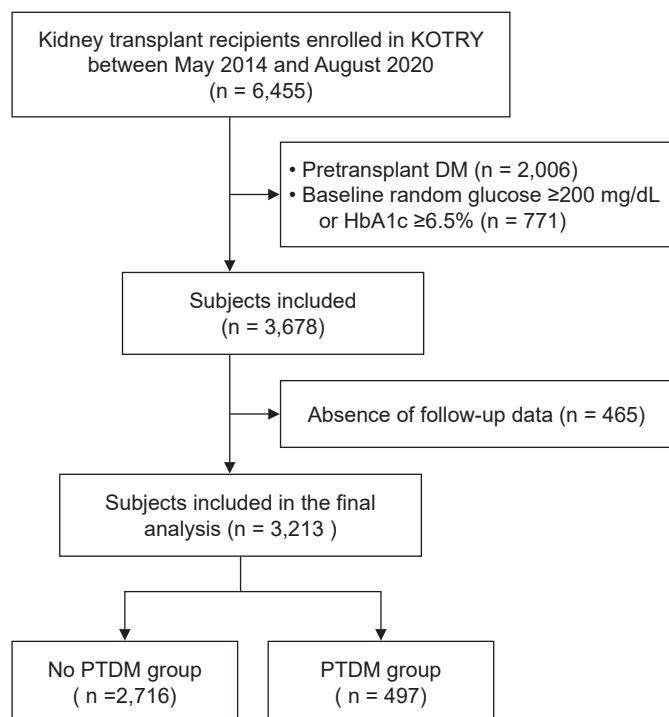


Figure 1. Flow chart of the study.

DM, diabetes mellitus; HbA1c, hemoglobin A1c; KOTRY, Korea Organ Transplantation Registry; PTDM, posttransplant diabetes mellitus.

Table 1. Features included in the machine learning and deep learning models

Baseline data (20 variables)	Baseline and follow-up data 6 months after KT (21 variables)	Follow-up data (10 variables)
Recipient sex	Use of vitamin D analog	Delayed graft function
Recipient age	Use of statin	Tacrolimus dosage at discharge and at 6 months
Donor age	Use of aspirin	Tacrolimus level at discharge and at 6 months
Donor sex	Body weight	Steroid dosage at discharge and at 6 months
Donor type (deceased or living)	SBP	Mycophenolate acid dosage at discharge and at 6 months
Retransplant	DBP	Rejection within 6 months
Desensitization	Heart rate	
Primary renal disease	eGFR	
Body mass index	WBC	
Height	Hemoglobin	
RRT before KT	Platelet count	
Recipient ABO blood type	Blood urea nitrogen	
Hemoglobin A1c	Serum albumin	
Intact PTH	Serum calcium	
PRA class I and II	Serum phosphorus	
Smoking	Serum uric acid	
Anti-thymocyte globulin	Total cholesterol	
IL-2 receptor antibody	Triglyceride	
History of CVD	LDL-C	
History of tumor	HDL-C	
	hsCRP	

CVD, cardiovascular disease; DBP, diastolic blood pressure; eGFR, estimated glomerular filtration rate; HDL-C, high-density lipoprotein cholesterol; hsCRP, high sensitivity C-reactive protein; IL-2, interleukin-2; KT, kidney transplantation; LDL-C, low density lipoprotein cholesterol; PRA, panel reactive antibody; PTH, parathyroid hormone; RRT, renal replacement therapy; SBP, systolic blood pressure; WBC, white blood cell.

incorrect predictions were false positive (FP) and false negative (FN). The four categories were used to calculate metrics for assessing model performance: accuracy, precision, recall (sensitivity), and F1 score.

$$\text{Accuracy} = (\text{TP} + \text{TN}) / (\text{TP} + \text{TN} + \text{FP} + \text{FN})$$

$$\text{Precision} = \text{TP} / (\text{TP} + \text{FP})$$

$$\text{Recall (sensitivity)} = \text{TP} / (\text{TP} + \text{FN})$$

$$\text{F1 core} = (\text{precision} \times \text{recall}) / (\text{precision} + \text{recall})$$

Training-deep learning

Fig. 3 shows the schema of deep learning. KTRs were classified into two classes based on the occurrence of PTDM within 1 year. Data set was split into training, validation, and test set as 7.5 to 1.5 to 1.5 ratio. On the baseline and rejection features, multi-layer perceptron (consisting of linear transformation and nonlinear activation function such as rectified linear unit) was done to generate the first output. On the follow-up features, long short-term memory

(used for dealing with dynamic data and sequential pattern learning) was applied to generate the second output. The outputs were passed through a linear layer to derive the final output. Weighted CrossEntropyLoss was used to calculate loss to ease the class imbalance problem, and weighted decay was added to optimizer Adam. Performances of algorithm were calculated by AUC of the ROC with 95% CI, accuracy, recall, precision, and F1 score.

Statistical analysis

Statistical analysis was performed using SAS version 9.4 (SAS Institute). Continuous variables were presented as mean and standard deviation for data with normal distribution and presented as median and interquartile ranges for data with nonparametric distribution. After distribution of data between PTDM group and No-PTDM group was determined, they were compared using independent t test or Wilcoxon rank sum test. Categorical data was presented

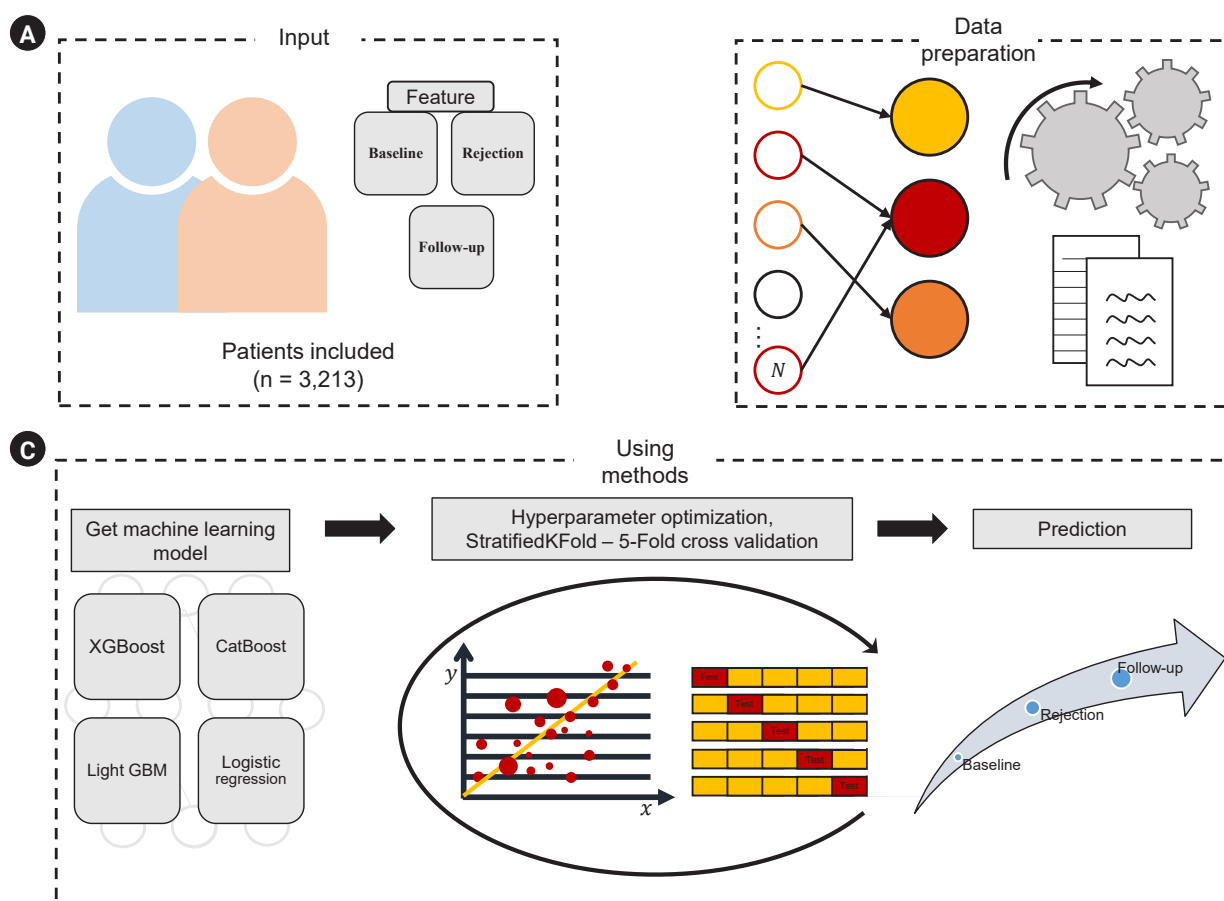


Figure 2. Schema of machine learning. (A) Baseline and follow-up variables including rejection data were included in the analysis. (B) Data preparation includes data cleaning, data transformation, feature engineering, data normalization and standardization, data splitting, and data augmentation. (C) Baseline and follow-up features were spread out and input into eXtreme Gradient Boosting (XGBoost), CatBoost, light gradient boosting machine (GBM), and logistic regression model. The model was optimized by hyperparameter optimization and StratifiedKFold. Five-fold cross validation was performed by splitting the data set into five validation sets and assessing the model on each.

as percentages and comparison between the two groups was performed using chi-square test and Fisher exact test. The *p*-values of <0.05 were considered significant.

Results

Baseline characteristics

Among 3,213 KTRs, 497 patients (15.5%) developed PTDM during 1 year after KT. Baseline characteristics of PTDM group versus No-PTDM group are shown in Table 2. The PTDM group was older and had a higher BMI and triglyceride level and a lower total cholesterol level at baseline than the No-PTDM group. The prevalence of hypertension and

cardiovascular disease and the proportion of hypertension or others as primary renal disease were higher in the PTDM group compared to the No-PTDM group.

Model performance of machine learning and deep learning

To analyze the statistical performance of the four machine learning models and deep learning model for PTDM prediction, we assessed AUC of ROC, accuracy, precision, recall (sensitivity), and F1 score (Table 3, Fig. 4). The XGBoost model showed the highest AUC (0.738; 95% CI, 0.677–0.798), which was followed by CatBoost (AUC = 0.727; 95% CI, 0.667–0.785), logistic regression (AUC = 0.719; 95% CI, 0.651–0.784), light GBM (AUC = 0.716; 95% CI, 0.654–0.777),

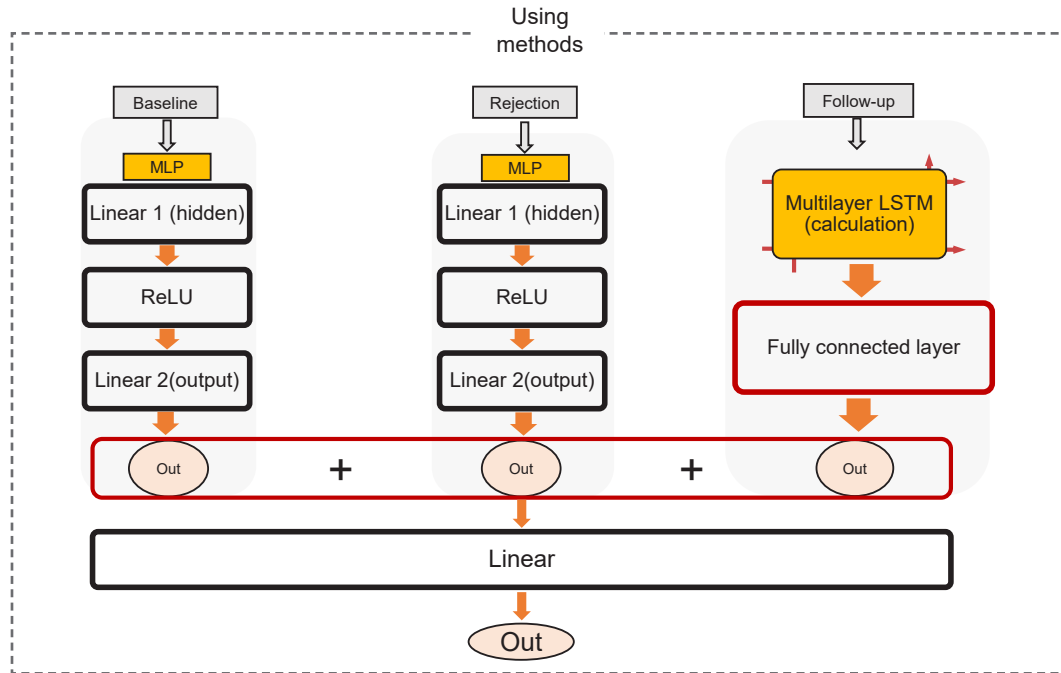


Figure 3. Schema of deep learning. On the baseline features and rejection features, a multilayer perceptron (MLP) was done to generate the first output. On the follow-up features, long short-term memory (LSTM) was applied to generate the second output. The three outputs were passed through a linear layer to derive the final output.

and deep learning (AUC = 0.699; 95% CI, 0.639–0.759).

While the model run on CatBoost exhibited the highest accuracy (0.87) and precision (0.79), the XGBoost model showed the highest AUC (0.738) and F1 score (0.42), and modest accuracy (0.86). Therefore, the XGBoost model showed the best performance among five models.

The incorporation of transplant-related risk factors was trialed, with hepatitis B and C virus infection and cytomegalovirus immunity additionally evaluated as machine learning variables. This approach led to an overall decline in performance, including reductions in AUC across all models, resulting in the exclusion of these variables from the final model. The performance metrics from this assessment are detailed in [Supplementary Table 1](#) (available online).

Feature importance of the XGBoost model

[Fig. 5](#) shows the importance of each feature of the XGBoost model determined by SHAP method. Feature importance was highest for recipient age, which was followed by baseline BMI, triglyceride level at post-KT 6 months, HbA1c

and high-density lipoprotein cholesterol (HDL-C) level at baseline, white blood cell count and serum uric acid level at 6 months, white blood cell count at baseline, tacrolimus trough level at discharge, and total cholesterol level at post-KT 6 months.

Discussion

We developed a machine learning-based prediction model for PTDM after 1 year of KT using a nationwide multicenter cohort of KTRs. The model incorporated a total of 72 variables, both at pre- and posttransplantation. These variables are used in clinical practice and can be extracted from electric medical records. Among four machine learning models and a deep learning model, XGBoost showed the best performance with an AUC of 0.738.

In this study, the incidence of PTDM within 1-year post-transplantation was 15.5%. The incidence of PTDM in KTRs is reported to range between 7% and 40% [1,2]. The variable incidence is related to variable diagnosis criteria and different timepoints of diagnosis after KT [13]. In contrast to the diagnosis criteria of type 2 DM utilizing the HbA1c, it

Table 2. Baseline characteristics of the kidney transplant recipients

Characteristic	Total	No-PTDM group	PTDM group	p-value
No. of recipients	3,213	2,716	497	
Recipient age (yr)	47.02 ± 11.69	46.02 ± 11.72	52.47 ± 9.87	<0.001
Recipient male sex	1,776 (55.3)	1,485 (54.7)	291 (58.6)	0.12
Donor age (yr)	47.30 ± 12.59	47.15 ± 12.48	48.12 ± 13.16	0.12
Donor male sex	1,762 (54.8)	1,493 (55.0)	269 (54.1)	0.77
Deceased donor	1,318 (41.0)	1,101 (40.5)	217 (43.7)	0.21
BMI (kg/m ²)	22.62 ± 3.41	22.45 ± 3.37	23.57 ± 3.43	<0.001
Smoking (%)	1.41 ± 0.62	1.41 ± 0.62	1.38 ± 0.57	0.06
Never	2,021 (62.9)	1,701 (62.6)	320 (64.4)	
Current	827 (25.7)	689 (25.4)	138 (27.8)	
Former	211 (6.6)	189 (7.0)	22 (4.4)	
Unknown	154 (4.8)	137 (5.0)	17 (3.4)	
Hypertension	2,807 (87.4)	2,350 (86.5)	457 (92.0)	0.001
History of CVD	230 (7.2)	177 (6.5)	53 (10.7)	0.001
Myocardial infarction	8 (0.2)	8 (0.3)	0 (0)	
Angina	16 (0.5)	12 (0.4)	4 (0.8)	
Heart failure	6 (0.2)	5 (0.2)	1 (0.2)	
Others	18 (0.6)	13 (0.5)	5 (1.0)	
Primary renal disease				<0.001
Hypertension	654 (20.4)	528 (19.4)	126 (25.4)	
Glomerulonephritis	1,423 (44.3)	1,239 (45.6)	184 (37.0)	
Others	361 (11.2)	294 (10.8)	67 (13.5)	
Unknown	775 (24.1)	655 (24.1)	120 (24.1)	
RRT before KT	1.59 ± 1.07	1.58 ± 1.05	1.63 ± 1.14	0.15
Hemodialysis	2,301 (71.6)	1,943 (71.5)	358 (72.0)	
Peritoneal dialysis	409 (12.7)	360 (13.3)	49 (9.9)	
KT	34 (1.1)	30 (1.1)	4 (0.8)	
Preemptive KT	469 (14.6)	383 (14.1)	86 (17.3)	
Dialysis vintage (mo)	4.53 ± 3.48	4.54 ± 3.49	4.47 ± 3.40	0.71
Retransplant	280 (8.7)	238 (8.8)	42 (8.5)	0.68
SBP (mmHg)	122.66 ± 17.46	122.71 ± 17.40	122.38 ± 17.82	0.70
DBP (mmHg)	75.64 ± 12.85	75.76 ± 12.77	75.03 ± 13.31	0.25
Hemoglobin (g/dL)	10.84 ± 1.59	10.84 ± 1.60	10.83 ± 1.55	0.98
Albumin (g/dL)	4.00 ± 0.51	4.00 ± 0.52	3.99 ± 0.49	0.67
Total cholesterol (mg/dL)	158.99 (29.0–384.0)	159.89 (29.0–384.0)	154.07 (52.0–333.0)	0.005
Triglyceride (mg/dL)	120.68 (4.0–855.0)	118.37 (4.0–855.0)	133.01 (32.0–715.0)	<0.001
Intact PTH (pg/mL)	346.58 ± 299.0	350.72 ± 260.0	323.83 ± 277.0	0.19
Vitamin D analogs	488 (15.2)	426 (15.7)	62 (12.5)	0.06
Statin	713 (22.2)	594 (21.9)	119 (23.9)	0.32
PRA ≥50%	179 (5.6)	149 (5.5)	30 (6.0)	0.70
HLA-DSA	660 (20.5)	558 (20.5)	102 (20.5)	0.99
Positive cross-match	287 (8.9)	240 (8.8)	47 (9.5)	0.72
ABO incompatibility	456 (14.2)	375 (13.8)	81 (16.3)	0.16
Desensitization	660 (20.5)	558 (20.6)	102 (20.5)	0.99
IL-2 receptor antibody	2,568 (79.9)	2,171 (79.9)	397 (79.9)	0.98
ATG	696 (21.7)	584 (21.5)	112 (22.5)	0.61
Tacrolimus	3,144 (97.9)	2,664 (98.3)	480 (97.2)	0.11
Cyclosporine	59 (1.8)	45 (1.7)	14 (2.8)	0.08
Mycophenolic acid	2,975 (92.6)	2,513 (92.5)	462 (93.0)	0.74
mTOR inhibitor	31 (1.0)	26 (1.0)	5 (1.0)	0.81
Corticosteroid	3,180 (99.0)	2,687 (98.9)	493 (99.2)	0.59
HBs Ag positivity	188 (5.9)	163 (6.0)	25 (5.0)	0.35
HCV Ab positivity	47 (1.5)	43 (1.6)	4 (0.8)	0.13
CMV IgG positivity	2,943 (91.6)	2,485 (91.5)	458 (92.2)	0.39

Data are expressed as number only, mean ± standard deviation, number (%), or median (interquartile range).

ATG, anti-thymocyte globulin; BMI, body mass index; CMV IgG, cytomegalovirus immunoglobulin G; CVD, cardiovascular disease; DBP, diastolic blood pressure; HBs Ag, hepatitis B virus surface antigen; HCV Ab, hepatitis C virus antibody; HLA-DSA, human leukocyte antigen-donor specific antibody; IL-2, interleukin-2; KT, kidney transplantation; mTOR, mammalian target of rapamycin; PRA, panel reactive antibody; PTDM, posttransplant diabetes mellitus; PTH, parathyroid hormone; RRT, renal replacement therapy; SBP, systolic blood pressure.

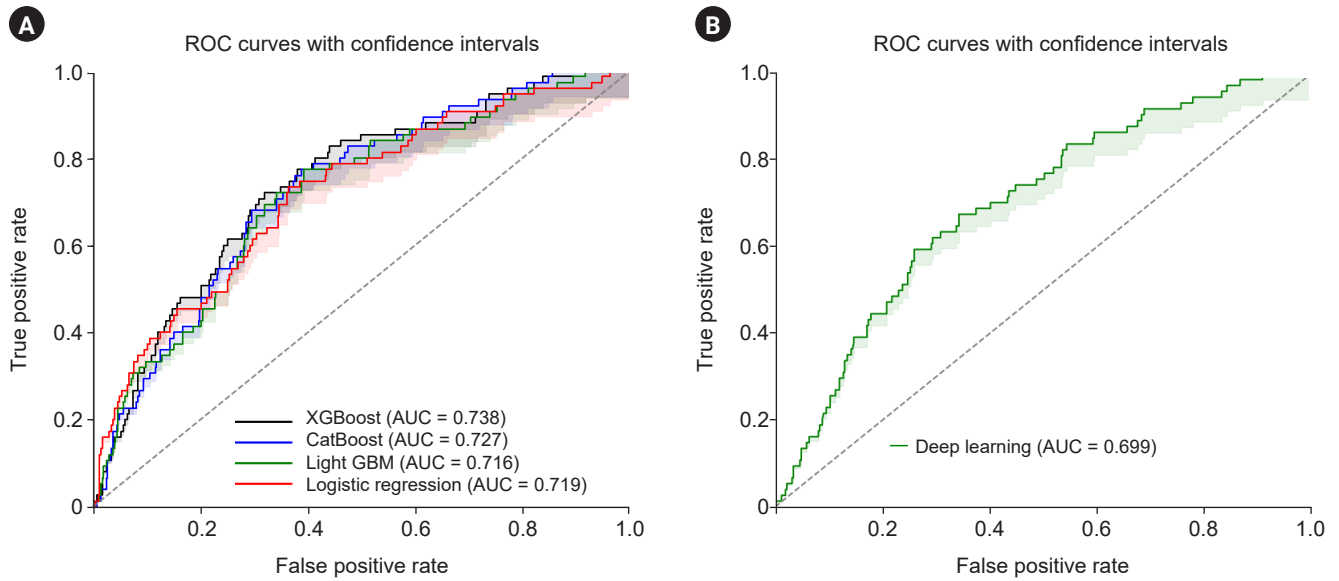


Figure 4. ROC of machine learning and deep learning models. (A) ROC of four machine learning models. (B) ROC of deep learning model.

AUC, area under the curve; GBM, gradient boosting machine; ROC, receiver operating characteristic; XGBoost, eXtreme Gradient Boosting;

Table 3. Performance metrics of machine learning and deep learning models

Analysis	AUC	Accuracy	Precision	Recall/sensitivity	F1 score
XGBoost	0.738	0.86	0.60	0.32	0.42
CatBoost	0.727	0.87	0.79	0.20	0.32
Light GBM	0.716	0.85	0.77	0.08	0.15
Logistic regression	0.719	0.85	0.63	0.14	0.23
Deep learning	0.699	0.68	0.27	0.64	0.19

AUC, area under the curve; GBM, gradient boosting machine; XGBoost, eXtreme Gradient Boosting.

is recommended not to use the thresholds of HbA1c for diagnosing PTDM [12]. It is because the diagnostic threshold for HbA1c ($\geq 6.5\%$) in type 2 DM is not related to the risk of diabetic retinopathy in KTRs [14], and because the HbA1c level is affected by high red blood cell turnover, anemia, and inhibition of red cell proliferation in the bone marrow due to immunosuppressants in the early posttransplant period [15]. Therefore, it is recommended to use the fasting plasma glucose level or the oral glucose tolerance test to diagnose PTDM [12]. Since fasting plasma glucose level is collected after 6 months and annually after KT in KOTRY, PTDM was diagnosed according to the fasting glucose criteria. The incidence of PTDM shows a biphasic pattern, with a peak in the first few months after transplantation

and a second surge over the next 2 to 3 years [16]. This study focused on the development of PTDM within 1 year of transplant because a majority of KTRs were not yet in the follow-up period of posttransplantation 2 or 3 years. An analysis of PTDM occurrence beyond 1 year posttransplantation was attempted but did not yield a significant conclusion. The insufficient number of additional PTDM cases identified during the 2- to 4-year follow-up period, coupled with the increased number of missing values, diminished the predictive power with the current sample size. Hence, these cases were not included in the analysis, with consideration for future studies contingent upon obtaining a larger sample size.

Various risk factors are known to contribute to the de-

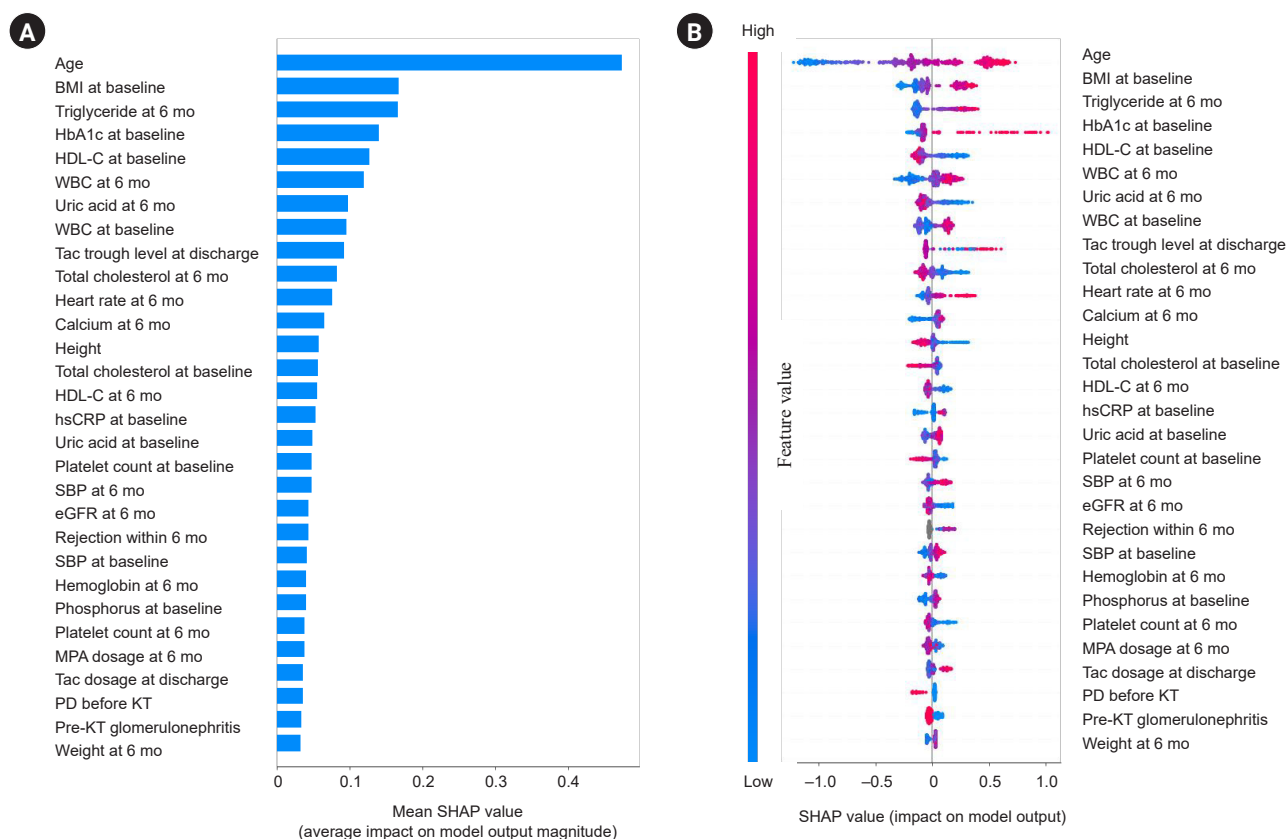


Figure 5. Feature importance of XGBoost model. (A) The average impact of each feature on the output of XGBoost model. The vertical axis shows the features included in the model. The horizontal axis depicts the mean SHAP (SHapley Additive exPlanations) value, which represents the average impact of each feature. (B) The impact of each feature on the output of XGBoost model. The vertical axis shows the features included in the model. The horizontal axis depicts the SHAP value, which represents the impact of each feature. The feature value is shown in blue to red colors; blue as low impact and red as high impact.

BMI, body mass index; eGFR, estimated glomerular filtration rate; HbA1c, hemoglobin A1c; HDL-C, high-density lipoprotein cholesterol; hsCRP, high sensitivity C-reactive protein; KT, kidney transplantation; MPA, mycophenolate acid; PD, peritoneal dialysis; SBP, systolic blood pressure; WBC, white blood cell; Tac, tacrolimus; XGBoost, eXtreme Gradient Boosting.

velopment of PTDM; age, obesity, acute rejection, virus infection including cytomegalovirus and hepatitis B and C, and immunosuppressants [4–7]. However, there are a few studies focusing on making a prediction model for PTDM. Chakkerla et al. [17] developed a risk prediction score including seven pretransplant variables and using multivariable regression models among 316 KTRs; the AUCs were 0.70 to 0.72. Rodrigo et al. [18] used the San Antonio diabetes prediction model and Framingham Offspring Study–Diabetes Mellitus algorithm to predict PTDM among 191 KTRs, which were originally developed in nontransplant population; each exhibited an AUC of 0.807 and 0.756, respectively. These two studies included a relatively small

number of subjects and did not consider factors related with immunosuppressive drugs. Recently, Cheng et al. [19] reported a risk prediction model among 495 KTRs using six variables in logistic regression. Variables included in their model were age, BMI, tacrolimus level, transient hyperglycemia, delayed graft function, and acute rejection, and the AUC was 0.916. The difference of this study is that we used machine learning algorithms and included a larger number of subjects and variables. The advantage of machine learning over traditional statistical methods is that it can analyze a large dataset and interpret complex, nonlinear relationships and interactions among many variables [8]. Therefore, machine learning may provide more accurate

prediction models by capturing hidden interactions between features. In this study, 41 pretransplant variables and 31 posttransplant variables were selected among 236 variables. A PTDM-prediction model was made by utilizing four machine learning algorithms and deep learning, a more advanced machine learning technique. Among five models, XGBoost demonstrated the highest performance with an AUC of 0.738 and accuracy of 0.86.

Although the highest recall/sensitivity (0.64) was achieved by the deep learning model, it exhibited significantly lower precision (0.27) compared to the other models. PTDM is not an acute emergency requiring immediate intervention but rather a chronic condition requiring long-term management of elevated blood glucose. Accurate diagnosis and gradual treatment are therefore essential for PTDM. Therefore, given the nature of PTDM, it is more important to ensure that diagnostic precision is maintained at an acceptable level rather than focusing solely on high sensitivity. The XGBoost model demonstrated the highest AUC, F1 score and high accuracy while also maintaining robust sensitivity and precision, making it the optimal choice.

In our XGBoost model, 10 features with highest importance were followed as; recipient age, baseline BMI, triglyceride level at post-KT 6 months, HbA1c and HDL-C level at baseline, white blood cell count and serum uric acid level at 6 months, white blood cell count at baseline, tacrolimus trough level at discharge, and total cholesterol level at post-KT 6 months. These features are consistent with previous reports; reflecting age, obesity, metabolic syndrome, and inflammation [6]. Among variables related to immunosuppression, tacrolimus level at discharge, rejection episode within 6 months, mycophenolate acid dosage at 6 months, and tacrolimus dosage at discharge were in the 30 top variables with importance. Calcineurin inhibitors and glucocorticoids are well known predisposing factors for PTDM, whereas mycophenolate acid has not been reported to increase the risk of PTDM [6]. Since cumulative dosage of corticosteroids was not collected in KOTRY database, the effect of corticosteroid could have been reduced. It is unclear why mycophenolate acid dosage at post-KT 6 months was in the model. It may be because variables with higher cardinality have the likelihood to have higher feature importance.

There are limitations to this study. First, PTDM was di-

agnosed only according to the fasting plasma glucose level and not by oral glucose tolerance test as recommended [12], which might have led to a lower incidence of PTDM. Furthermore, HbA1c, which could serve as a valuable criterion for diagnosing PTDM along with fasting plasma glucose level, especially for patients with unknown oral glucose tolerance, was not included in the analysis due to its follow-up data being unavailable in the KOTRY dataset. The inclusion of this variable could potentially have resulted in a higher AUC score. Second, factors that might have affected hyperglycemia, such as the cumulative dosage of corticosteroids, time-averaged tacrolimus levels or viral infection, were not included in the analysis because of lack of data. We attempted to include hepatitis B and C virus infection and cytomegalovirus immunity in the models, but it did not improve the models' performance. Third, external validation in different countries, race, and ethnicity is needed to make the prediction model generalizable. However, our study has advantages in that it used a nationwide multicenter database and employed machine learning and deep learning in making the prediction model. Variables before and after transplantation were included in the model, which makes our model useful in clinical practice. Moreover, immunosuppressant used in current era were included in the analysis, which makes our model more practical than previous models in literature.

In conclusion, we implemented machine learning to predict the development of PTDM after 1 year in KTRs. Our model could aid clinicians in early diagnosis, prevention, and counseling of PTDM. Risk factors for PTDM should be evaluated and individualized risk assessment should be done during pre-transplantation work-up. This can lead to early diagnosis and prompt management of PTDM, thereby improving clinical outcomes of KTRs.

Additional information

¹Department of Medicine, Seoul National University College of Medicine, Seoul, Republic of Korea

²Department of Internal Medicine, Incheon St. Mary's Hospital, College of Medicine, The Catholic University of Korea, Seoul, Republic of Korea

³Center for Artificial Intelligence in Healthcare, Seoul National University Bundang Hospital, Seongnam, Republic of Korea

⁴Department of Internal Medicine, Seoul National University Bundang Hospital, Seongnam, Republic of Korea

⁵Division of Nephrology, Department of Internal Medicine, CHA Bundang Medical Center, CHA University, Seongnam, Republic of Korea

Korea

⁶Department of Surgery, SMG-SNU Boramae Medical Center, Seoul, Republic of Korea

⁷Department of Surgery, Myongji Hospital, Goyang, Republic of Korea

⁸Department of Internal Medicine, Yonsei University College of Medicine, Seoul, Republic of Korea

⁹Department of Surgery, Yonsei University College of Medicine, Seoul, Republic of Korea

Conflicts of interest

All authors have no conflicts of interest to declare.

Funding

This research was supported by a grant from the Korea Health Technology R&D Project through the Korea Health Industry Development Institute (KHIDI), funded by the Ministry of Health and Welfare, Republic of Korea (grant number: HI23C047600). Additional funding was provided by a grant from the Patient-Centered Clinical Research Coordinating Center (PACEN), also supported by the Ministry of Health and Welfare, Republic of Korea (grant numbers: HI19C0481 and HC20C0054). Furthermore, this research was supported by the National Institute of Health (NIH) research project (2014-ER6301-00, 2014-ER6301-01, 2014-ER6301-02, 2017-ER6301-00, 2017-ER6301-01, 2017-ER6301-02, 2020-ER7201-00, 2020-ER7201-01, 2020-ER7201-02, 2023-ER0805-00, and 2023-ER0805-01).

Acknowledgments

The authors thank the members of the KOTRY Study Group ([Appendix](#), available online) for their contribution.

Data sharing statement

The data presented in this study are available from the corresponding author upon reasonable request.

Authors' contributions

Conceptualization, Project administration: HEY, Sejoong Kim

Data curation, Formal analysis: SC, Sangwoong Kim

Funding acquisition: Sangwoong Kim, HEY, Sejoong Kim

Investigation: JCJ, HEY, Sejoong Kim

Methodology: Sangwoong Kim, HEY, Sejoong Kim

Resources: YHL, HM, JHL, JY, MSK

Software: Sangwoong Kim

Writing—original draft: SC, MRP, Sangwoong Kim, HEY

Writing—review & editing: All authors

All authors read and approved the final manuscript.

ORCID

Seoyoung Choi, <https://orcid.org/0009-0009-0286-1510>

Mi Ryung Pyo, <https://orcid.org/0009-0003-1466-4676>

Sangwoong Kim, <https://orcid.org/0009-0004-0818-1261>

Jong Cheol Jeong, <https://orcid.org/0000-0003-0301-7644>

Yu Ho Lee, <https://orcid.org/0000-0001-5231-0551>

Hyejin Mo, <https://orcid.org/0000-0002-5383-0863>

Jeong-Hoon Lee, <https://orcid.org/0000-0003-0152-3172>

Jaeseok Yang, <https://orcid.org/0000-0002-5378-7797>

Myoung Soo Kim, <https://orcid.org/0000-0002-8975-8381>

Hye Eun Yoon, <https://orcid.org/0000-0002-6347-7282>

Sejoong Kim, <https://orcid.org/0000-0002-7238-9962>

References

1. Montori VM, Basu A, Erwin PJ, Velosa JA, Gabriel SE, Kudva YC. Posttransplantation diabetes: a systematic review of the literature. *Diabetes Care* 2002;25:583-592.
2. Conte C, Secchi A. Post-transplantation diabetes in kidney transplant recipients: an update on management and prevention. *Acta Diabetol* 2018;55:763-779.
3. Sharif A, Baboolal K. Complications associated with new-onset diabetes after kidney transplantation. *Nat Rev Nephrol* 2011;8:34-42.
4. Sharif A, Baboolal K. Risk factors for new-onset diabetes after kidney transplantation. *Nat Rev Nephrol* 2010;6:415-423.
5. Sharif A, Cohn S. Post-transplantation diabetes-state of the art. *Lancet Diabetes Endocrinol* 2016;4:337-349.
6. Jenssen T, Hartmann A. Post-transplant diabetes mellitus in patients with solid organ transplants. *Nat Rev Endocrinol* 2019;15:172-188.
7. Xia M, Yang H, Tong X, Xie H, Cui F, Shuang W. Risk factors for new-onset diabetes mellitus after kidney transplantation: a systematic review and meta-analysis. *J Diabetes Investig* 2021;12:109-122.
8. Kim KJ, Tagkopoulos I. Application of machine learning in rheu-

- matic disease research. *Korean J Intern Med* 2019;34:708–722.
9. Jang HJ, Cho KO. Applications of deep learning for the analysis of medical data. *Arch Pharm Res* 2019;42:492–504.
 10. Inker LA, Eneanya ND, Coresh J, et al. New creatinine- and cystatin C-based equations to estimate GFR without race. *N Engl J Med* 2021;385:1737–1749.
 11. American Diabetes Association. 2. Classification and Diagnosis of Diabetes: Standards of Medical Care in Diabetes-2021. *Diabetes Care* 2021;44(Suppl 1):S15–S33.
 12. Sharif A, Hecking M, de Vries AP, et al. Proceedings from an international consensus meeting on posttransplantation diabetes mellitus: recommendations and future directions. *Am J Transplant* 2014;14:1992–2000.
 13. Jenssen T, Hartmann A. Emerging treatments for post-transplantation diabetes mellitus. *Nat Rev Nephrol* 2015;11:465–477.
 14. American Diabetes Association. Diagnosis and classification of diabetes mellitus. *Diabetes Care* 2011;34(Suppl 1):S62–S69.
 15. Hare MJ, Shaw JE, Zimmet PZ. Current controversies in the use of haemoglobin A1c. *J Intern Med* 2012;271:227–236.
 16. Porrini EL, Díaz JM, Moreso F, et al. Clinical evolution of post-transplant diabetes mellitus. *Nephrol Dial Transplant* 2016;31:495–505.
 17. Chakkeri HA, Weil EJ, Swanson CM, et al. Pretransplant risk score for new-onset diabetes after kidney transplantation. *Diabetes Care* 2011;34:2141–2145.
 18. Rodrigo E, Santos L, Piñera C, et al. Prediction at first year of incident new-onset diabetes after kidney transplantation by risk prediction models. *Diabetes Care* 2012;35:471–473.
 19. Cheng F, Li Q, Wang J, Wang Z, Zeng F, Zhang Y. Analysis of risk factors and establishment of a risk prediction model for post-transplant diabetes mellitus after kidney transplantation. *Saudi Pharm J* 2022;30:1088–1094.