



A Prediction Model for Tacrolimus Daily Dose in Kidney Transplant Recipients With Machine Learning and Deep Learning Techniques

Qiwen Zhang^{1,2†}, Xueke Tian^{1,2†}, Guang Chen^{1,2}, Ze Yu³, Xiaojian Zhang^{1,2}, Jingli Lu^{1,2}, Jinyuan Zhang³, Peile Wang^{1,2}, Xin Hao⁴, Yining Huang⁵, Zeyuan Wang³, Fei Gao^{3*} and Jing Yang^{1,2*}

¹ Department of Pharmacy, The First Affiliated Hospital of Zhengzhou University, Zhengzhou, China, ² Henan Key Laboratory of Precision Clinical Pharmacy, Zhengzhou University, Zhengzhou, China, ³ Beijing Medicinovo Technology Co. Ltd, Beijing, China, ⁴ Dalian Medicinovo Technology Co. Ltd, Dalian, China, ⁵ McCormick School of Engineering, Northwestern University, Evanston, IL, United States

OPEN ACCESS

Edited by:

Surasak Saokaew,
University of Phayao, Thailand

Reviewed by:

Khanita Duangchaemkarn,
University of Phayao, Thailand
Ekkarat Boonchieng,
Chiang Mai University, Thailand

*Correspondence:

Jing Yang
jingyang_0101@163.com
Fei Gao
gaofei9000@163.com

[†]These authors have contributed
equally to this work

Specialty section:

This article was submitted to
Precision Medicine,
a section of the journal
Frontiers in Medicine

Received: 11 November 2021

Accepted: 22 April 2022

Published: 27 May 2022

Citation:

Zhang Q, Tian X, Chen G, Yu Z,
Zhang X, Lu J, Zhang J, Wang P,
Hao X, Huang Y, Wang Z, Gao F and
Yang J (2022) A Prediction Model for
Tacrolimus Daily Dose in Kidney
Transplant Recipients With Machine
Learning and Deep Learning
Techniques. *Front. Med.* 9:813117.
doi: 10.3389/fmed.2022.813117

Tacrolimus is a major immunosuppressor against post-transplant rejection in kidney transplant recipients. However, the narrow therapeutic index of tacrolimus and considerable variability among individuals are challenges for therapeutic outcomes. The aim of this study was to compare different machine learning and deep learning algorithms and establish individualized dose prediction models by using the best performing algorithm. Therefore, among the 10 commonly used algorithms we compared, the TabNet algorithm outperformed other algorithms with the highest R^2 (0.824), the lowest prediction error [mean absolute error (MAE) 0.468, mean square error (MSE) 0.558, and root mean square error (RMSE) 0.745], and good performance of overestimated (5.29%) or underestimated dose percentage (8.52%). In the final prediction model, the last tacrolimus daily dose, the last tacrolimus therapeutic drug monitoring value, time after transplantation, hematocrit, serum creatinine, aspartate aminotransferase, weight, CYP3A5, body mass index, and uric acid were the most influential variables on tacrolimus daily dose. Our study provides a reference for the application of deep learning technique in tacrolimus dose estimation, and the TabNet model with desirable predictive performance is expected to be expanded and applied in future clinical practice.

Keywords: prediction model, tacrolimus, daily dose, kidney transplant, machine learning, genetic polymorphism

INTRODUCTION

Tacrolimus, a calcineurin inhibitor, is a widely acceptable used immunosuppressive drug in kidney transplant recipients (KTRs). The narrow therapeutic index of tacrolimus leads to a very close edge between therapeutic and toxic blood concentration (1). An inappropriate regimen of tacrolimus in the early phase after transplantation is associated with acute rejections if the therapeutic range is not achieved, or adverse events (nephrotoxicity, infections as well as the development of new-onset diabetes or hypertension) if the therapeutic range is exceeded (2–9). Therefore, even small dose adjustments should be carefully chosen, and an accurate estimation of the relationship between tacrolimus dose and dose-response is important for postoperative outcomes.

The predictable suitable dose of tacrolimus is critical for the health status of postoperative KTRs. The Kidney Disease: Improving Global Outcomes guidelines suggested that 5–15 ng/ml of tacrolimus trough levels should be used during the first 2–4 months post-transplantation, and reduced thereafter in KTRs to minimize toxicity and adverse effects (10). More effective immunosuppressive therapy protocol during the early phase after solid organ transplantation would not only enhance the quality and length of life but would also reduce the need for transplantation (11). How to translate technological innovations into a workable tacrolimus dosing tool for clinicians has been explored in recent studies.

A consensus report on therapeutic drug monitoring (TDM) of tacrolimus-personalized therapy was updated in 2019. The report considered the most relevant advances in tacrolimus pharmacokinetics, pharmacogenetics, pharmacodynamics, and immunologic biomarkers to assist professionals to individualize tacrolimus treatment (12). Clinicians usually made dosage adjustments based on TDM in the early post-transplantation phase, until it reached long-term maintenance immunosuppression levels. However, the dose required to achieve the targeted whole blood concentrations of tacrolimus varies among individuals (13, 14). Numerous factors have been reported that are associated with the expression of metabolic enzyme and transporter to affect the pharmacokinetics of tacrolimus, including patient demographics, gene polymorphisms, time after transplantation, concomitant medications, comorbidities, hematocrit, hepatic function, and so forth (15). Thus, how to synthesize these potential factors and translate them into the tacrolimus dose adjustment strategy for immunosuppressive therapy in KTRs is still necessary to be explored and confirmed.

To evaluate the relevance of individual factors to influencing tacrolimus effect, researchers attempted to develop clinically practicable dosing models for tacrolimus stable dose. Methods of multiple linear regression (MLR) and population pharmacokinetics (PPK) have been utilized in tacrolimus dosing models (12, 16–19). With the development of machine learning and deep learning techniques, more studies applied these techniques to enhance the model expression of the complicated relationship between individual factors and medication dose. Tang et al. first applied multiple machine learning algorithms to establish a stable dosing prediction model of tacrolimus in Chinese KTRs, and proved the method was superior to the MLR (20, 21). In addition to the algorithms used in Tang's research, some algorithms with more sophisticated principles were developed, such as eXtreme Gradient Boosting (XGBoost), light gradient boosting machine (LightGBM), Categorical Boosting (CatBoost), random forest (RF), and TabNet (a deep learning technique), which were highly recognized in the algorithm competitions (22–26). For dose prediction of continuous data, these algorithms based on the principle of decision trees and deep neural networks might be more suitable. Herein, we conducted this study to investigate the individualized factors that markedly affect tacrolimus stable dose as well as to identify the most workable algorithm for prediction of the dose regimen of Chinese KTRs.

Taking the tacrolimus dose prediction model in this study as an example, with the increasing data amount of input subjects, the machine learning model could constantly optimize the parameters to achieve better accuracy and practicality. It is prospective to revolutionize the prescription and promote individualized medication in the future.

METHODS

Study Population

A retrospective analysis of recipients who received kidney allografts between April 2012 and June 2019 was performed. Clinical and demographic data were collected from the medical records in the First Affiliated Hospital of Zhengzhou University, China. The study protocol was approved by the Ethics Committee of the First Affiliated Hospital of Zhengzhou University (2020-KY-147). Eligible for enrollment were patients aged younger than 75 years and were treated with tacrolimus as part of their immune-suppressive regimen within 3 months post-transplantation. All recipients went through kidney transplantation for the first time and the donors were relatives and non-relatives. Patients who did not meet the criteria aforementioned and those with more missing values (missing rate over 20%) were excluded. Some variables showing the remarkable influence on tacrolimus dose even with a high percentage of missing values were not deleted and were interpolated subsequently (such as CYP3A5 and ABCB1). In the end, 584 patients with 5,439 cases of tacrolimus daily dose were included. The enrollment of patients is displayed in **Figure 1**.

Data Collection and Processing

The workflow of collecting and processing data is shown in **Figure 2**. Data of laboratory tests, medication, and diagnosis were collected from the patients who met the criteria. We collected the same type of drug combinations, including glucocorticoid (GC), calcium channel blockers (CCB), PPI (proton pump inhibitor), Wuzhi softgel, enzyme inducer, and mycophenolic acid (MPA) (**Supplementary Table S1**). The pathological status included a diagnosis of diarrhea, abdominal pain, abdominal distension, belching, anorexia, and gastrointestinal motility recovery. The daily tacrolimus dose equaled the sum of the doses prescribed by each order on that day. All results of successive TDM tests were retained because the order of tacrolimus that was closest to these tests was a long-term order, which considered that the corresponding medication conditions of these TDM tests were the same. For initial medication cases, days after transplantation = time of tacrolimus administration - time of kidney transplantation. For adjusting medication cases, days after transplantation = time of the last TDM test - time of kidney transplantation. The laboratory data were taken from the closest test results before the first administration of tacrolimus after kidney transplant surgery. The concomitant medication data were taken from results of 3 days before the first administration of tacrolimus to the time before the first TDM test.

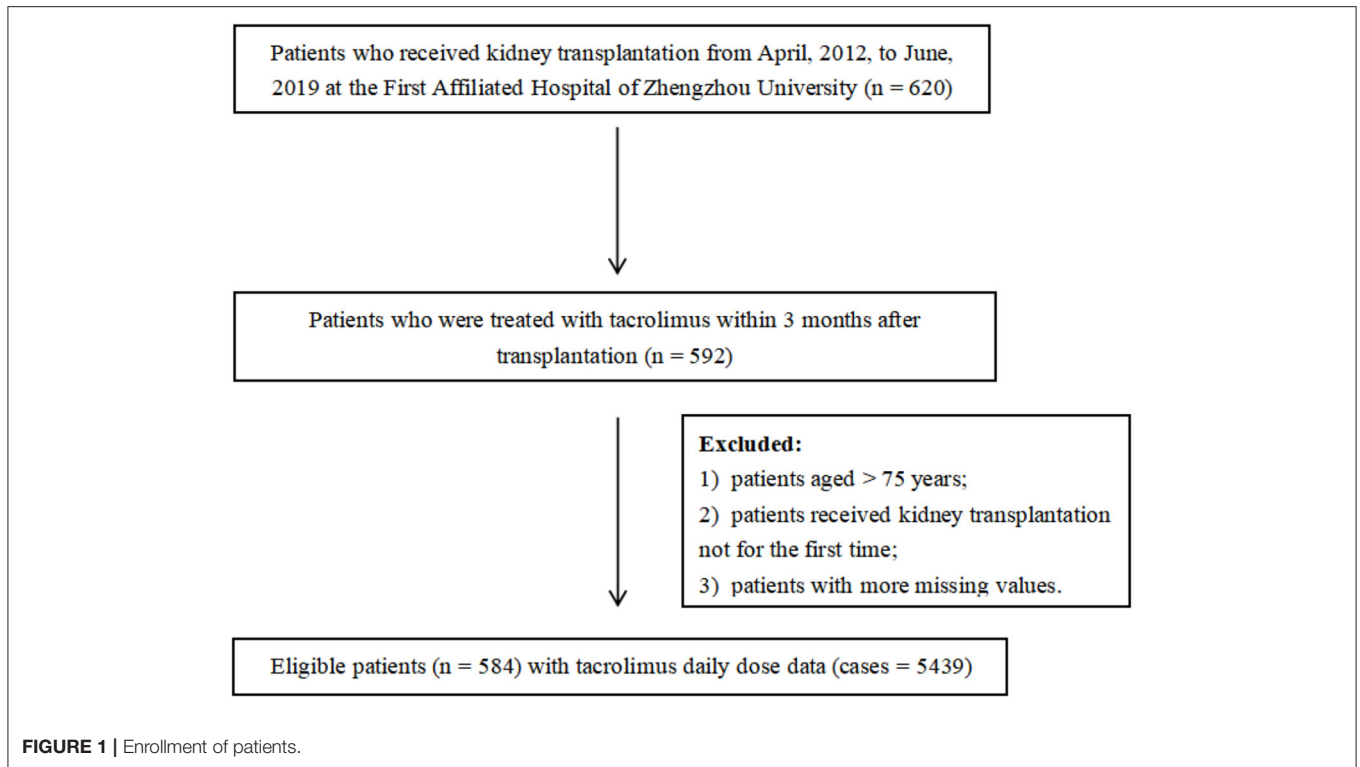


FIGURE 1 | Enrollment of patients.

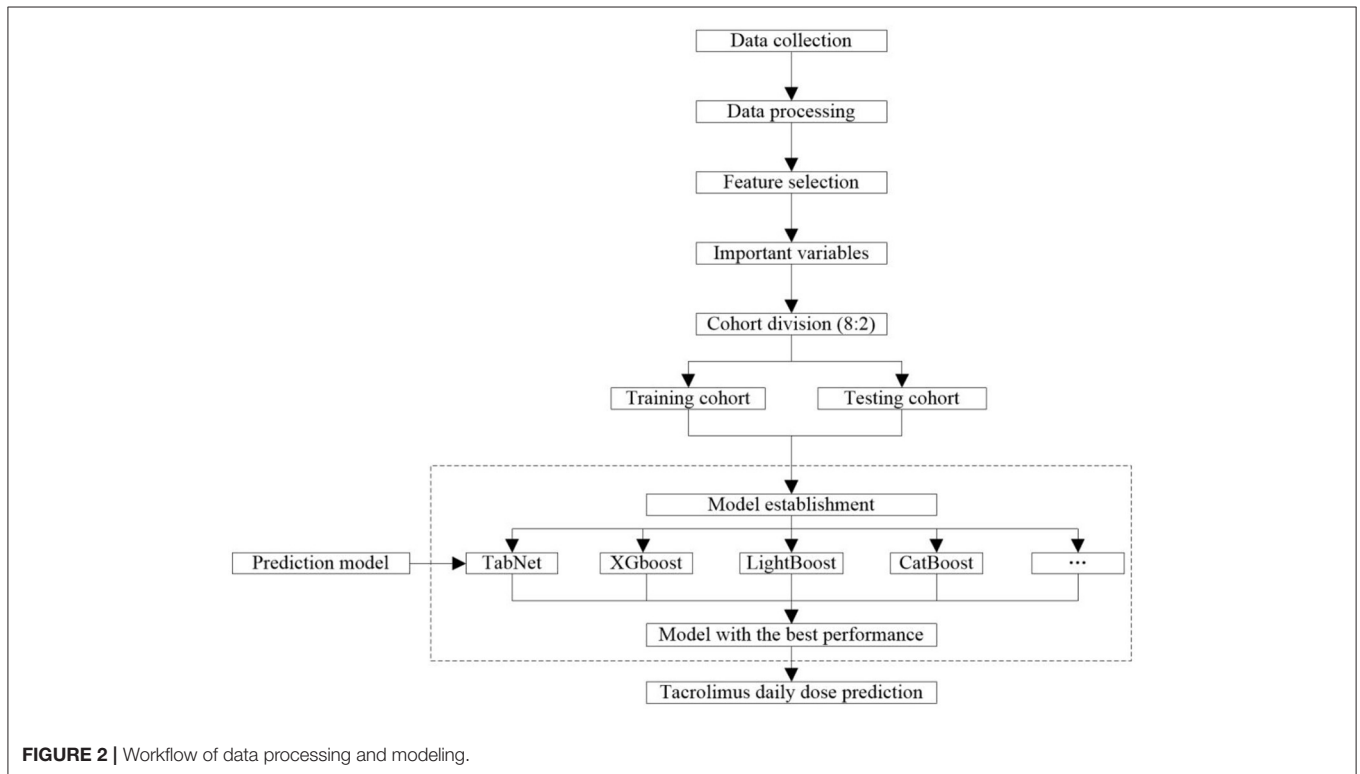


FIGURE 2 | Workflow of data processing and modeling.

Variables Selection

A univariate correlation testing procedure was implemented to remove all variables that were not directly related to tacrolimus

daily dose. The Pearson Correlation test was applied for continuous variables by investigating the association between the variable and outcome. In addition, the Mann–Whitney U test was

used for binary variables, and Analysis of Variance (ANOVA) was used for multiclass variables. This screening procedure adopted the criteria of p -value ≤ 0.05 to recruit impact variables in the following stages. Subsequently, the univariates were further screened by machine learning, and their importance scores were calculated and ranked. Those ranked at the top of the list were selected as the important variables to construct the model.

Before the machine learning algorithms, RF was applied to interpolate missing data (27). The missing values were then imputed using multiple regression sequentially based on different types of missing covariates. After imputation of RF, the remaining data set was used for the modeling stage, and the missing percentages after interpolation are summarized in **Supplementary Table S2**.

Model Construction

The workflow of modeling is shown in **Figure 2**. A total of 10 frequently used algorithms including deep learning method and machine learning methods were selected for generating the predictive models using the processed feature set. These algorithms were XGBoost, LightGBM, CatBoost, Gradient Boosted Decision Tree (GBDT), RF, support vector regression (SVR), K-nearest neighbor (KNN), Least Absolute Shrinkage and Selection Operator (LASSO) regression, ridge regression (RR), linear regression, and TabNet (22–26, 28–34). The data set was randomly allocated to the training cohort used to train the model and the testing cohort used to verify the model according to 8:2 ratio. The predictive performance of these algorithms is compared in this study. To minimize overfitting and obtain reliable results, 5-fold cross-validation was used to refine model prediction performance. The final model was constructed using the algorithm with the best predictive capacity based on the selected important variables.

Model Evaluation

To evaluate and compare the predictability of models, we primarily assessed the R-square (R^2), mean absolute error (MAE), mean square error (MSE), root mean square error (RMSE), and percentage of overestimated or underestimated dose in the testing cohort. More specifically, R^2 is the squared correlation between predicted and actual dose value (square root of tacrolimus dose). The MAE is the average of the absolute difference between the actual and the predicted dose values in this case. The RMSE is the square root of the mean square error between the predicted dose and actual dose, and MSE is the average value of the distance between each predicted dose and the average actual dose.

RESULTS

Baseline Information

Based on the whole data set, a total of 584 KTRs with 5,439 cases of dose data were identified in this study. The baseline information of the study population is shown in **Table 1**, including information regarding tacrolimus treatments, patient demographic information, combinations, genetic polymorphism, assay index, kidney transplantation information, and some

TABLE 1 | Baseline characteristic of study population.

Variables	Total cohort (Cases = 5,439)	Missing rate (%)
Age (yrs), median (IQR)	32 (26–41)	2.89
Sex, n (%)		3.44
Female	1456 (26.8%)	
Male	3796 (69.8%)	
Height (cm), median (IQR)	170 (160–173)	3.05
Weight (kg), median (IQR)	59 (50–65)	3.05
BMI (kg/m^2), median (IQR)	20.6 (18.7–22.4)	3.05
Tacrolimus dose (mg/d), median (IQR)	5.0 (3.5–6.0)	0
Last tacrolimus dose (mg/d), median (IQR)	4.5 (3.5–6.0)	0
Last tacrolimus TDM (ng/mL), median (IQR)	9.5 (7.5–12.0)	0
GC dose (mg/d), median (IQR)	18.8 (12.0–25.0)	0
Nifedipine, n (%)	1,548 (28.46%)	0
Levamlodipine besylate, n (%)	868 (15.96%)	0
Valsartan amlodipine, n (%)	1,413 (25.98%)	0
Nikadine, n (%)	147 (2.70%)	0
Felodipine, n (%)	367 (6.75%)	0
PPI, n (%)	5,347 (98.31%)	0
Wuzhi softgel, n (%)	5,065 (90.5%)	0
Enzyme inducer, n (%)	11 (0.2%)	0
MPA, n (%)	3,905 (71.8%)	0
CYP3A5*3, n (%)		52.07
A/A	231 (4.2%)	
A/G	1,161 (21.3%)	
G/G	1,215 (22.3%)	
ABCB1, n (%)		52.07
C/C	1,016 (18.7%)	
C/T	1,129 (20.8%)	
T/T	462 (8.5%)	
UA ($\mu\text{mol}/\text{L}$), median (IQR)	343 (261–429)	3.92
SCr ($\mu\text{mol}/\text{L}$), median (IQR)	262 (145–571)	3.92
AST (U/L), median (IQR)	13 (10–18)	4.04
TBIL ($\mu\text{mol}/\text{L}$), median (IQR)	8.2 (5.9–11.7)	4.03
RBC ($10^{12}/\text{L}$), median (IQR)	3.27 (2.87–3.76)	1.62
Hb (g/L), median (IQR)	99 (87.9–111)	1.47
HCT (L/L), median (IQR)	0.30 (0.27–0.35)	1.47
NEU (%), median (IQR)	89.1 (76.6–95.3)	1.47
LYM (%), median (IQR)	4.9 (1.7–13.7)	1.47
Time after transplantation (d), median (IQR)	10 (5–20)	0
Living relative's organ transplantation, n (%)	788 (14.5%)	0
Hypertension, n (%)	3,783 (69.6%)	0
Diabetes, n (%)	23 (0.4%)	0
Other pathological status, n (%)	515 (9.5%)	0

BMI, body mass index; GC, glucocorticoid; PPI, proton pump inhibitor; MPA, mycophenolic acid; UA, uric acid; SCr, serum creatinine; AST, aspartate aminotransferase; TBIL, total bilirubin; RBC, red blood cells; Hb, hemoglobin; HCT, hematocrit; NEU, neutrophil; LYM, lymphocyte.

other body conditions. The continuous variables were described by “median (interquartile range, IQR)”, and the classification variables were described by “frequency (%)”. The median age of patients in this study was 32 (IQR 26–41), and

TABLE 2 | Univariate analysis results.

Variable	Statistic (r/U/F)	p-value
Last tacrolimus TDM	-0.146	<0.001
Last tacrolimus daily dose	0.542	<0.001
Time after transplantation	-0.352	<0.001
Age	0.143	<0.001
Sex	1870835.5	<0.001
Height	0.397	<0.001
Weight	0.406	<0.001
BMI	0.296	<0.001
GC dose	0.127	<0.001
PPI	1,91,491	<0.001
Wuzhi softgel	7,06,284	<0.001
Levamlodipine besylate	19,64,818	0.651
Valsartan amlodipine	2789975.5	0.28
Nifedipine	2908343.5	0.046
Nikadine	4,48,043	0.002
Felodipine	857026	0.01
Enzyme inducer	27028.5	0.584
MPA	2568291.5	<0.001
CYP3A5*3	359811.5	<0.001
ABCB1	1.464	0.232
UA	-0.065	<0.001
SCr	0.180	<0.001
AST	-0.110	<0.001
TBIL	-0.027	0.05
RBC	0.020	0.142
Hb	-0.020	0.146
HCT	0.051	<0.001
NEU%	0.182	<0.001
LYM%	-0.171	<0.001
Hypertension	28,87,309	<0.001
Diabetes	51,447	0.146
Living donor kidney transplantation from relatives	15,86,384	<0.001
Pathological status	11,80,315	0.009

BMI, body mass index; GC, glucocorticoid; PPI, proton pump inhibitor; MPA, mycophenolic acid; UA, uric acid; SCr, serum creatinine; AST, aspartate aminotransferase; TBIL, total bilirubin; RBC, red blood cells; Hb, hemoglobin; HCT, hematocrit; NEU, neutrophil; LYM, lymphocyte.

the proportion of female patients was 26.8%. In terms of drug combinations, nifedipine occupied the highest proportion of CCB (28.46 %), a median daily dose of GC was 18.8 (IQR 12.0–25.0) mg, and PPI (98.31 %), Wuzhi softgel (90.5 %), and MPA (71.8 %) had very high proportion in the study population.

Univariate Analysis

We determined the significant associations between univariate and tacrolimus daily dose (p -value ≤ 0.05). Through this screening procedure, 25 variables were found significant, and entered the subsequent processing and modeling stages (Table 2).

Comparison of Predictive Algorithms

In this subsection, 10 machine learning methods were compared in terms of R^2 , MAE, MSE, and RMSE in the testing cohort. A large R^2 indicated that the established model had a good fitting, and small errors (i.e., MAE, MSE, and RMSE) indicated higher predictive accuracy overall. Strikingly, the TabNet algorithm outperformed all other methods in R^2 (0.813), MAE (0.464 mg/d), MSE (0.615 mg/d), and RMSE (0.784 mg/d). The percentage of the overestimated dose (5.29 %) ranked in the top 4 and the percentage of underestimated dose (8.52 %) in the testing cohort of TabNet ranked in the top 1, demonstrating a robust predictive performance. The specific results of the model evaluation are listed in Table 3. The model prediction performance after the 5-fold cross-validation is displayed in Figure 3. It could be seen that after 5-fold cross-validation, TabNet had the highest R^2 and 20% accuracy value in the testing cohort, indicating its best fitness and accuracy among 10 algorithms. Regarding the prediction accuracy of 10 models, a boxplot is illustrated in Figure 4, indicating that TabNet had the highest prediction accuracy among 10 models.

Dose Prediction Model Using TabNet

Based on the comparison of algorithms, the TabNet model was verified as the best model for predicting the dose of tacrolimus in KTRs. As shown in Table 4, the accuracy of the predicted dose within ± 20 , ± 30 , and $\pm 40\%$ of the actual dose was 86.19, 91.33, and 93.48%, respectively. It automatically recruited the most influential variables referring to the sum of their importance scores. Table 5 lists the top 10 important variables to predict tacrolimus dose and their importance scores in the TabNet model, including the last tacrolimus daily dose (0.316), the last tacrolimus TDM (0.219), time after transplantation (0.083), hematocrit (HCT; 0.079), SCr (0.068), aspartate aminotransferase (AST; 0.058), weight (0.037), CYP3A5 (0.037), BMI (0.036), and uric acid (UA; 0.021).

DISCUSSION

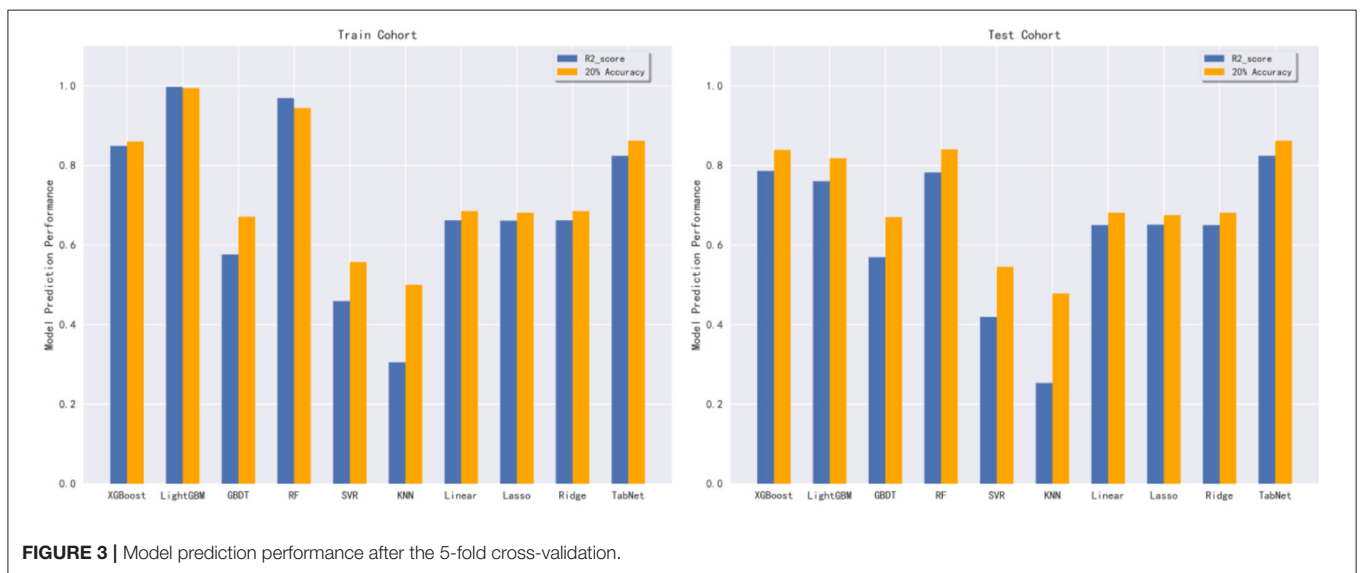
Tacrolimus dose adjustment is generally reactive administration, which has a wide variability in the early phase after transplantation (10, 35). Our research was designed based on the data of appropriate tacrolimus administration among Chinese, which comprehensively analyzed the widely varying multivariate and provided an accurately predicted dose during the early postoperative period. Overall, we found different performances of the 10 algorithms that were used to predict tacrolimus dosing in the Chinese KTRs. When all the cases were investigated, the TabNet algorithm achieved the best performance.

As a novel deep learning technique, whether it is unsupervised learning for filling missing features or supervised learning for actual decision-making, the TabNet encoder is used to encode the input features first; then the missing features are filled separately with decoder connections according to different uses, or connected with the full connection layer to achieve the final decision. The TabNet encoder architecture is mainly composed of a feature transformer, an attentive transformer, and feature masking. The decoder uses the feature transformer layer to

TABLE 3 | R^2 , MAE, MSE, RMSE results, and percentage of overestimated or underestimated dose in the testing cohort of each predictive algorithms.

Predictive algorithm	R^2	MAE (mg/d)	MSE (mg/d)	RMSE (mg/d)	Underestimated dose (%)	Overestimated dose (%)
XGBoost	0.786	0.488	0.681	0.821	4.60	11.51
LightGBM	0.760	0.523	0.761	0.868	7.21	10.97
GBDT	0.569	0.923	1.368	1.169	7.75	25.28
RF	0.782	0.461	0.693	0.829	4.91	11.09
SVR	0.419	1.010	1.829	1.343	31.34	14.12
KNN	0.253	1.241	2.365	1.537	20.56	31.68
Linear regression	0.650	0.756	1.110	1.050	12.28	19.64
LASSO regression	0.651	0.756	1.108	1.050	12.50	20.02
Ridge regression	0.650	0.756	1.110	1.050	12.28	19.64
TabNet	0.824	0.468	0.558	0.745	5.29	8.52

GBDT, Gradient Boosted Decision Tree; RF, random forest; SVR, support vector regression; KNN, K-nearest neighbor; LASSO, Least Absolute Shrinkage and Selection Operator.

**FIGURE 3** | Model prediction performance after the 5-fold cross-validation.

reconstruct the encoded features to the raw data table features (26). Moreover, TabNet uses a sequential attention mechanism to choose a subset of meaningful features to process at each decision step, enabling interpretability and more efficient learning as the learning capacity used for the most salient features. TabNet employs a single deep learning architecture for feature selection and reasoning (26). Additionally, based on retaining the end-to-end and representation learning characteristics of deep neural networks, TabNet also has the advantages of tree model interpretability and sparse feature selection (36). Other studies based on real-world data showed that TabNet outperformed ensemble tree-based algorithms since it could process highly non-linear relationships with its depth, without overfitting due to instance-wise feature selection (26). However, there are some limitations of implementing or translating TabNet compared to other machine learning algorithms, including (i) TabNet runs slowly, consumes more time and resources, and even runs several times longer than XGBoost; (ii) TabNet could not give full play to its advantages when the amount of data was not very large.

Furthermore, compared with conventional modeling methods, machine learning and deep learning techniques have indubitable advantages in dealing with real-world data, such as (i) machine learning and deep learning can deal with more complex, high-dimensional, and interactive variables from the clinical environment, which is lacking in conventional models; (ii) machine learning and deep learning models have a stronger generalization and better accuracy than conventional models (37–39). Recently, the application of machine learning and deep learning techniques on individualized dose prediction models has been approbatory, such as a novel vancomycin dose prediction model through XGBoost and warfarin maintenance dose prediction through LightGBM (40, 41). With the increasing number of input subject data, machine learning and deep learning models can continually optimize parameters to achieve better performance and practicality. Herein, we calculated and compared model predictive performance among all algorithms. The R^2 of the testing cohort of our tacrolimus dose prediction model using TabNet reached 0.824 in the overall

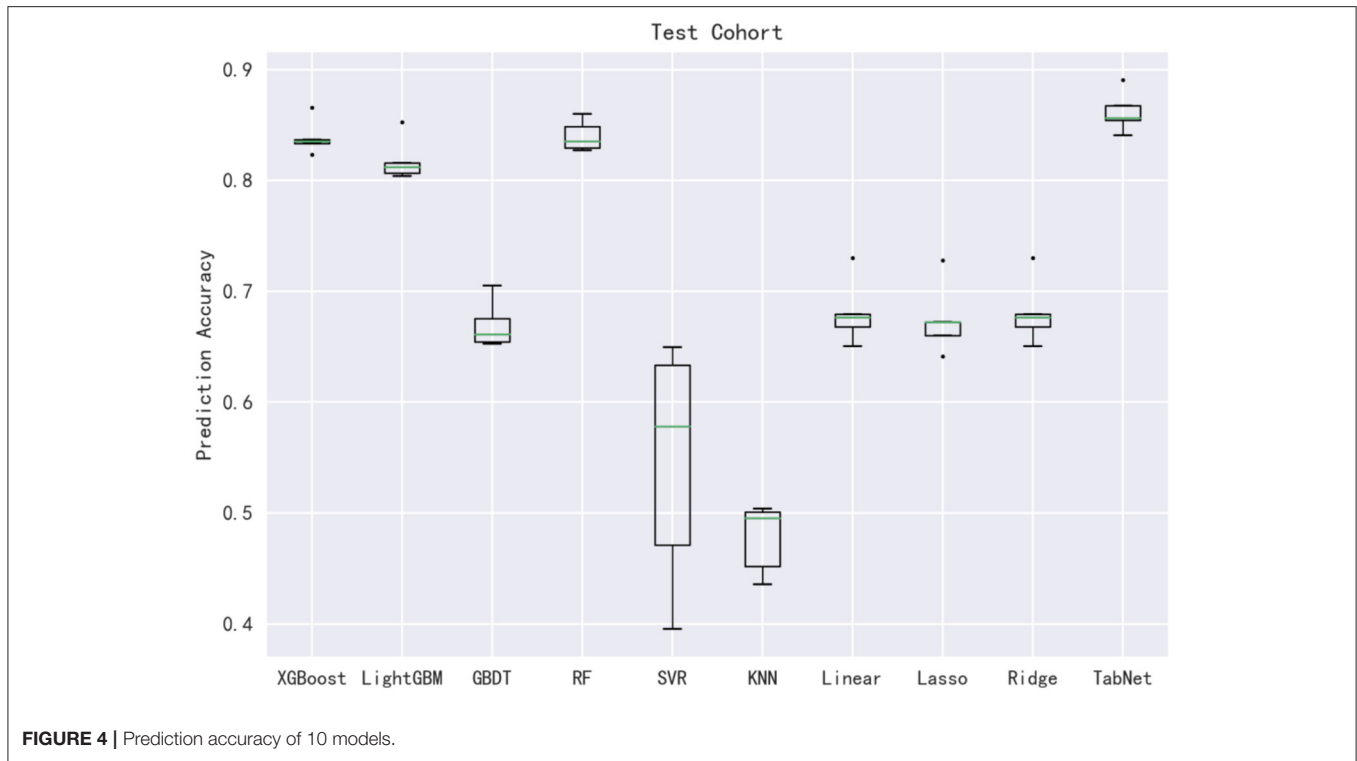


FIGURE 4 | Prediction accuracy of 10 models.

TABLE 4 | The prediction accuracy of the predicted value in different confidence intervals.

Model	Accuracy of the predicted dose within $\pm 20\%$ of the actual dose	Accuracy of the predicted dose within $\pm 30\%$ of the actual dose	Accuracy of the predicted dose within $\pm 40\%$ of the actual dose
XGBoost	83.89%	89.37%	92.41%
LightGBM	81.82%	88.61%	91.87%
GBDT	66.97%	75.87%	81.24%
RF	84.01%	89.49%	92.36%
SVR	54.54%	71.11%	83.35%
KNN	47.76%	64.87%	75.18%
Linear regression	68.08%	80.51%	86.07%
Lasso regression	67.47%	80.59%	86.03%
Ridge regression	68.08%	80.51%	86.07%
TabNet	86.19%	91.33%	93.48%

TABLE 5 | Ranking of importance scores.

Variable	Importance score
Last tacrolimus daily dose	0.316
Last tacrolimus TDM	0.219
Time after transplantation	0.083
HCT	0.079
SCr	0.068
AST	0.058
Weight	0.037
CYP3A5	0.037
BMI	0.036
UA	0.021

HCT, hematocrit; SCr, serum creatinine; AST, aspartate aminotransferase; BMI, body mass index; UA, uric acid.

comparison. Compared to the results of Tang et al., which applied eight machine learning algorithms, including multiple linear regression, artificial neural network, regression tree, multivariate adaptive regression splines, boosted regression tree, SVR, RF, LASSO regression, and Bayesian additive regression trees (RF achieved the lowest MAE of 0.73 in the validation cohort), our prediction model could achieve a better outcome (TabNet achieved the MAE of 0.464) (20).

After estimating the importance of variables through TabNet, the last tacrolimus daily dose, the last tacrolimus TDM, time after

transplantation, HCT, SCr, AST, weight, CYP3A5, BMI, and UA were identified as the most influential factors in dose prediction. In the previous PPK models, among the transplant recipients, the factors most associated with whole blood apparent clearance variation for tacrolimus included cytochrome CYP3A5, weight, HCT, postoperative days, and hepatic function (e.g., AST), SCr, and the CYP3A5 polymorphism was the most frequently included variable (12, 17–19, 42). In our deduction, the important variables were consistent with the result of previous studies. Due to the extensibility of the machine learning model, the model could be further improved after the addition of learnable data and more unrecognized novel influencing variables could be mined. Additionally, the machine learning model had more

flexibility in that the recommended dose varies with the updated influencing variables, thus it was useful for the next medication prediction. Nevertheless, the PPK model generally just predicts the initial medication. Third, according to the second consensus report about TDM of tacrolimus-personalized therapy issued by the Immunosuppressive Drugs Scientific Committee of the International Association of Therapeutic Drug Monitoring and Clinical Toxicity in 2019, the range of R^2 was unstable varying from 0.27 to 0.99 in the PPK models, whereas our model had a good performance of $R^2 = 0.813$ (12).

The last tacrolimus daily dose and TDM value were closely related to the prediction of the next tacrolimus dose, and the time after transplantation had been chosen which was beneficial for considering relevant factors that change with time (19). In this study, the height, weight, and gender of patients were regarded as control factors, since we considered the integrity of their data in our study population and their fixed effects. Above all, the physiology existed a wide discrepancy between obese and non-obese individuals (43). Multiple studies verified in humans and rodents, that obesity changes the mRNA or protein expression levels of hepatic CYP3A in the liver and intestine, which was a major oxidatively metabolizing enzyme of tacrolimus (43–45). Sawamoto et al. reported that obese patients could maintain tacrolimus concentration well at lower doses compared with non-obese patients (46). Hence, weight and BMI were important in predicting tacrolimus dose. Patients expressing at least one *CYP3A5*1* allele need a higher tacrolimus dose than those not carrying this allele (*CYP3A5*3/*3*) to reach the same blood concentration (12).

As mentioned above, tacrolimus is mainly metabolized in the liver and intestinal wall through isoforms of CYP3A enzymes (CYP3A4 and CYP3A5). The major enzyme involved in its biological transformation is CYP3A5, and the catalytic efficiency of CYP3A4 is lower (41). The association between CYP3A5 and tacrolimus dose has been observed in KTRs (16, 18). Asberg A et al. found that the *CYP3A5* genotype led to higher clearance and lower bioavailability of tacrolimus, and its inclusion in the model could improve dose prediction (16). The marked pharmacokinetic variability of tacrolimus was partly due to the *CYP3A5* genotype, therefore, CYP3A5 was important for our predictive results. According to the consensus report, the *ABCB1* rs1045642 (3435C>T) was reported to affect the transporter activity which could influence the concentration of tacrolimus (12). However, this factor was excluded due to the univariate analysis results in this research.

In terms of biochemical parameters affecting the tacrolimus dose, a PPK model proved that lower SCr and lower HCT levels were identified as important factors to enhance tacrolimus clearance (36). The remarkable impact of SCr on tacrolimus dose herein was consistent with the results of previous research, which illustrated that decreased SCr has a significant association with tacrolimus dose (47). Furthermore, HCT explained 4~14% of the variability in tacrolimus dose requirements and clearance *in vivo* (48). A high percentage of tacrolimus in blood was binding to erythrocytes, which could be influenced by HCT and red blood cell count significantly (49). Multiple studies point out that HCT

varied hugely in transplant recipients and increased widely after transplantation, and the concentration of tacrolimus associated with erythrocytes tended to increase along with the augment of HCT in post-transplant patients (49, 50). Additionally, UA, as an indicator of kidney function, was demonstrated to be increased after tacrolimus therapy in kidney or liver transplant recipients (51, 52). The increasing UA level may imply a decline in kidney function, which may need a change in the tacrolimus regimen, hence, UA could be regarded as a posterior factor that affects the tacrolimus dose.

Additionally, drug–drug interactions with tacrolimus dose-response were also considered in the study. We included several kinds of concomitant medications, such as GC, CCB, PPI, Wuzhi softgel, enzyme inducer, and MPA. Of which, GC, Wuzhi softgel, and MPA were administered in the majority of KTRs, and unbalanced between positive and negative samples, hence, there was no significant difference among the participants. However, based on clinical experience, these drugs could influence tacrolimus concentration. For instance, according to Yan et al. and Wang et al., Wuzhi softgel could influence tacrolimus elimination and noticeably increase its drug concentration in Chinese organ transplant patients (53, 54). The effect of these concomitant drugs on tacrolimus might be further investigated based on balanced clinical data in the future.

In conclusion, this study was rigorously designed to mine in-depth the clinical and genetic influencing factors in KTRs treated with tacrolimus after kidney transplantation. In our study, a deep learning technique, TabNet, was utilized based on the multidimensional database to select the important factors relating to tacrolimus dose, which contributed to the practicable prediction accuracy of tacrolimus dosage in KTRs. The application of our proposed model for estimating the appropriate dose of tacrolimus provides recommendations for a clinically required therapeutic regimen.

DATA AVAILABILITY STATEMENT

The original contributions presented in the study are included in the article/**Supplementary Material**, further inquiries can be directed to the corresponding authors.

ETHICS STATEMENT

Written informed consent was obtained from the individual(s) for the publication of any potentially identifiable images or data included in this article.

AUTHOR CONTRIBUTIONS

QZ and JY designed the study. ZY and FG provided medical support. XZ and JL guided the research. GC extracted the data. XT and PW collected and assembled the data. JZ, YH, and ZW analyzed the data. XH wrote the manuscript. All authors read and approved the final manuscript for publishing.

FUNDING

This work was supported by the National Key R&D Program of China (Grant No. 2020YFC2008304), the Youth Program of National Natural Science Foundation of China (No. 81903704).

REFERENCES

- Johnston A. Equivalence and interchangeability of narrow therapeutic index drugs in organ transplantation. *Eur J Hosp Pharm.* (2013) 20:302–7. doi: 10.1136/ejpharm-2012-000258
- Staatz C, Taylor P, Tett S. Low tacrolimus concentrations and increased risk of early acute rejection in adult renal transplantation. *Nephrol Dial Transplant.* (2001) 16:1905–9. doi: 10.1093/ndt/16.9.1905
- Gaynor JJ, Ciancio G, Guerra G, Sageshima J, Roth D, Goldstein MJ, et al. Lower tacrolimus trough levels are associated with subsequently higher acute rejection risk during the first 12 months after kidney transplantation. *Transpl Int.* (2016) 29:216–26. doi: 10.1111/tri.12699
- Arreola-Guerra JM, Serrano M, Morales-Buenrostro LE, Vilatobá M, Alberú J. Tacrolimus trough levels as a risk factor for acute rejection in renal transplant patients. *Ann Transplant.* (2016) 21:105–14. doi: 10.12659/AOT.895104
- Ericson JE, Zimmerman KO, Gonzalez D, Melloni C, Gupta JJ, Hill KD, et al. A Systematic Literature Review Approach to Estimate the Therapeutic Index of Selected Immunosuppressant Drugs After Renal Transplantation. *Ther Drug Monit.* (2017) 39:13–20. doi: 10.1097/FTD.0000000000000364
- Hoorn EJ, Walsh SB, McCormick JA, Furstenberg A, Yang CL, Roeschel T, et al. The calcineurin inhibitor tacrolimus activates the renal sodium chloride cotransporter to cause hypertension. *Nat Med.* (2011) 17:1304–9. doi: 10.1038/nm.2497
- Penninga L, Moller CH, Gustafsson F, Steinbruchel DA, Glud C. Tacrolimus versus cyclosporine as primary immunosuppression after heart transplantation: systematic review with meta-analyses and trial sequential analyses of randomised trials. *Eur J Clin Pharmacol.* (2010) 66:1177–87. doi: 10.1007/s00228-010-0902-6
- Shen CL, Yang AH, Lien TJ, Tarng DC, Yang CY. Tacrolimus blood level fluctuation predisposes to coexisting bk virus nephropathy and acute allograft rejection. *Sci Rep.* (2017) 7:1986. doi: 10.1038/s41598-017-02140-1
- Sulanc E, Lane JT, Puumala SE, Groggel GC, Wrenshall LE, Stevens RB. New-onset diabetes after kidney transplantation: an application of 2003 international guidelines. *Transplantation.* (2005) 80:945–52. doi: 10.1097/01.TP.0000176482.63122.03
- Egeland EJ, Robertsen I, Hermann M, Midtvedt K, Storset E, Gustavsen MT, et al. High tacrolimus clearance is a risk factor for acute rejection in the early phase after renal transplantation. *Transplantation.* (2017) 101:e273–e9. doi: 10.1097/TP.0000000000001796
- Cowan J, Hutton B, Fergusson N, Bennett A, Tay J, Cameron DW, et al. Clinical outcomes of immunoglobulin use in solid organ transplant recipients: protocol for a systematic review and meta-analysis. *Syst Rev.* (2015) 4:167. doi: 10.1186/s13643-015-0156-6
- Brunet M, van Gelder T, Asberg A, Haufroid V, Hesselink DA, Langman L, et al. Therapeutic Drug Monitoring of Tacrolimus-Personalized Therapy: Second Consensus Report. *Ther Drug Monit.* (2019) 41:261–307. doi: 10.1097/FTD.0000000000000640
- Saint-Marcoux F, Woillard JB, Jurado C, Marquet P. Lessons from routine dose adjustment of tacrolimus in renal transplant patients based on global exposure. *Ther Drug Monit.* (2013) 35:322–7. doi: 10.1097/FTD.0b013e318285e779
- Staatz CE, Tett SE. Clinical pharmacokinetics and pharmacodynamics of tacrolimus in solid organ transplantation. *Clin Pharmacokinet.* (2004) 43:623–53. doi: 10.2165/00003088-200443100-00001
- Scott LJ, McKeage K, Keam SJ, Plosker GL. Tacrolimus: a further update of its use in the management of organ transplantation. *Drugs.* (2003) 63:1247–97. doi: 10.2165/00003495-200363120-00006
- Asberg A, Midtvedt K, van Guilder M, Storset E, Bremer S, Bergan S, et al. Inclusion of CYP3A5 genotyping in a nonparametric population model improves dosing of tacrolimus early after transplantation. *Transpl Int.* (2013) 26:1198–207. doi: 10.1111/tri.12194
- Andreu F, Colom H, Elens L, van Gelder T, van Schaik RHN, Hesselink DA, et al. A new CYP3A5*3 and CYP3A4*22 cluster influencing tacrolimus target concentrations: a population approach. *Clin Pharmacokinet.* (2017) 56:963–75. doi: 10.1007/s40262-016-0491-3
- Bergmann TK, Hennig S, Barraclough KA, Isbel NM, Staatz CE. Population pharmacokinetics of tacrolimus in adult kidney transplant patients: impact of CYP3A5 genotype on starting dose. *Ther Drug Monit.* (2014) 36:62–70. doi: 10.1097/FTD.0b013e31829f1ab8
- Woillard JB, Saint-Marcoux F, Debord J, Asberg A. Pharmacokinetic models to assist the prescriber in choosing the best tacrolimus dose. *Pharmacol Res.* (2018) 130:316–21. doi: 10.1016/j.phrs.2018.02.016
- Tang J, Liu R, Zhang YL, Liu MZ, Hu YF, Shao MJ, et al. Application of machine-learning models to predict tacrolimus stable dose in renal transplant recipients. *Sci Rep.* (2017) 7:42192. doi: 10.1038/srep42192
- Tang J, Xu J, Zhang YL, Liu R, Liu MZ, Hu YF, et al. Incorporation of gene-environment interaction terms improved the predictive accuracy of tacrolimus stable dose algorithms in chinese adult renal transplant recipients. *J Clin Pharmacol.* (2019) 59:890–9. doi: 10.1002/jcph.1379
- Chen T, Guestrin C. XGBoost: A Scalable Tree Boosting System. In: *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining.* San Francisco, CA (2016).
- Ke G, Meng Q, Finley T, Wang T, Chen W, Ma W, et al. LightGBM: A Highly Efficient Gradient Boosting Decision Tree. In: *Proceedings of 31st Conference on Neural Information Processing Systems.* Long Beach, CA (2017).
- Prokhorenkova L, Gusev G, Vorobev A, Dorogush AV, Gulina A. CatBoost: unbiased boosting with categorical features. *arXiv [Preprint].* (2017). doi: 10.48550/arXiv.1706.09516
- Janßen R, Zabel J, von Lukas U, Labrenz M. An artificial neural network and Random Forest identify glyphosate-impacted brackish communities based on 16S rRNA amplicon MiSeq read counts. *Mar Pollut Bull.* (2019) 149:110530. doi: 10.1016/j.marpolbul.2019.110530
- Arik SO, Pfister T. TabNet: attentive interpretable tabular learning. *arXiv [Preprint].* (2019). doi: 10.48550/arXiv.1908.07442
- Shah AD, Bartlett JW, Carpenter J, Nicholas O, Hemingway H. Comparison of random forest and parametric imputation models for imputing missing data using MICE: a CALIBER study. *Am J Epidemiol.* (2014) 179:764–74. doi: 10.1093/aje/kwt312
- Ye J, Chow JH, Chen J, Zheng Z, editors. Stochastic gradient boosted distributed decision trees. *Acm Conference on Information & Knowledge Management.* (2009) 2061: 1646301. doi: 10.1145/1645953.1646301
- Cosgun E, Limdi NA, Duarte CW. High-dimensional pharmacogenetic prediction of a continuous trait using machine learning techniques with application to warfarin dose prediction in African Americans. *Bioinformatics.* (2011) 27:1384–9. doi: 10.1093/bioinformatics/btr159
- Almansour NA, Syed HF, Khayat NR, Altheeb RK, Juri RE, Alhiyafi J, et al. Neural network and support vector machine for the prediction of chronic kidney disease: a comparative study. *Comput Biol Med.* (2019) 109:101–11. doi: 10.1016/j.combiomed.2019.04.017
- Goshvarpour A, Goshvarpour A. Verhulst map measures: new biomarkers for heart rate classification. *Phys Eng Sci Med.* (2022). doi: 10.1007/s13246-022-01117-3
- Tibshirani RJ. Regression shrinkage and selection via the LASSO: a retrospective. *J R Stat Soc Series B Stat Methodol.* (2011) 73:273–82. doi: 10.1111/j.1467-9868.2011.00771.x

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fmed.2022.813117/full#supplementary-material>

33. McDonald GC. Ridge regression. *WIREs Comput Stat.* (2010) 1:93–100. doi: 10.1002/wics.14
34. Richardson AM, Joshy G, D'Este CA. Understanding statistical principles in linear and logistic regression. *Med J Aust.* (2018) 208:332–4. doi: 10.5694/mja17.00222
35. Kidney Disease: Improving Global Outcomes Transplant Work G. KDIGO clinical practice guideline for the care of kidney transplant recipients. *Am J Transplant.* (2009) 9(Suppl. 3):S1–155. doi: 10.1111/j.1600-6143.2009.02834.x
36. Yan J, Xu T, Yu Y, Xu H. Rainfall forecast model based on the tabnet model. *Water.* (2021) 13:1272. doi: 10.3390/w13091272
37. Mo X, Chen X, Li H, Li J, Zeng F, Chen Y, et al. Early and accurate prediction of clinical response to methotrexate treatment in juvenile idiopathic arthritis using machine learning. *Front Pharmacol.* (2019) 10:1155. doi: 10.3389/fphar.2019.01155
38. Lee HC, Yoon SB, Yang SM, Kim WH, Ryu HG, Jung CW, et al. Prediction of acute kidney injury after liver transplantation: machine learning approaches vs. logistic regression model. *J Clin Med.* (2018) 7:11. doi: 10.3390/jcm7110428
39. Kruppa J, Ziegler A, König IR. Risk estimation and risk prediction using machine-learning methods. *Hum Genet.* (2012) 131:1639–54. doi: 10.1007/s00439-012-1194-y
40. Huang X, Yu Z, Wei X, Shi J, Wang Y, Wang Z, et al. Prediction of vancomycin dose on high-dimensional data using machine learning techniques. *Expert Rev Clin Pharmacol.* (2021) 14:761–71. doi: 10.1080/17512433.2021.1911642
41. Liu Y, Chen J, You Y, Xu A, Li P, Wang Y, et al. An ensemble learning based framework to estimate warfarin maintenance dose with cross-over variables exploration on incomplete data set. *Comput Biol Med.* (2021) 131:104242. doi: 10.1016/j.compbiomed.2021.104242
42. Andrews LM, Hesselink DA, van Schaik RHN, van Gelder T, de Fijter JW, Lloberas N, et al. A population pharmacokinetic model to predict the individual starting dose of tacrolimus in adult renal transplant recipients. *Br J Clin Pharmacol.* (2019) 85:601–15. doi: 10.1111/bcp.13838
43. Yoshinari K, Takagi S, Yoshimasa T, Sugatani J, Miwa M. Hepatic CYP3A expression is attenuated in obese mice fed a high-fat diet. *Pharm Res.* (2006) 23:1188–200. doi: 10.1007/s11095-006-0071-6
44. Ulvestad M, Skottheim IB, Jakobsen GS, Bremer S, Molden E, Asberg A, et al. Impact of OATP1B1, MDR1, and CYP3A4 expression in liver and intestine on interpatient pharmacokinetic variability of atorvastatin in obese subjects. *Clin Pharmacol Ther.* (2013) 93:275–82. doi: 10.1038/clpt.2012.261
45. Lam JL, Jiang Y, Zhang T, Zhang EY, Smith BJ. Expression and functional analysis of hepatic cytochromes P450, nuclear receptors, and membrane transporters in 10- and 25-week-old db/db mice. *Drug Metab Dispos.* (2010) 38:2252–8. doi: 10.1124/dmd.110.034223
46. Sawamoto K, Huong TT, Sugimoto N, Mizutani Y, Sai Y, Miyamoto K. Mechanisms of lower maintenance dose of tacrolimus in obese patients. *Drug Metab Pharmacokinet.* (2014) 29:341–7. doi: 10.2133/dmpk.DMPK-13-RG-110
47. Yagmurdur MC, Sevmis S, Emiroglu R, Moray G, Bilgin N, Haberal M. Tacrolimus conversion in kidney transplant recipients: analysis of 107 patients. *Transplant Proc.* (2004) 36:144–7. doi: 10.1016/j.transproceed.2003.11.044
48. de Jonge H, de Looor H, Verbeke K, Vanrenterghem Y, Kuypers DR. *In vivo* CYP3A4 activity, CYP3A5 genotype, and hematocrit predict tacrolimus dose requirements and clearance in renal transplant patients. *Clin Pharmacol Ther.* (2012) 92:366–75. doi: 10.1038/clpt.2012.109
49. Zahir H, McCaughan G, Gleeson M, Nand RA, McLachlan AJ. Factors affecting variability in distribution of tacrolimus in liver transplant recipients. *Br J Clin Pharmacol.* (2004) 57:298–309. doi: 10.1046/j.1365-2125.2003.02008.x
50. Storset E, Holford N, Midtvedt K, Bremer S, Bergan S, Asberg A. Importance of hematocrit for a tacrolimus target concentration strategy. *Eur J Clin Pharmacol.* (2014) 70:65–77. doi: 10.1007/s00228-013-1584-7
51. Kanbay M, Akcay A, Huddam B, Usluogullari CA, Arat Z, Ozdemir FN, et al. Influence of cyclosporine and tacrolimus on serum uric acid levels in stable kidney transplant recipients. *Transplant Proc.* (2005) 37:3119–20. doi: 10.1016/j.transproceed.2005.08.042
52. Tumgor G, Arikan C, Kilic M, Aydogdu S. Frequency of hyperuricemia and effect of calcineurin inhibitors on serum uric acid levels in liver transplanted children. *Pediatr Transplant.* (2006) 10:665–8. doi: 10.1111/j.1399-3046.2006.00556.x
53. Yan L, Yang ZQ, Shi YY, Ren J, Yang CL, Wan ZL, et al. Effects of wuzhi capsules on blood concentration of tacrolimus in renal transplant recipients. *Ann Transplant.* (2019) 24:594–604. doi: 10.12659/AOT.918980
54. Wang DD, Chen X, Li ZP. Wuzhi capsule and haemoglobin influence tacrolimus elimination in paediatric kidney transplantation patients in a population pharmacokinetics analysis: a retrospective study. *J Clin Pharm Ther.* (2019) 44:611–7. doi: 10.1111/jcpt.12828

Conflict of Interest: ZY, JZ, ZW, and FG are employed by Beijing Medicinovo Technology Co. Ltd, Beijing, China. XH is employed by Dalian Medicinovo Technology Co. Ltd, Dalian, China.

The remaining authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Copyright © 2022 Zhang, Tian, Chen, Yu, Zhang, Lu, Zhang, Wang, Hao, Huang, Wang, Gao and Yang. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.