Construction and Interpretation of Prediction Model of Teicoplanin Trough Concentration *via* Machine Learning

Pan Ma[1], Ruixiang Liu[1], Wenrui Gu[1], Qing Dai[1], Yu Gan[1], Jing Cen[1], Shenglan Shang[2]*, Fang Liu[1]* and Yongchuan Chen[1]*

[1] Department of Pharmacy, The First Affiliated Hospital of Third Military Medical University (Army Medical University), Chongqing, China, [2] Department of Clinical Pharmacy, General Hospital of Central Theater Command of PLA, Wuhan, China

Objective: To establish an optimal model to predict the teicoplanin trough concentrations by machine learning, and explain the feature importance in the prediction model using the SHapley Additive exPlanation (SHAP) method.

Methods: A retrospective study was performed on 279 therapeutic drug monitoring (TDM) measurements obtained from 192 patients who were treated with teicoplanin intravenously at the First Affiliated Hospital of Army Medical University from November 2017 to July 2021. This study included 27 variables, and the teicoplanin trough concentrations were considered as the target variable. The whole dataset was divided into a training group and testing group at the ratio of 8:2, and predictive performance was compared among six different algorithms. Algorithms with higher model performance (top 3) were selected to establish the ensemble prediction model and SHAP was employed to interpret the model.

Results: Three algorithms (SVR, GBRT, and RF) with high R^2 scores (0.676, 0.670, and 0.656, respectively) were selected to construct the ensemble model at the ratio of 6:3:1. The model with $R^2 = 0.720$, MAE = 3.628, MSE = 22.571, absolute accuracy of 83.93%, and relative accuracy of 60.71% was obtained, which performed better in model fitting and had better prediction accuracy than any single algorithm. The feature importance and direction of each variable were visually demonstrated by SHAP values, in which teicoplanin administration and renal function were the most important factors.

Conclusion: We firstly adopted a machine learning approach to predict the teicoplanin trough concentration, and interpreted the prediction model by the SHAP method, which is of great significance and value for the clinical medication guidance.

Keywords: machine learning, SHAP, precision medicine, prediction model, model explanation, algorithm, teicoplanin

INTRODUCTION

Teicoplanin is a glycopeptide antibiotic for the treatment of severe infections caused by Gram-positive bacteria, including methicillin-resistant *Staphylococcus aureus* (MRSA) (1). As an alternative to vancomycin, teicoplanin shows comparable clinical outcomes with fewer occurrences of nephrotoxicity, ototoxicity, and red man syndrome (2). However, with a very highly bound to plasma albumin, teicoplanin has a very long terminal elimination half-life (ranging from 100 to 170 h) and even takes several days to achieve the effective plasma concentration, which results in a great individual variability and permitting once daily dose (3). Therefore, an initial loading dose is required to achieve effective plasma concentration rapidly (3). Teicoplanin is highly bioavailable (>90%) and eventually excreted in urine as a prototype. Because of these pharmacokinetic characteristics, the fixed dosing regimens of teicoplanin administered to patients suffering from hypoalbuminemia (3), and/or renal insufficiency, and/or an expansion of the extracellular fluids might lead to the wide variations and fluctuations of concentrations (4).

The plasma trough concentration of teicoplanin is closely associated with its therapeutic efficacy. A large number of studies have shown that treated with the conventional regimen, many patients may fail to reach therapeutic targets that lead to clinical failure. However, repeated exposure to suboptimal concentrations increases the risk factor of teicoplanin resistance (5, 6). According to previous studies, 10–30 mg/l was regarded as the target teicoplanin trough level for successful treatment (5, 6). The teicoplanin trough concentrations are mainly influenced by the teicoplanin administration regimen and the patient's pathophysiological conditions (such as age, weight, serum albumin, renal function, concomitant therapy, concomitant diseases, etc.) (4).

Customization of the antimicrobial dosing regimen is continuously gaining more relevance in the antimicrobial stewardship programs (7, 8). In this regard, therapeutic drug monitoring (TDM), by measuring drug exposure in plasma, may be helpful in individual therapies (3). TDM is an effective method that assures adequate trough concentration for maximum efficacy and thus, prevents adverse effects resulting from overexposure (8–10). Based on the daily monitoring of teicoplanin concentration on our TDM platform, individual variation is evident, with low concentrations of teicoplanin, most of which are unable to reach an effective plasma trough concentration. However, some hospitals have no TDM platform due to the limited medical conditions, and sampling and testing of TDM cost time and money. In order to bring convenience to clinicians and save time and money for patients, more than TDM, more powerful drug concentration prediction tools are needed.

Machine learning algorithms, as a subdiscipline of artificial intelligence, take advantage of large-scale complex algorithms and datasets to uncover useful patterns, that can evaluate data-driven estimation when forecasting from multiple variables and obtain nonlinear variable relations to deliver predicted clinical outcomes with high accuracy (11, 12). The rapidly developing machine learning has been widely applied in the biomedicine field, such as clinical diagnostics, precision treatments, and health

monitoring (13). However, population pharmacokinetic (PPK) models are adopted by the ongoing research on teicoplanin trough concentration. It includes certain criteria such as age, weight, and creatinine/creatinine clearance rate (8, 14). Few studies on the prediction of teicoplanin trough concentration have adopted machine learning to model. In this study, the machine learning approach was employed to establish an optimal ensemble model to predict the teicoplanin trough concentrations, which can assist clinicians in guiding the dosage of medication. Furthermore, the SHapley Additive exPlanation (SHAP) method was used to explain the feature importance in our ensemble prediction model, so that our study could also provide a reasonable explanation for the prediction, which demonstrated how the relevant factors influenced the teicoplanin trough concentrations.

METHODS

Patients and Data

A retrospective study was conducted among patients who underwent teicoplanin intravenously at the First Affiliated Hospital of Army Medical University from November 2017 to July 2021. Patients were enrolled in this study according to the following inclusion criteria: (1) age > 14 years; (2) > 2–3 days of treatment with teicoplanin (steady-state concentration); and (3) underwent TDM of teicoplanin in which the trough blood samples were collected immediately before administering the next dose. The following exclusion criteria were applied: (1) pregnant women and (2) failed to reach the lower limit of quantification (LLOQ) for teicoplanin through concentration assay.

Ethics Approval

This study was approved by the Hospital Ethics Committee of the Southwest Hospital of Army Medical University ([B]KY2021095) and performed in accordance with the Declaration of Helsinki. In the ethical approval documents, the informed consent has been exempted. The procedures in this study are fully compliant with the ethical standards in accordance with the Institutional Research Committees.

Measurement of Teicoplanin Trough Concentration

The teicoplanin plasma trough concentration was measured by high-performance liquid chromatography (HPLC) (1200 Series, Agilent Technologies Incorporation). Determination was performed using the Innoval-C_{18} column (5 μm, 4.6 mm × 250 mm, Dikma Technologies). The mobile phase was 76% sodium dihydrogen phosphate (0.01 mmol/L) and 24% acetonitrile (pH 2.9). The UV detection wavelength was 240 nm. The trough plasma concentration linear range was 3.125–100.000 mg/l (correlation coefficient $R^2 = 0.9998$). Both the intra- and interday precisions were within 7%.

Data Collection and Processing

The teicoplanin dataset includes teicoplanin administration (loading dose, time of loading dose, loading intervals,

maintenance intervals, and total duration of treatment), demographic information (age, height, weight, gender, and APACHE II), laboratory parameters [albumin (ALB), estimated glomerular filtration rate (eGFR), cystatin C (Cys-C), creatinine clearance rate (CLcr), aspartate aminotransferase (AST), alanine aminotransferase (ALT), TBIL, NEU%, and PLT], concomitant therapy (ECMO, CRRT, and co-medication), and concomitant diseases (AML, hyperproteinemia, sepsis) were obtained from the hospital's electronic medical record system (EMRS). After cleaning up of teicoplanin dataset, the target variable and relevant crucial covariates were screened subsequently. The rate of missing data is 3.32%. The mean filling method in Python (version 3.6, Python Software Foundation) was employed to fill the missing data, resulting in a dataset of 279 × 27. The teicoplanin trough concentrations were selected as the target variable, while the whole dataset was randomly divided into a training group and testing group at the ratio of 8:2.

Modeling and Validation

The linear correlation between the teicoplanin trough concentrations and the relevant covariates was evaluated (**Supplementary Table S1**). According to the correlation coefficient, the linear correlation among them was poor. Therefore, six nonlinear machine learning algorithms for modeling were employed to predict the teicoplanin trough concentrations, including support vector regression (SVR), random forest (RF), Adaptive Boosting (Adaboost), Boostrap aggregating (Bagging), Gradient Boosted Regression Trees (GBRT), and eXtreme Gradient BoostingX (XGBoost).

In order to evaluate the single algorithm predictive performance, the metrics of R-squared (R^2), mean square error (MSE), and mean absolute error (MAE) were used. R^2 indicates the explanation degree of the independent variable to the dependent variable. The proportion of a single algorithm in the final model was determined through the prediction of different algorithms. The final result of the ensemble model is the weighted average based on the ranking of the top three algorithms. The calculating formulas are as follows:

$$MAE(y^o, y^p) = \frac{1}{N} \sum_{i=1}^{N} \left| y_i^o - y_i^p \right|$$

$$MSE(y^o, y^p) = \frac{1}{N} \sum_{i=1}^{N} \left(y_i^o - y_i^p \right)^2$$

$$R^2 \left(y^o, y^p \right) = 1 - \frac{\sum_{i=1}^{N} \left(y_i^o - y_i^p \right)^2}{\sum_{i=1}^{N} \left(y_i^o - \overline{y^o} \right)}$$

$$\overline{y^o} = \frac{1}{N} \sum_{i=1}^{N} y_i^o$$

R^2 represents the goodness of fit of the model, and the value range is 0–1. The closer R^2 gets to 1, the better the goodness of fit of the model becomes. y^o represents the observed value; y^p represents the predicted value. With reference to MSE and MAE, when their values decrease, the model has improved the goodness of fit. In addition, the accuracy of predicted trough concentration compared with the observed concentration was investigated. The absolute accuracy represented the accuracy of

the predicted trough concentration to be within ± 5 mg/L of the observed trough concentration, while the relative accuracy showed that the predicted trough concentration was within ± 30% of the observed trough concentration.

The top three algorithms were selected to establish the ensemble prediction model of teicoplanin trough concentrations. In addition, another dataset of 20 patients were collected as the validation group to corroborate the performance of the prediction models. The workflow of data processing, algorithm selection, and modeling were displayed in **Figure 1**.

Model Interpretation

SHapley Additive exPlanation, is a game-theoretic method that provides information to machine learning outputs. It determines and allocates credit for model outputs by means of Shapley values coming from game theory including all related covariants (15). As an additive feature attribution method, SHAP value represents contributions of each feature in a certain sample, in which each feature is regarded as a "contributor." A feature with a positive SHAP value improves the output value, and those larger numerical values make greater contributions (16, 17). SHAP values were used to provide the interpretation of our ensemble prediction model (18), in which the SHAP summary plot, the importance ranking, and the SHAP dependence plot of the relevant covariates were demonstrated based on the permutation explainer provided by the SHAP Python package (version 0.39.0).

Statistical Analysis

Statistical analysis was performed using IBM SPSS version 25.0 (IBM Corporation, Armonk, New York, USA). The Kolmogorov–Smirnov test was used to evaluate whether the measurement data were normally distributed. Measurement data were presented as the median and interquartile range (IQR) for nonnormal distribution variables and mean ± SD for normal distribution variables. Measurement data were analyzed by Mann-Whitney U test (non-normal distribution) and independent t-test (normal distribution). Categorical data were expressed as n (%) and analyzed by the chi-squared test ($n \geq 5$) or Fisher's exact test ($n < 5$). The tests were two-sided with a $p < 0.05$ which deemed statistically significant.

RESULTS

Baseline Patient Characteristics

This study was performed on 279 TDM measurements obtained from 192 patients who underwent teicoplanin treatment. The whole dataset was randomly divided into training group and testing group at the ratio of 8:2, which were 223 and 56 cases, respectively. The baseline information of 27 variables and the comparison between the training and testing groups were shown in **Table 1**, without any significant difference between variables of the two groups ($p > 0.05$).

Algorithm Selection

According to the linear correlation result (**Supplementary Table S1**), the linear correlation between

FIGURE 1 | The workflow of data processing and algorithm selection.
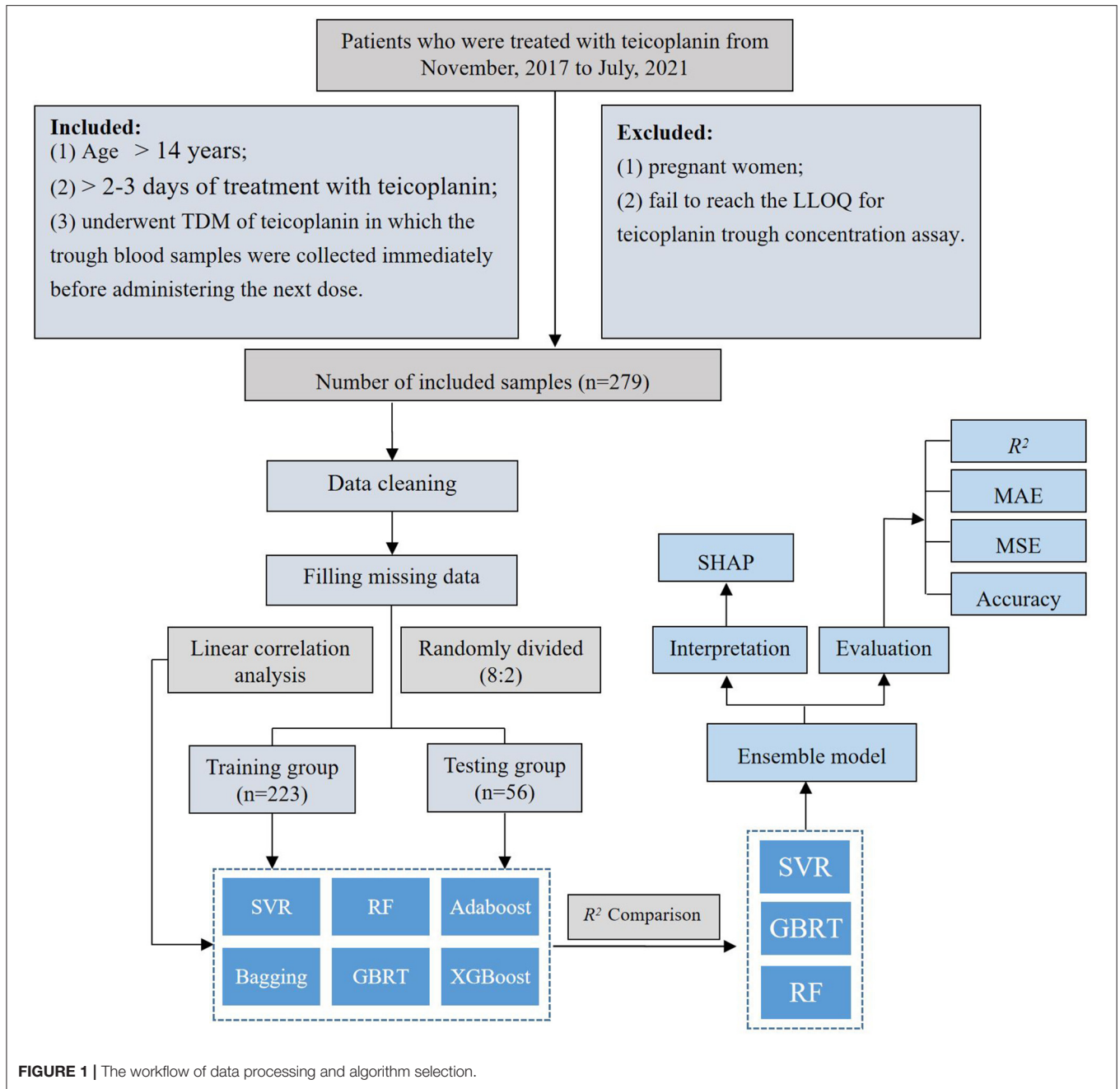
the teicoplanin trough concentrations and the relevant covariates was poor. Thus, six nonlinear algorithms were included for the algorithm selection. The performance metrics of six different algorithms including R^2, MAE, MSE, and accuracy were shown in **Table 2**. Among the six algorithms, SVR has the best predictive performance of prediction, with the highest R^2, accuracy, and lowest MAE, MSE. To select the algorithms to establish the ensemble prediction model for further promoting stability and accuracy, R^2 was chosen to evaluate the goodness-of-fit of the model. Among the six algorithms, SVR, GBRT, and RF had high goodness-of-fit, which is 0.676, 0.670, and 0.656, respectively. As

a result, the top three performing algorithms (SVR, GBRT, and RF) were chosen to predict teicoplanin trough concentration and for a subsequent experiment.

Modeling and Validation

To establish the ensemble prediction model of teicoplanin trough concentration, the target parameters were set as the highest R^2, absolute accuracy, and relative accuracy, then the weight proportion of three candidate algorithms (SVR, GBRT, and RF) with a high R^2 score was adjusted. Based on the automatic calculations of machine learning, the ensemble model composed

TABLE 1 | The description of the study samples.

Variables	Values		P value
	Training (n = 223)	**Testing (n = 56)**	
Teicoplanin trough concentration (mg/L)	13.32 (8.87,19.12)	11.291 (7.62,22.11)	0.603 [a]
Teicoplanin administration			
Loading dose (mg/kg)	7.91 (6.67, 12.18)	7.45 (6.09, 12.23)	0.149 [a]
Times of loading dose			0.747 [d]
< 3	68 (30.49)	14 (25)	
3-5	149 (66.82)	41 (73.21)	
> 5	6 (2.69)	1 (1.79)	
Loading intervals (h)			0.924 [d]
12	201 (90.13)	52 (92.86)	
24	14 (6.28)	3 (5.36)	
Others	8 (3.59)	1 (1.79)	
Maintenance dose (mg/kg)	8 (6.09, 12.18)	7.08 (5.64, 12.23)	0.086 [a]
Maintenance intervals (h)			0.249 [d]
12	13 (5.83)	2 (3.57)	
24	196 (87.89)	47 (83.93)	
Others	14 (6.28)	7 (12.5)	
Total duration of treatment (day)	5 (3, 8)	4 (3, 6.75)	0.079 [a]
Demographic information			
Age (years)	53 (40, 66)	52.5 (38.25, 65.75)	0.823 [a]
Height (cm)	165 (155.07, 168.27)	165 (155.07, 170)	0.782 [a]
Weight (kg)	59 (53.44, 65.7)	65 (53.44, 73.75)	0.059 [a]
Gender, male (n, %)	80 (35.87%)	25 (44.64%)	0.226 [c]
APACHE II	24 (20, 28)	23 (19, 26)	0.053 [a]
Laboratory parameters			
ALB (g/L)	32.32 ± 4.91	32.40 ± 4.59	0.911 [b]
eGFR (ml/min/L)	97.06 (61.84, 120.57)	86.63 (50.61, 111.41)	0.063 [a]
Cys-C (mg/L)	1.52 (0.97, 1.7)	1.6 (1.05, 1.76)	0.453 [a]
CLcr [e] (mL/min/1.73m^2)	85.55 (55.43, 125.25)	78.23 (40.22, 123.97)	0.317 [a]
AST (IU/L)	41.7 (22.5, 79.9)	34.2 (19.63, 81.6)	0.754 [a]
ALT (IU/L)	26.5 (14, 49.2)	24.8 (10.78, 69.25)	0.695 [a]
TBIL (umol/L)	20.4 (12.4, 48.1)	19.25 (13.93, 50.05)	0.984 [a]
NEU%	77.1 (64.5,87.5)	79.5 (67.53,89.1)	0.457 [a]
PLT (10^9/L)	83 (37, 196)	135.5 (40, 252.25)	0.142 [a]
Concomitant therapy			
ECMO (n, %)	11 (4.93%)	2 (3.57%)	1.000 [d]
CRRT (n, %)	61 (27.35%)	16 (28.57%)	0.855 [c]
Co-medication [f] (n, %)	17 (7.62%)	3 (5.36%)	0.774 [d]
Concomitant diseases			
AML (n, %)	30 (13.45%)	8 (14.29%)	0.871 [c]
Hypoproteinemia (n, %)	143 (64.13%)	34 (60.71%)	0.636 [c]
Sepsis, (n, %)	80 (35.87%)	14 (25%)	0.110 [c]

APACHE II, acute physiology and chronic health evaluation II; ALB, albumin; Cys-C, cystatin C; eGFR, estimated glomerular clearance; CLcr, creatinine clearance rate; AST, aspartate aminotransferase; ALT, alanine aminotransferase; TBIL, total bilirubin; NEU%, the percentage of neutrophils; PLT, platelet count; ECMO, extracorporeal membrane oxygenation; CRRT, continuous renal replacement therapy, AML, acute myeloid leukemia.

Measurement data were presented as median and interquartile range (IQR) for non-normal distribution variables and mean ± SD for normal distribution variables. Categorical data were expressed as n (%).

[a] Mann–Whitney U test.

[b] Independent t-test.

[c] Chi-squared test.

[d] Fisher's exact test.

[e] Creatinine clearance was calculated by the Cockcroft formula. CLcr = (140 – age [years]) ×weight (WT, kg) × 0.85 (if female)/0.818 × SCr (μmol/L).

[f] Comedication included Furosemide, Amikacin Sulfate, Cyclosporine, Isepamicin Sulfate, Amphotericin B liposome and Colistin Sulfate.

TABLE 2 | The model performance metrics of six different algorithms.

Model	R^2	MAE	MSE	Accuracy-1 [a]	Accuracy-2 [b]
SVR	0.676	3.868	26.071	76.79%	62.50%
GBRT	0.670	4.054	26.568	69.64%	62.50%
RF	0.656	4.410	27.683	62.50%	48.21%
Bagging	0.652	4.440	28.059	64.29%	44.64%
Adaboost	0.610	4.743	31.386	55.36%	48.21%
XGBoost	0.551	4.630	36.186	60.71%	55.36%

SVR, Support Vector Regression; GBRT, Gradient Boosted Regression Trees; RF, Random Forest; Bagging, Boostrap aggregating; Adaboost, Adaptive Boosting; XGBoost, eXtreme Gradient Boosting.

[a] *Absolute accuracy, the predict trough concentration was within ± 5 mg/l of the observed trough concentration.*

[b] *Relative accuracy, the predict trough concentration was within ± 30% of the observed trough concentration.*

TABLE 3 | The model performance metrics of the ensemble model.

Group	R^2	MAE	MSE	Accuracy-1 [a]	Accuracy-2 [b]
Testing group	0.720	3.628	22.571	83.93%	60.71%
Validation group	0.686	3.196	18.260	77.78%	88.89%

[a] *Absolute accuracy, the predict trough concentration was within ± 5 mg/l of the observed trough concentration.*

[b] *Relative accuracy, the predict trough concentration was within ± 30% of the observed trough concentration.*

of SVR, GBRT, and RF (6:3:1) was determined. Compared to any single algorithm, the ensemble model had the best performance with the highest R^2, absolute accuracy and lowest MAE, MSE (**Table 3**). Based on the testing group's data, the absolute accuracy (± 5 mg/l) of the ensemble model was 83.93%, and the relative accuracy (± 30%) was 60.71%. To validate the ensemble model, another dataset of 20 patients were collected from the hospital as the validation group. The results showed that validation group had higher relative accuracy and lower MAE, MSE than the testing group (**Table 3**), indicating that the model has quite good generalization ability. The exact distribution of predicted and observed values for teicoplanin trough concentration was shown in **Figure 2**.

Interpretation of the Ensemble Model

Based on the selected relevant variables, the SHAP figures demonstrated the positive or negative correlations between the relevant variables and the teicoplanin trough concentrations. The SHAP summary plot of the top 20 relevant variables in the ensemble model was displayed in **Figure 3A**. The feature values ranked the importance of the prediction model, with loading dose and maintenance dose on the top two. The dot color represents the feature values of each variable, which is redder when the feature value gets higher and bluer when the feature value gets lower. Each feature value of a certain variable corresponds to a SHAP value (x-axis). For one sample, the aggregation of the SHAP values of each variable equals to the predicted teicoplanin trough concentration. To identify the features that influenced the

ensemble model the most, the average of absolute SHAP values of each relevant variable (top 20) was calculated, the top 12 of which included loading dose, maintenance dose, eGFR, duration of teicoplanin treatment, weight, CLcr, age, ALB, maintenance intervals, Cys-C, gender, and sepsis in a descending order. Among them, the SHAP value of loading dose has the highest score (0.200), followed by the SHAP value of maintenance dose (0.199), and eGFR (0.182) demonstrating their importance in predicting the teicoplanin trough concentration (**Figure 3B**).

The SHAP dependence plot of the top 12 relevant variables was displayed in **Figure 4**. Our results showed higher loading dose, maintenance dose, duration of teicoplanin treatment, weight, ALB, Cys-C, as well as lower eGFR, CLcr and age were related to higher teicoplanin trough concentration. Female patients and patients with sepsis comorbidities may have higher teicoplanin trough concentration.

DISCUSSION

Herein, we constructed an optimal prediction model of teicoplanin trough concentration, and used SHAP method to interpret of the prediction model. We selected the algorithms through R^2 comparison and continuously debug the ratio to optimize the ensemble model. Ultimately, SVR, GBRT, and RF (6:3:1) were determined, of which the R^2 and the absolute accuracy exceeded any single algorithm, and the MAE, MSE were lower than any single algorithm. The SHAP values demonstrated the feature importance and direction of each variable, and clarified the correlation between the target variable and the relevant important covariates, which is of great significance and value for the clinical medication guidance.

Machine learning is used broadly in the biomedicine field. Its main ability is to gather and interpret any relevant data even on a large scale and thus, transforms medicine to a data-driven approach. Precision treatment is one of the top applications of machine learning, where a patient receives tailored medical care, such as personalized dose adjustment, plasma concentration prediction, and adverse drug events prediction (19–22). Ensemble learning, one of the key features of machine learning, comes from a combination of various models that is capable of producing a final prediction. Random forests, gradient boosting, and stacking/meta-ensembles are some of the approaches available in this feature (13). In this study, the ensemble model performed better than any single algorithm included by contrasting the goodness-of-fit and accuracy.

The traditional pharmacokinetic analysis is based on mathematically simple techniques with poor applicability and high requirements for data quality (23). PPK analysis, a new statistical approach, combines the traditional pharmacokinetic model with population statistics model, of which nonlinear mixed-effects modeling (NONMEM) is the most widely used program (23). However, owing to the explicit analytical model used, PPK model is relatively rigid to apply, where adding or removing a parameter may be complicated (24). In contrast, self-organization is what makes up machine learning. It enables computers to access previous data without being
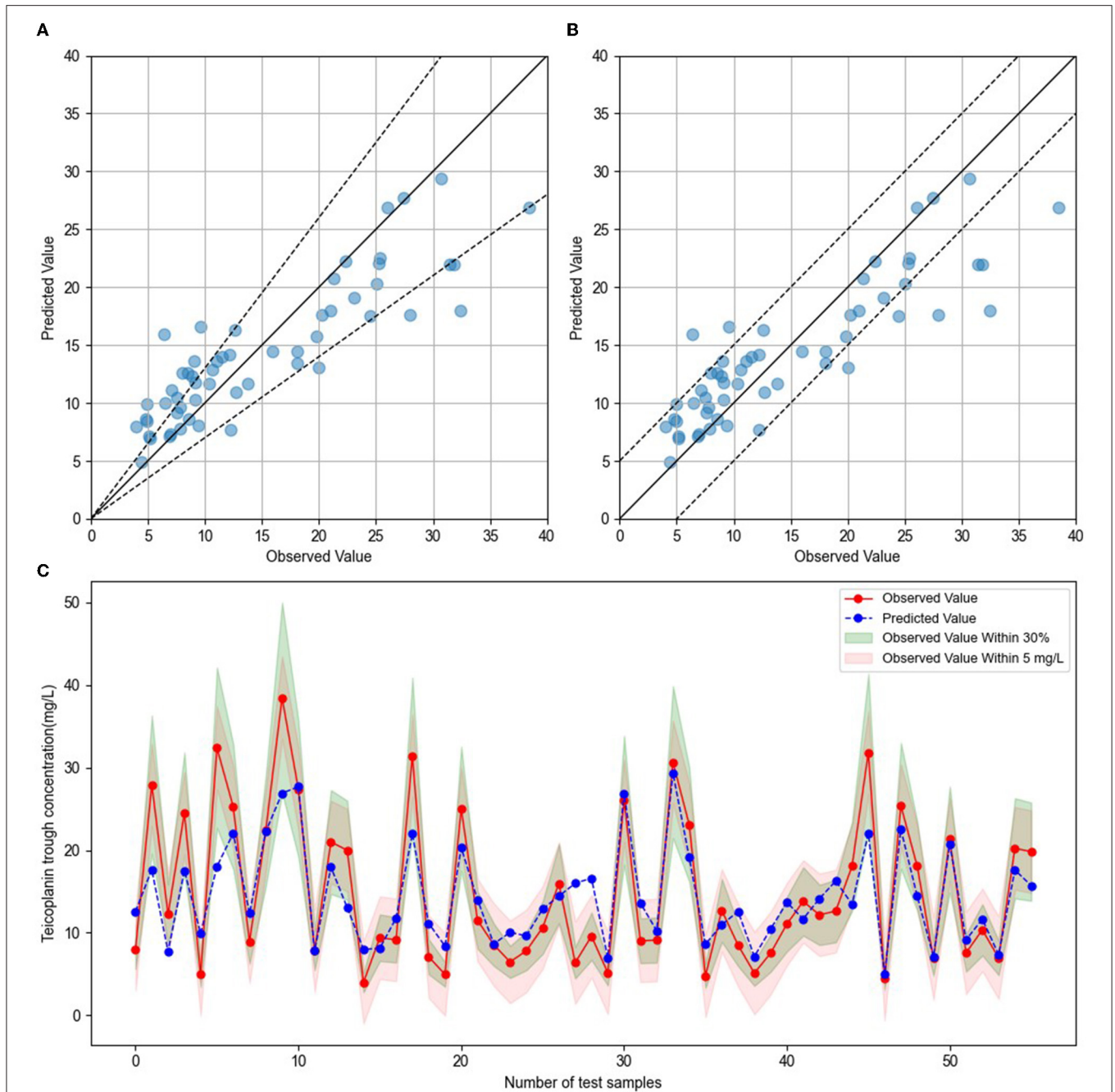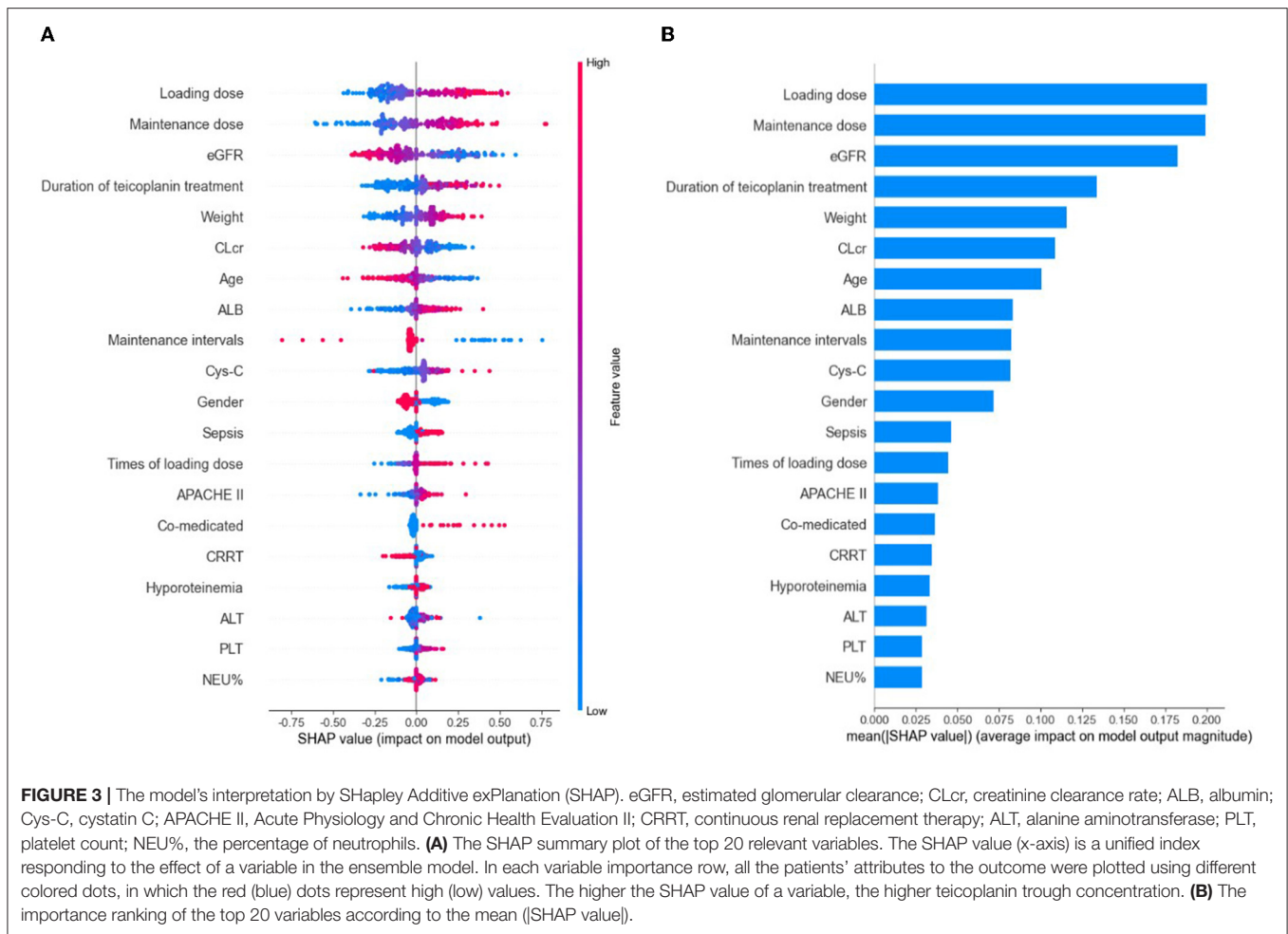
FIGURE 2 | Comparison of predicted and observed value. **(A)** The blue dots represented testing sample, with observed values on the x-axis and predicted values on the y-axis. The blue dots between the dotted lines indicated that the predict values were within ± 30% of the observed values. **(B)** The blue dots represented testing sample, with observed values on the x-axis and predicted values on the y-axis. The blue dots between the dotted lines indicated that the predict values were within ± 5 mg/l of the observed values. **(C)** The red dots indicated the observed values, and blue dots indicated the predicted values. The green shade represented within ± 30% of the observed values, and the red shade represented within ± 5 mg/l of the observed values.

explicitly programmed. Many researches have reported that the predicting accuracy of machine learning approach exceeded the PPK method (20, 25). Huang et al. constructed an ensemble prediction model of vancomycin trough concentrations, and co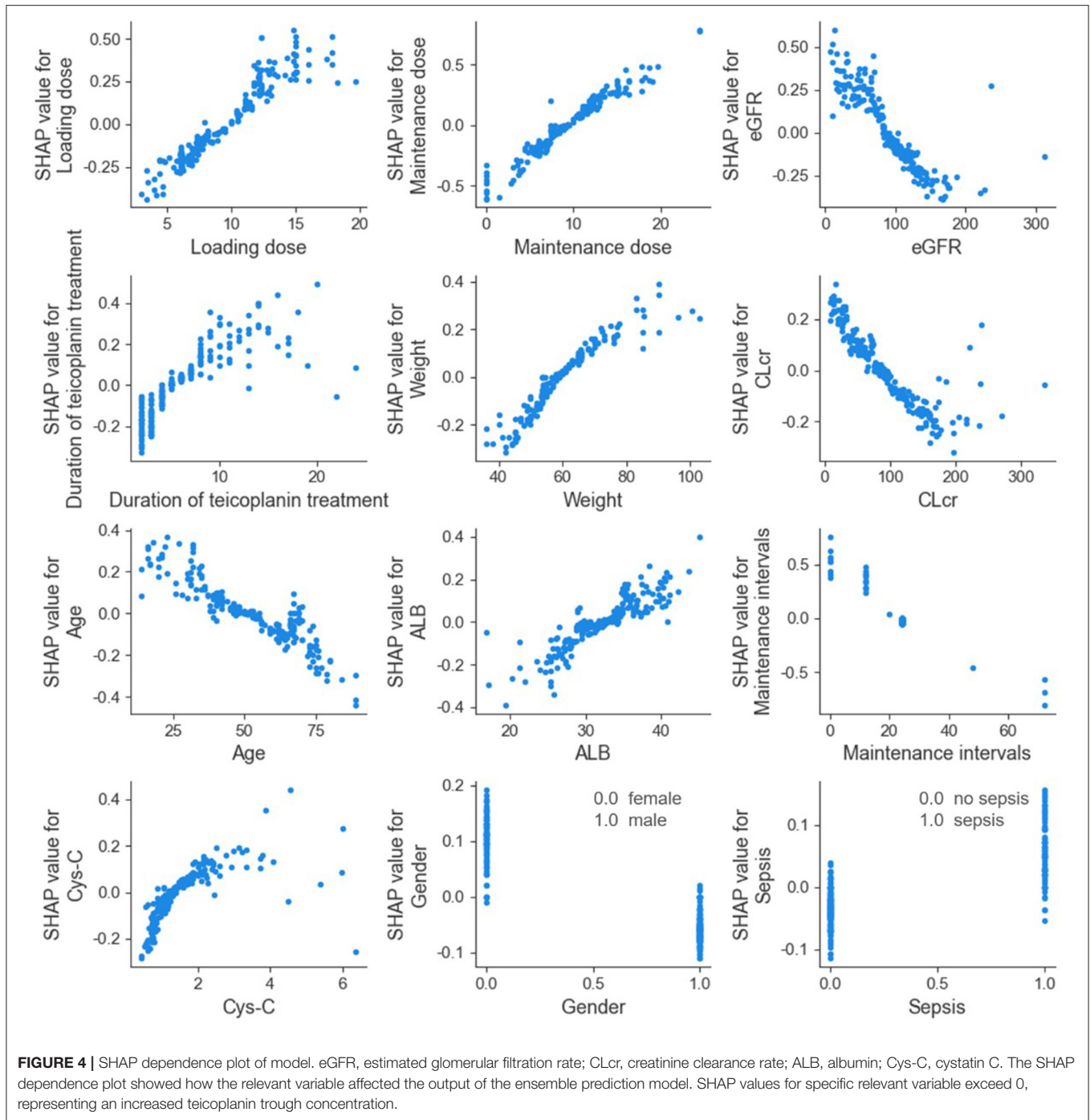mpared with PPK model. Their findings showed that machine learning model works better with higher accuracy of prediction (20). The evaluation parameters (R^2 and accuracy) of our ensemble predicting model have surpassed its vancomycin counterpart, suggesting that our model has a good prediction effect and prospect of clinical application.

FIGURE 3 | The model's interpretation by SHapley Additive exPlanation (SHAP). eGFR, estimated glomerular clearance; CLcr, creatinine clearance rate; ALB, albumin; Cys-C, cystatin C; APACHE II, Acute Physiology and Chronic Health Evaluation II; CRRT, continuous renal replacement therapy; ALT, alanine aminotransferase; PLT, platelet count; NEU%, the percentage of neutrophils. **(A)** The SHAP summary plot of the top 20 relevant variables. The SHAP value (x-axis) is a unified index responding to the effect of a variable in the ensemble model. In each variable importance row, all the patients' attributes to the outcome were plotted using different colored dots, in which the red (blue) dots represent high (low) values. The higher the SHAP value of a variable, the higher teicoplanin trough concentration. **(B)** The importance ranking of the top 20 variables according to the mean (|SHAP value|).

The interpretation of predictions from a complex statistic model might make equal sense to the model prediction itself in healthcare (26). As a classic *posthoc* interpretation method, SHAP identifies the significant influencing factors with its effect magnitude (27). In this study, the distribution of SHAP values of a relevant covariate, and also its importance and direction were measured. The averages of absolute SHAP values indicated that teicoplanin administration was the most important factor, for which the loading dose, maintenance dose, duration of teicoplanin treatment and maintenance intervals ranked first, second, fourth, and ninth, respectively. Due to its long elimination half-life, teicoplanin requires ample time for the concentration to achieve constant state. As a result, loading doses are required to exhibit the same concentration promptly. It has been reported that increase of loading doses is beneficial for the clinical outcomes, but significant teicoplanin underexposure onset of the therapy is imminent if insufficient dosing persists (28, 29), which were consistent with our study. The SHAP dependence plot showed that the teicoplanin trough concentration was positively correlated with loading dose, maintenance dose, duration of teicoplanin treatment, and negatively correlated with

maintenance intervals. It indicated that sufficient loading dose should be ensured first to rapidly achieve the effective plasma concentration, and on this basis, adequate maintenance dose, treatment duration and appropriate maintenance intervals were also necessary.

Since teicoplanin is mainly eliminated as prototype through the kidney, renal dysfunction causes a prolongation of the elimination half-life and an elevated plasma concentration of teicoplanin (28). A large number of studies have demonstrated that renal function-related parameters including eGFR, CLcr, and Cys-C were the significant covariate influencing teicoplanin elimination (9, 14, 30, 31). The concomitant diseases and medication that affect the renal function can also influence teicoplanin trough concentration. For example, sepsis is often accompanied by multiple organ dysfunction, including renal insufficiency, leading to plasma accumulation of teicoplanin due to the reduced elimination. Co-medication with drugs that are explicitly warned by instructions with a high risk of exacerbating renal toxicity, also increases the metabolic burden of renal function and affects the elimination of teicoplanin. Consistent with our findings, our results showed that low level of eGFR and CLcr, as well as high level of Cys-C were closely related to

FIGURE 4 | SHAP dependence plot of model. eGFR, estimated glomerular filtration rate; CLcr, creatinine clearance rate; ALB, albumin; Cys-C, cystatin C. The SHAP dependence plot showed how the relevant variable affected the output of the ensemble prediction model. SHAP values for specific relevant variable exceed 0, representing an increased teicoplanin trough concentration.

higher teicoplanin trough concentration, with the importance ranking third, sixth, and tenth, respectively. Moreover, patients with sepsis comorbidities and comedication might have higher teicoplanin trough concentration. Furthermore, the level of plasma ALB was another important factor that affects the teicoplanin trough concentration. With a high-binding rate of plasma ALB (90–95%), most teicoplanin combine with plasma ALB as teicoplanin-ALB complex (32). Our results demonstrated that ALB was positively related with the teicoplanin trough

concentration, ranking eighth in importance. For patients with hypoalbuminemia, ALB supplementation should be the first priority, which matters not only for the drug treatment, but for maintaining the normal physical function. Meanwhile, shortening the loading interval and appropriately increasing the loading dose can be a feasible measure. Researches have shown that the concomitant therapy such as continuous renal replacement therapy (CRRT) and extracorporeal membrane oxygenation (ECMO) may interfere with the pharmacokinetics of

teicoplanin (33), for which drugs may be cleared during *in vitro* CRRT or adhere to the fibers and catheters of oxygenator during ECMO (34, 35). Consistently, our results indicated that the teicoplanin trough concentrations of patients with ECMO and CRRT therapy showed a downward trend. In our study, pediatric patients (aged < 14 years) were excluded because of their diverse pharmacokinetics (36). According to the medication instruction, no dose adjustment is required for the elderly patients. However, our SHAP values showed that age was positively related with the teicoplanin trough concentration, which might result from the commonly concomitant therapy for elders. Fan et al. found that gender affected the tigecycline trough plasma concentration in ICU patients, and women were independent risk factors for high-tigecycline exposure (37). Similar results were obtained in our SHAP analysis that female patients have higher teicoplanin trough concentration compared with male. Thus, we suggest to take all the aforementioned factors into account in the teicoplanin administration regimen.

Despite the promising results, there is room to optimize our ensemble prediction model overall. Considerable limitations of this study should be taken into account. First, due to limited samples on hand, accuracy may be compromised. Construction of the model itself calls for a modest number of samples, let alone further modeling that the study may deem necessary. Second, since retrospective data rather than prospective data were used in the study, some uncontrollable factors were inevitable. For instance, the fluctuation in blood collection time point might lead to changes in the teicoplanin plasma concentration. Third, an external validation should be performed in the future studies to improve the applicability of this model.

Our study primarily aims to encourage the application of machine learning methods in biomedicine. To the best of our knowledge, scarcely any study has adopted machine learning approach to predict the teicoplanin trough concentration yet, and we firstly used SHAP values to interpret of the ensemble algorithm model. Therefore, our study fills the gap in this research field. In the future, we plan to further establish an easy-to-use web application based on the presented prediction model, which then could serve as a real-time support tool in clinical decision by self-learning and optimizing, and to help with the personalized dose adjustment of teicoplanin.

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.

AUTHOR CONTRIBUTIONS

PM: conception and design of the study, acquisition of data, and drafting the article. RL, QD, and YG: acquisition of data. WG and JC: analysis and interpretation of data. SS: drafting the article and analysis and interpretation of data. FL: conception and design of the study and revising it critically for important intellectual content. YC: conception and design of the study and final approval of the version to be submitted. All authors contributed to the article and approved the submitted version.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fmed.2022.808969/full#supplementary-material

REFERENCES

1. Hirai T, Hosohata K, Ogawa Y, Iwamoto T. Clinical predictors of nephrotoxicity associated with teicoplanin: Meta-analysis and meta-regression. *Basic Clin Pharmacol Toxicol.* (2021) 130:110–21. doi: 10.1111/bcpt.13679

2. Kim BK, Kim JH, Sohn KH, Kim JY, Chang YS, Kim SH. Incidence of teicoplanin adverse drug reactions among patients with vancomycin-associated adverse drug reactions and its risk factors. *Korean J Intern Med.* (2020) 35:714–22. doi: 10.3904/kjim.2018.404

3. Pea Teicoplanin F. and therapeutic drug monitoring: An update for optimal use in different patient populations. *J Infect Chemother.* (2020) 26:900–7. doi: 10.1016/j.jiac.2020.06.006

4. Emoto C, Johnson TN, Yamada T, Yamazaki H, Fukuda T. Teicoplanin physiologically based pharmacokinetic modeling offers a quantitative assessment of a theoretical influence of serum albumin and renal function on its disposition. *Eur J Clin Pharmacol.* (2021) 77:1157–68. doi: 10.1007/s00228-021-03098-w

5. McKenzie C. Antibiotic dosing in critical illness. *J Antimicrob Chemother.* (2011) 66:ii25–31. doi: 10.1093/jac/dkq516

6. Nakamura A, Takasu O, Sakai Y, Sakamoto T, Yamashita N, Mori S, et al. Development of a teicoplanin loading regimen that rapidly achieves target serum concentrations in critically ill patients with severe infections. *J Infect Chemother.* (2015) 21:449–55. doi: 10.1016/j.jiac.2015.02.002

7. Pope SD, Dellit TH, Owens RC, Hooton TM. Results of survey on implementation of Infectious Diseases Society of America and Society for Healthcare Epidemiology of America guidelines for developing an institutional program to enhance antimicrobial stewardship. *Control Hosp Epidemiol.* (2009) 30:97–8. doi: 10.1086/592979

8. Zhou L, Gao Y, Cao W, Liu J, Guan H, Zhang H, et al. Retrospective analysis of relationships among the dose regimen, trough concentration, efficacy, and safety of teicoplanin in Chinese patients with moderate-severe Gram-positive infections. *Infect Drug Resist.* (2018) 11:29–36. doi: 10.2147/IDR.S146961

9. Byrne CJ, Roberts JA, McWhinney B, Ryder SA, Fennell JP, O'Byrne P, et al. Population pharmacokinetics of teicoplanin and attainment of pharmacokinetic/pharmacodynamic targets in adult patients

with haematological malignancy. *Clin Microbiol Infect.* (2017) *23*
674.e677–674.e613. doi: 10.1016/j.cmi.2017.02.032

10. Ueda T, Takesue Y, Nakajima K, Ichiki K, Doita A, Wada Y, et al.
Enhanced loading regimen of teicoplanin is necessary to achieve therapeutic
pharmacokinetics levels for the improvement of clinical outcomes in patients
with renal dysfunction. *Eur J Clin Microbiol Infect Dis.* (2016) 35:1501–9.
doi: 10.1007/s10096-016-2691-z

11. Gautier T, Ziegler LB, Gerber MS, Campos-Náñez E, Patek SD. Artificial
intelligence and diabetes technology: a review. *Metabolism.* (2021) 154872.
doi: 10.1016/j.metabol.2021.154872

12. Rani P, Kotwal S, Manhas J, Sharma V, Sharma S. Machine learning and
deep learning based computational approaches in automatic microorganisms
image recognition: methodologies, challenges, and developments. *Arch
Comput Methods.* (2021) 1–37. doi: 10.1007/s11831-021-09639-x

13. Goecks J, Jalili V, Heiser LM, Gray JW. How machine learning will transform
biomedicine. *Cell.* (2020) 181:92–101. doi: 10.1016/j.cell.2020.03.022

14. Zhao W, Zhang D, Storme T, Baruchel A, Declèves X, Jacqz-Aigrain E.
Population pharmacokinetics and dosing optimization of teicoplanin in
children with malignant haematological disease. *Br J Clin Pharmacol.* (2015)
80:1197–207. doi: 10.1111/bcp.12710

15. Bach S, Binder A, Montavon G, Klauschen F, Müller KR, Samek
W. On Pixel-Wise Explanations for Non-Linear Classifier Decisions
by Layer-Wise Relevance Propagation. *PLoS ONE.* (2015) 10:e0130140.
doi: 10.1371/journal.pone.0130140

16. Lundberg SM, Nair B, Vavilala MS, Horibe M, Eisses MJ, Adams T,
et al. Explainable machine-learning predictions for the prevention
of hypoxaemia during surgery. *Nat Biomed Eng.* (2018) 2:749–60.
doi: 10.1038/s41551-018-0304-0

17. Tseng PY, Chen YT, Wang CH, Chiu KM, Peng YS, Hsu SP, et al. Prediction of
the development of acute kidney injury following cardiac surgery by machine
learning. *Crit Care.* (2020) 24:478. doi: 10.1186/s13054-020-03179-9

18. Liu C, Liu X, Mao Z, Hu P, Li X, Hu J, et al. Interpretable machine
learning model for early prediction of mortality in ICU patients
with rhabdomyolysis. *Med Sci Sports Exerc.* (2021) 53:1826–34.
doi: 10.1249/MSS.0000000000002674

19. Anastopoulos IN, Herczeg CK, Davis KN, Dixit AC. Multi-Drug Featurization
and Deep Learning Improve Patient-Specific Predictions of Adverse Events.
Int J Environ Res Public Health. (2021) 18. doi: 10.3390/ijerph18052600

20. Huang X, Yu Z, Bu S, Lin Z, Hao X, He W, et al. An ensemble model for
prediction of vancomycin trough concentrations in pediatric patients. *Drug
Des Devel Ther.* (2021) 15:1549–59. doi: 10.2147/DDDT.S299037

21. Woillard JB, Labriffe M, Debord J, Marquet P. Tacrolimus exposure
prediction using machine learning. *Clin Pharmacol Ther.* (2021) 110:361–9.
doi: 10.1002/cpt.2123

22. Zhu X, Huang W, Lu H, Wang Z, Ni X, Hu J, et al. A machine learning
approach to personalized dose adjustment of lamotrigine using noninvasive
clinical parameters. *Sci Rep.* (2021) 11:5568. doi: 10.1038/s41598-021-85157-x

23. You W, Widmer N, De Micheli G. Example-based support vector machine
for drug concentration analysis. *Conf Proc IEEE Eng Med Biol Soc.* (2011)
2011:153–7. doi: 10.1109/IEMBS.2011.6089917

24. Sibieude E, Khandelwal A, Girard P, Hesthaven JS, Terranova N. *Population
pharmacokinetic model selection assisted by machine learning. J Pharmacokinet
Pharmacodyn.* (2021). doi: 10.1007/s10928-021-09793-6

25. Poynton MR, Choi BM, Kim YM, Park IS, Noh GJ, Hong SO, et al. Machine
learning methods applied to pharmacokinetic modelling of remifentanil
in healthy volunteers: a multi-method comparison. *J Int Med Res.* (2009)
37:1680–91. doi: 10.1177/147323000903700603

26. Wang C, Feng L, Qi Y. Explainable deep learning predictions for illness
risk of mental disorders in Nanjing, China. *Environ Res.* (2021) 202:111740.
doi: 10.1016/j.envres.2021.111740

27. Stojić A, Stanić N, Vuković G, Stanišić S, Perišić M, Šoštarić A, et al.
Explainable extreme gradient boosting tree-based prediction of toluene,

ethylbenzene and xylene wet deposition. *Sci Total Environ.* (2019) 653:140–
147. doi: 10.1016/j.scitotenv.2018.10.368

28. Ogawa R, Kobayashi S, Sasaki Y, Makimura M, Echizen H. Population
pharmacokinetic and pharmacodynamic analyses of teicoplanin in Japanese
patients with systemic MRSA infection. *Int J Clin Pharmacol Ther.* (2013)
51:357–66. doi: 10.5414/CP201739

29. Sato M, Chida K, Suda T, Muramatsu H, Suzuki Y, Hashimoto H,
et al. Recommended initial loading dose of teicoplanin, established by
therapeutic drug monitoring, and outcome in terms of optimal trough
level. *J Infect Chemother.* (2006) 12:185–9. doi: 10.1007/s10156-006-
0446-Y

30. Gao L, Xu H, Ye Q, Li S, Wang J, Mei Y, et al. Population pharmacokinetics
and dosage optimization of teicoplanin in children with different renal
functions. *Front Pharmacol.* (2020) 11:552. doi: 10.3389/fphar.2020.
00552

31. Kasai H, Tsuji Y, Hiraki Y, Tsuruyama M, To H, Yamamoto Y. Population
pharmacokinetics of teicoplanin in hospitalized elderly patients using cystatin
C as an indicator of renal function. *J Infect Chemother.* (2018) 24:284–91.
doi: 10.1016/j.jiac.2017.12.002

32. Brink AJ, Richards GA, Lautenbach EE, Rapeport N, Schillack V, van
Niekerk L, et al. Albumin concentration significantly impacts on
free teicoplanin plasma concentrations in non-critically ill patients
with chronic bone sepsis. *Int J Antimicrob Agents.* (2015) 45:647–51.
doi: 10.1016/j.ijantimicag.2015.01.015

33. Cunio CB, Uster DW, Carland JE, Buscher H, Liu Z, Brett J, et
al. *Towards precision dosing of vancomycin in critically ill patients:
an evaluation of the predictive performance of pharmacometric models
in ICU patients. Clin Microbiol Infect.* (2020) S1198-743X(20)30388-8.
doi: 10.1016/j.cmi.2020.07.00

34. Lindberg BR, Videm V, Dahl T, Sørensen G, Fiane AE, Thiara AS. Influence of
the ECMO circuit on the concentration of nutritional supplements. *Sci Rep.*
(2020) 10:19275. doi: 10.1038/s41598-020-76299-5

35. Schetz M. Drug dosing in continuous renal replacement therapy: general
rules. *Curr Opin Crit Care.* (2007) 13:645–51. doi: 10.1097/MCC.0b013e3282f
0a3d3

36. Zhang T, Sun D, Shu Z, Duan Z, Liu Y, Du Q, et al. Population
pharmacokinetics and model-based dosing optimization of
teicoplanin in pediatric patients. *Front Pharmacol.* (2020) 11:594562.
doi: 10.3389/fphar.2020.594562

37. Fan G, Jin L, Bai H, Jiang K, Xie J, Dong Y. Safety and efficacy of tigecycline
in intensive care unit patients based on therapeutic drug monitoring.
Ther Drug Monit. (2020) 42:835–40. doi: 10.1097/FTD.0000000000
000784