# Postoperative MPA-AUC$_{0\text{-}12h}$ Prediction for Kidney Transplant Recipients based on Few-shot Learning

Pan Qiao
Donghua University
Shanghai, China
panqiao@dhu.edu.cn

Yu Xinyu
Donghua University
Shanghai, China
974581233@qq.com

Li Xinyu
Donghua University
Shanghai, China
13918802469@163.com

Shaokun*
Ruijin Hospital affiliated to Shanghai Jiao Tong University Medical College
Shanghai, China
*Correspondence: shaokuntrue@hotmail.com

***Abstract*—Mycophenolic acid (MPA) is a commonly used immunosuppressive drug. The anti-immune rejection effect of mycophenolic acid is closely related to its exposure level in the body. In clinical practice, mycophenolate acid drug exposure level is usually reflected by monitoring the area under the drug-time curve MPA-AUC$_{0\text{-}12h}$. Calculating the MPA-AUC$_{0\text{-}12h}$ requires numerous blood sampling time points. Not only does the medical staff have more work, but patients suffer more as well. Limited sampling strategies (LSS) are generally used to reduce the number of time points. Nevertheless, this method involves complicated calculations and the predictive accuracy is very low for small sample data. A new method of predicting the MPA-AUC$_{0\text{-}12h}$ value is proposed based on the selection of SHAP features with an improved neural network for small sample data. The experimental results show that the average prediction errors of the MPA-AUC$_{0\text{-}12h}$ values on different data sets by our method are better than that of the baseline models.**

***Keywords-kidney transplantation; MPA-AUC$_{0\text{-}12h}$; SHAP; affinitynet***

## I. INTRODUCTION

Mycophenolic acid (MPA) is a kind of immunosuppressant commonly used in the clinic. It is widely used in the prevention and treatment of acute rejection of transplanted organs[1]. Its strong immunosuppressive effect can significantly reduce the incidence of rejection after transplantation[2]. Clinically, the Area Under Curve (AUC) of postoperative mycophenolic acid administration MPA-AUC$_{0\text{-}12h}$ in renal transplantation patients is often monitored to evaluate the postoperative mycophenolic acid exposure in renal transplantation patients[3]. Too low drug exposure level (low MPA-AUC$_{0\text{-}12h}$) will lead to an increase in acute rejection, and too high drug exposure level (high MPA-AUC$_{0\text{-}12h}$) will lead to an increase in the incidence of gastrointestinal reactions and other adverse reactions.

However, the clinical detection of the MPA-AUC$_{0\text{-}12h}$ value is complicated. Medical staff often need to draw peripheral venous blood from patients at 10 time points. Then the plasma MPA concentration was determined by high-pressure liquid chromatography. According to the linear trapezoidal method, doctors calculated the area under the curve during 0-12 hours (MPA-AUC$_{0\text{-}12h}$) after medication[4]. Due to a number of blood sampling time points, the workload of medical staff is heavy, the cost of testing is high, and the patients are also very painful.

To solve the problem of too many clinical blood sampling time points, limited sampling strategies (LSS)[5] are usually used to calculate the value of MPA-AUC$_{0\text{-}12h}$. The method of limited sampling is divided into two steps. The first step is to determine the time point of blood collection. The second step is to estimate the value of MPA-AUC$_{0\text{-}12h}$ by Multiple Linear Regression (MLR). Although this method solves the problem of too many times of blood collection in clinical practice, the calculation process of choosing blood sampling time points is complicated. In the second step, the main method of existing research is to predict the value of MPA-AUC$_{0\text{-}12h}$ with multiple linear regression and an artificial neural network. The problem with these methods is that when the number of blood sampling time points is very small, the dimension of data in the calculation process is low. Moreover, due to the difficulty in obtaining clinical medical sample data, the data quantity is small. Therefore, the accuracy of the finite sampling method based on linear regression and a simple artificial neural network is not high. There are many research methods for this kind of small sample data. However, most of these few-shot learning methods need to be based on similar large data sets. And the model cannot guarantee the stability of prediction under the condition that there are deviations in the sample distribution of the training set and test set. The prediction effect on the test set is poor.

Given the problems existing in the current methods, this paper proposed a method for selecting blood sampling time points based on the SHAP method and a method for predicting the value of MPA-AUC$_{0\text{-}12h}$ after kidney transplantation based on a few-shot learning model. Then improve the AffinityNet model to predict the MPA-AUC$_{0\text{-}12h}$ value by combining the method of causal reweighting. The contributions of this paper are as follows.

DOI reference number:10.18293/SEKE2022-091

(1) We propose a SHAP-based method to choose blood sampling time points. The method reflects the contribution degree of feature to model output by calculating the marginal contribution of each feature, so as to select the most important feature combination for model training. In this paper, the SHAP-based method is used to select 3 or 4 blood sampling time points from 10 points for calculating the MPA-AUC$_{0\text{-}12h}$.

(2) We propose an improved AffinityNet to predict the MPA-AUC$_{0\text{-}12h}$ value for small sample data. Causal reweighting is used to deal with the deviation in the distribution of small sample data on training and test sets. The sample weight obtained by the causality weight algorithm is integrated into the AffinityNet model, which improves the stability of model prediction on the test set. Due to the low data dimension, the gaussian function is used as an attention kernel to reduce the number of aggregation nodes in the attention layer. By improving the feature extraction effect of the attention pooling layer, the KNN model becomes more suitable for low-dimensional vector aggregation.

## II. Related Work

### A. Selection Method of Blood Sampling Point

Ratain and Vogelzang[6] first proposed the use of a limited sampling strategy combined with linear regression analysis to predict the MPA-AUC$_{0\text{-}12h}$ value. Yichen Jia[7] et al. used the limited sampling strategy to establish a model to predict the MPA value on the data of 36 kidney transplant patients. They finally got the best blood sampling time points for the model at 0h, 3h, 4h, and 8h. Shao Kun et al.[8] used the limited sampling method and multiple linear regression model to predict the MPA-AUC$_{0\text{-}12h}$ based on the data of 108 patients in the early stage. Although this method can reduce the number of blood sampling time points, the calculation process is complicated, and the accuracy of the linear regression method is not high after the data dimension is reduced.

### B. Few-sample Learning Method

Wang YX et al.[18] generated virtual data through the data generation method and trained the classification model with the meta-learning method. Santoro A et al.[9]proposed MANN neural network based on the meta-learning method in 2016. In 2018, Howard J et al.[10] proposed ULMFit fine-tuning language model, which fine-tunes the model by changing the learning rate. However, both meta-learning methods and model fine-tuning methods need a large base data set.

Currently, Tianle Ma[11] proposed AffinityNet small sample neural network for medical disease type prediction. However, AffinityNet is suitable for data with higher feature dimensions. In addition, due to the deviation of the distribution of the training set and test set in the small sample data, the prediction of the model on the test set is not stable.

## III. Method

We propose a prediction method of MPA-AUC$_{0\text{-}12h}$ value based on SHAP feature selection. The AffinityNet is improved with causal weights to make it more effective for predicting small sample data. The sample weight obtained by the causal weight algorithm on global data is combined with the loss function of the AffinityNet model. The overall flow chart of the proposed method is shown in Figure 1.


Figure1. the Overall framework of the model.

### A. Blood Collection Time-point Selection Module

In this paper, a SHAP-based method is proposed to select the features of blood sampling time points data of kidney transplantation patients. Features with larger SHAP values have a greater influence on results [12].

We used the method to select features of blood collection point data. The input data were blood MPA concentration of kidney transplantation patients before $(0h)$ and after $0.5h$, $1h$, $1.5h$, $2h$, $4h$, $6h$, $8h$, $10h$ and $12h$ in one medication cycle. The output is the feature ranking of MPA-AUC$_{0\text{-}12h}$ values predicted by features at different blood sampling time points. SHAP method measures the importance of blood sampling point features by calculating the marginal contribution of each feature to the model prediction. For all feature sets of blood sampling time points $F = \{X_0, X_{0.5}, X_1, X_{1.5}, X_2, X_4, X_6, X_8, X_{10}, X_{12}\}$ and one feature $X_i$ of blood sampling time points, the SHAP-based method trains two models $f(F)$ and $f(F/X_i)$ with feature $X_i$ and without feature $X_i$ respectively. The predicted values $E[f(F)]$ and $E[f(F/X_i)]$ were obtained by the two models respectively. Then, the marginal contribution $W_i$ of the blood sampling point features on all features was

$$W_i = E[f(F)] - E[f(F/\{X_i\})] \quad (1)$$

In actual prediction, we calculated the marginal contribution of feature $X_i$ on all subsets of feature set excluding feature $X_i$ and calculated the average value to obtain the final SHAP value of the feature of a blood collection point. The calculation method is shown in Formula 2

$$w_i = \sum_{S \epsilon F\{x_i\}} \frac{|S|!(|F|-|S|-1)!}{|F|!}(f(S \cup \{x_i\}) - f(S)) \quad (2)$$

Where $S$ is the full subset of the sum of the feature sets excluding $X_i$ features. $f(S \cup X_i) - f(S)$ is the marginal contribution of characteristic $X_i$ to a subset. Then the marginal contribution of the $X_i$ feature on all subsets was averaged to obtain the SHAP value of the feature $X_i$ at blood sampling time points. Finally, SHAP values of all blood sampling time points were ranked. Therefore, the features of the highest blood sampling time points were selected for the construction of the Prediction model of the MPA-AUC$_{0\text{-}12h}$ value.

### *B. MPA-AUC$_{0\text{-}12h}$ Prediction Module*

The model prediction consists of the AffinityNet and causal sample weight modules.

#### 1) AffinityNet Model Structure

AffinityNet model is composed of a Feature Attention Layer and several stacked KNN Attention Pooling layers. The feature attention layer assigns corresponding weight values to the features of each blood collection point. Before these blood sampling time points enter the attention feature extraction module of the graph, attention calculation is carried out for blood sampling time points through the attention mechanism.

Let $H_i$ be the vector composed of $p$ blood sampling time points data of patient I after taking medicine. $W_q$ is the attention weight of the qth blood collection point. The sum of weights of $p$ features is 1, which satisfies the following formula

$$\sum_{q=1}^{p} w_q = 1, w_q \geq 0, W = (w_1, w_2, \dots, w_q) \quad (3)$$

Unlike the usual transforms, the feature attention layer performs element-by-element multiplication. The data of $P$ blood sampling time points and corresponding feature weights of patient $i$ after taking medication were used for element counterpoint multiplication. The data $f(h_i)$ of patient $i$ can be represented by Formula 4. Where $x$ is the parallel multiplication of elements. $W$ still satisfies the constraint in formula 3.

$$f(h_i) = W \times h_i \quad (4)$$

The distance $d'_{ij}$ between patients after transformation can be expressed as Formula 5. It can better measure the similarity between patients and achieve a better aggregation effect.

$$d'_{ij} = ||f(h_i) - f(h_j)|| = ||w \times h_i - w \times h_j||^2 \quad (5)$$

The KNN attention pooling layer can stack a large number of graph attention networks (GAT) together. The kidney transplant patient data vector is represented in the form of nodes in AffinityNet. The neighbor nodes of this node are calculated by the vector similarity. The high-dimensional representation of the data vector is represented by the neighbor nodes of the node at this level through the attention mechanism.

Figure 2 describes how the KNN attention pooling layer extracts the data features of patient $i$. The data vector of mycophenolic acid after the patient $i$ takes the medicine is $h_i$. In the generated graph network architecture, other patient data vectors similar to patient $i$ are $h_{i-1}, h_{i+1}, h_{i+2}$. These vectors are called $h'_i$ neighbors in the graph structure. Then the high-dimensional data feature $h'_i$ of the *i-th* patient can be represented by $h_{i-1}, h_{i+1}, h_{i+2}$.

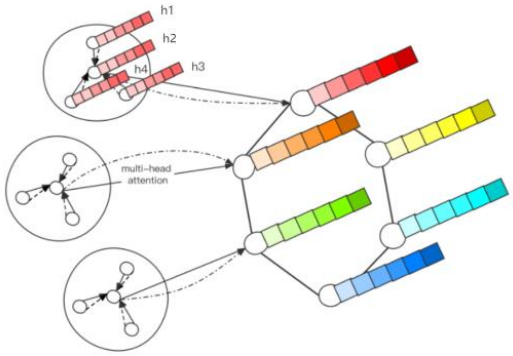


Figure2. The component unit of the attention pooling layer in KNN.

In the graph structure, a node and its nearest m neighbors should have similar feature representations. We employ the GAT module and attention-based pooling to represent vector high-dimensional features. The expression formula is as follows:

$$h'_i = f(\sum_{j \in N(i)} a(h_i, h_j).h_j) \quad (6)$$

Among them, $h_i$ represents the data vector representation of the kidney transplant patient $i$. $h'_i$ represents the data transformation feature representation of the kidney transplant the $i$-th patient, and $N(i)$ represents the neighbor of the kidney transplant patient $i$ in the figure. In the KNN attention pooling layer, $k$ is a hyperparameter used to determine how many neighbors of a node will calculate its high-dimensional vector. $f(.)$ is the ReLU() nonlinear activation function combining weight $W$ and bias $b$:

$$f(h) = \max(Wh + b) \quad (7)$$

$a_{ij} = a(h_i, h_j)$ is the attention calculation after normalization of the $i$-th and $j$-th patients. The goal is to calculate the similarity between patients. For introducing the graph attention layer, we use the Gaussian kernel function, which is more suited for low-dimensional vector aggregation. The calculation formula of attention kernel after quoting the Gaussian function is as follows:

$$a_{ij} = a(h_i, h_j) = \frac{e^{-\frac{||h_i - h_j||^2}{2\sigma}}}{\sum_{j \in N(i)} e^{-\frac{||h_i - h_j||^2}{2\sigma}}} \quad (8)$$

#### 2) Causal Sample Weight Model

In causal studies, collinearity between features is an important cause of prediction instability[13]. So that the model cannot learn true causality between features and predicted outcomes. It has been proved in causal studies that under ideal conditions there exists a set of sample weight values that make the original eigenmatrix nearly orthogonal and minimize the collinearity between input variables. Zheyan Shen et al.[14] in 2020 proposed a sample reweighted de-correlation operator to reduce collinearity of input variables to improve collinearity between input matrix features. A Decorrelated Weighting Regression (DWR) algorithm was proposed by Kun Kuang et al. [15] in 2020.

We use the Sample Reweighted Decorrelation Operator (SRDO) algorithm to reduce collinearity between different features of the input matrix. First of all, a design matrix $X$ is used to create an unrelated transformation matrix $\tilde{X}$ according to the column random resampling method. The resulting matrix $\tilde{X}$ breaks the co-distribution among variables in the original matrix $X$. Then the sample weight is learned by the density ratio estimation method. Study a set of sample weights that make the variable distribution D of $X$ matrix close to distribution $\tilde{X}$ of $\tilde{D}$ matrix.

Specifically, samples in $\tilde{X}$ matrix are labeled as positive samples (Z=1) and samples in the $X$ matrix are labeled as negative samples (Z=0). Fit a probability classifier. According to The Bayesian theory, the density ratio, namely the sample weight, is:

$$w(x) = \frac{p_{\tilde{D}}(x)}{P_D(x)} = \frac{p(x|\tilde{D})}{p(x|D)} = \frac{p(\tilde{D})}{p(D)} \frac{p(Z=1|x)}{p(Z=0|x)} \tag{9}$$

Where $x$ is constant in all samples, so it can be ignored. To find the unit mean of $w(x)$, we can further divide $w(x)'$, the mean of $w(x)$.

$$w(x)' = \frac{1}{n}\sum_{i=1}^{n} w(x_i) \tag{10}$$

$$w(x) = \frac{w(x)}{w(x)'} \tag{11}$$

After the sample weight is obtained, it is counterbalanced by the loss value of the model to correct the loss value of the model. MSE is used as the Loss function in our regression task. The final combination of weights is as follows:

$$loss = \sum_{i=1}^{n}((y_i - y_i^{pred})^2 \times w_i) \tag{12}$$

The specific causal sample weight and model integration process are shown in Figure 3:

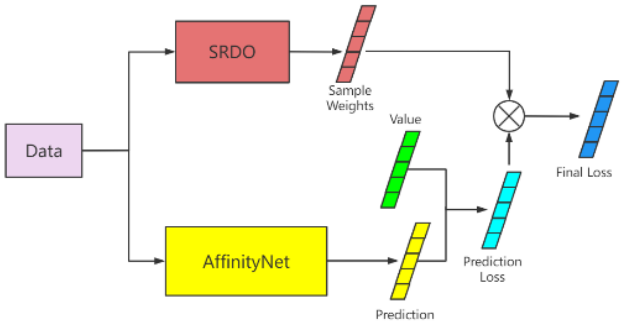


Figure3. The combination of causal sample weight and AffinityNet

## IV. EXPERIMENT

### A. Dataset

The dataset contains 152 kidney transplant patients. The data of each patient included the features of 10 blood sampling points and the MPA-AUC$_{0\text{-}12h}$ values. The 10 blood sampling time points were the blood mycophenolic acid concentrations collected before medication ($0h$) and $0.5h$, $1h$, $1.5h$, $2h$, $4h$, $6h$, $8h$, $10h$ and $12h$ after medication within one medication interval. The patient's MPA-AUC$_{0\text{-}12h}$ value was the outcome index to be predicted.

### B. Blood Collection Time-point Selection

*1) Select blood sampling point by SHAP:*

SHAP values and importance rankings of different blood sampling time points are shown in Table 1. According to the results in the table, the importance of blood sampling time points from high to low is $2h$, $6h$, $4h$, $1.5h$, $8h$, $1h$, $10h$, $0.5h$, $12h$, $0h$.

TABLE I RANKING OF THE FEATURE IMPORTANCE

| *Rank* | *Feature* | *SHAP Value* |
|---|---|---|
| 1 | $T_2$ | 6.55 |
| 2 | $T_6$ | 5.07 |
| 3 | $T_4$ | 4.53 |
| 4 | $T_{1.5}$ | 3.30 |
| 5 | $T_8$ | 3.28 |
| 6 | $T_1$ | 3.18 |
| 7 | $T_10$ | 3.15 |
| 8 | $T_{0.5}$ | 2.91 |
| 9 | $T_{12}$ | 1.51 |
| 10 | $T_0$ | 0.62 |

The features of different blood sampling times are selected. The MPA-AUC$_{0\text{-}12h}$ values were predicted by the original AffinityNet. In each group, 20% of the data were randomly selected as the training set. The MPE(Mean Predict Error) values are shown in Table 2.

TABLE II FEATURE SELECTION COMPARISON

| | $MPA-AUC_{0-12h}$**(portion = 0.2)** | |
|---|---|---|
| Data | train set/31 | test set/121 |
| T2,T6 | 17.042 | 25.390 |
| T2,T4,T6 | 14.021 | 21.587 |
| T1.5,T2,T4,T6 | 13.014 | 20.794 |
| T1.5,T2,T4,T6,T8 | 12.176 | 18.253 |
| T1,T1.5,T2,T4,T6,T8 | 12.098 | 18.638 |

The MPE of the few-shot learning model decreases with the continuous addition of new blood collection point features. The final blood sampling time points T2, T4 and T6 were chosen for the purposes of ensuring accuracy and reducing the number of features as much as possible.

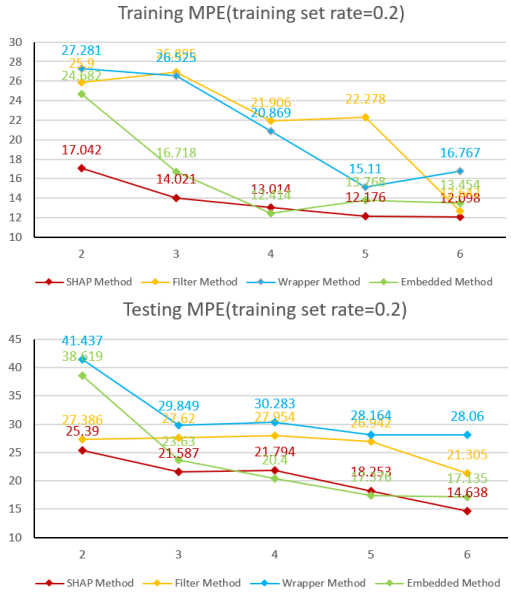


Figure4. MPE comparison of different feature selection methods

*2) Comparison of different feature selection methods:*

The SHAP-based method was compared with traditional feature selection methods including the Pearson, the Recursive Feature Elimination (RFE), and the Random Forest. As shown in Figure 4, features extracted by SHAP have better effects through model training. The prediction effect is better than other feature selection methods in most cases. When the features increase, the model MPE shows a steady decline.

## C. MPA-AUC$_{0\text{-}12h}$ Prediction

*1) Model parameter tuning:*

**a. Aggregate function experiment**

Given the reduced dimension of kidney transplantation data after feature selection, this model uses a vector similarity calculation method that is more suitable for low-latitude data as the core of attention. Observe the influence of different aggregation functions on the prediction effect of the original AffinityNet model. The features used in the experiment are the features of $T2$, $T4$ and $T6$ blood sampling time points selected in the previous experiment. The specific comparison results are shown in Table 3.

TABLE III MPE WITH DIFFERENT AGGREGATION FUNCTIONS

| | MPE (portion = 0.2) | | MPE (portion = 0.1) | |
|---|---|---|---|---|
| Similarity calculation function | train set | test set | train set | test set |
| affine | **14.021** | 21.587 | 15.192 | 38.456 |
| gaussian | 14.087 | **18.693** | **14.265** | **24.272** |
| inner-product | 18.076 | 24.645 | 19.032 | 31.662 |
| $avg_pool$ | 19.653 | 29.435 | 21.386 | 36.498 |
| cosine | 19.545 | 30.094 | 26.767 | 43.018 |
| key-value | 20.417 | 32.984 | 25.477 | 37.950 |

It can be seen from Table 3 that when 10% of the data is used as training data, the model that uses Gaussian as the attention core realizes the lowest MPE on both the training set and the test set. When 20% of the data is used as training data, using Gaussian as the attention core model has the lowest MPE on the test set, but it is slightly worse than Affine as the attention core on the training set. Therefore, in general, the Gaussian kernel function is the best attention aggregation function of the AffinityNet model in this dataset.

**b. Aggregate node number experiment**

In order to improve the attention pooling layer's ability to aggregate similar functions, this paper proposes to reduce the KNN attention layer to aggregate neighbor nodes. An experiment is designed to compare the model prediction results of different aggregation nodes, as shown in Figure 5.

When the $K$ value is 3, the model effect is the best. As the $K$ value increases, the model effect gradually becomes worse. This is because when the amount of data samples is small, the KNN attention pooling layer is extracting the value of a certain node. In the feature, the high-dimensional features of the node can be better represented by fewer neighbor nodes, and the learning efficiency can be improved.

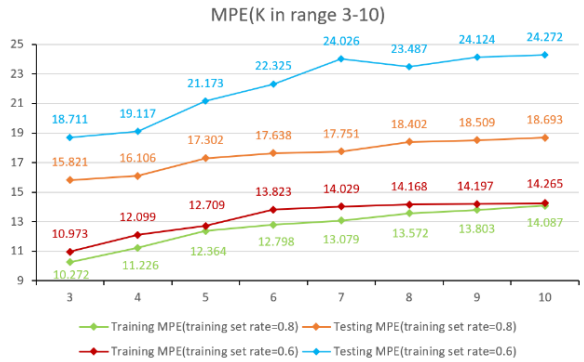


Figure5. The influence of K value selection on the model

*2) Comparison of model prediction results:*

We use traditional machine learning methods, basic neural networks, and the original AffinityNet model as the baseline model. The comparison results of different models are shown in Table 4.

TABLE IV COMPARISON OF PREDICTION MODELS

| | MPE (20% data for training) | | MPE (10% data for training) | |
|---|---|---|---|---|
| predictive model | train set | test set | train set | test set |
| Logic Regression | 26.84 | 45.54 | 29.88 | 49.62 |
| Random Forest | 25.24 | 49.06 | 28.37 | 52.33 |
| SVM | 23.99 | 43.09 | 30.01 | 51.84 |
| ANN | 17.81 | 32.96 | 21.57 | 36.48 |
| CNN | 48.59 | 79.16 | 60.60 | 79.12 |
| AffinityNet | 10.27 | 15.82 | 10.97 | 18.71 |
| Ours | **8.59** | **10.65** | **8.98** | **14.01** |

As can be seen from Table 4, when the number of samples is small, the prediction effect of traditional machine learning methods on test sets is poor. The effect is not good when the sample size is small and the distribution of the training set and test set is different. The prediction effect of the simple ANN and CNN model is not as good as the AffinityNet model. The features of the data can not be better extracted and learned. Our improved AffinityNet has the best prediction effect among all models. Especially in the test set, the prediction effect was significantly improved. Compared with the AffinityNet model, the model with causal weight has no obvious improvement in the training set. When 20% of the samples were used as training samples, the MPE value decreased from 10.27% to 8.59%. When 10% of the samples were used as training samples, the MPE value decreased from 10.97% to 8.98%. However, the effect on the test set is significantly improved. When 20% samples were used as training samples, the MPE of the original model test set was reduced from 15.82% to 10.65%. When 10% samples were used as training samples, the MPE of the original model was reduced from 18.71% to 14.01%. This shows that the model achieves a better learning effect after adding causal weight. The stability of prediction is greatly improved when the distribution of the training set and test set is different.

## D. External Data Verification

In order to prove the generalization of our proposed method of predicting MPA-AUC$_{0\text{-}12h}$ after kidney transplantation based on SHAP feature selection and improved AffinityNet model, this section uses the kidney transplant patient data set provided by

another tertiary hospital as an external the data undergoes model validation.

Data from the second hospital dataset included blood mycophenolic acid concentrations at 10 time points and patients' MPA-AUC$_{0\text{-}12h}$ values of 40 kidney transplant patients. The 10 blood sampling time points included before medication ($0h$) and after medication $0.5h$, $1h$, $1.5h$, $2h$, $4h$, $6h$, $8h$, $10h$ and $12h$. MPA-AUC$_{0\text{-}12h}$ is the target value to be predicted.

We used the SHAP-based method for feature selection of blood sampling time points from the data of kidney transplant patients in the second hospital. Then the improved AffinityNet model proposed in this paper is used for prediction. The selection results are shown in Table 5:

TABLE V RANKING OF BLOOD SAMPLING TIME POINTS

| Rank | Feature Name | SHAP Value |
|---|---|---|
| 1 | $T_4$ | 5.795 |
| 2 | $T_6$ | 5.678 |
| 3 | $T_8$ | 4.914 |
| 4 | $T_3$ | 3.052 |
| 5 | $T_{2.5}$ | 1.278 |
| 6 | $T_2$ | 1.150 |
| 7 | $T_12$ | 1.082 |
| 8 | $T_{1.5}$ | 1.008 |
| 9 | $T_1$ | 0.488 |
| 10 | $T_0.5$ | 0.426 |
| 11 | $T_0$ | 0.376 |

It can be seen from the results in the table that the feature importance ranking of 10 blood sampling time points for MPA-AUC$_{0\text{-}12h}$ value prediction is $4h$, $6h$, $8h$, $3h$, $2.5h$, $2h$, $12h$, $1.5h$, $1h$, $0.5h$, $0h$ from high to low. We also selected the features of the three most important blood sampling time points at 4h, 6h, and 8h to construct the prediction model. Since the data set of the second hospital had less data, we used 80% of the data as a training set and 20% as a test set. The prediction effect of the improved AffinityNet model proposed in this paper is compared with that of other baseline models. The comparison results are shown in Table 6:

TABLE VI MODEL PREDICTION EFFECT COMPARISON

| Prediction Model | Train Set MPE | Test Set MPE |
|---|---|---|
| Logic Regression | 29.73 | 48.55 |
| Random Forest | 24.25 | 46.96 |
| SVM | 28.90 | 49.04 |
| ANN | 19.89 | 36.90 |
| CNN | 44.55 | 59.06 |
| AffinityNet | 12.68 | 18.36 |
| Ours | **10.54** | **14.22** |

As can be seen from the results in Table 6, the improved AffinityNet neural network proposed in this paper can also achieve good prediction results in other data sets. And the prediction effect of the model is better than all baseline models. The training set MPE reached 10.54%. The MPE can reach 14.22%. Compared with the original AffinityNet, the MPE on the test set is reduced by 22.55%.

## V. CONCLUSION

In this paper, a blood collection point selection method based on SHAP is proposed, and the AffinityNet model is improved with causal weight to complete the prediction of MPA-AUC$_{0\text{-}12h}$ The experimental results show that compared with the previous method, the method presented in this paper can achieve better prediction effect, and can effectively reduce the number of clinical blood sampling. It reduces the workload of clinicians and makes the predicted MPA-AUC$_{0\text{-}12h}$ values have greater clinical reference value.


## ACKNOWLEDGEMENTS

This work was supported by the National Key RD Program of China under Grant 2019YFE0190500.


## REFERENCES


[1] Scott LJ, McKeage K, Keam SJ, et al. Tacrolimus: a further update of its use in the management of organ transplantation[J]. Drugs, 2003, 63(12): 1247-1297

[2] Lu Y P, Zhu Y C, Liang M Z, et al. Therapeutic drug monitoring of mycophenolic acid can be used as predictor of clinical events for kidney transplant recipients treated with mycophenolate mofetil[J]. Transplantation Proceedings, 2006, 38(7): 2048-2050.

[3] Le Meur Y, Buchler M, Thierry A, et al. Individualized mycophenolate mofetil dosing based on drug exposure significantly improves patient outcomes after renal transplantation[J]. Am J Transplant, 2007, 7(1): 2496-2503.

[4] Ren B, Li MW, Tang L, et al. Rapid determination of mycophenolic acid in plasma by HPLC[J]. Chin Hosp Pharm, 2008, 28: 407−408.

[5] Willis C, Taylor PJ, Salm P, et al. Evaluation of limited sampling strategies for estimation of 12-hour mycophenolic acid area under the plasma concentration-time curve in adult renal transplant patients[J]. Ther Drug Monit, 2000, 22: 549-554.

[6] Ratain,M.J. and Vogelzang,N.J.(1987). Limited sampling model for vinblastine pharmacokinetics. Cancer Treat. Rep. 71, 935–939.

[7] Jia Y, Peng B, Li L, et al. Estimation of mycophenolic acid area under the curve with limited-sampling strategy in Chinese renal transplant recipients receiving enteric-coated mycophenolate sodium[J]. Ther Drug Monit, 2017, 39(1): 29-36.

[8] Shao Kun,et al. The relationship between the pharmacokinetics of mycophenolate mofetil and the polymorphism of multidrug resistance gene 1[J]. Chinese Journal of Organ Transplantation, 2009(02) : 81-84.

[9] Santoro A, Bartunov S, Botvinick M, et al. One-shot learning with memory-augmented neural networks. arXiv preprint arXiv: 1605.06065,2016.

[10] Howard J, Ruder S. Universal language model fine-tuning for text classification. arXiv:1801.06146,2018.

[11] Ma T, Zhang A. AffinityNet: semi-supervised few-shot learning for disease type prediction[J]. 2018, 33(01): 1069-1076.

[12] Futagami Katsuya, Fukazawa Yusuke, Kapoor Nakul, KIto Tomomi. Pairwise acquisition prediction with SHAP value interpretation[J]. The Journal of Finance and Data Science. 10.1016/J.JFDS.2021.02.001.

[13] Martin Arjovsky, Léon Bottou, Ishaan Gulrajani, and David Lopez-Paz. Invariant risk minimization. arXiv preprintarXiv: 1907.02893,2019.

[14] Zheyan Shen, Peng Cui, Tong Zhang, and Kun Kuang. Stable learning via sample reweighting. AAAI. 5692–5699,2020.

[15] K. Kuang, R. Xiong, P. Cui, S. Athey, and B. Li. Stable prediction with model misspecification and agnostic distribution shift. Proceedings of the AAAI Conference on Artificial Intelligence, vol. 34, no. 4, 2020:4485–4492