

A novel method to predict white blood cells after kidney transplantation based on machine learning

DIGITAL HEALTH
Volume 10: 1–20
© The Author(s) 2024
Article reuse guidelines:
sagepub.com/journals-permissions
DOI: 10.1177/20552076241288107
journals.sagepub.com/home/dhj



Songping He¹, Xiangxi Li², Zunyuan Zhao², Bin Li¹, Xin Tan³, Hui Guo⁴,
Yanyan Chen⁵ and Xia Lu⁴ 

Abstract

Background: Abnormal white blood cell count after kidney transplantation is an important adverse clinical outcome. The abnormal white blood cell count in patients after surgery may be caused by the use of immunosuppressive agents and other factors. A lower white blood cell count than normal will greatly increase the probability of adverse outcomes such as infection and reduce the success rate of surgery.

Objective: To establish a machine learning prediction model of leukocyte drop to abnormal level after kidney transplantation, and provide reference for clinical treatment.

Methods: A total of 546 kidney transplant patients were selected as the study subjects. The time correlation feature of the ratio of the duration time of each variable to the total time in different intervals was innovatively introduced. Least absolute shrinkage and selection operator algorithm was used for correlation analysis of 85 candidate variables, and the top 20 variables were retained in the end. Eight machine learning algorithms, including Logistic-L1, Logistic-L2, support vector machine, decision tree, random forest, multilayer perceptron, extreme gradient boosting and light gradient boosting machine, were used for the five-fold cross-validation on all data sets, and the algorithm with the best performance was selected as the final prediction algorithm based on the average area under the curve.

Results: As the final prediction model, the accuracy, sensitivity, specificity and area under the curve values of the multilayer perceptron model in test set were 71.34%, 61.18%, 82.28% and 77.30%, respectively. The most important factors affecting leukopenia after surgery were the proportion of time of lymphocyte less than normal, blood group AB, gender, and platelet CV.

Conclusions: The multilayer perceptron model explored in this study shows significant potential in predicting abnormal white blood cell counts after kidney transplantation. This model can help stratify risk following transplantation, subject to external and/or prospective validation.

Keywords

Clinical prediction, multilayer perceptron, feature extraction, kidney transplant, machine learning

Submission date: 14 January 2024; Acceptance date: 13 September 2024

¹Digital Manufacturing Equipment National Engineering Research Center, Huazhong University of Science and Technology, Wuhan, China

²National NC System Engineering Research Center, Huazhong University of Science and Technology, Wuhan, China

³Wuhan Intelligent Equipment Industrial Institute Co., Ltd, Wuhan, China

⁴Institute of Organ Transplantation, Tongji Hospital, Tongji Medical College, Huazhong University of Science and Technology; Key Laboratory of Organ Transplantation, Ministry of Education; NHC Key Laboratory of Organ Transplantation; Key Laboratory of Organ Transplantation, Chinese Academy of Medical Sciences, Wuhan, China

⁵Big Data and Artificial Intelligence Office, Tongji Hospital, Tongji Medical College, Huazhong University of Science and Technology, Wuhan, China

Corresponding author:

Xia Lu, Institute of Organ Transplantation, Tongji Hospital, Tongji Medical College, Huazhong University of Science and Technology; Key Laboratory of Organ Transplantation, Ministry of Education; NHC Key Laboratory of Organ Transplantation; Key Laboratory of Organ Transplantation, Chinese Academy of Medical Sciences, Wuhan, China.
Email: xialu@tjh.tjmu.edu.cn



Introduction

Kidney transplant surgery is the most effective treatment for end-stage renal disease today¹ with the continuous progress of modern medical technology. The success rate of kidney transplantation is also increasing, and the quality of life of patients with end-stage renal disease has been greatly improved.² A more common problem in the field of kidney transplantation today is that postoperative infection of kidney transplant recipients has become the main cause of eventual death of patients due to the use of various immunosuppressive drugs.³ Lung infection is one of the main types of infection after kidney transplantation.⁴ White blood cell level is an important laboratory standard to determine bacterial or viral infection. The normal range of white blood cell count in the human body should be between 4.0 and 10.0*10⁹/L. Bacterial or viral infection is likely to occur in the human body beyond this range. This indicator also needs to be given attention in clinical practice. The occurrence of infection is highly correlated with abnormal count of white blood cells.⁵

With the development of artificial intelligence, common machine learning models have been developed and applied in healthcare.⁶ Some scholars have compared the prediction performance of machine learning models and traditional regression methods, and found that the prediction performance of machine learning methods is better than that of traditional regression methods.⁷ At present, in the field of kidney transplantation, some prediction models based on artificial intelligence algorithm based on machine learning are available to predict and diagnose the occurrence of risk outcomes after kidney transplantation; however, they mainly focus on the occurrence of serious outcomes such as acute rejection (AR), delayed graft function (DGF), and death.⁸⁻¹⁴ Studies on the decrease in white blood cells below the normal level in patients after kidney transplantation are few. In addition, most of the feature indicators included in the existing studies are category-type features that can be directly extracted from the real-world data. For clinical variables monitored for a long time, statistics such as mean and variance are usually extracted directly; a more comprehensive and reasonable feature extraction method is lacking.^{15,16}

Therefore, the short-term clinical characteristics after surgery need to be fully explored to determine the post-operative leukocyte level of kidney transplant patients. Moreover, a machine learning model should be established to predict whether the adverse outcome of leukocyte number will fall below the normal level in the long term after surgery. The other important tasks are to assist doctors in conducting fast identification and improve the prognostic effect of kidney transplant surgery.

In order to meet the above needs, this study proposed a time-dependent feature extraction method for some features

repeatedly detected in a period of time, focusing on the proportion of the duration of each feature in different value ranges to the total time, and incorporated the longitudinal temporal features into the prediction algorithm modeling in this way. Aiming at the goal of abnormal white blood cell count in patients after kidney transplantation, least absolute shrinkage and selection operator (LASSO) algorithm was used to get 20 features that had the greatest impact on the target from 85 candidate features into the modeling. Finally, from Logistic L1 (Logic-L1), Logistic L2 (Logic-L2), support vector machine (SVM), decision tree (DT), random forest (RF), multilayer perceptron (MLP), extreme gradient boosting (XGBoost) and light gradient boosting machine (LightGBM). Among the eight machine learning models, the MLP model with the best prediction effect was selected as the final model. The area under the curve (AUC) on the test set reached 77.30%.

Methods

Sample establishment

This study was conducted in a real-world data set.¹⁷ In this retrospective study, all kidney transplantation operations were performed in a third-class A hospital in Wuhan. The Organ Transplantation Institute of this hospital is the largest comprehensive medical service and research institution that specializes in the clinical research of organ transplantation in China. To date, more than 6000 cases of kidney transplantation have been conducted, which ranked first in China. This study initially planned to include the electronic medical record data of a total of 958 kidney transplant patients in this hospital from January 1, 2017 to June 31, 2019. Only the previous clinical data were used, and the identification number and other information were desensitized. Thus, the privacy of patients could be fully protected. Accordingly, the requirement of patients' informed consent was exempted. Each patient included in the study was also given a unique identification number to ensure anonymity and accuracy.

Due to the great differences between relative kidney transplantation and donation of cardiac death (DCD) kidney transplantation, child kidney transplantation and adult kidney transplantation, single kidney transplantation and multiple kidney transplantation, it is necessary to ensure the unity of research objects. In addition, because what this paper needs to predict is the level of white blood cell count within one year after surgery, the follow-up period of blood routine must be more than one year in order to meet our observation time requirements. Finally, considering the integrity of the data, the patient data with missing values in the candidate features were deleted to avoid the impact of missing values on the prediction results and improve the prediction effect of the model. To sum up,

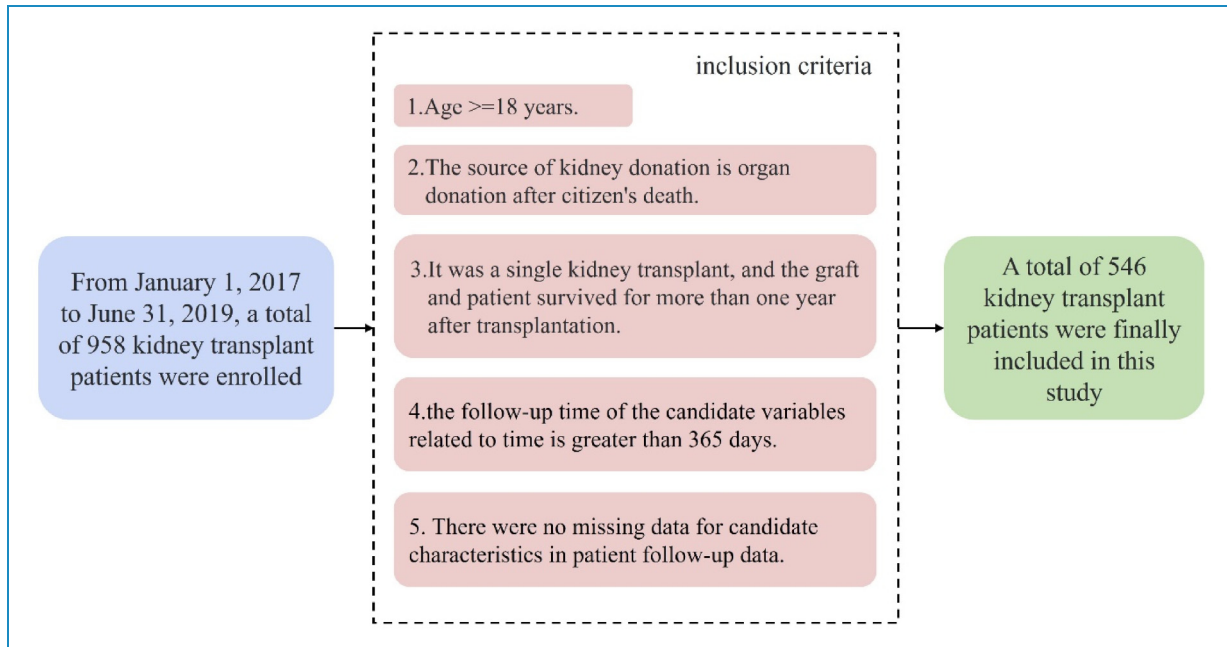


Figure 1. Sample establishment flowchart.

we determined five inclusion criteria for the study samples in this paper, as shown in Figure 1. Real-world data of a total of 546 patients were finally included in this study.

Included variables

The patient data in this study were all real-world data, which were exported and extracted from the hospital's information system. The prospective collection method was adopted. The included characteristic variables are shown in Table 1.

The abovementioned variables are all clinical data of recipients. The electronic information level of kidney donor and donor data in domestic hospitals is imperfect. Thus, this project included clinical variables of recipients except donor blood type, which is more in line with our clinical application scenario.

Feature extraction

Among all the variables included, some variables were repeatedly detected and had long observation period, low recording frequency, and unfixed characteristics, such as blood routine, liver function, kidney function, and tacrolimus blood concentration, of the test data within 30 days after surgery. In the existing studies, only the occurrence times of extreme, mean, or outlier value in a period of research time are usually extracted as the characteristic variable to be included at last, which is obviously incomprehensive and lacks the characteristics reflecting the changes in the variable in this period.

Table 1. List of candidate characteristic variables.

Category of variables	Variables
General condition of the patient	Gender, age, height, weight, BMI, number of weight changes during hospitalization (kg), and blood type.
Previous history and disease history	Smoking, alcohol use, high blood pressure, diabetes, blood transfusion history, vaccines, infectious diseases, and viral hepatitis.
Preoperative initial blood routine examination	White blood cell count, neutrophil count, lymphocyte count, red blood cell count, hemoglobin count, and platelet count.
Test and check within 30 days after surgery	Blood routine, liver function, and kidney function.
Medication status	Blood concentration of tacrolimus within 30 days and total initial doses of mycophenolate mofetil and mycophenolate sodium enteric-coated tablets.

For this type of variable, this study focuses on the proportion of the duration of each feature in different value ranges in the total time. In this way, the longitudinal temporal features are incorporated into the modeling of the

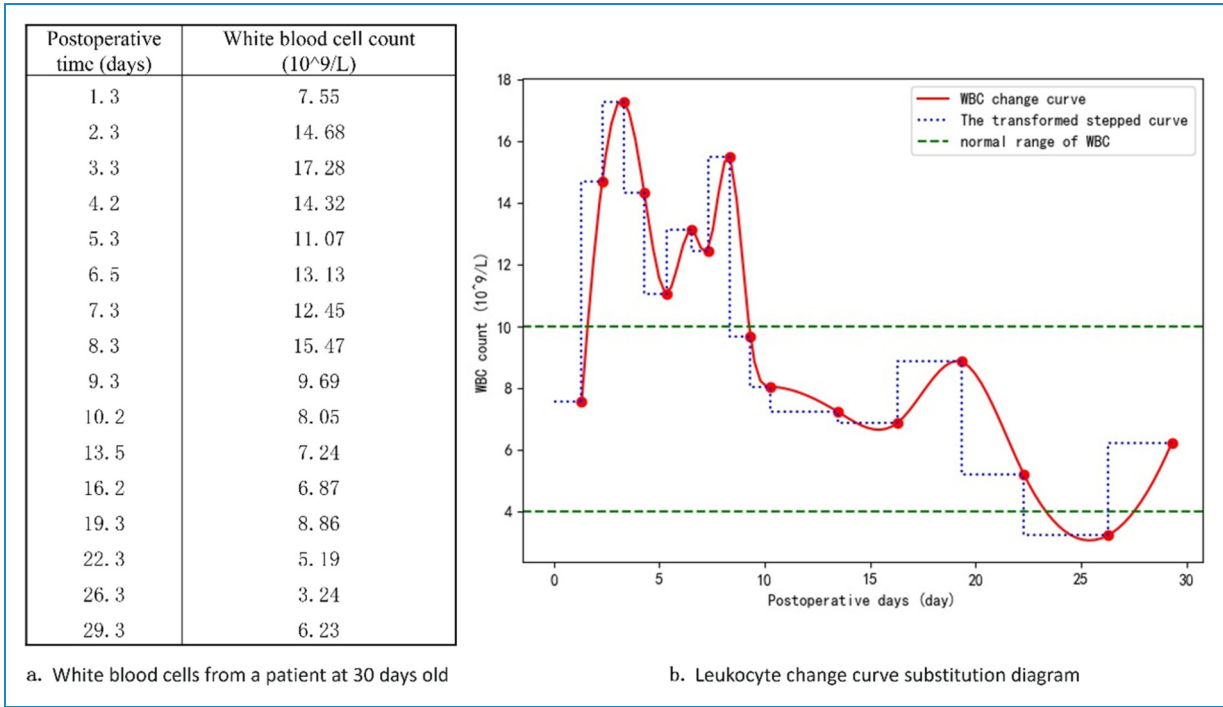


Figure 2. Schematic of extraction of duration proportion of each count of leukocyte.

prediction algorithm. For the routine white blood cell count, the normal range as mentioned above is between 4.0 and $10.0 \times 10^9/L$, and the original record of the white blood cell count of a patient numbered 2017350 within 30 days after surgery is shown in Figure 2a. Through interpolation, smooth curves were used to plot the general trend of changes. We used a step-shaped curve to replace the original curve for obtaining the percentage of different durations of white blood cell count in the total time of the patient. Accordingly, the proportion of the sum of each duration of white blood cell count in different ranges in the total time can be calculated, as shown in Figure 2b.

The specific calculation process is as follows.

The duration of each white blood cell count T_i is calculated, as shown in Formula (1):

$$T_i = t_i - t_{i-1} \quad (1)$$

where t_i represents the number of days between the i time postoperative record and the operation time particularly when $i = 1$, $t_{i-1} = 0$.

$$R_{0-4} = \frac{\sum_l^a T_l}{t_i} \times 100\% \quad (2)$$

$$R_{4-10} = \frac{\sum_m^b T_m}{t_i} \times 100\% \quad (3)$$

$$R_{10+} = \frac{\sum_n^c T_n}{t_i} \times 100\% \quad (4)$$

In Formulas (2), (3), and (4), R_{0-4} , R_{4-10} , and R_{10+} respectively represent the proportion of the duration in the total time when the white blood cell counts are less than 4, 4–10, and greater than 10; T_l means T_1, T_2, \dots, T_a : a duration periods are less than 4; T_m means T_1, T_2, \dots, T_b : b duration periods range from 4 to 10; T_n means T_1, T_2, \dots, T_c : c duration periods are in the range greater than 10; t_i indicates the total time, that is, the number of days between the last record within 30 days, $T_l + T_m + T_n = t_i$, $a + b + c = i$. The duration of each count of leukocyte in each patient is finally obtained, as shown in Table 2.

The features in the abovementioned table are taken as the timing features of the white blood cell count, and they are included in the next study together with the mean value of the white blood cell count, the coefficient of variation CV, and the maximum value. Other variables for feature extraction using the same method include neutrophil count, lymphocyte count, blood glucose value, uric acid value, and tacrolimus concentration, which will be mentioned later.

Statistical analysis of normal and abnormal groups

This study focused on the white blood cell levels of all patients from 30 days to 365 days after surgery. With $4000/mm^3$ as the dividing line, all samples were divided into two groups: the abnormal group with white blood cell count that dropped below $4000/mm^3$, and the normal

Table 2. Extracted time proportion characteristics of different intervals.

ID	$R_{0-4}(\%)$	$R_{6-10}(\%)$	$R_{10+}(\%)$
.....
2017350	13.63	62.42	23.95
2017352	9.69	84.13	6.18
2017356	22.87	42.84	34.29
2017357	0	87.48	12.52
2017359	0	77.96	22.04
.....

group with white blood cell count that remained above $4000/\text{mm}^3$. SPSS 26.0 software was used for statistical analysis of the two groups of data. Table 3 shows the results of statistical comparison between the two groups of patients.

In Table 3, the weight change is the patient's discharge weight minus admission weight. WBC, NEU, LYM, RBC, HGB, and PLT respectively, represent the white blood cell count, neutrophil count, lymphocyte count, red blood cell count, hemoglobin, and platelet count in blood routine. Pre represents preoperative. The most recent test results before the time point of operation were taken. Max, Min, Mean, and CV represent the maximum value, minimum value, average value, and coefficient of variation of each variable within 30 days after surgery. Inc represents the difference between the maximum value after surgery and the preoperative value. Rdc represents the difference between the minimum value after surgery and the preoperative value. The values with the mathematical comparison symbol "<" and ">" are the proportion of the duration of different value ranges of this variable in the total time within 30 days. CR, eGFR, UREA, UA, AST, and ALT, respectively, represent creatinine, glomerular filtration rate, urea, uric acid, glutamic-oxalacetic transaminase, and alanine transaminase. After the operation of these variables will have a large fluctuation, the closest to the post-operative 30 days of the test results was taken due to the short time. TAC represents the concentration of tacrolimus. TDS represents the sum of the initial dose of mycophenolate mofetil and mycophenolate sodium enteric-coated tablets.

In Table 3, the data for numerical variables are expressed as mean (standard deviation), and the data for categorical variables are expressed as counts in each group (percentage of the group). For numerical variables, Shapiro–Wilk test was first used for normality test. If P value < .05, then the variable did not obey normal distribution. In this study,

except for preoperative red blood cells and preoperative platelets, all other numerical variables did not obey normal distribution. For continuous variables with non-normal distribution, Mann–Whitney rank sum test was used, and $P < .001$ indicated significant difference. For comparison of continuous variables with normal distribution, T test was used, and $P < .05$ indicated significant difference. For class variables, Chi-squared test was used, and $P < .05$ indicated significant difference.

As shown in Table 3, significant differences are observed between the normal and abnormal groups in gender, height, maximum value of blood routine, average value, and proportion of each time.

Feature transformation

For numerical data, standard deviation standardization is adopted for preprocessing. After processing by this method, the values of different variables are processed into the form of mean value 0 and standard deviation 1. For categorical data, a categorical variable with m possible values can be converted into m binary features, and these features are mutually exclusive.

The abovementioned characteristic changes will result in more reasonable distance calculation between different types of variables.

Feature screening

With the aforementioned treatment, the initial number of variables included in this study was 85. Necessary feature screening was required to determine the risk variables most associated with the reduction in white blood cells to below normal levels after kidney transplantation. The minimum absolute contraction and selection (LASSO) algorithm commonly used in feature selection algorithms was used in this study,^{18,19} "lambda.min" variable was used as the final λ value. LASSO regression algorithm forces the coefficient of non-significant variables to shrink to 0 through L1 regularization. In this process, potential confounding variables are excluded, and only the part of variables with the greatest effect on the outcome are reserved for its prediction.

A total of 85 candidate variables were screened using R version 3.6.2. Figure 3 shows the coefficient variation curve of each candidate variable with the value of λ . The x axis represents the value of $\log(\lambda)$, and the y axis represents the coefficient of each variable in LASSO regression.

Figure 4 shows the number of retained variables and classification error change chart. x axis represents the coefficient that has not been reduced to 0 variable number, the lower x axis represents the value of $\log(\lambda)$, and y axis represents 95% CI binary error. The dashed line with the smallest dichotomous error corresponds to the optimal lambda.min.

Table 3. Statistical analysis and comparison of all candidate variables in the normal and abnormal groups.

Variable name	Total sample number (n = 546)	Normal group (n = 281)	Abnormal group (n = 265)	P-Value
Gender				<.001
Male	393 (71.9)	225 (80.1)	168 (63.4)	
Female	153 (28.1)	56 (19.9)	97 (36.6)	
Age, years	40.4 (10.8)	40.7 (11.0)	40.0 (10.5)	.489
Height, cm	167.7 (7.4)	168.9 (6.6)	166.5 (7.9)	.001
Weight, kg	60.5 (11.0)	61.4 (10.1)	59.4 (11.7)	.015
BMI	21.4 (3.2)	21.5 (2.9)	21.3 (3.4)	.293
Weight change, kg	-4.9 (4.2)	-4.8 (3.9)	-4.9 (4.4)	.652
Donor blood type				.067
A	178 (32.6)	87 (31.0)	91 (34.3)	
B	128 (23.4)	73 (25.9)	55 (20.7)	
AB	54 (9.9)	20 (7.1)	34 (12.8)	
O	186 (34.1)	101 (36.0)	85 (32.1)	
Recipient's blood type				.034
A	181 (33.1)	90 (32.0)	91 (34.3)	
B	128 (23.4)	74 (26.3)	54 (20.3)	
AB	59 (10.8)	21 (7.5)	38 (14.3)	
O	178 (32.6)	96 (34.1)	82 (30.9)	
Hypertension				.745
Yes	354 (64.8)	184 (65.4)	170 (64.1)	
No	192 (35.2)	97 (34.6)	95 (35.9)	
Diabetes				.195
Yes	45 (8.2)	19 (6.8)	26 (9.8)	
No	501 (91.8)	262 (93.2)	239 (90.2)	
History of blood transfusion				.727
Yes	161 (29.5)	81 (28.8)	80 (30.2)	
No	385 (70.5)	200 (71.2)	185 (69.8)	

(continued)

Table 3. Continued.

Variable name	Total sample number (n = 546)	Normal group (n = 281)	Abnormal group (n = 265)	P-Value
Vaccinations				.666
Yes	82 (15.0)	44 (15.7)	38 (14.3)	
No	464 (85.0)	237 (84.3)	227 (85.7)	
Smoking				.236
Yes	174 (31.9)	96 (34.2)	78 (29.4)	
No	372 (68.1)	185 (65.8)	187 (70.6)	
Drinking				.840
Yes	95 (17.4)	48 (17.1)	47 (17.7)	
No	451 (82.6)	233 (82.9)	218 (82.3)	
Infectious diseases				.877
Yes	77 (14.1)	39 (13.9)	38 (14.3)	
No	469 (85.9)	242 (86.1)	227 (85.7)	
Viral hepatitis				.712
Yes	53 (9.7)	26 (9.3)	27 (10.2)	
No	493 (90.3)	255 (90.7)	238 (89.8)	
WBC-Pre, *10 ⁹ /L	6.5 (2.0)	6.7 (2.2)	6.1 (1.7)	.001
NEU-Pre, *10 ⁹ /L	4.5 (1.8)	4.6 (2.0)	4.3 (1.5)	.078
LYM-Pre, *10 ⁹ /L	1.34 (0.48)	1.43 (0.50)	1.26 (0.48)	<.001
RBC-Pre, *10 ¹² /L	3.8 (0.7)	3.8 (0.7)	3.7 (0.7)	.033
HGB-Pre, g/L	111.8 (21.3)	112.9 (21.7)	110.5 (20.8)	.179
PLT-Pre, *10 ⁹ /L	186.6 (64.4)	190.2 (64.1)	182.7 (64.5)	.123
WBC-Max, *10 ⁹ /L	16.4 (4.6)	16.9 (4.5)	15.7 (4.6)	<.001
NEU-Max, *10 ⁹ /L	15.1 (4.3)	15.6 (4.2)	14.5 (4.3)	.001
LYM-Max, *10 ⁹ /L	2.02 (0.88)	2.27 (0.89)	1.76 (0.81)	<.001
LYM-Min, *10 ⁹ /L	0.28 (0.17)	0.31 (0.17)	0.24 (0.15)	<.001
RBC-Max, *10 ¹² /L	3.9 (0.6)	4.0 (0.6)	3.8 (0.6)	<.001
HGB-Max, g/L	118.5 (16.1)	120.9 (16.0)	115.8 (15.8)	<.001

(continued)

Table 3. Continued.

Variable name	Total sample number (n = 546)	Normal group (n = 281)	Abnormal group (n = 265)	P-Value
PLT-Max, *10 ⁹ /L	256.5 (80.7)	258.7 (78.8)	254.2 (82.7)	.299
WBC-Inc, *10 ⁹ /L	9.9 (4.3)	10.1 (4.1)	9.6 (4.4)	.018
NEU-Inc, *10 ⁹ /L	10.6 (4.0)	10.9 (3.9)	10.2 (4.0)	.015
LYM-Rdc, *10 ⁹ /L	1.06 (0.46)	1.12 (0.48)	1.01 (0.41)	.008
WBC-Mean, *10 ⁹ /L	9.6 (2.7)	10.2 (2.9)	8.9 (2.3)	<.001
NEU-Mean, *10 ⁹ /L	7.8 (2.1)	8.2 (2.0)	7.3 (2.1)	<.001
LYM-Mean, *10 ⁹ /L	1.0 (0.4)	1.17 (0.45)	0.90 (0.39)	<.001
RBC-Mean *10 ¹² /L	3.3 (0.5)	3.4 (0.5)	3.2 (0.5)	<.001
HGB-Mean, g/L	98.6 (13.7)	101.3 (14.3)	95.6 (12.5)	<.001
PLT-Mean, *10 ⁹ /L	164.7 (51.2)	171.0 (51.1)	157.9 (50.4)	.003
WBC-CV, %	34.9 (15.8)	33.9 (19.6)	36.0 (10.5)	.001
NEU-CV, %	45.4 (12.2)	44.4 (11.9)	46.5 (12.4)	.064
LYM-CV, %	54.4 (12.5)	55.5 (12.6)	53.2 (12.3)	.033
RBC-CV, %	10.4 (4.1)	10.1 (3.9)	10.8 (4.2)	.042
HGB-CV, %	10.6 (4.0)	10.3 (3.9)	10.9 (4.2)	.111
PLT-CV, %	28.3 (13.5)	25.9 (11.5)	30.8 (14.8)	<.001
WBC < 4, %	4.8 (10.7)	2.3 (7.1)	7.5 (13.1)	<.001
4 < WBC < 10, %	68.4 (22.3)	66.3 (23.4)	70.7 (20.9)	.032
WBC > 10, %	26.7 (22.8)	31.3 (24.1)	21.7 (20.2)	<.001
NEU < 1.8, %	1.0 (4.7)	0.6 (3.5)	1.5 (5.7)	.015
1.8 < NEU < 6.3, %	58.0 (24.3)	54.2 (24.8)	62.0 (23.0)	<.001
NEU > 6.3, %	40.9 (24.6)	45.1 (25.1)	36.4 (23.4)	<.001
LYM < 1.1, %	47.3 (30.7)	36.9 (26.1)	58.4 (31.5)	<.001
1.1 < LYM < 3.2, %	49.9 (29.7)	58.8 (25.9)	40.4 (30.6)	<.001
LYM > 3.2, %	2.6 (10.0)	4.2 (12.6)	1.0 (5.8)	<.001
RBC < 3.8, %	73.9 (29.8)	67.9 (31.6)	80.3 (26.5)	<.001
3.8 < RBC < 5.1, %	25.2 (28.8)	30.9 (30.4)	19.1 (25.6)	<.001

(continued)

Table 3. Continued.

Variable name	Total sample number (n = 546)	Normal group (n = 281)	Abnormal group (n = 265)	P-Value
RBC > 5.1, %	0.8 (6.6)	1.1 (7.8)	0.5 (5.0)	.929
HGB < 115, %	78.9 (27.1)	73.9 (29.8)	84.3 (22.6)	<.001
115 < HGB < 150, %	20.6 (26.3)	25.3 (28.6)	15.5 (22.6)	<.001
HGB > 150, %	0.4 (3.0)	0.7 (4.0)	0.1 (0.9)	.151
PLT < 125, %	24.1 (26.3)	20.8 (24.9)	27.6 (27.3)	.002
125 < PLT < 350, %	73.2 (26.4)	76.4 (25.3)	69.8 (27.2)	.003
PLT > 350, %	2.6 (9.9)	2.8 (10.4)	2.5 (9.4)	.930
CR, $\mu\text{mol/L}$	131.2 (115.1)	117.2 (83.7)	146.0 (139.6)	.093
eGFR, mL/min	49.0 (26.8)	51.0 (26.5)	46.8 (26.8)	.074
UREA, mmol/L	9.6 (6.7)	8.8 (4.9)	10.4 (8.0)	.074
UA, $\mu\text{mol/L}$	277.0 (152.6)	281.6 (149.0)	272.1 (156.4)	.321
UA < 142, %	3.5 (10.6)	3.5 (11.6)	3.4 (9.4)	.242
142 < UA < 340, %	47.6 (25.8)	47.2 (26.4)	48.1 (25.3)	.771
UA > 340, %	48.8 (27.0)	49.1 (27.4)	48.4 (26.7)	.818
AST, U/L	21.7 (16.9)	22.0 (19.9)	21.4 (13.0)	.240
ALT, U/L	28.7 (27.6)	31.6 (32.4)	25.7 (20.8)	.196
GLU < 4, %	3.4 (8.6)	4.3 (10.5)	2.4 (5.8)	.204
4 < GLU < 6, %	60.6 (25.4)	62.5 (23.8)	58.6 (26.9)	.216
GLU > 6, %	36.0 (25.7)	33.0 (23.8)	38.9 (27.3)	.033
TAC-Mean, ng/mL	8.9 (1.8)	8.9 (1.7)	8.9 (1.9)	.671
TAC-CV, %	34.4 (13.6)	33.6 (12.9)	35.2 (14.3)	.090
TAC < 8, %	43.5 (23.0)	43.3 (22.7)	43.7 (23.3)	.838
8 < TAC < 10, %	23.5 (20.0)	24.6 (19.8)	22.3 (20.2)	.112
TAC > 10, %	33.0 (22.9)	32.1 (21.5)	33.9 (24.3)	.740
Rabbit anti-human thymocyte immunoglobulin				.116
Yes	156 (28.6)	72 (25.6)	84 (31.7)	
No	390 (71.4)	209 (74.4)	181 (68.3)	

(continued)

Table 3. Continued.

Variable name	Total sample number (n = 546)	Normal group (n = 281)	Abnormal group (n = 265)	P-Value
Anti-human T cell porcine immunoglobulin				.156
Yes	69 (12.6)	30 (10.7)	39 (14.7)	
No	477 (87.4)	251 (89.3)	226 (85.3)	
Basiliximab injection				.192
Yes	361 (66.1)	193 (68.7)	168 (63.4)	
No	185 (33.9)	88 (31.3)	97 (36.6)	
Mycophenolate mofetil				.908
Yes	403 (73.8)	208 (74.0)	195 (73.6)	
No	143 (26.2)	73 (26.0)	70 (26.4)	
Mycophenolate sodium enteric-coated tablets				.507
Yes	166 (30.4)	89 (31.7)	77 (29.1)	
No	380 (69.6)	192 (68.3)	188 (70.9)	
TDS, Pills/day	5.4 (1.4)	5.4 (1.4)	5.4 (1.4)	.812

The top 20 variables ranked by the absolute value of the coefficient are shown in Figure 5.

As shown in Figure 5a, variables such as the proportion of lymphocyte duration less than 1.1 in the total time, blood type AB of recipient, gender, coefficient of variation of platelet, height, mean value of white blood cells at 30 days, and proportion of neutrophil duration greater than 6.3 in the total time are the most significant variables that affect the decline in postoperative white blood cell count.

Machine learning model selection

On the whole data set, the top 20 sorted variables were used as the final characteristic variables, and the normal group label to was set to 0 and the abnormal group label to was set 1. With the average AUC as the criterion, the model with the best performance was selected as the final prediction model by using the method of five-fold cross-validation. The candidate machine learning models included 8 types: L1 regularized logistic regression (Logic-L1), L2 regularized logistic regression (Logic-L2), SVM, DT, RF, MLP, XGBoost tree, and distributed gradient lift tree (LightGBM). The layout and training of the model were conducted in the sklearn machine learning library supported by Python 3.8.

The average evaluation indexes of the five-fold cross-validation of each machine learning model in all data sets are shown in Table 4. The average ROC curves of eight machine learning models are plotted in Figure 6. Table 4 and Figure 6 show that MLP model has the best comprehensive effect on this task, and it is selected as the final prediction model.

Model building and optimization

A total of 546 patients were randomly divided into the training and test sets according to a ratio of 7:3. A total of 382 patients were included in the training set: 198 normal samples and 184 abnormal samples. A total of 164 patients were included in the test set: 83 normal samples and 81 abnormal samples.

The MLP model is a neural network model with at least one hidden layer of neurons between the input and output layers, and it adopts full connection between layers. The error backpropagation (BP) algorithm is used for training. The training process adjusts the connection weight between neurons and the threshold value of each neuron according to the data of the training set. The structural schematic diagram of the MLP model is shown in Figure 7.

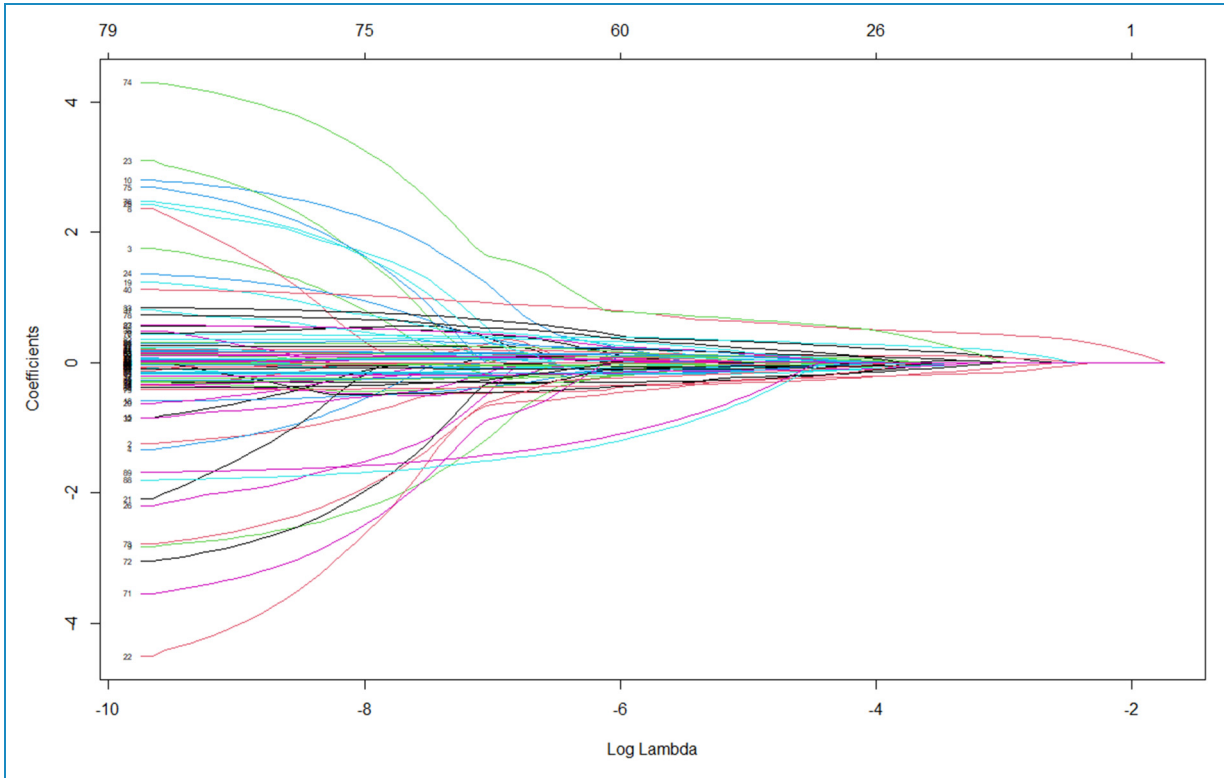


Figure 3. Variation chart of coefficients of variables in LASSO regression.

On the training set, the super parameters of the model are determined by means of half-fold cross-validation, and the value of AUC is selected as the evaluation index. If grid search is conducted for all parameters simultaneously, then the training time will increase exponentially. Therefore, this study adopts heuristic grid search to adjust parameters individually. The loss function adopts the cross entropy loss function, as shown in Formula (5):

$$-\log(y_t \log(y_p) + (1 - y_t) \log(1 - y_p)) \quad (5)$$

where y_t represents the real label of a certain sample, and its value can only be set in this dichotomy task. It represents the probability that the sample takes 1. For the whole model, its loss function is the average value of the loss function of all samples.

Activation function of neurons adopts Sigmoid function, as shown in Formula (6):

$$f(x) = \frac{1}{1 + e^{-x}} \quad (6)$$

First, the number of neurons in the first hidden layer is determined. According to experience, the number of hidden neurons should be between the sizes of the input and output layers. Too many neurons may cause overfitting. In this study, the input was 20-dimensional and the output was dichotomous. Accordingly, the

number of hidden layer neurons was assumed to be between 3 and 22. The average AUC of cross verification corresponding to the number of different neurons in the first hidden layer is shown in Figure 8. The number of neurons in the first hidden layer is finally determined to be 18, and the optimal average AUC is 77.41%.

Then, the number of neurons in the second hidden layer is determined. The same method is adopted. The corresponding model results of different numbers of neurons in the second hidden layer are shown in Figure 9. The figure shows that the maximum average AUC is 77.23%, but the effect of the model will decline after adding another hidden layer. Thus, the model is finally determined to be a hidden layer with 18 neurons. Adam algorithm is used as the optimization algorithm. Adam algorithm is based on stochastic gradient descent. However, it is different from stochastic gradient descent in that the learning rate can be adjusted adaptively in the training process. Thus, it can achieve better training effect. The initial learning rate needs to be determined when modeling. In this study, the list of candidate learning rates is set as {0.0001, 0.0005, 0.001, 0.005, 0.01, 0.1}. The number of neurons and layers that have been determined are taken to get the corresponding model effects of different learning rates, as shown in Figure 10. The optimal initial learning rate is still the

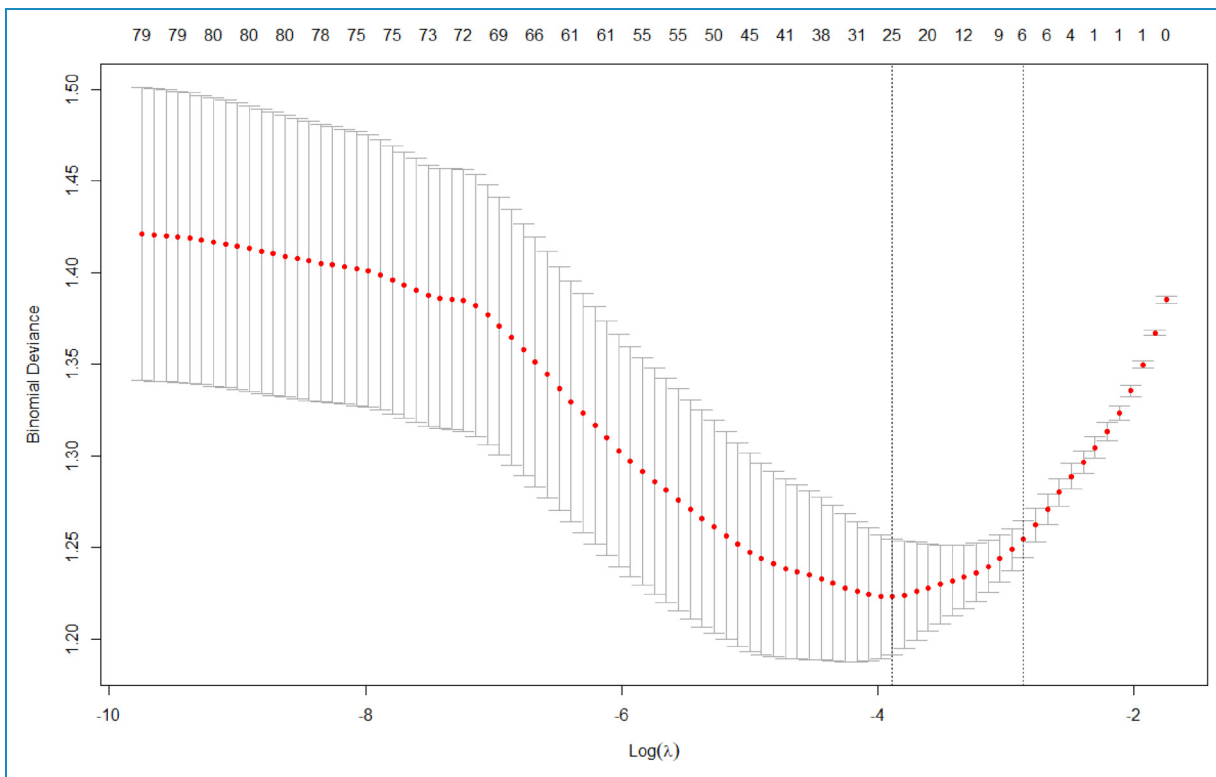


Figure 4. Changes in the number of retained variables and classification errors.

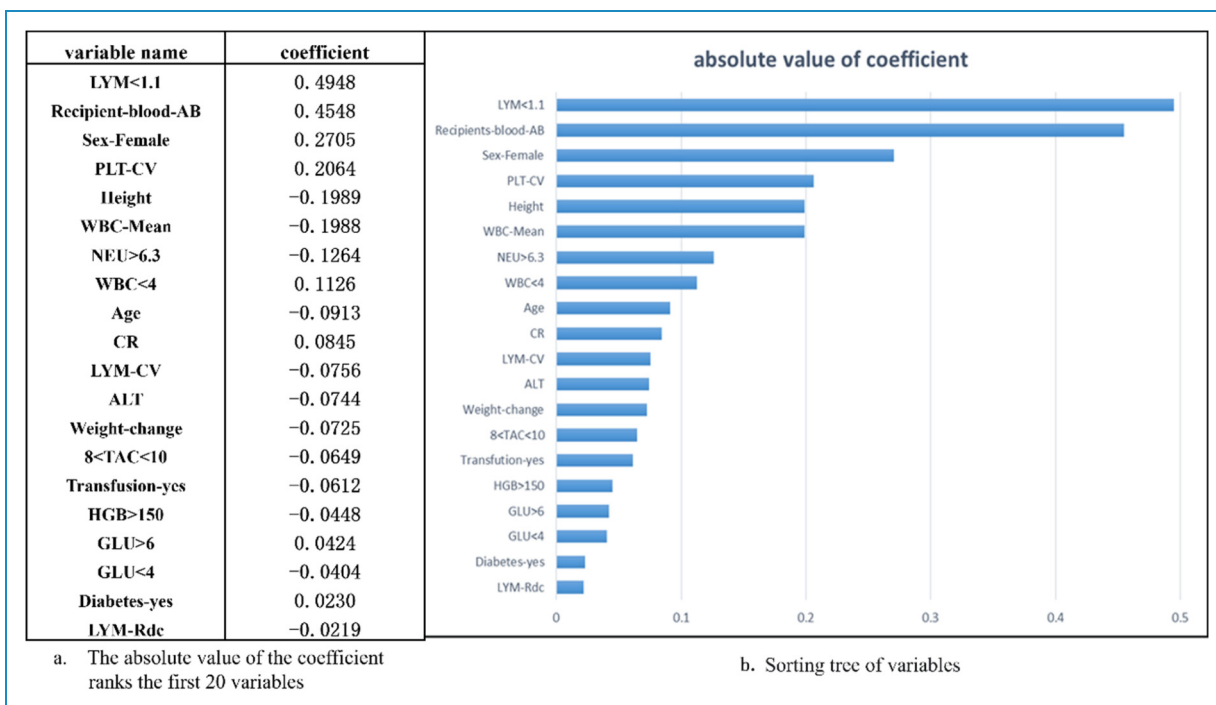


Figure 5. Ranking of the top 20 variables.

Table 4. Comparison table of average evaluation indexes for cross-validation of each model.

Model name	Accuracy %	Precision %	Recall rate %	F1 value %	AUC value %
Logistic-L1	69.04	69.56	65.28	66.97	75.93
Logistic-L2	69.22	70.17	64.52	66.81	76.16
SVM	69.41	71.43	62.26	66.27	75.64
DT	60.62	59.07	66.03	61.58	63.89
RF	66.30	66.84	67.16	64.13	70.78
MLP	70.14	71.34	64.90	67.60	76.82
XGBoost	61.90	60.83	59.62	59.98	66.54
LightGBM	62.64	62.54	58.11	59.99	67.61

default learning rate of 0.001, and the average AUC value for cross-validation remains at 77.41%.

The key parameters of the MLP model established in this study were set as follows: the number of hidden layers: 1; the number of hidden layer neurons: 18; neuron activation function: sigmoid function; optimization algorithm: Adam algorithm; initial learning rate: 0.001; all other parameters are default.

Results

After the steps of sample establishment, feature selection, model selection, model establishment and optimization, the final model was trained, and five-fold cross validation was used to avoid over-fitting of the model. After training, the average evaluation index on the training set and on the test set are shown in Table 5, and the ROC curve is shown in Figure 11.

In order to visualize the prediction effect of the model more intuitively, Figure 12 shows the confusion matrix of the average prediction effect of the model on the test set (the number of cases of each type is taken as an integer). From this figure, it can be seen that the prediction model has a clear tendency to predict correctly on both types of labels.

To verify the robustness of the results of this study, the results of this study were compared with those of other published studies. AUC value is often used as one of the most important evaluation indicators of binary classification model, which is widely used because it can distinguish the false high accuracy of the model due to randomness. Therefore, this study uses AUC value as a comparison

standard. In recent years, the prediction model for abnormal white blood cell count after kidney transplantation has not been developed by other scholars. This study uses the research in the field of kidney transplantation that uses real world data to predict other targets as a comparison object. In 2021, Xue developed and verified the donor risk score system using the medical records of 1807 donors and 3599 recipients of kidney transplants from 29 medical centers in China, and the AUC value reached 0.7552.²⁰ In 2023, Quinino used seven commonly used machine learning methods to build a machine learning prediction model for the immediate transplant function after deceased donor kidney transplantation. Among them, the XGBoost classification prediction model obtained the best prediction effect with an AUC value of 0.78.²¹ The AUC value of the final prediction result of this study reaches 0.7730, which is better than the final prediction result of Xue's study and roughly equal to the final prediction result of Quinino's study, which can verify the usability of the final results of this study. In addition, this study used five-fold cross validation in the training of the model, which effectively avoided overfitting. This allows the robustness of the results of this study to be further verified.

Discussion

Infection is a major cause of postoperative morbidity and mortality among kidney transplant recipients.²² Based on Medicare claims data collected by the American Kidney Data System, the initial infection rate is estimated to be 45% in the first 3 years after kidney transplantation.²³ The most common type of infection is lung infection, followed by urinary system infection, upper respiratory tract infection, and surgical incision infection. Infection can be prevented to some extent. The way to effectively reduce infection after kidney transplantation is an important prognostic problem of kidney transplantation. According to past experience, whether the infection will occur after the operation is closely related to the patient's own basic physical condition, the dosage of immunosuppressive drugs, and other factors.²⁴ In this study, a total of 546 patients with kidney transplantation were finally included, and the diagnostic reports in their electronic medical records were obtained. The two-dimensional correlation table of whether the white blood cell count was less than $4 \times 10^9/L$ 30 days after kidney transplantation and whether infection occurred 30 days after the electronic diagnosis report was drawn, as shown in Table 6.

Chi-squared test was used to determine the correlation between the two factors, and the final Chi-squared value was 6.199, and the P value was $.013 < .05$. In conclusion, a significant correlation exists between the white blood cell count of patients 30 days after the operation below

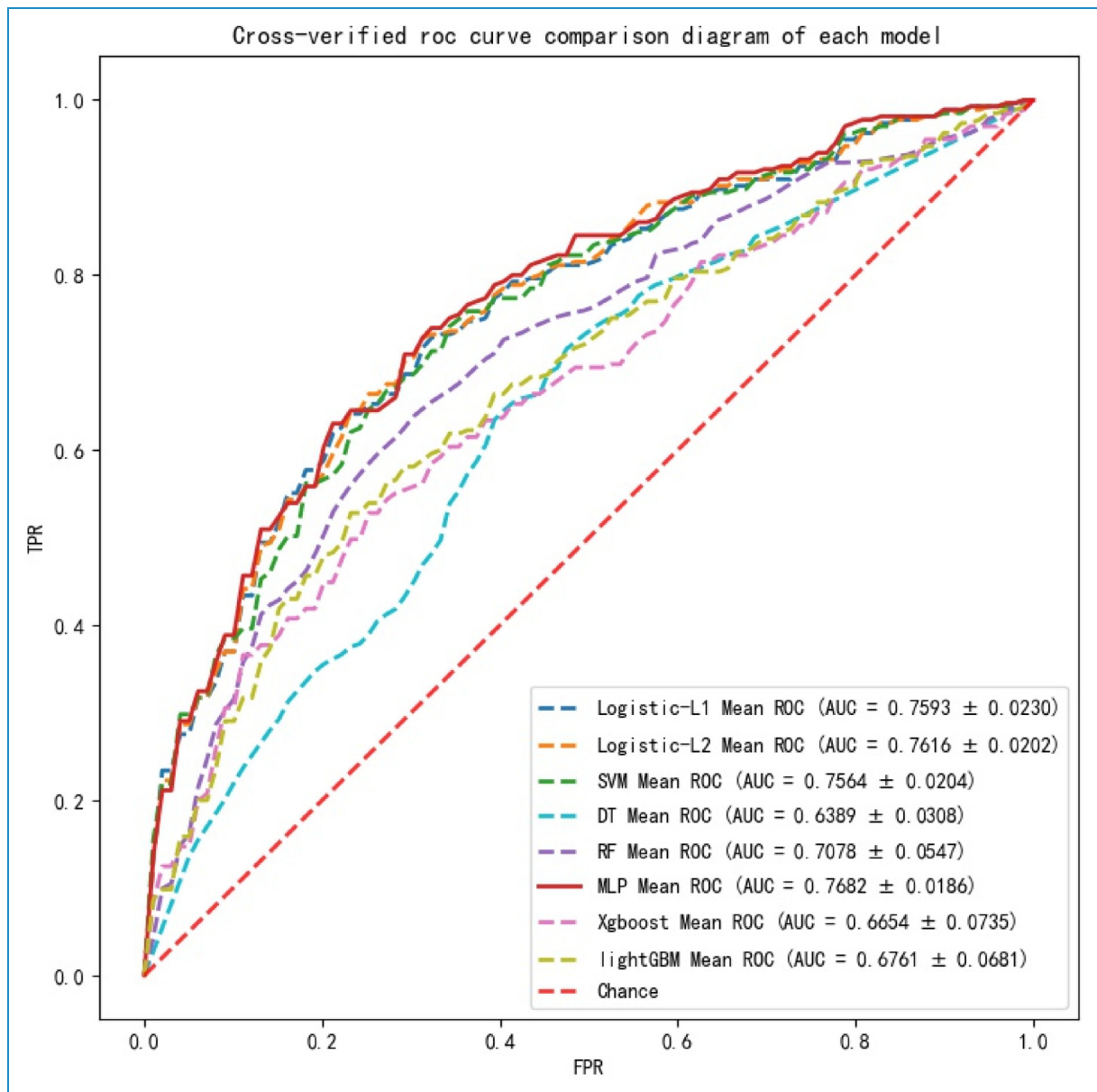


Figure 6. Comparison of average ROC curves of all models.

the lower normal limit and the occurrence of infection. Furthermore, patients with lower white blood cell than the lower normal limit are more likely to develop infection. If the perioperative examination index can be used to predict the long-term postoperative white blood cell level, then it can also effectively predict the possibility of infection.

In this study, LASSO algorithm was used to determine the variables that had the greatest effect on lower than normal count of white blood cells. The variables were as follows: the proportion of lymphocyte less than 1.1 within 30 days, blood type AB of recipient, gender, platelet CV, height, the mean value of white blood cells within 30 days, the proportion of neutrophils greater than 3.3, the

proportion of white blood cells less than 4, age, and creatinine.

1. The variable characteristics of lymphocytes, neutrophils, white blood cells directly reflect the quality of the patient's own blood function, and it is highly correlated with the level of white blood cells 30 days after surgery. Among them, the proportion of lymphocytes less than 1.1 is significantly higher in the abnormal group, with an average ratio of 58.4%. In this study, lymphocytes are the first important factor to predict the lower than normal level of white blood cells after kidney transplantation.

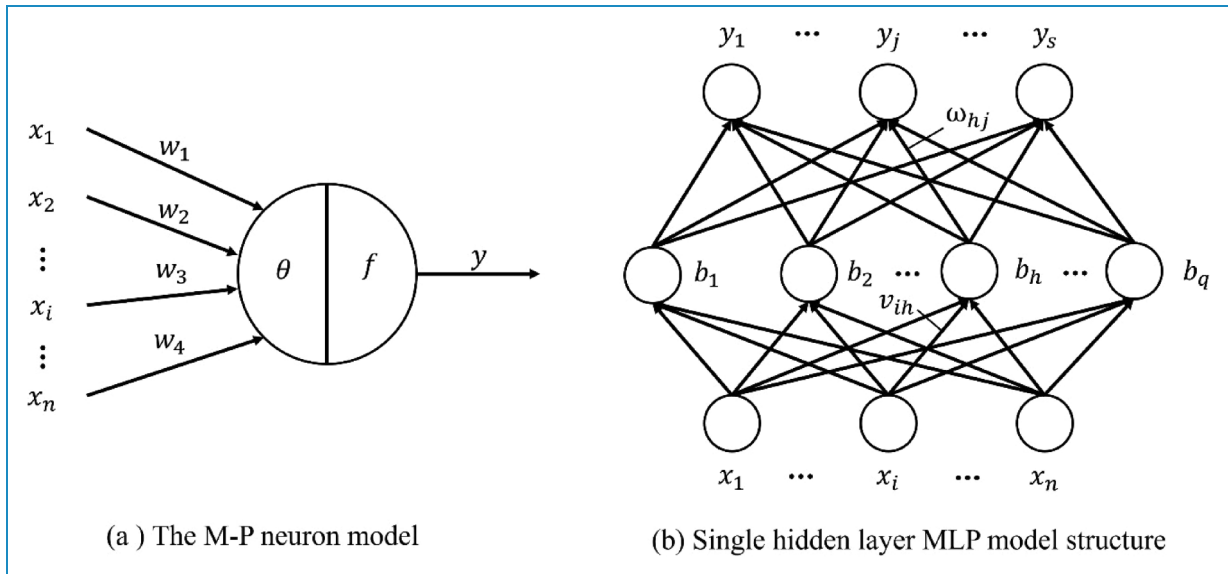


Figure 7. The structural schematic diagram of the MLP model.

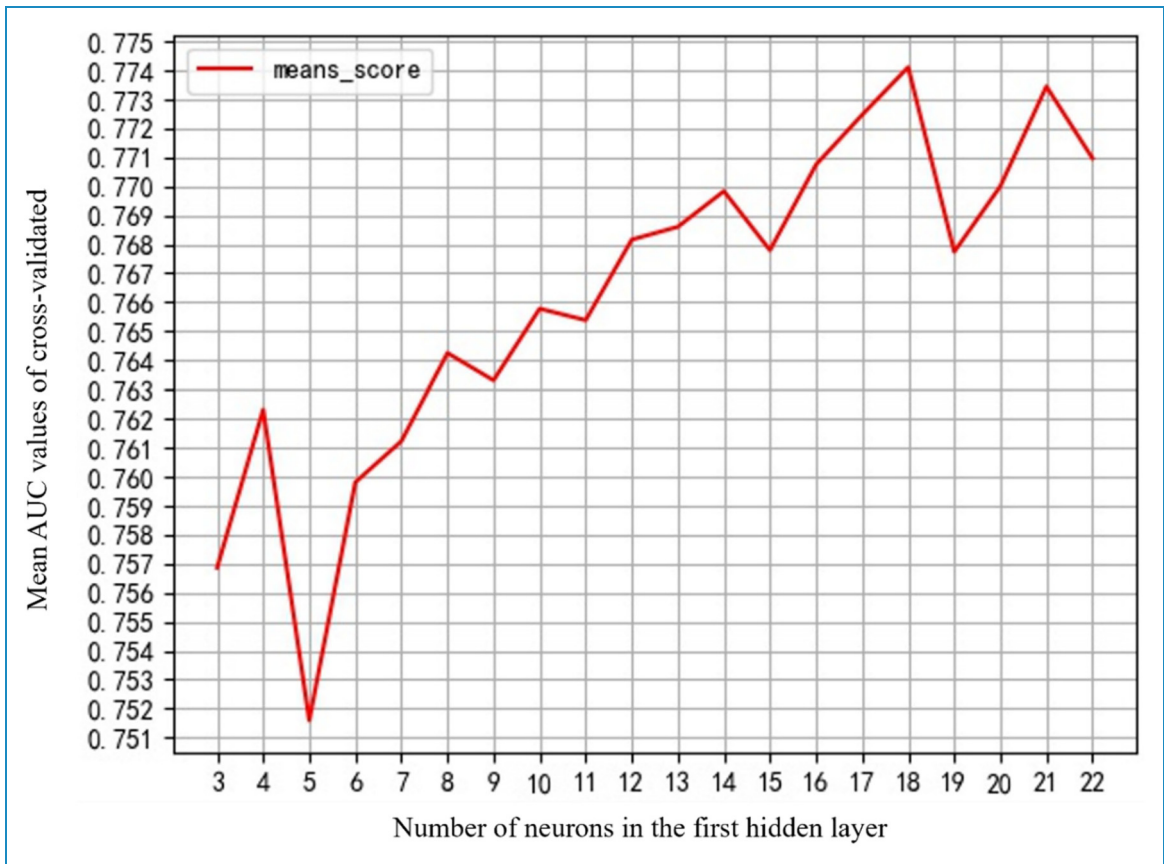


Figure 8. Model effect of different numbers of neurons in the first hidden layer.

2. Recipients with blood group AB are more likely to have decreased white blood cell count. The proportion of the recipients with blood group AB is small in the total sample

group, and they can accept all four blood types of kidney donors. This result is rarely mentioned in previous studies, and it is worthy of further study and discussion.

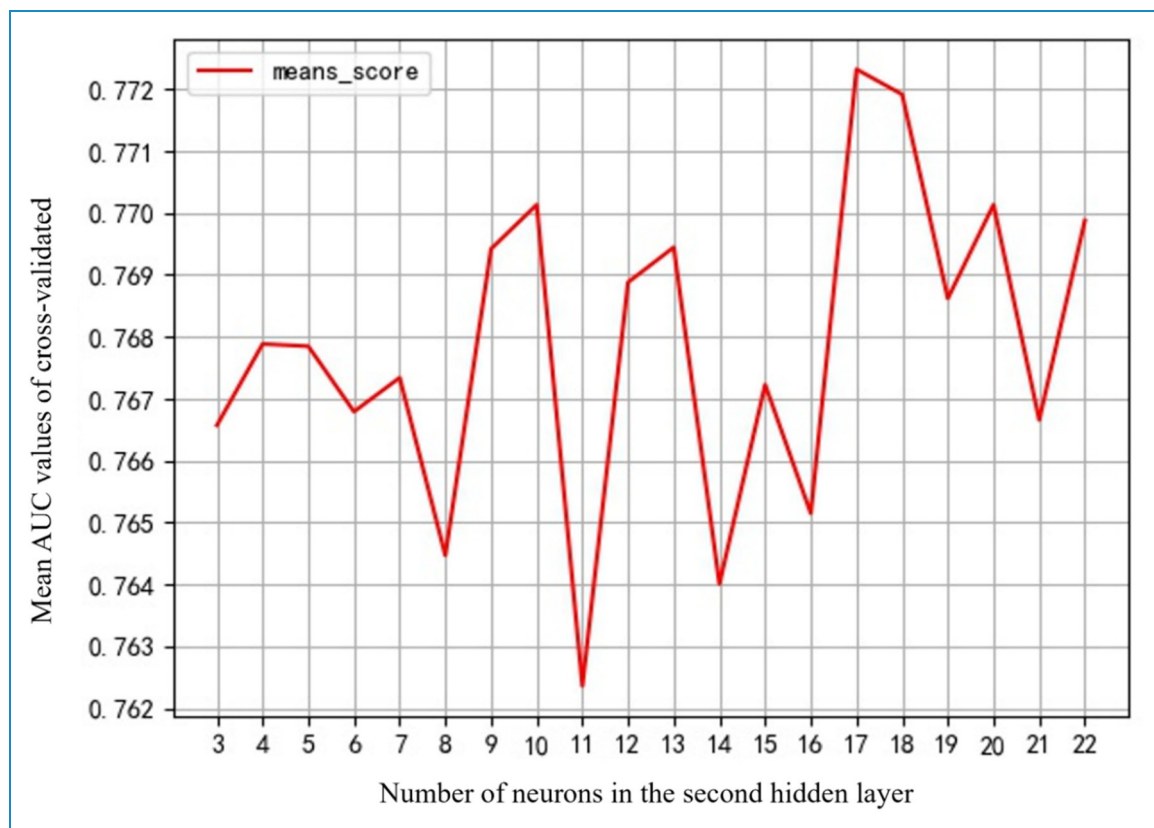


Figure 9. Model effects of different numbers of neurons in the second hidden layer.

- Female patients are more likely to suffer from decreased white blood cell count after kidney transplantation. Previous studies mentioned that females are more likely to suffer from delayed graft recovery after surgery.^{25,26} In the total sample group of the current study, the proportion of female kidney transplant recipients is significantly less than that of male patients. However, the proportion of female white blood cells decreasing to normal levels is greater. The average height and weight of men are significantly higher than those of women. Thus, the larger proportion of women in the abnormal group accounts for the difference in height between the normal and abnormal groups. Thus, height is a higher ranking variable as well. One possible explanation is that no obvious difference exists in drug selection and dosage of postoperative immunosuppressive agents between male and female at present. Under the same drug dosage, female is more strongly inhibited due to the smaller size, and the decrease in white blood cells is more obvious. In future clinical treatment, different drug dosages can be arranged for male and female.
- Creatinine, the ratio of duration of tacrolimus within the effective concentration to the total time, whether a

history of blood transfusion exists, and other variables are also among the top of all candidate variables. Blood glucose content and whether diabetes exists also have a certain effect on the drop of postoperative white blood cell to the normal level. Therefore, considerable attention should be paid to the changes in patients' creatinine and blood glucose concentration in subsequent clinical treatment.

In this study, a total of 546 cases of data were included after screening. For the sample size of input data for machine learning models, it is generally considered that the data size of several hundred to several thousand is acceptable, but generally not less than 300. But sometimes there are examples of using small sample sizes to build machine learning models, For example, in 2020, Kyle R. Jackson used the data of 140 marginal deceased donor kidneys and transplant recipients to compare the prediction accuracy of the two donor kidney scoring systems.²⁷ There are also examples of using a large sample size to construct a machine learning predictive model. For example, Yunwei Zhang used the data of 3624 patients in 2023 to develop a prediction model for the graft survival rate of kidney transplantation.²⁸ In 2018, Richard D Riley mentioned a calculation method to meet the minimum sample size

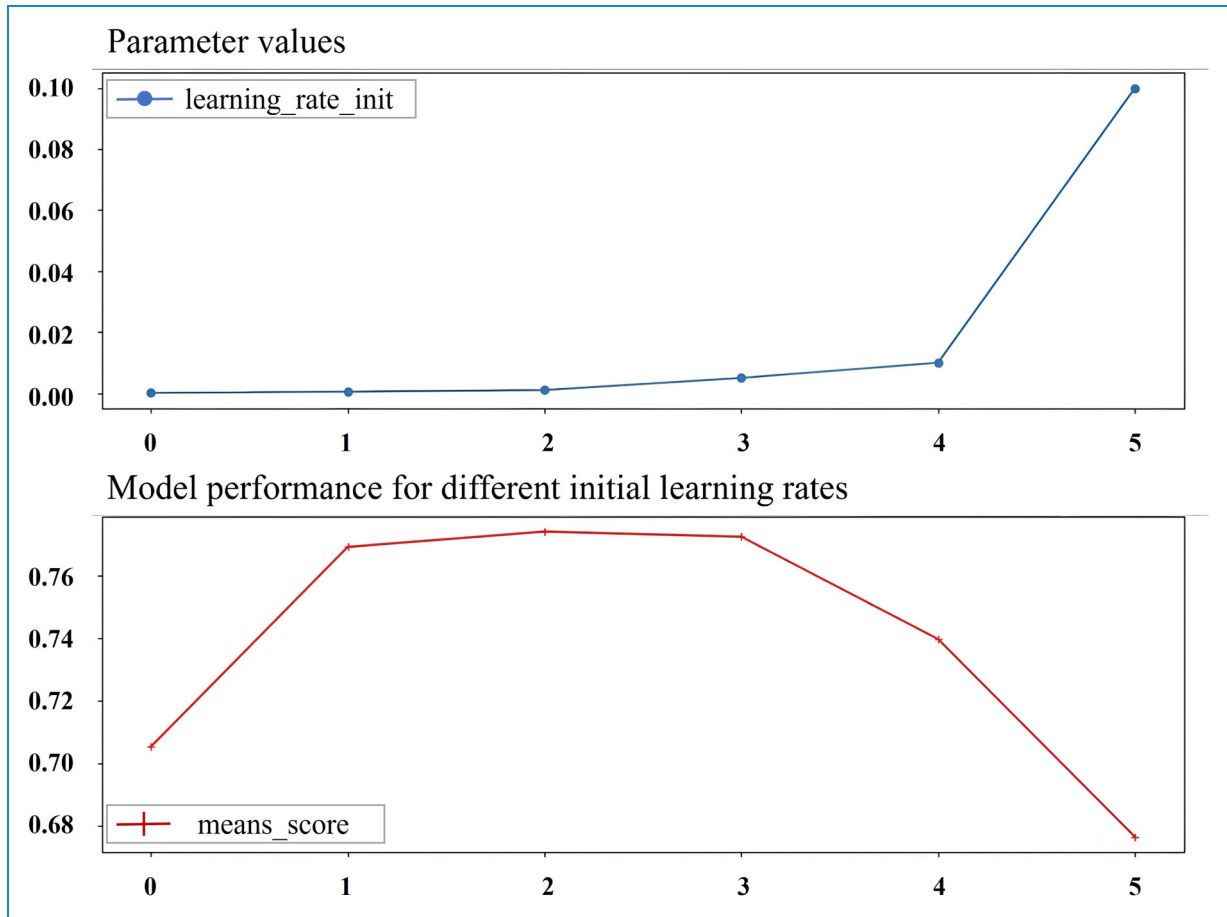


Figure 10. Model effects corresponding to different initial learning rates.

Table 5. Shows the effect of the model on the training and test sets.

Data set	Accuracy %	Sensitivity %	Specificity %	AUC value %
Training set (Five-fold)	71.67	64.44	83.56	77.41
Test set	71.34	61.18	82.28	77.30

requirements in his study.²⁹ In this study, the total follow-up number was 546 people, the follow-up time was 1 year, the abnormal patients were 265, the abnormal rate was 46.89%, and the number of features included in the modeling was 20. The calculated EPP of this study is $(546 \times 1 \times 46.89\%) / 20 = 12.8$, which conforms to the rule of thumb that EPP is at least 10. Therefore, the sample size of this study is sufficient.

In the stage of sample establishment, considering the integrity of the data and the prediction effect of the

model, the processing of missing values in this study will delete the data of cases with missing values in the candidate features. The reason for this is that the data used in this study is sufficient and complete. Even if the cases with missing values are directly deleted, the number of cases finally included in the model will still be 546. It can fully support the operation of the prediction model. However, for the data set with relatively small amount of data and poor data integrity, it is necessary to develop a machine learning missing value filling method with better effect than the traditional missing value filling method to ensure sufficient data amount.

Some reports on kidney transplantation and machine learning have pointed out the superior performance of RF, SVM, and XGBoost in this respect, and this study found that after sufficient feature engineering processing, including the feature extraction method based on duration proportion proposed in this study and LASSO regression feature screening, compared with the above several machine learning methods, MLP has a better prediction effect. The following is a discussion on this phenomenon from the perspective of algorithm characteristics: Compared with

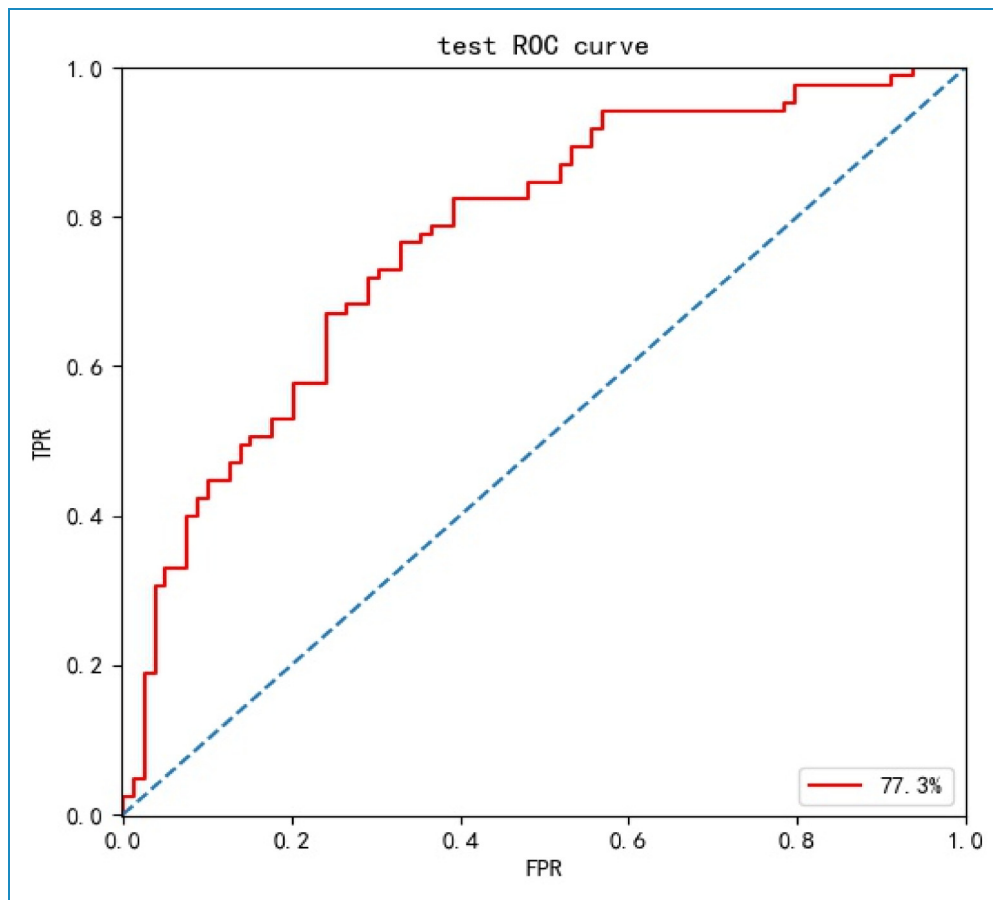


Figure 11. ROC curve of the model on the test set.

the above machine learning models, MLP has major disadvantages such as high overfitting risk, large training complexity and weak interpretability. However, this study did not directly input all features into the model when building the model, but pre-used LASSO regression to sort and screen the importance of features. The number of features decreased from 85 to 20, which removed redundant information for the MLP model in advance, reduced the overfitting risk and training complexity of the model, and at the same time, the obtained feature ranking of the importance of the outcome clearly gave the 20 features with the highest importance and their importance scores, which improved the interpretability of the entire model. At the same time, the information obtained by the feature extraction method based on duration ratio proposed in this paper will bring more additional information to the model after being processed by MLP with more complex structure than other methods. Several major shortcomings of the MLP model are remedied, while its advantages such as strong feature learning ability and strong generalization ability for large-scale data sets and high-dimensional inputs compared with other machine learning models are maintained, making it obtain better prediction results.

It should be limited that all cases in this study were first-time kidney transplantation, and the causes of kidney failure and induction immunosuppressive agent were not recorded, which resulted in the fact that the causes of kidney failure, induction immunosuppressive agent, previous kidney replacement methods, and previous transplant history were not included in the study. The influence of these features on leukocyte abnormalities in patients after kidney transplantation is expected to be further studied.

Conclusion

The decrease in leukocyte level below the normal level is an adverse outcome after kidney transplantation. Thus, the feature that the duration of different variables in different intervals accounted for the proportion of the total observation time was introduced in this study. It could provide more predictive information than only extracting the mean and extreme values. Through statistical analysis and LASSO algorithm for feature selection, the time proportion of lymphocyte less than 1.1 within 30 days, blood group AB of recipient, gender, platelet CV, height, and other variables were found to have the greatest effect on the postoperative

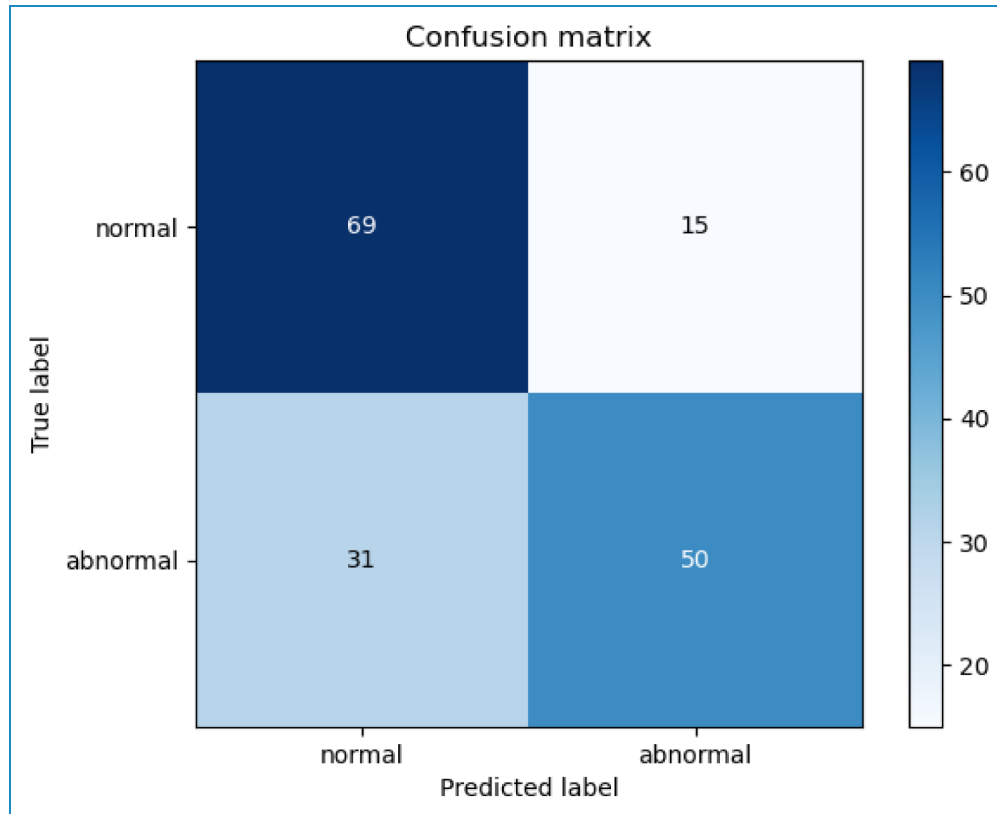


Figure 12. Test set confusion matrix of the classification prediction model.

Table 6. Contingency table of leukocyte level and infection occurrence.

		Electronic diagnosis reports if an infection has developed after 30 days		
		No	Yes	Total
Whether the white blood cell count is less than $4 \times 10^9/L$ 30 days after surgery	No	237	44	281
	Yes	201	64	265
Total		438	108	546

level of white blood cells. The results can be used as a realistic basis to remind doctors to study the internal relationship between these features and abnormal white blood cell count after kidney transplantation. Eight machine learning algorithms were used in all data sets for comparison, and the MLP model was determined to have the best effect. After the parameters of the divided training set were adjusted, a good predictive effect was achieved in the test set. Specifically, the AUC value of the model reaches 77.30%,

which has a certain auxiliary effect on clinical treatment, can help doctors predict the risk of surgery, and pay more attention to patients with high risk of surgery in advance.

Among the features obtained in this study with the high ranking correlation with abnormal white blood cell count after kidney transplantation, the internal reasons and medical principles of some features related to the target need to be more comprehensive and specific experimental and modeling studies, such as recipient AB blood type, patient gender, and height.

Contributorship: All authors substantially contributed to preparation of the manuscript and gave final approval for publication.

Data availability: The data underlying this article cannot be shared publicly due to the data used in this paper concerns the personal privacy of hospital patients. The data will be shared on reasonable request to the corresponding author.


Declaration of conflicting interests: The authors declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Ethics approval and patient consent: This study was approved by the Medical Ethics Committee of Tongji Hospital Affiliated to

Tongji Medical College of Huazhong University of Science and Technology (REC number: EC-20211103-1002) with the approval document number TJ-IRB20231274, and the approval document also contains the approval of the request to exempt the subject from informed consent.

Funding: The authors disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: This work was supported by the National Natural Science Foundation of China (grant number No. 72104085).

Guarantor: XL.

ORCID ID: Xia Lu  <https://orcid.org/0009-0001-0263-9030>

References

- Manook M, Koeser L, Ahmed Z, et al. Post-listing survival for highly sensitised patients on the UK kidney transplant waiting list: a matched cohort analysis. *Lancet* 2017; 389: 727–734.
- Weinrauch LA and D’Elia JA. Solid organ transplantation. *JACC: Heart Failure* 2018; 6: 348–349.
- Kumar R and Ison MG. Opportunistic infections in transplant patients. *Infect Dis Clin* 2019; 33: 1143–1157.
- Kinnunen S, Karhapää P, Juutilainen A, et al. Secular trends in infection-related mortality after kidney transplantation. *Clin J Am Soc Nephrol* 2018; 13: 755–762.
- Honda T, Uehara T, Matsumoto G, et al. Neutrophil left shift and white blood cell count as markers of bacterial infection. *Clin Chim Acta* 2016; 457: 46–53.
- Ravindhran B, Chandak P, Schafer N, et al. Machine learning models in predicting graft survival in kidney transplantation: meta-analysis. *BJS Open* 2023 Mar 7; 7: zrad011.
- Senanayake S, White N, Graves N, et al. Machine learning in predicting graft failure following kidney transplantation: a systematic review of published predictive models. *Int J Med Inform* 2019 Oct; 130: 103957.
- Brier ME, Ray PC and Klein JB. Prediction of delayed renal allograft function using an artificial neural network. *Nephrol Dial Transplant* 2003; 18: 2655–2659.
- Decruyenaere A, Decruyenaere P, Peeters P, et al. Prediction of delayed graft function after kidney transplantation: comparison between logistic regression and machine learning methods. *BMC Med Inform Decis Mak* 2015; 15: 1–10.
- Yoo KD, Noh J, Lee H, et al. A machine learning approach using survival statistics to predict graft survival in kidney transplant recipients: a multicenter cohort study. *Sci Rep* 2017; 7: 1–12.
- Senanayake S, White N, Graves N, et al. Machine learning in predicting graft failure following kidney transplantation: a systematic review of published predictive models. *Int J Med Inf* 2019; 130: 103957.
- Peng B, Gong H, Tian H, et al. The study of the association between immune monitoring and pneumonia in kidney transplant recipients through machine learning models. *J Transl Med* 2020; 18: 1–11.
- Luo Y, Tang Z, Hu X, et al. Machine learning for the prediction of severe pneumonia during posttransplant hospitalization in recipients of a deceased-donor kidney transplant. *Ann Transl Med* 2020; 8: 82.
- Bae S, Massie AB, Caffo BS, et al. Machine learning to predict transplant outcomes: helpful or hype? A national cohort study. *Transpl Int* 2020; 33: 1472–1480.
- Badrouchi S, Ahmed A, Bacha MM, et al. A machine learning framework for predicting long-term graft survival after kidney transplantation. *Expert Syst Appl* 2021; 182: 115235.
- Topuz K, Zengul FD, Dag A, et al. Predicting graft survival among kidney transplant recipients: a Bayesian decision support model. *Decis Support Syst* 2018; 106: 97–109.
- Makady A, de Boer A, Hillege H, et al. What is real-world data? A review of definitions based on literature and stakeholder interviews. *Value Health* 2017; 20: 858–865.
- Li Y, Yan L, Li Y, et al. Development and validation of routine clinical laboratory data derived marker-based nomograms for the prediction of 5-year graft survival in kidney transplant recipients. *Ageing (Albany NY)*. 2021; 13: 9927.
- Chen RY, Zhang S, Zhuang SY, et al. A simple nomogram for predicting infectious diseases in adult kidney transplantation recipients. *Front Public Health* 2022; 10: 2989.
- Wujun X, Changxi W, Jianghua C, et al. A prediction model of delayed graft function in deceased donor for renal transplant: a multi-center study from China. *Renal Fail* 2021; 43: 520–529.
- Quinino RM, Fabiana A, de Andrade LGM, et al. A machine learning prediction model for immediate graft function after deceased donor kidney transplantation. *Transplantation* 2023; 107: 1380–1389.
- Karuthu S and Blumberg EA. Common infections in kidney transplant recipients. *Clin J Am Soc Nephrol* 2012; 7: 2058–2070.
- Snyder JJ, Israni AK, Peng Y, et al. Rates of first infection following kidney transplant in the United States. *Kidney Int* 2009; 75: 317–326. <http://dx.doi.org/10.1038/ki.2008.580>
- Alangaden GJ, Thyagarajan R, Gruber SA, et al. Infectious complications after kidney transplantation: current epidemiology and associated risk factors. *Clin Transplant* 2006; 20: 401–409.
- Augusto JF, Garnier AS, Demiselle J, et al. Hypogammaglobulinemia and risk of severe infection in kidney transplant recipients. *Transpl Infect Dis* 2016; 18: 741–751.
- Florescu DF, Kalil AC, Qiu F, et al. What is the impact of hypogammaglobulinemia on the rate of infections and survival in solid organ transplantation? A meta-analysis. *Am J Transplant* 2013; 13: 2601–2610.
- Jackson KR, Munivenkatappa RB, Wesson RN, et al. What’s the score? A comparison of deceased donor kidney scoring systems and correlation with graft outcome. *Clin Transplant* 2020; 34: 13802.
- Zhang Y, Deng D, Muller S, et al. A multi-step precision pathway for predicting allograft survival in heterogeneous cohorts of kidney transplant recipients. *Transpl Int* 2023; 36: 11338.
- Riley RD, Snell KI, Ensor J, et al. Minimum sample size for developing a multivariable prediction model: PART II - binary and time-to-event outcomes. *Stat Med* 2019; 38: 1276–1296.