

Interpretable SVM Model for Predicting CMV Infection in Seropositive Kidney Transplant Recipients: A Single-Center Retrospective Study

Guangli Zhong¹, Yujie Tang¹, Runtao Feng¹, Xujun Xu¹, Ya Zhang¹, Junpeng Wang², Song Zhou¹, Ming Zhao¹

¹Department of Organ Transplantation, Zhujiang Hospital, Southern Medical University, Guangzhou, 510282, People's Republic of China; ²Department of Urology, Henan Provincial People's Hospital, Zhengzhou University People's Hospital, Zhengzhou, 450003, People's Republic of China

Correspondence: Ming Zhao; Song Zhou, Email zhaoming02@hotmail.com; 23493313@qq.com

Background: Cytomegalovirus (CMV) infection is a serious complication after kidney transplantation. Although most recipients are CMV-seropositive (R+), preventive strategies for this group remain controversial, whereas they are relatively well established for CMV-seronegative recipients (R-). Conventional serostatus-based classification alone is insufficient to accurately assess infection risk in R+ individuals. Therefore, we aimed to develop machine learning models that integrate clinical and immune variables to provide a precise risk prediction tool for CMV infection in R+ recipients.

Methods: This study included patients from June 2023 to December 2024, and were randomly divided into training and validation cohorts in a 7:3 ratio. Feature selection was performed in the training cohort using the Boruta algorithm. Six machine learning models were applied to identify the best model for predicting CMV infection risk in R+ patients, and model interpretability was assessed using SHAP.

Results: Of 162 R+ patients, 51.2% developed CMV DNAemia. Seven key predictors were identified, including T-cell subsets (CD8+, CD4+, CD4+CD27-), recipient age, cold ischemia time, donor type, and prevention strategy. Among these, CD4+ and CD8+ T-cell subset counts were the most influential predictors, with lower counts associated with a higher risk of CMV infection. The support vector machine (SVM) achieved the best discrimination in the validation cohort (AUC, 0.821; 95% CI, 0.692–0.932).

Conclusion: The interpretable SVM model showed promising performance for identifying R+ recipients at high risk of CMV infection and potentially individualized prophylactic and monitoring strategies. External validation in prospective cohorts is warranted.

Keywords: machine learning, kidney transplantation, cytomegalovirus, infection, predictive model

Introduction

Kidney transplantation (KTx) is the preferred treatment for end-stage renal diseases. Although immunosuppressants reduce rejection rates, they increase the risk of opportunistic infections.^{1,2} Cytomegalovirus (CMV), a herpesvirus, usually remains latent in immunocompetent hosts but can cause severe complications in transplant recipients,³ CMV may manifest as asymptomatic viraemia, a syndrome, or tissue-invasive disease,^{4,5} and has also been linked to rejection, cardiovascular events, and post-transplant diabetes.^{6–8}

CMV infection is influenced by multiple factors, with donor and recipient CMV serostatus serving as key determinants for risk stratification.^{9,10} CMV-seropositive recipients (R+) were considered to have a moderate infection risk, regardless of donor serostatus (D+ or D-).¹¹ Moreover, the majority of kidney transplant recipients are CMV-seropositive,¹² yet the optimal preventive strategy for this group has not been clearly established.^{13,14} Universal prophylaxis and preemptive therapy represent the two main strategies for CMV infection, both of which are effective when patients are closely monitored.^{15,16} If a preemptive therapy strategy is adopted, a reduced frequency of CMV

monitoring may increase the risk of infection and adversely affect graft outcomes. Despite reducing CMV infections, prophylaxis is associated with high costs, drug resistance, and leukopenia.^{17,18} In addition, dose reduction of prophylactic agents appears to be a feasible approach, and several studies have shown that low-dose prophylaxis can be effective in R+ recipients at relatively low risk of infection. Therefore, accurate assessment of individual CMV risk is crucial for selecting the optimal preventive strategy in R+ patients.^{19,20}

Conventional CMV serostatus-based stratification is insufficient for accurately assessing infection risk in R+ recipient, especially for guiding individualized prophylaxis.²¹ Assessing whether a recipient is over- or mildly immunosuppressed may therefore provide additional value.²² This immune status of transplant recipients is closely associated with CMV infection, and peripheral blood lymphocyte subsets reflect the overall immune condition.²³ Non-specific lymphocyte subsets have limited accuracy in infection risk prediction, and research using CMV-specific lymphocyte subsets has shown improved predictive performance.²⁴ However, its clinical application remains restricted owing to high costs, technical complexity, and limited availability in some centers. Therefore, there is an urgent need to develop a risk assessment tool based on simple and readily available indicators with high accuracy.

In recent years, machine learning (ML) has shown great potential in the field of transplantation, with successful applications in predicting postoperative infections, assessing graft function, and other clinical aspects.^{25,26} Machine learning algorithms can process high-dimensional data and capture complex nonlinear relationships, which can enhance the predictive accuracy of clinical models.^{27,28} Moreover, recent studies have increasingly emphasized supervised and explainable machine learning frameworks, demonstrating their value in improving predictive accuracy and model interpretability in complex prediction tasks.^{29,30} Combining interpretable machine learning with non-specific immune indicators holds the potential to further enhance the predictive capability and interpretability of the model.

In this study, we applied six machine-learning algorithms to develop a risk model for CMV infection in R+ recipients to support individualized prevention and management. Because post-transplant CMV infection occurs mainly during the early post-transplant period and the prophylactic treatment window, we focused on predicting CMV DNAemia within 6 months after transplantation. Clinical variables and non-specific lymphocyte subsets were included in the model. SHapley Additive exPlanations (SHAP) was used to improve model transparency. Ultimately, we will obtain an interpretable tool for assessing the risk of R+ receptor infection.

Materials and Methods

Data Source and Study Population

This study applied a machine-learning model to predict CMV infection within 6 months after kidney transplantation (KTx). Data were collected from multiple sources to improve completeness and transparency. Donor characteristics, recipient demographics, and transplant-related variables were obtained from the China Organ Transplant Response System (COTRS). Laboratory test results and treatment-related information were extracted from the electronic medical record system of Zhujiang Hospital. Follow-up and outcome data were supplemented through telephone follow-up when necessary. The final dataset included donor and recipient demographics, transplant-related characteristics, laboratory findings, medication regimens, and clinical outcomes.

A total of 162 adult patients who underwent KTx between June 2023 and December 2024 were included in this study. Patients were included (1) aged ≥ 18 years, (2) CMV-seropositive status before transplantation, and (3) negative CMV-DNA test before transplantation. Patients were excluded (1) death or allograft nephrectomy within a month post-transplant, (2) prior history of other solid organ transplantation, (3) loss to follow-up.

Study Variables and Data Preprocessing

Feature selection plays a crucial role before model training because it effectively reduces noise and alleviates overfitting by eliminating irrelevant and redundant features. According to clinical experience and previous reports,^{10,31} common clinical indicators and immune parameters measured approximately a month post-transplantation were included in the analysis, while variables with more than 30% missing data were excluded to ensure model robustness. Finally, we included the remaining 36 features in further analysis. Additionally, we employed multiple imputation to reduce

accidental bias by excluding participants with missing data, addressing missing data. Subsequently, we applied the Z-score algorithm to normalize the input values, achieving consistent evaluation of attributes and preventing model overfitting. The final variables included the following:

Recipient: sex, age, height, weight, body mass index (BMI), anti-CMV IgG titer kidney transplant history, CMV prevention strategy, panel reactive antibody (PRA), and dialysis modality.

Donor: sex, age, height, weight, BMI, serum creatinine level, and donor type.

Transplant-related: cold ischemia time, warm ischemia time, total perioperative intravenous steroid dose, HLA mismatch, and induction therapy.

Laboratory indicators (around a month post-op): CD4+, CD8+, CD4+CD27-, CD4+CD28- CD8+CD27-, CD8+CD28- counts, CD4+/CD8- ratio, and recipient's scr.

Clinical events included: delayed graft function (DGF), BK virus infection, and rejection before infection.

Dataset Splitting and Variable Selection

Patients were randomly divided into training and validation sets in a 7:3 ratio. The Boruta algorithm was used for feature selection in the training set, which compared the Z-score of each variable against that of randomly permuted “shadow features.”³² After 500 iterations with a significance threshold of 0.05, the features were classified as: important (green area, consistently higher Z-scores than shadow features), unimportant (red area, consistently lower), or tentative (yellow area, intermediate). Both important and tentative features were retained for subsequent modeling. To ensure stability and avoid multicollinearity, the variance inflation factor (VIF) was calculated,³³ with a VIF <10 considered acceptable.

Definitions

Microbiological and Serology Tests

Quantitative surveillance of serum CMV DNA was performed using a commercial PCR kit (DaAn Gene Co., Ltd., China) on an ABI 7500 system (Thermo Fisher Scientific, USA). The assay had a lower limit of quantification (LLOQ) of 500 copies/mL. Samples with a detectable signal below the LLOQ were reported as “positive, <500 copies/mL”, whereas samples ≥ 500 copies/mL were quantified and reported as an integer value. Pretransplant CMV serostatus of donors and recipients was assessed using an electrochemiluminescence immunoassay (ECLIA) on a Cobas e801 analyzer (Roche Diagnostics, Germany).

CMV Infection and Monitoring Strategy

In this study, the primary outcome (“CMV infection”) was defined as CMV DNAemia, ie, any detectable serum CMV DNA by PCR, including results reported as “positive, <500 copies/mL” CMV syndrome and CMV disease were diagnosed according to established guideline definitions.³⁴ Following a preoperative baseline assessment, serum CMV DNA was monitored weekly during the first postoperative month and subsequently at months 1, 3, and 6 after transplantation. All included patients had at least 6 months of follow-up.

Prophylaxis and Preemptive Therapy

CMV management at our center follows a stepwise dose-escalation approach. For prophylaxis, valganciclovir is initiated at a low dose (112.5 mg daily) after renal function recovery—a dose lower than that reported in previous studies,^{35,36} and maintained for at least three months; if CMV DNA replication occurs, the dose is escalated until clearance and then reduced back. For preemptive therapy, low-dose valganciclovir (225–450 mg daily) is started upon confirmed replication, escalated until control, and discontinued after two consecutive negative tests. Some patients do not receive prophylaxis for economic or other reasons.

Rejection

Rejection was defined based on the clinical criteria and/or confirmed by pathological findings from a graft kidney biopsy.³⁷

Delayed Graft Function

DGF was defined as the need for at least one dialysis treatment within one week after kidney transplantation.³⁸

Model Construction

After feature selection using the Boruta algorithm only in the training set, six machine learning models were developed: logistic regression (LR),³⁹ support vector machine (SVM),⁴⁰ XGBoost (XGB),⁴¹ random forest (RF),⁴² LightGBM (LGBM),⁴³ and multilayer perceptron (MLP).⁴⁴ CMV infection within 6 months of KTx was defined as a binary outcome. Hyperparameters for all models were optimized by grid search with 10-fold cross-validation within the training set to maximize the area under the receiver operating characteristic curve (AUC).⁴⁵ For the SVM model, the tuning process specifically included kernel type, regularization parameter C, and gamma. The final optimal hyperparameter settings for all models, including the SVM parameters, are provided in [Table S1](#). The optimal parameter configuration obtained through grid search effectively reduces the risk of overfitting and improves the predictive performance of each machine learning model on the dataset.⁴⁶

Model Explainability

The SHAP (Shapley Additive Explanations) framework, based on the principles of cooperative game theory,⁴⁷ was used to interpret the best-performing model. The SHAP values quantify the contribution of each feature, allowing for the ranking of feature importance and overall model interpretation. Mean absolute SHAP values and Beeswarm plots were used to illustrate the relative importance of the features and their directional effects on CMV infection risk, respectively.

Statistical Analysis

All analyses were performed using the R software (version 4.2.7). Continuous variables were expressed as mean \pm SD or median (IQR), and compared using Student's *t* test or Mann–Whitney *U*-test, respectively. Categorical variables were summarized as frequencies (%) and analyzed using the Chi-square or Fisher's exact test. A two-tailed $p < 0.05$ was considered statistically significant.

Results

Patient Characteristics

After applying the exclusion criteria, 162 patients were included in this study. A total of 78 patients were excluded: 70 patients were under 18 years of age, 1 patient was CMV-seronegative, 1 patient had a history of liver transplantation, 2 patients died within a month post-transplant, 2 patients underwent graft nephrectomy within one month, and 4 patients were lost to follow-up. The flow of the study is shown in [Figure 1](#).

Among the 162 CMV-seropositive recipients, 83 (51.2%) developed CMV DNAemia. Comparisons between the infected and non-infected groups in terms of donor and recipient characteristics, transplantation-related variables and laboratory variables around day 30 are presented in [Table 1](#). Within the infected group, the median recipient age was significantly higher than that in the non-infected group (37.0% vs. 46.0%, $P = 0.007$), suggesting that increasing age elevates the risk of infection. Donor serum creatinine levels were significantly higher in the infected group (92.0 vs 131.0, $P = 0.005$), and the proportion of living donor kidney transplants was significantly lower, suggesting that donor-related and perioperative factors may contribute to CMV risk, possibly through differences in graft condition, ischemia exposure, and early post-transplant inflammatory burden. No significant differences were observed in other recipient or donor variables between the two groups. Among transplantation-related variables, cold ischemia time was associated with infection, with prolonged cold ischemia time being a risk factor for infection, whereas no significant differences were found in other variables such as warm ischemia time or induction therapy. In post-transplantation laboratory examinations, CD4+ and CD8+ T-lymphocyte counts were significantly elevated in the non-infected group, suggesting that lymphocyte counts may reflect the risk of infection. The training and validation sets were well-balanced, are documented in [Table S2](#).

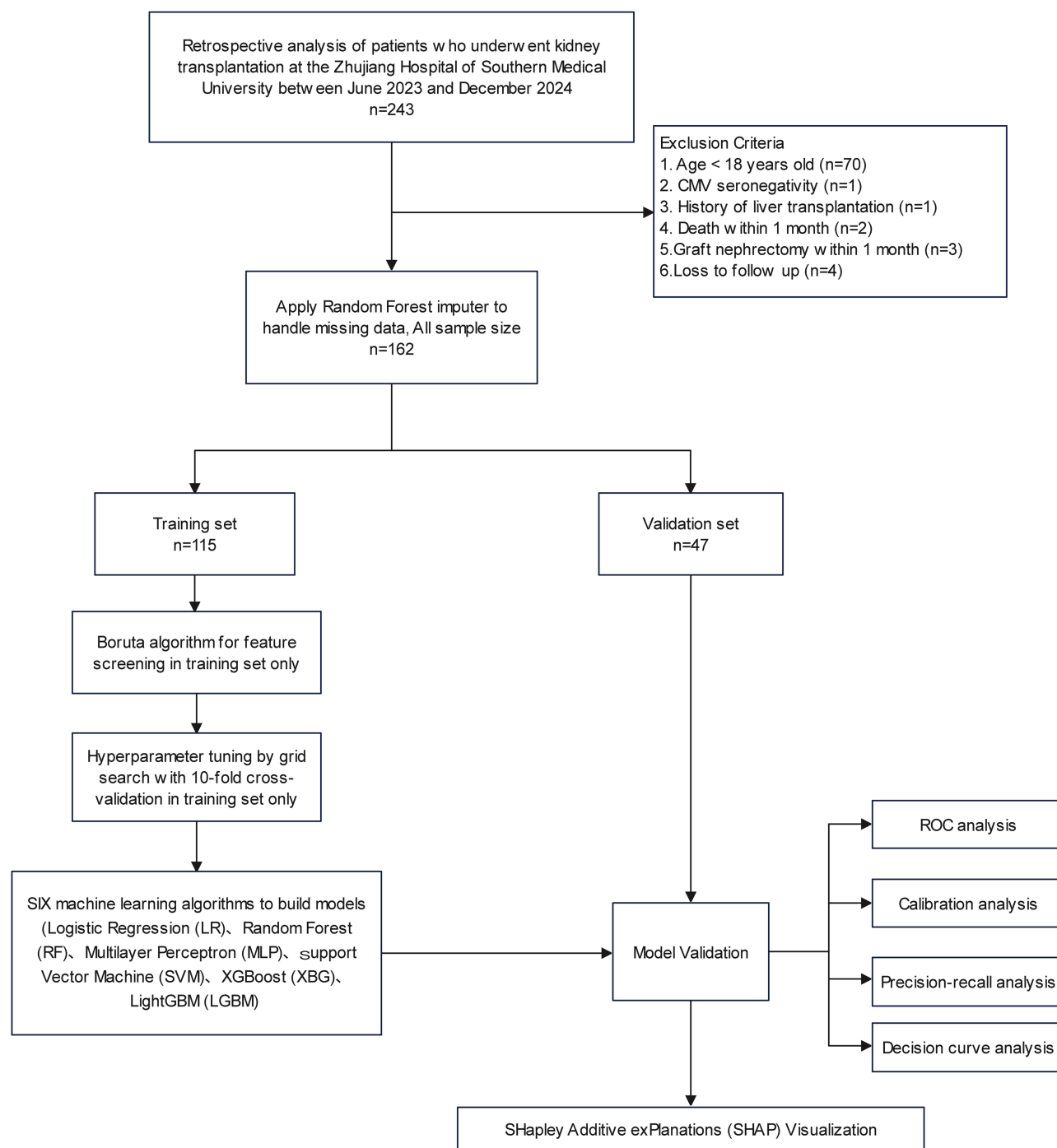


Figure 1 Flow chart of the study. A total of 243 kidney transplant recipients were screened (June 2023–December 2024). After applying the exclusion criteria, 162 recipients were included and split into a training set (n=115) and a testing set (n=47).

To further assess the impact of different preventive strategies, patients were divided into a universal prophylaxis group and a preemptive therapy group (Table 2). The median time to infection was 33 days post-transplantation, and 48.2% of infected patients exhibited a viral load exceeding 500 copies/mL. Fourteen patients developed CMV syndrome, while three were diagnosed with tissue-invasive CMV disease confirmed by biopsy, including two cases of gastrointestinal involvement and one case of CMV hepatitis. In the preemptive therapy group, 65.3% of the patients developed CMV infection, and approximately 37.9% required immediate antiviral treatment. In contrast, the incidence of CMV infection was significantly lower in the universal prophylaxis group (31.3%), with only four patients (6.0%) showing

Table 1 Characteristics of Non-Infected and Infected Recipients

Characteristics	Total (n = 162)	Non-Infected (n = 79)	Infected (n = 83)	Statistic	p
Recipient					
Sex, n (%)				$\chi^2=0.31$	0.579
Male	99 (61.1)	50 (63.3)	49 (59)		
Female	63 (38.9)	29 (36.7)	34 (41)		
Age, year, Median (IQR)	42.0 (33.2, 52.0)	37.0 (31.0, 50.5)	46.0 (36.0, 54.5)	Z=-2.71	0.007
Height, cm, Mean \pm SD	164.9 \pm 8.0	165.5 \pm 8.4	164.3 \pm 7.6	t=0.96	0.341
Weight, kg, Median (IQR)	61.1 (53.0, 70.2)	60.7 (52.4, 69.7)	61.3 (53.9, 71.8)	t=-1.21	0.356
BMI, kg/m ² , Median (IQR)	22.4 (20.0, 24.9)	22.0 (19.9, 23.7)	23.1 (20.5, 25.7)	t=-2.06	0.044
Pretransplant anti-CMV IgG titer, IU/mL, Median (IQR)	369.0 (170.0, 500.0)	399.0 (173.0, 500.0)	308.0 (158.5, 500.0)	Z=-0.43	0.663
PRA status, n (%)				$\chi^2=0.57$	0.451
Positive	24 (14.8)	10 (12.7)	14 (16.9)		
Negative	138 (85.2)	69 (87.3)	69 (83.1)		
Previous kidney transplant, n (%)				$\chi^2=2.84$	0.092
Yes	16 (9.9)	11 (13.9)	5 (6)		
No	146 (90.1)	68 (86.1)	78 (94)		
Dialysis modality, n (%)					0.437
PD	35 (21.6)	18 (22.8)	17 (20.5)		
HD	122 (75.3)	60 (75.9)	62 (74.7)		
PD and HD	3 (1.9)	0 (0)	3 (3.6)		
Pre-emptive	2 (1.2)	1 (1.3)	1 (1.2)		
Donor					
Sex, n (%)				$\chi^2=0.01$	0.911
Male	93 (57.4)	45 (57)	48 (57.8)		
Female	69 (42.6)	34 (43)	35 (42.2)		
Age, year, Median (IQR)	49.0 (37.0, 56.0)	50.0 (38.0, 58.0)	47.0 (37.0, 55.0)	Z=-1.59	0.111
Height, cm, Median (IQR)	165.0 (159.0, 170.0)	165.0 (158.5, 170.0)	168.0 (160.0, 170.0)	Z=-0.73	0.465
Weight, kg, Median (IQR)	63.2 (57.2, 70.0)	61.8 (56.2, 68.0)	65.0 (59.4, 70.0)	Z=-1.51	0.132
BMI, kg/m ² , Mean \pm SD	23.1 \pm 3.0	22.8 \pm 2.9	23.3 \pm 3.0	t=-1.15	0.253
Scr, μ mol/L, Median (IQR)	108.0 (70.2, 166.0)	92.0 (64.5, 142.0)	131.0 (76.0, 216.2)	Z=-2.79	0.005
Donor type, n (%)				$\chi^2=10.10$	0.006
DBD	95 (58.6)	39 (49.4)	56 (67.5)		
DCD	22 (13.6)	9 (11.4)	13 (15.7)		
LKD	45 (27.8)	31 (39.2)	14 (16.9)		
Transplantation					
WIT, h, Median (IQR)	0.0 (0.0, 2.0)	0.0 (0.0, 2.2)	0.0 (0.0, 1.8)	Z=-1.84	0.066
CIT, h, Median (IQR)	10.0 (4.3, 13.4)	7.9 (3.8, 12.5)	11.0 (5.8, 13.8)	Z=-2.48	0.013
^a Perioperative IV steroid dose, mg, Median (IQR)	1875.0 (1750.0, 2405.0)	1950.0 (1750.0, 2395.0)	1870.0 (1750.0, 2407.5)	Z=-0.07	0.946
A mismatch, n (%)				$\chi^2=4.37$	0.112
0	58 (35.8)	34 (43)	24 (28.9)		
1	86 (53.1)	39 (49.4)	47 (56.6)		
2	18 (11.1)	6 (7.6)	12 (14.5)		
B mismatch, n (%)				$\chi^2=6.04$	0.049
0	22 (13.6)	16 (20.3)	6 (7.2)		
1	85 (52.5)	37 (46.8)	48 (57.8)		
2	55 (34.0)	26 (32.9)	29 (34.9)		

(Continued)

Table 1 (Continued).

Characteristics	Total (n = 162)	Non-Infected (n = 79)	Infected (n = 83)	Statistic	p
DRBI mismatch, n (%)				$\chi^2=4.88$	0.087
0	26 (16.0)	13 (16.5)	13 (15.7)		
1	95 (58.6)	52 (65.8)	43 (51.8)		
2	41 (25.3)	14 (17.7)	27 (32.5)		
ABDR mismatch, n (%)					0.165
0	5 (3.1)	3 (3.8)	2 (2.4)		
1	18 (11.1)	13 (16.5)	5 (6)		
2	30 (18.5)	16 (20.3)	14 (16.9)		
3	46 (28.4)	20 (25.3)	26 (31.3)		
4	40 (24.7)	19 (24.1)	21 (25.3)		
5	19 (11.7)	8 (10.1)	11 (13.3)		
6	4 (2.5)	0 (0)	4 (4.8)		
Induced therapy, n (%)					0.086
ATG	117 (72.2)	54 (68.4)	63 (75.9)		
BX	37 (22.8)	18 (22.8)	19 (22.9)		
None	8 (4.9)	7 (8.9)	1 (1.2)		
^b Laboratory variables around day 30 and clinical event					
BK infected, n (%)				$\chi^2=0.60$	0.438
0	149 (92.0)	74 (93.7)	75 (90.4)		
1	13 (8.0)	5 (6.3)	8 (9.6)		
Rejection before infected n (%)				$\chi^2=3.10$	0.059
0	157 (96.9)	79 (100)	78 (94)		
1	5 (3.1)	0 (0)	5 (6)		
DGF, n (%)				$\chi^2=0.67$	0.414
0	129 (79.6)	65 (82.3)	64 (77.1)		
1	33 (20.4)	14 (17.7)	19 (22.9)		
^b Recipient's scr around d30, $\mu\text{mol/L}$, Median (IQR)	142.5 (112.0, 183.0)	138.0 (107.5, 165.0)	147.0 (114.5, 219.5)	Z=-1.63	0.104
CD4+ count, μL , Median (IQR)	383.0 (179.2, 713.2)	520.0 (251.5, 786.5)	265.0 (155.5, 523.5)	Z=-3.17	0.002
CD8+ count, μL , Median (IQR)	298.0 (178.0, 485.8)	381.0 (216.0, 568.0)	251.0 (147.5, 363.5)	Z=-3.62	< 0.001
CD4+/CD8+ ratio, Median (IQR)	1.3 (0.8, 1.9)	1.2 (0.9, 1.8)	1.3 (0.7, 2.0)	Z=-0.08	0.937
CD4+CD27- count μL , Median (IQR)	35.0 (18.2, 63.0)	40.0 (20.0, 73.5)	29.0 (16.0, 57.5)	Z=-2.18	0.029
CD4+CD28- count μL , Median (IQR)	10.0 (4.0, 40.0)	13.0 (5.0, 54.0)	8.0 (3.0, 24.5)	Z=-1.98	0.048
CD8+CD27- count μL , Median (IQR)	54.5 (26.0, 103.0)	62.0 (27.0, 106.5)	48.0 (25.0, 97.0)	Z=-1.26	0.208
CD8+CD28- count μL , Median (IQR)	71.5 (29.2, 127.2)	88.0 (36.5, 153.0)	58.0 (24.0, 103.5)	Z=-2.27	0.023

Notes: ^a Perioperative was defined as from transplantation to hospital discharge; IV steroid dose indicates cumulative intravenous methylprednisolone (Solu-Medrol) dose (mg); ^b "day 30" or "d30" were defined as within 30 ± 7 days after transplantation. Continuous variables: values are expressed as Median (IQR) (for non-normally distributed variables) or Mean \pm SD (for normally distributed variables); Categorical variables: values are expressed as n (%).

Abbreviations: BMI, body mass index; CMV, cytomegalovirus; Pre-Tx, pre-transplant; IgG, immunoglobulin G; PRA, panel reactive antibody; PD, peritoneal dialysis; HD, hemodialysis; BKV, BK virus; DGF, delayed graft function; DBD, donation after brain death; DCD, donation after circulatory death; LKD, living kidney donor; WIT, warm ischemia time; CIT, cold ischemia time; ATG, anti-thymocyte globulin; BX, basiliximab; HLA, human leukocyte antigen; CD, cluster of differentiation.

viral loads exceeding 500 copies/mL. The incidence of CMV syndrome and tissue-invasive CMV disease was also lower in the prophylaxis group, although the difference was not statistically significant. Overall, in our cohort, low-dose valganciclovir prophylaxis was associated with a lower incidence of CMV DNAemia and fewer cases with viral load >500 copies/mL, while severe CMV manifestations were numerically less frequent but not statistically significantly different. These findings suggest that, when combined with risk stratification, prophylactic strategies may be further individualized in moderate-risk recipients.

Table 2 Comparison of CMV Infection Outcomes Between the Prophylaxis and Preemptive Therapy Groups

Outcomes	Total (n = 162)	Prophylaxis (n = 67)	Preemptive (n = 95)	p
CMV Infected, n (%)	83 (51.2)	21 (31.3)	62 (65.3)	<0.001
CMV viral load >500 copies/mL, n (%)	40 (24.7)	4 (6.0)	36 (37.9)	<0.001
Infected time after KT-x, (day)	33.00[25.00, 53.50]	30.00[25.00, 75.00]	34.00[25.00, 48.50]	0.317
CMV syndrome, n (%)	14 (8.6)	3 (4.5)	11 (11.6)	0.194
CMV disease, n (%)	3 (1.9)	0 (0.0)	3 (3.2)	0.381

Notes: Categorical variables: values are expressed as n (%).
Abbreviations: CMV, Cytomegalovirus; KT-x, kidney transplantation.

Model Development and Model Performance

Boruta selected seven predictive features in the training set: CD8⁺ T cell count, CD4⁺ T cell count, CD4⁺CD27⁻ cell count, recipient age, cold ischemia time, donor type, and CMV prevent strategy (Figure 2). No multicollinearity was detected (VIF < 10, Table S3). Six machine learning models were constructed and optimized.

Model performance was assessed using receiver operating characteristic (ROC) curve analysis. Figure 3A and B present the ROC curves for the training and validation cohorts, respectively. Among six models, the SVM model achieved an AUC of 0.9463 (0.898–0.983) in the training set and the highest AUC of 0.821 (0.692–0.932) in the validation set. Although the SVM model has a slight overfitting, it has the highest AUC value in the validation sets of all models, and the degree of overfitting is less than that of other models. In addition, this model showed the highest specificity (85.00%) and PPV (86.96%), while also maintaining good clinical applicability, with a sensitivity of 68.97% and an F1 score of 76.92%. The performance metrics of all models are presented in Table 3, including AUC, Accuracy, Sensitivity, Specificity, PPV, NPV, and F1 score. Furthermore, the SVM model showed good calibration, with a Brier

Variable Importance

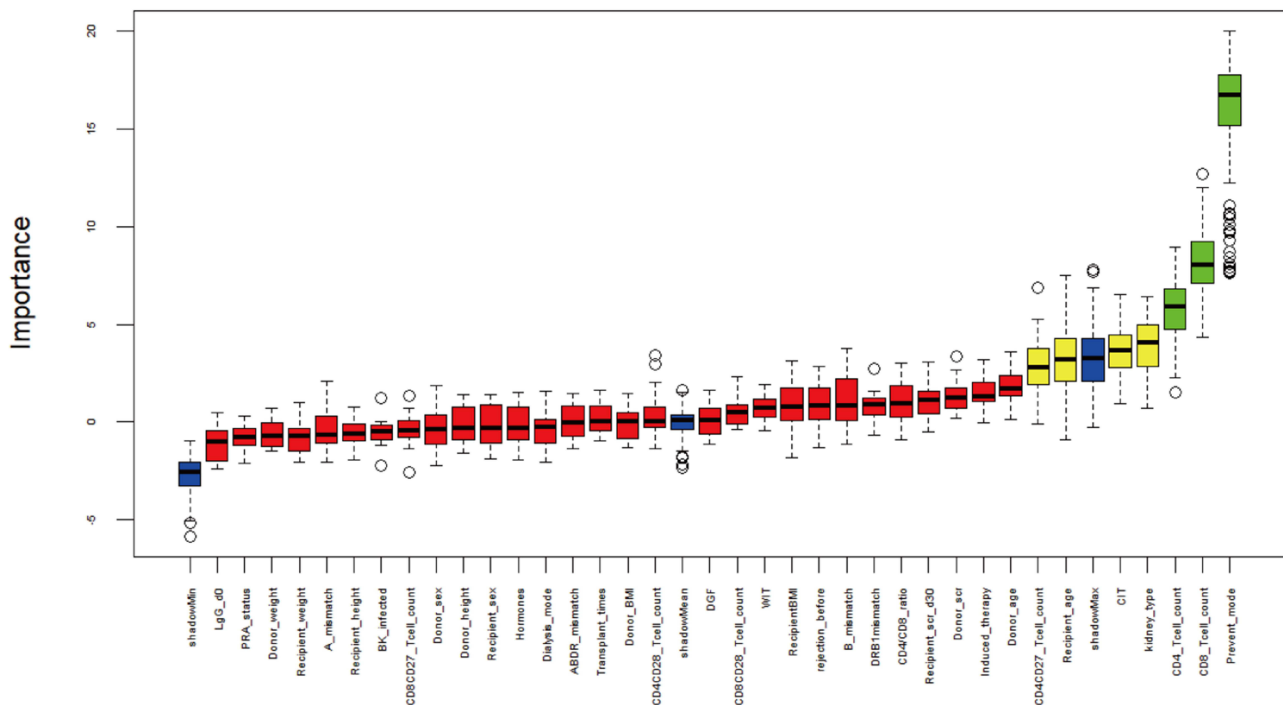


Figure 2 Feature selection based on the Boruta algorithm. The horizontal axis is the name of each variable, and the vertical axis is the Z value of each variable. The box plot shows the Z value of each variable during model calculation. The green boxes represent important variables, yellow boxes represent tentative variables, and red boxes represent unimportant variables. Blue boxes represent shadow features variables, which serve as reference importance.

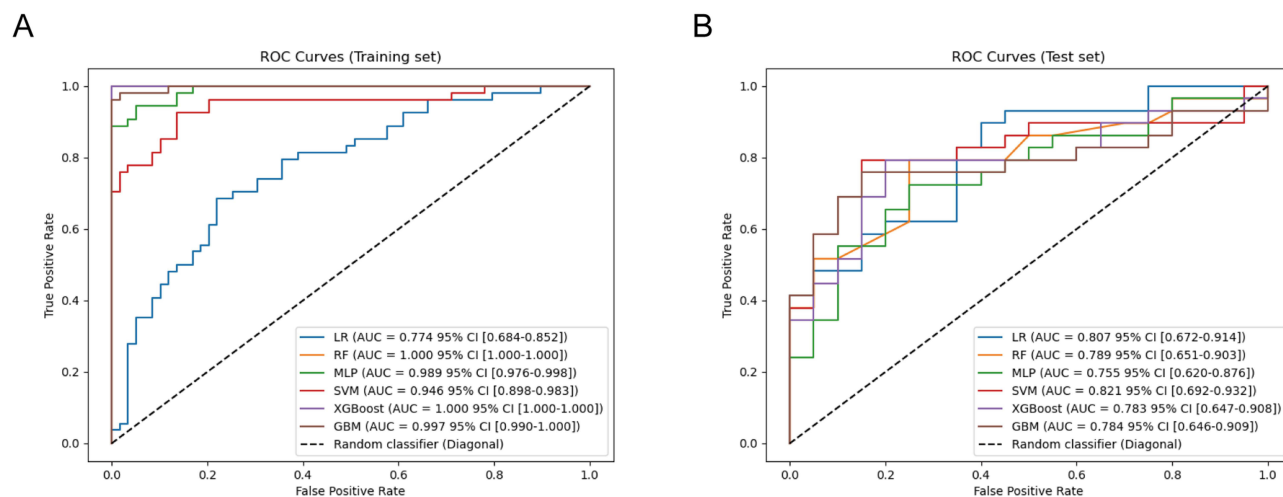


Figure 3 Receiver operating characteristic (ROC) curves of the six prediction models. **(A)** Training cohort. **(B)** Validation cohort. The area under the ROC curve (AUC) with 95% confidence intervals for each model is shown in the legend.

score of 0.135 in the training set (Figure 4A) and 0.160 in the validation set (Figure 4B), and provided greater net benefit across a clinically relevant threshold range of approximately 0.10–0.70 in the validation cohort (Figure 5). Therefore, the SVM model was identified as the optimal model for predicting the post-transplant infection risk. This may be because SVM is well suited to relatively small datasets with mixed clinical and immunological variables and can model nonlinear class boundaries while maintaining better generalization.⁴⁸ In contrast, several tree-based models showed near-perfect training performance but weaker validation performance, suggesting a greater tendency toward overfitting in the present dataset.

Model Interpretability Based on SHAP

The SHapley Additive exPlanations (SHAP) method was employed to interpret the output of the SVM model. This game theory-based approach quantifies the contribution of each feature to the model’s predictions. The SHAP summary plots

Table 3 Detailed Performance Metrics of Various Machine Learning Models for Predicting CMV Infection Risk in CMV-Seropositive Patients Across Training and Test Sets

		SVM	XGBoost	LightGBM	LR	RF	MLP
Training	Accuracy	85.84%	100.00%	97.35%	73.45%	100.00%	92.92%
	Sensitivity	81.48%	100.00%	96.30%	68.52%	100.00%	90.74%
	PPV	88.00%	100.00%	98.11%	74.00%	100.00%	94.23%
	NPV	84.62%	100.00%	96.77%	73.85%	100.00%	92.06%
	Specificity	89.83%	100.00%	98.31%	77.97%	100.00%	94.92%
	FI Score	84.62%	100.00%	97.20%	71.15%	100.00%	92.45%
	AUC	0.9463 (0.898–0.983)	1.0000 (1.000–1.000)	0.9975 (0.990–1.000)	0.7743 (0.692–0.932)	1.0000 (1.000–1.000)	0.9893 (0.976–0.998)
Test	Accuracy	75.51%	77.55%	75.51%	69.39%	71.43%	73.47%
	Sensitivity	68.97%	75.86%	75.86%	62.07%	68.97%	72.41%
	PPV	86.96%	84.62%	81.48%	81.82%	80.00%	80.77%
	NPV	62.50%	69.23%	67.50%	59.26%	62.50%	65.22%
	Specificity	85.00%	80.00%	75.00%	80.00%	75.00%	75.00%
	FI Score	76.92%	80.00%	78.57%	70.59%	74.07%	76.36%
	AUC	0.8207 (0.692–0.932)	0.7828 (0.647–0.908)	0.7845 (0.646–0.909)	0.8069 (0.672–0.914)	0.7888 (0.651–0.903)	0.7552 (0.620–0.876)

Abbreviations: AUC, area under the receiver operating characteristic curve; LightGBM, Light Gradient Boosting Machine; LR, logistic regression; MLP, multilayer perceptron; SVM, support vector machine; XGBoost, eXtreme Gradient Boosting; PPV, positive predictive value; NPV, negative predictive value.

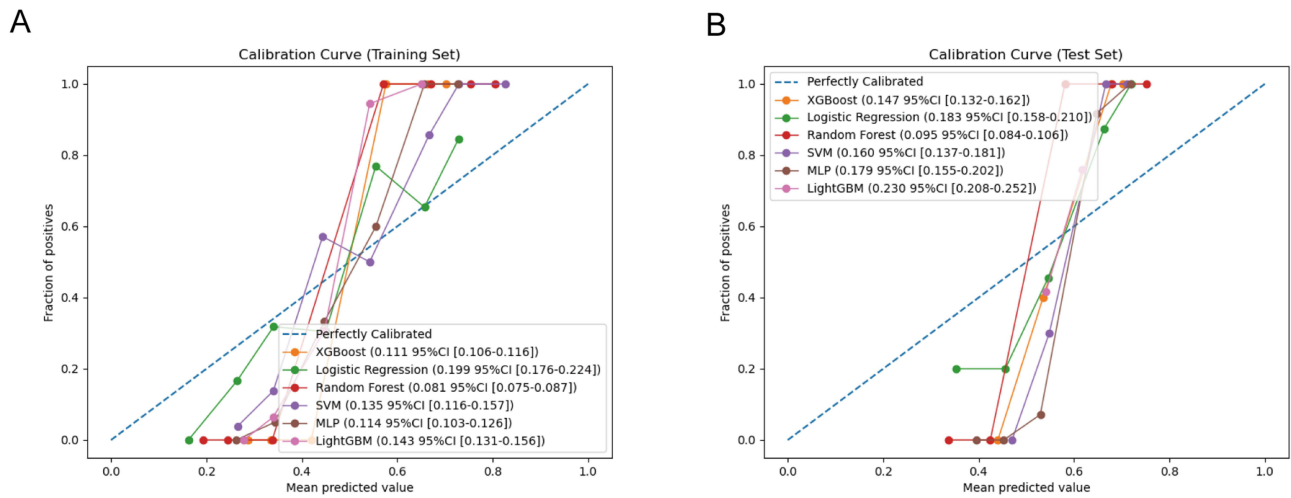


Figure 4 Calibration curves of the six prediction models. (A) Training cohort. (B) Validation cohort. The dashed diagonal line indicates perfect calibration. Brier scores (95% CI) for each model are shown in the legend.

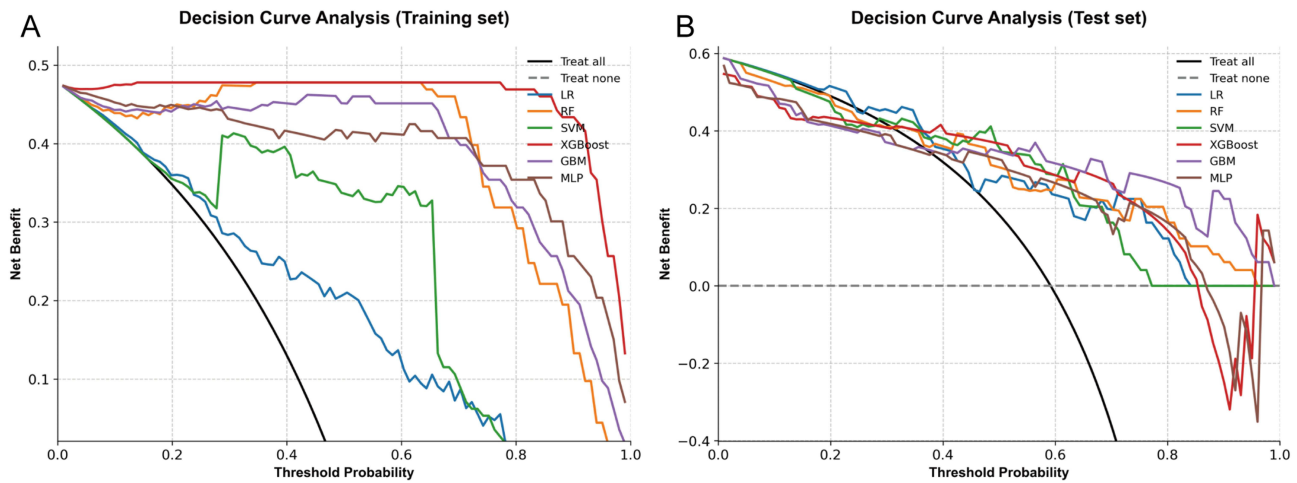


Figure 5 Decision curve analysis of the six prediction models. (A) Training cohort. (B) Validation cohort. Net benefit is plotted against threshold probability.

(Figure 6). Feature importance was ranked in descending order based on mean absolute SHAP values, as illustrated in Figure 6A. Among all features, CD8+ T cell count, CD4+ T cell count, and CD4+CD27- T cell count contributed the most substantially to the prediction. This pattern is biologically plausible because post-transplant CMV control depends predominantly on cellular immunity, and these T-cell subsets may more directly reflect the host’s antiviral immune competence than broader demographic or perioperative variables.⁴⁹ Figure 6B shows the SHAP beeswarm plot, which visualizes the impact of individual feature values on the model predictions. Positive SHAP values indicated an increased risk of post-transplant infection in our model. Specifically, lower levels of CD8+, CD4+, and CD4+CD27- T cells were associated with positive SHAP values, thus pushing the prediction of infection. The remaining features made relatively small contributions. To further evaluate the structure of the model, we developed an exploratory model using only three immunological features. This resulted in a marked decline in the predictive performance, with AUC values decreasing in both the training and validation cohorts (Figure 7). This finding suggests that the inclusion of clinical variables improves predictive performance in internal validation. Clinical variables provide complementary prognostic information beyond immune-cell subsets alone, because they capture non-immunological contributors to CMV risk such as recipient vulnerability, graft-related injury, donor characteristics, and preventive management strategy. Therefore, the seven-feature SVM model may serve as a promising internally validated tool for predicting the risk of post-transplant CMV infection.

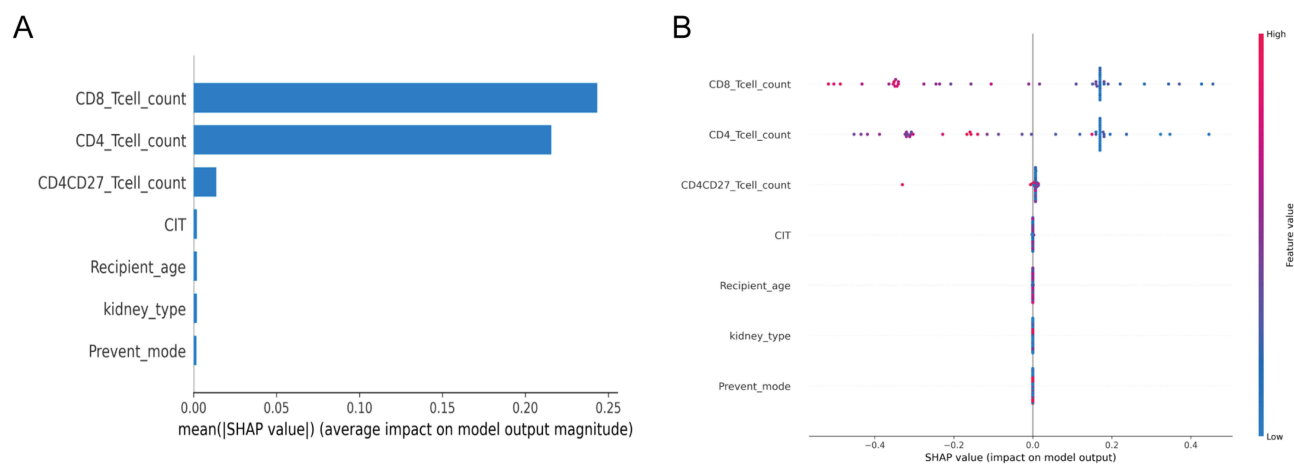


Figure 6 SHAP-based interpretation of the final model. **(A)** Importance ranking plot of features of the logistic regression model. **(B)** Characteristic attributes in SHAP. The abscissa is the SHAP value, and each line denotes a feature. Higher eigenvalues are indicated by red dots, and lower eigenvalues are indicated by blue dots.

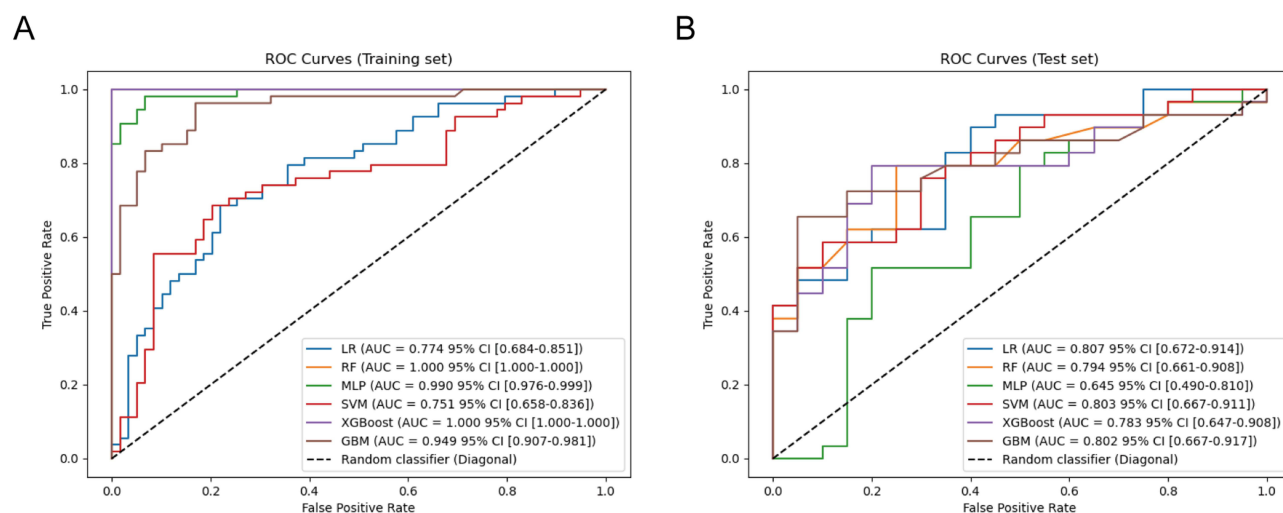


Figure 7 Receiver operating characteristic (ROC) curves of prediction models built using the top three predictors. **(A)** Training cohort. **(B)** Validation cohort. The area under the ROC curve (AUC) for each model is shown in the legend.

Discussion

In this single-center study, we developed and validated a predictive SVM model for cytomegalovirus infection after kidney transplantation. The model incorporated routine clinical variables and nonspecific immune parameters, which were selected using the Boruta algorithm, and constructed using machine learning techniques. Compared with previously reported approaches based on CMV-specific immune assays,^{50,51} our model relies on routinely available clinical and nonspecific immune variables, which may improve feasibility and accessibility in routine transplant care. Although direct performance comparisons across studies should be interpreted cautiously because of differences in study populations, endpoints, monitoring strategies, and validation methods, our findings suggest that a model based on readily available variables can still achieve encouraging predictive performance while offering greater practical applicability.

Boruta is a random-forest-based feature selection algorithm capable of handling both linear and nonlinear variables.⁵² By applying this method, we were able to accurately identify the true risk factors associated with post-transplant CMV infection. The use of the Boruta algorithm effectively reduced the number of input features, thereby simplifying the model while maintaining high predictive performance. This streamlined model can easily be implemented to assist clinicians in optimizing antiviral prophylaxis and post-transplant management strategies.

Machine learning offers a powerful computational approach capable of handling complex datasets and uncovering the nonlinear relationships among variables. In the present study, among the six tested machine learning algorithms, the support vector machine (SVM) model achieved the highest AUC and demonstrated superior predictive accuracy and clinical net benefit. The utility of SVM has been supported by several studies^{53,54} In this study, we developed an SVM-based model incorporating seven easily accessible variables to predict the risk of post-transplant CMV infection.

This study aimed to predict early CMV infection after kidney transplantation. In the SVM model, seven variables (CD8+ T cell count, CD4+ T cell count, CD4+CD27- T cell count, recipient age, cold ischemia time, donor type, and CMV prophylaxis strategy) were identified as key predictors of infection risk. Immune induction therapy was completed before one month after surgery, but lymphocytes remained stable for a longer period of time under the action of cytodepleting agents,⁵⁵ thus cellular immune responses play a central role in controlling CMV infection.⁵⁶ CD8+ T cells are the major effector cells responsible for viral clearance.⁵⁷ Therefore, higher CD8+ T cell counts were associated with a lower risk of infection in our model. CD4+ T cells are also essential for regulating antiviral immunity and guiding immunosuppressive therapy. While high CD4+ T cell levels have been linked to an increased risk of rejection, low levels are associated with a greater susceptibility to infection.⁵⁸ By secreting cytokines such as IFN- γ and TNF- α , to modulate CD8+ T cells, NK cells, and B cells, CD4+ T cells contribute to viral control. This association is biologically plausible because CMV control after transplantation relies heavily on cellular immunity; lower CD4+ and CD8+ T-cell counts may indicate impaired antiviral immune competence under post-transplant immunosuppression. Loss of CD27 indicates terminal differentiation, which enhances IFN- γ secretion and antiviral activity. CMV infection drives the expansion of CD27-CD45RA+ terminally differentiated subsets within V δ 2- γ δ T cells, particularly in seropositive recipients.

Recipient age has consistently been identified as a major risk factor for CMV reactivation. Older recipients showed a significantly higher incidence of infection,⁵⁹ likely due to immunosenescence and frailty-related immune dysfunction. Prolonged cold ischemia time has also been recognized as a high-risk factor for CMV infection reactivation,⁶⁰ as ischemia-reperfusion injury can trigger inflammatory responses that promote viral activation.³ CMV prophylaxis strategy plays a crucial role in infection prevention,⁶¹ this is clinically plausible because prophylaxis modifies the early post-transplant viral replication environment and may affect both the timing and magnitude of CMV DNAemia during the period of highest immunosuppressive intensity.⁶² The inclusion of this variable in our model allowed for individualized risk assessment across diverse clinical settings. For instance, among patients not receiving universal prophylaxis, those identified as low risk by the model may be safe with a reduced monitoring frequency. Conversely, for patients under universal prophylaxis, those classified as high risk may benefit from enhanced antiviral dosing or prolonged prophylaxis duration. In addition, induction therapy is an important determinant of CMV infection risk.⁶³ Although the induction regimen was not retained in the final model and did not differ significantly between groups, we observed a slightly higher proportion of anti-thymocyte globulin (ATG) use in the infected group, while the number of patients without induction was small. This may reflect limited statistical power due to the modest sample size. Future studies with larger cohorts are needed to better evaluate the contribution of induction therapy to CMV risk prediction.

From a clinical perspective, the main problem addressed in this study is the insufficient risk stratification of CMV-seropositive kidney transplant recipients. Although these recipients are generally classified as having intermediate CMV risk, they remain clinically heterogeneous, and serostatus-based classification alone is often insufficient to guide individualized prevention and monitoring strategies.⁶⁴ Our model was therefore developed to address a real-world management problem: how to identify, early after transplantation, which R+ recipients may require closer virologic surveillance or a more intensive preventive approach, and which recipients may be suitable for a less intensive management strategy. In real-world practice, the model may serve as a clinical decision-support tool rather than a purely hypothetical prediction model. Because all seven predictors are routinely available in standard post-transplant care, the model could be applied during the early post-transplant period without requiring specialized CMV-specific immune assays. Patients classified as higher risk may benefit from closer CMV DNA surveillance, earlier intervention, or a more individualized prophylactic strategy, whereas those classified as relatively lower risk may be considered for less

intensive monitoring under appropriate clinical supervision. In this way, the model may help bridge the gap between uniform serostatus-based management and personalized CMV prevention.

The relatively high incidence observed in our study may be explained by several factors. First, the study endpoint was detectable CMV DNAemia rather than CMV disease alone, and therefore included both asymptomatic CMV replication and CMV disease. Second, the prophylactic dose used in our center was relatively low in a subset of patients, which is consistent with the current tendency to reduce prophylactic intensity in intermediate-risk recipients in order to minimize treatment-related complications and drug exposure; notably, not all patients received universal prophylaxis. Third, patients underwent relatively close virologic monitoring during follow-up, partly because of concern that lower-dose prophylaxis might miss clinically relevant infection, and even low-level DNAemia was captured.

In this study, pediatric recipients, as well as high-risk (D+/R-) and low-risk (D-/R-) populations, were not included due to differences in immune maturity, disease spectrum, and management strategies, as well as a high proportion of missing clinical data.^{65,66} Nevertheless, the model may be applicable to these populations in future studies, as it relies on routinely available and clinically interpretable variables. It should be noted that this is an exploratory study with a relatively small sample size, and issues such as overfitting may exist; therefore, the model requires further validation in large, multicenter cohorts. Predictive models based on artificial intelligence can support clinical decision-making, but should be used as a reference tool and interpreted in conjunction with comprehensive clinical judgment.

In summary, we successfully established an interpretable machine learning model to predict CMV infection in CMV-seropositive kidney transplant recipients based on routinely available clinical and non-CMV-specific immune variables. The final SVM model demonstrated promising predictive performance in this single-center cohort. When integrated with a precision risk-prediction model, low-dose valganciclovir prophylaxis may offer a more individualized preventive approach for moderate-risk recipients.

Limitation

Although this study developed a machine learning-based model with promising predictive performance, several important limitations should be acknowledged. First, the relatively small sample size is a major limitation of the present study and may reduce the stability and robustness of model estimation and cause overfitting. Therefore, the current findings should be validated in prospective and multicenter cohorts. Second, owing to the retrospective design, several variables are frequently unavailable, such as donor CMV serostatus and recipient diabetes history. Including these variables in future analyses may further enhance predictive accuracy and support more precise risk stratification. Third, this study focused on early post-transplant CMV infection within the 6 months, and late-onset CMV infection beyond this period was not assessed. Studies with larger cohorts and longer follow-up designs are needed to further validate the model and assess its generalizability. Finally, differences in CMV prophylactic regimens across centers may affect model applicability, external validation in diverse clinical settings is essential before clinical implementation.

Conclusion

We developed an interpretable SVM model based on routinely available clinical variables and immune parameters to predict early CMV infection in CMV-seropositive kidney transplant recipients. In this single-center cohort, the model showed promising predictive performance and may support clinically meaningful risk stratification. In practice, such risk stratification may help identify recipients who require closer CMV DNA surveillance or a more individualized prophylactic strategy during the early post-transplant period. However, the model should be regarded as a decision-support tool rather than a substitute for clinical judgment. Prospective external validation in multicenter cohorts and direct comparison with CMV-specific immune diagnostics are required before clinical implementation.

Abbreviations

AUC, area under the receiver operating characteristic curve; ATG, anti-thymocyte globulin; BMI, body mass index; BX, basiliximab; CD, cluster of differentiation; CIT, cold ischemia time; CMV, cytomegalovirus; DBD, donation after brain death; DCA, decision curve analysis; DCD, donation after circulatory death; DGF, delayed graft function; HD, hemodialysis; HLA, human leukocyte antigen; IQR, interquartile range; KTx, kidney transplantation; LGBM, light gradient boosting machine; LKD, living kidney donor; LLOQ, lower limit of quantification; LR, logistic regression; ML, machine learning; MLP, multilayer perceptron; PD, peritoneal dialysis; PRA, panel reactive antibody; R+, CMV-seropositive recipient; RF, random forest; ROC, receiver operating characteristic; Scr, serum creatinine; SD, standard deviation; SHAP, Shapley additive explanations; SVM, support vector machine; WIT, warm ischemia time; XGB, extreme gradient boosting.

Data Sharing Statement

The datasets generated and/or analyzed during the current study are not publicly available due to privacy and ethical restrictions. However, de-identified data supporting the findings of this study are available upon reasonable request from either of the two corresponding authors, Dr. Ming Zhao or Dr. Song Zhou. Requests will be reviewed in accordance with institutional policies and applicable ethical and legal requirements.

Compliance with Ethics Guidelines

The study was conducted according to the guidelines of the Declaration of Helsinki. All kidney donations were voluntary, with written informed consent obtained from donors (and/or their authorized representatives, as applicable), and the donation and transplantation procedures adhered to the principles of the Declaration of Istanbul. And this study was approved by the Ethics Committee of Zhujiang Hospital of Southern Medical University (approval 2023-KY-083-02). The ethics committees waived the need for informed patient consent, in compliance with local legislation on retrospective analyses of de-identified health data.

Consent for Publication

All authors have consented to publication.

Author Contributions

All authors made a significant contribution to the work reported, whether that is in the conception, study design, execution, acquisition of data, analysis and interpretation, or in all these areas; took part in drafting, revising or critically reviewing the article; have agreed on the journal to which the article will be submitted; reviewed and agreed on all versions of the article before submission, during revision, the final version accepted for publication, and any significant changes introduced at the proofing stage; and agree to take responsibility and be accountable for the contents of the article.

Funding

This work was partially supported by the National Natural Science Foundation of China (82270781), Natural Science Foundation of Guangdong Province (2023A1515220245) and Medical Science and Technology Research Project of Henan Province (Grant No. LHGJ20250024).

Disclosure

The author(s) report no conflicts of interest in this work.

References

1. Agrawal A, Ison MG, Danziger-Isakov L. Long-term infectious complications of kidney transplantation. *Clin J Am Soc Nephrol.* 2022;17(2):286–295. doi:10.2215/CJN.15971020

2. Bharati J, Anandh U, Kotton CN, Mueller T, Shingada AK, Diagnosis RR. Prevention, and Treatment of Infections in Kidney Transplantation. *Semin Nephrol.* 2023;43(5):151486. doi:10.1016/j.semnephrol.2023.151486
3. Heald-Sargent TA, Forte E, Liu X, et al. New insights into the molecular mechanisms and immune control of cytomegalovirus reactivation. *Transplantation.* 2020;104(5):e118–e124. doi:10.1097/TP.0000000000003138
4. Taksinwarajarn T, Sobhonslidsuk A, Kantachuvesiri S, et al. Role of highly sensitive nucleic acid amplification testing for plasma cytomegalovirus DNA load in diagnosis of cytomegalovirus gastrointestinal disease among kidney transplant recipients. *Transpl Infect Dis.* 2021;23(4):e13635. doi:10.1111/tid.13635
5. Scherger SJ, Molina KC, Palestine AG, Pecan PE, Bajrovic V. Cytomegalovirus retinitis in the modern era of solid organ transplantation. *Transplant Proc.* 2024;56(7):1696–1701. doi:10.1016/j.transproceed.2024.08.007
6. Ponticelli C, Favi E, Ferrareso M. New-onset diabetes after kidney transplantation. *Medicina.* 2021;57(3):250. doi:10.3390/medicina57030250
7. Rodríguez-Goncer I, Fernández-Ruiz M, Aguado JM. A critical review of the relationship between post-transplant atherosclerotic events and cytomegalovirus exposure in kidney transplant recipients. *Expert Rev Anti Infect Ther.* 2020;18(2):113–125. doi:10.1080/14787210.2020.1707079
8. Ruenroengbun N, Sapankaew T, Chaiyakittisopon K, Phoompong P, Ngamprasertchai T. Efficacy and safety of antiviral agents in preventing allograft rejection following CMV prophylaxis in high-risk kidney transplantation: a systematic review and network meta-analysis of randomized controlled trials. *Front Cell Infect Microbiol.* 2022;12:865735. doi:10.3389/fcimb.2022.865735
9. Leeaphorn N, Garg N, Thamcharoen N, Khankin EV, Cardarelli F, Pavlakis M. Cytomegalovirus mismatch still negatively affects patient and graft survival in the era of routine prophylactic and preemptive therapy: a paired kidney analysis. *Am J Transplant.* 2019;19(2):573–584. doi:10.1111/ajt.15183
10. Raval AD, Kistler KD, Tang Y, Murata Y, Snyderman DR. Epidemiology, risk factors, and outcomes associated with cytomegalovirus in adult kidney transplant recipients: a systematic literature review of real-world evidence. *Transpl Infect Dis.* 2021;23(2):e13483. doi:10.1111/tid.13483
11. Grossi PA, Peghin M. Recent advances in cytomegalovirus infection management in solid organ transplant recipients. *Curr Opin Organ Transplant.* 2024;29(2):131–137. doi:10.1097/MOT.0000000000001139
12. Ajani JA, Barthel JS, Bentrem DJ, et al. Esophageal and esophagogastric junction cancers. *J Natl Compr Canc Netw.* 2011;9(8):830–887. doi:10.6004/jnccn.2011.0072
13. Choi MC, Kang M, Koh HH, et al. Cytomegalovirus infection in seropositive kidney transplant recipients with diverse immunological risks under preemptive strategy. *J Med Virol.* 2025;97(7):e70474. doi:10.1002/jmv.70474
14. Chung MC, Chen CH, Chang SS, et al. Prevention and management of cytomegalovirus infection and disease in kidney transplant: a consensus statement of the transplantation society of Taiwan. *J Formos Med Assoc = Taiwan Yi Zhi.* 2025;124(2):104–111. doi:10.1016/j.jfma.2024.05.009
15. Ruenroengbun N, Numthavaj P, Sapankaew T, et al. Efficacy and safety of conventional antiviral agents in preventive strategies for cytomegalovirus infection after kidney transplantation: a systematic review and network meta-analysis. *Transpl Int.* 2021;34(12):2720–2734. doi:10.1111/tri.14122
16. Witzke O, Nitschke M, Bartels M, et al. Valganciclovir prophylaxis versus preemptive therapy in cytomegalovirus-positive renal allograft recipients: long-term results after 7 years of a randomized clinical trial. *Transplantation.* 2018;102(5):876. doi:10.1097/TP.0000000000002024
17. Reischig T, Vlas T, Kacer M, et al. A randomized trial of valganciclovir prophylaxis versus preemptive therapy in kidney transplant recipients. *J Am Soc Nephrol.* 2023;34(5):920–934. doi:10.1681/ASN.0000000000000090
18. Raval AD, Kistler KD, Tang Y, Vincenti F. Burden of neutropenia and leukopenia among adult kidney transplant recipients: a systematic literature review of observational studies. *Transpl Infect Dis.* 2023;25(1):e14000. doi:10.1111/tid.14000
19. Hellemans R, Abramowicz D. Cytomegalovirus after kidney transplantation in 2020: moving towards personalized prevention. *Nephrol Dial Transplant.* 2022;37(5):810–816. doi:10.1093/ndt/gfaa249
20. Razonable RR, Humar A. Cytomegalovirus in solid organ transplant recipients—guidelines of the American Society of Transplantation Infectious Diseases Community of Practice. *Clin Transplant.* 2019;33(9):e13512. doi:10.1111/ctr.13512
21. Prakash K, Chandorkar A, Saharia KK. Utility of CMV-specific immune monitoring for the management of CMV in solid organ transplant recipients: a clinical update. *Diagnostics.* 2021;11(5):875. doi:10.3390/diagnostics11050875
22. Gardiner BJ, Lee SJ, Robertson AN, Snell GI, Westall GP, Peleg AY. Global immune biomarkers and donor serostatus can predict cytomegalovirus infection within seropositive lung transplant recipients. *Transplantation.* 2025;109(10):1656–1664. doi:10.1097/TP.0000000000005422
23. Kim HD, Bae H, Yun S, et al. Impact of induction immunosuppressants on T lymphocyte subsets after kidney transplantation: a prospective observational study with focus on anti-thymocyte globulin and basiliximab induction therapies. *IJMS.* 2023;24(18):14288. doi:10.3390/ijms241814288
24. Gliga S, Fiedler M, Dornieden T, et al. Comparison of three cellular assays to predict the course of CMV infection in liver transplant recipients. *Vaccines.* 2021;9(2):88. doi:10.3390/vaccines9020088
25. Liu XY, Feng RT, Feng WX, et al. An integrated machine learning model enhances delayed graft function prediction in pediatric renal transplantation from deceased donors. *BMC Med.* 2024;22(1):407. doi:10.1186/s12916-024-03624-4
26. Xiang X, Liu H, Wang T, et al. Prediction of postoperative infection through early-stage salivary microbiota following kidney transplantation using machine learning techniques. *Ren Fail.* 2025;47(1):2519816. doi:10.1080/0886022X.2025.2519816
27. Senanayake S, White N, Graves N, Healy H, Baboolal K, Kularatna S. Machine learning in predicting graft failure following kidney transplantation: a systematic review of published predictive models. *Int J Med Inform.* 2019;130:103957. doi:10.1016/j.ijmedinf.2019.103957
28. Paquette FX, Ghassemi A, Bukhtiyarova O, et al. Machine learning support for decision-making in kidney transplantation: step-by-step development of a technological solution. *JMIR Med Inform.* 2022;10(6):e34554. doi:10.2196/34554
29. Asif D, Arif MS, Mukheimer A. A data-driven approach with explainable artificial intelligence for customer churn prediction in the telecommunications industry. *Results Eng.* 2025;26:104629. doi:10.1016/j.rineng.2025.104629
30. Arif MS, Rehman AU, Asif D. Explainable machine learning model for chronic kidney disease prediction. *Algorithms.* 2024;17(10). doi:10.3390/a17100443
31. Gardiner BJ, Nierenberg NE, Chow JK, Ruthazer R, Kent DM, Snyderman DR. Absolute lymphocyte count: a predictor of recurrent cytomegalovirus disease in solid organ transplant recipients. *Clin Infect Dis.* 2018;67(9):1395–1402. doi:10.1093/cid/ciy295
32. Zhou H, Xin Y, Li S. A diabetes prediction model based on Boruta feature selection and ensemble learning. *BMC Bioinf.* 2023;24(1):224. doi:10.1186/s12859-023-05300-5

33. Kim JH. Multicollinearity and misleading statistical results. *Korean J Anesthesiol.* 2019;72(6):558–569. doi:10.4097/kja.19087
34. Ljungman P, Boeckh M, Hirsch HH, et al. Definitions of cytomegalovirus infection and disease in transplant patients for use in clinical trials. *Clin Infect Dis.* 2017;64(1):87–91. doi:10.1093/cid/ciw668
35. Shi Y, Lerner AH, Rogers R, et al. Low-dose valganciclovir prophylaxis is safe and cost-saving in CMV-seropositive kidney transplant recipients. *Prog Transplant.* 2021;31(4):368–376. doi:10.1177/15269248211046037
36. Lee JH, Lee H, Lee SW, Hwang SD, Song JH. Efficacy and safety according to the dose of valganciclovir for cytomegalovirus prophylaxis in transplantation: network meta-analysis using recent data. *Transplant Proc.* 2021;53(6):1945–1950. doi:10.1016/j.transproceed.2021.05.006
37. Becker JU, Seron D, Rabant M, Roufosse C, Naesens M. Evolution of the definition of rejection in kidney transplantation and its use as an endpoint in clinical trials. *Transpl Int.* 2022;35:10141. doi:10.3389/ti.2022.10141
38. Wu WK, Famure O, Li Y, Kim SJ. Delayed graft function and the risk of acute rejection in the modern era of kidney transplantation. *Kidney Int.* 2015;88(4):851–858. doi:10.1038/ki.2015.190
39. Zhang W, Huang G, Zheng K, et al. Application of logistic regression and machine learning methods for idiopathic inflammatory myopathies malignancy prediction. *Clin Exp Rheumatol.* 2023;41(2):330–339. doi:10.55563/clinexprheumatol/8ievtq
40. Burges CJC. A tutorial on support vector machines for pattern recognition. *Data Min Knowl Disco.* 1998;2(2):121–167. doi:10.1023/A:1009715923555
41. Hou N, Li M, He L, et al. Predicting 30-days mortality for MIMIC-III patients with sepsis-3: a machine learning approach using XGboost. *J Transl Med.* 2020;18(1):462. doi:10.1186/s12967-020-02620-5
42. Gao J, Liu Y. Prediction and the influencing factor study of colorectal cancer hospitalization costs in China based on machine learning-random forest and support vector regression: a retrospective study. *Front Public Health.* 2024;12:1211220. doi:10.3389/fpubh.2024.1211220
43. Yang X, Dou F, Tang G, Xiu R, Zhao X. Interpretable machine learning model for predicting anastomotic leak after esophageal cancer surgery via LightGBM. *BMC Cancer.* 2025;25(1):976. doi:10.1186/s12885-025-14387-3
44. Yu J, Li Q, Zhang H, et al. Contrast-enhanced computed tomography radiomics and multilayer perceptron network classifier: an approach for predicting CD20+ B cells in patients with pancreatic ductal adenocarcinoma. *Abdom Radiol.* 2022;47(1):242–253. doi:10.1007/s00261-021-03285-4
45. Adnan M, Alarood AAS, Uddin MI, Ur Rehman I. Utilizing grid search cross-validation with adaptive boosting for augmenting performance of machine learning models. *PeerJ Comput Sci.* 2022;8:e803. doi:10.7717/peerj-cs.803
46. Soper DS. Greed is good: rapid hyperparameter optimization and model selection using greedy k-fold cross validation. *Electronics.* 2021;10(16). doi:10.3390/electronics10161973
47. Li M, Sun H, Huang Y, Chen H. Shapley value: from cooperative game to explainable artificial intelligence. *Auton Intell Syst.* 2024;4(1):2. doi:10.1007/s43684-023-00060-8
48. Liu Z, Elashoff D, Piantadosi S. Sparse support vector machines with L0 approximation for ultra-high dimensional omics data. *Artif Intell Med.* 2019;96:134–141. doi:10.1016/j.artmed.2019.04.004
49. Bestard O, Kaminski H, Couzi L, Fernández-Ruiz M, Manuel O. Cytomegalovirus cell-mediated immunity: ready for routine use? *Transpl Int: Off J Eur Soc OrganTransplant.* 2023;36:11963. doi:10.3389/ti.2023.11963
50. Reusing JO, Agena F, Kotton CN, Campana G, Pierrotti LC, David-Neto E. QuantiFERON-CMV as a predictor of CMV events during preemptive therapy in CMV-seropositive kidney transplant recipients. *Transplantation.* 2024;108(4):985–995. doi:10.1097/TP.0000000000004870
51. Lee H, Park KH, Ryu JH, et al. Cytomegalovirus (CMV) immune monitoring with ELISPOT and QuantiFERON-CMV assay in seropositive kidney transplant recipients. *PLoS One.* 2017;12(12):e0189488. doi:10.1371/journal.pone.0189488
52. Dong W, Lal T, Liu F, Pronovost P, Bora S, Hoehn RS. Methodological considerations for optimal variable selection in machine learning for health services research. *Health Serv Outcomes Res Method.* 2025;25(4):474–486. doi:10.1007/s10742-025-00347-8
53. Guo K, Zhu B, Zha L, et al. Interpretable prediction of stroke prognosis: SHAP for SVM and nomogram for logistic regression. *Front Neurol.* 2025;16:1522868. doi:10.3389/fneur.2025.1522868
54. Zhou X, Wang H, Xu C, et al. Application of kNN and SVM to predict the prognosis of advanced schistosomiasis. *Parasitol Res.* 2022;121(8):2457–2460. doi:10.1007/s00436-022-07583-8
55. Fabrizio VA, Rodriguez-Sanchez MI, Mauguen A, et al. Adoptive therapy with CMV-specific cytotoxic T lymphocytes depends on baseline CD4+ immunity to mediate durable responses. *Blood Adv.* 2021;5(2):496–503. doi:10.1182/bloodadvances.2020002735
56. Jacobs SE, Ibrahim U, Vega AB, et al. Dynamics of cytomegalovirus-specific T-cell recovery in allogeneic hematopoietic cell transplant recipients using a commercially available flow cytometry assay: a pilot study. *Transpl Infect Dis off J Transplant Soc.* 2024;26(3):e14290. doi:10.1111/tid.14290
57. Kotton CN. CMV: prevention, Diagnosis and Therapy. *Am J Transplant.* 2013;13(Suppl 3):24–40;quiz40. doi:10.1111/ajt.12006
58. Freiwald T, Büttner S, Cheru NT, et al. CD4+ T cell lymphopenia predicts mortality from Pneumocystis pneumonia in kidney transplant patients. *Clin Transplant.* 2020;34(9):e13877. doi:10.1111/ctr.13877
59. Hemmersbach-Miller M, Alexander BD, Pieper CF, Schmader KE. Age matters. Older age as a risk factor for CMV reactivation in the CMV serostatus positive kidney transplant recipient. *Eur J Clin Microbiol Infect Dis.* 2020;39(3):455–463. doi:10.1007/s10096-019-03744-3
60. Kirisri S, Vongsakulyanon A, Kantachuesiri S, Razonable RR, Bruminhent J. Predictors of CMV infection in CMV-seropositive kidney transplant recipients: impact of pretransplant CMV-specific humoral immunity. *Open Forum Infect Dis.* 2021;8(6):ofab199. doi:10.1093/ofid/ofab199
61. Carbone J. The immunology of posttransplant CMV infection: potential effect of CMV immunoglobulins on distinct components of the immune response to CMV. *Transplantation.* 2016;100(Suppl 3):S11–S18. doi:10.1097/TP.0000000000001095
62. Magid M, Byrns J, Gommer J, Yang Z, Lee HJ, Harris M. Early versus delayed initiation of cytomegalovirus prophylaxis after liver transplant. *Pharmacotherapy.* 2022;42(8):634–640. doi:10.1002/phar.2714
63. Tang Y, Guo J, Li J, Zhou J, Mao X, Qiu T. Risk factors for cytomegalovirus infection and disease after kidney transplantation: a meta-analysis. *Transpl Immunol.* 2022;74:101677. doi:10.1016/j.trim.2022.101677
64. Kumar L, Murray-Krezan C, Singh N, et al. A systematic review and meta-analysis of optimized CMV preemptive therapy and antiviral prophylaxis for CMV disease prevention in CMV high-risk (D+R-) kidney transplant recipients. *Transplant Direct.* 2023;9(8):e1514. doi:10.1097/TXD.0000000000001514

65. Cho MH. Pediatric kidney transplantation is different from adult kidney transplantation. *Korean J Pediatr.* 2018;61(7):205–209. doi:10.3345/kjp.2018.61.7.205
66. Huang JG, Tan MYQ, Quak SH, Aw MM. Risk factors and clinical outcomes of pediatric liver transplant recipients with post-transplant lymphoproliferative disease in a multi-ethnic asian cohort. *Transpl Infect Dis off J Transplant Soc.* 2018;20(1). doi:10.1111/tid.12798

Infection and Drug Resistance

Publish your work in this journal

Infection and Drug Resistance is an international, peer-reviewed open-access journal that focuses on the optimal treatment of infection (bacterial, fungal and viral) and the development and institution of preventive strategies to minimize the development and spread of resistance. The journal is specifically concerned with the epidemiology of antibiotic resistance and the mechanisms of resistance development and diffusion in both hospitals and the community. The manuscript management system is completely online and includes a very quick and fair peer-review system, which is all easy to use. Visit <http://www.dovepress.com/testimonials.php> to read real quotes from published authors.

Submit your manuscript here: <https://www.dovepress.com/infection-and-drug-resistance-journal>

Dovepress
Taylor & Francis Group