Construction of Machine Learning Models to Predict Changes in Immune Function Using Clinical Monitoring Indices in HIV/AIDS Patients After 9.9-Years of Antiretroviral Therapy in Yunnan, China

Bingxiang Li1†, Mingyu Li2†, Yu Song3,4,5†, Xiaoning Lu6†, Dajin Liu7†, Chenglu He3,4,5†, Ruixian Zhang8, Xinrui Wan3,4,5, Renning Zhang3,4,5, Ming Sun1*, Yi-Qun Kuang9,10*†, and Ya Li3,4,5*†

1 Institute of Medical Biology, Chinese Academy of Medical Sciences and Peking Union Medical College, Kunming, China, 2 Department of Laboratory Medicine, Yunnan Provincial Institute of Infectious Diseases, Kunming, China, 3 Department of Laboratory Medicine, First Affiliated Hospital of Kunming Medical University, Kunming, China, 4 Yunnan Key Laboratory of Laboratory Medicine, First Affiliated Hospital of Kunming Medical University, Kunming, China, 5 Yunnan Innovation Team of Clinical Laboratory and Diagnosis, First Affiliated Hospital of Kunming Medical University, Kunming, China, 6 Department of Gynecology, First Affiliated Hospital of Kunming Medical University, Kunming, China, 7 Department of Medical Records and Statistics, First Affiliated Hospital of Kunming Medical University, Kunming, China, 8 Department of Disease Control and Prevention, The First People’s Hospital of Yunnan Province, The Affiliated Hospital of Kunming University of Science and Technology, Kunming, China, 9 National Health Commission (NHC) Key Laboratory of Drug Addiction Medicine, First Affiliated Hospital of Kunming Medical University, Kunming Medical University, Kunming, China, 10 Scientific Research Laboratory Center, First Affiliated Hospital of Kunming Medical University, Kunming, China

*Correspondence: Ya Li, 13708710060@139.com, yq610433@hotmail.com, Ming Sun, sunming@imbcams.com.cn
†These authors have contributed equally to this work

Objective: To investigate trends in clinical monitoring indices in HIV/AIDS patients receiving antiretroviral therapy (ART) at baseline and after treatment in Yunnan Province, China and to provide the basis for guiding clinical treatment to obtain superior clinical outcomes.

Methods: A total of 96 HIV/AIDS patients who had started and persisted in highly active ART treatment from September 2009 to September 2019 were selected. Of these, 54 had a CD4 cell count < 200 cells/μl while 42 had a CD4 cell count ≥ 200 cells/μl. Routine blood tests, liver and renal function, and lipid levels were measured before and 3, 6, 9, and 12 months after treatment. Lymphocyte subset counts and viral load were measured once per year, and recorded for analysis and evaluation. Three machine learning models (support vector machine [SVM], random forest [RF], and multi-layer perceptron [MLP]) were constructed that used the clinical indicators above as parameters. Baseline and follow-up results of routine blood and organ function tests were used to analyze and predict CD4+ T cell data after treatment during long-term follow-up. Predictions of the three models were preliminarily evaluated.
INTRODUCTION

AIDS remains a serious public health problem in China (Xu et al., 2014; Chen et al., 2015; Chen et al., 2018). Yunnan province is located on China’s southwestern border with Vietnam, Myanmar, and Laos and has a large cross-border population. Yunnan is also close to the Golden Triangle, China–Myanmar–Laos and has a large cross-border population. In 1989, the first outbreak of human immunodeficiency virus type 1 (HIV-1) among injecting drug users occurred in Dehong Prefecture, Yunnan (Jia et al., 2011). Since then, Yunnan has become the center of an HIV-1 epidemic in China and the country’s worst-hit region for AIDS (Wang et al., 2015; Li et al., 2016). Studies show that nearly 25% of new HIV cases in China come from Yunnan (Xiao et al., 2007; Duan et al., 2008; Chow et al., 2013), of which 92.6% are caused by unprotected sex (Su et al., 2016; Li et al., 2017; Zhu et al., 2018).

HIV is a retrovirus that primarily infects CD4+ T lymphocytes, leading to a progressive decline in their number, gradually weakening the host’s immune system leading to acquired immune deficiency syndrome (AIDS). In untreated infected patients, the numbers of CD4 cells decline progressively (Fèvrier et al., 2011), and so the CD4 cell count has become an important indicator for the selection of treatment plans and measurement of the effectiveness of antiretroviral therapy (ART) (Gazzard and Moyle, 1998; Carpenter et al., 2000; Dybul et al., 2002). In addition, the number of CD4+ T lymphocytes is an important indicator by which to judge the progression of the disease and evaluate patient prognosis. After receiving antiviral treatment, patients infected with HIV undergo a period of immune reconstruction of variable duration.

Conclusion: By the incorporation of clinical indicators in SVM, RF, and MLP machine learning models, the immune function and recuperation of HIV/AIDS patients can be predicted and evaluated, thereby better guiding clinical treatment.

Keywords: HIV/AIDS, RF, MLP, SVM, machine learning model, CD4/CD8 ratio, prediction, ART
type can be used to build predictive models. Since the inconsistent prediction accuracy in different models, the prediction results shared by multiple models are more accurate (Huang et al., 2018; Renganathan, 2019; Blanchet et al., 2020). This study used three machine learning methods to build the predictive model.

In Yunnan province, a relatively underdeveloped frontier province, it is not feasible to count CD4 or other lymphocyte subsets because these parameters depend on a flow cytometry platform. For confirmed HIV/AIDS patients and others during follow-up, indicators such as routine blood and liver function tests, etc. are more readily available, thus, the present study aims to construct three different models based on different baseline levels of CD4, CD8, the CD4/CD8 ratio and other follow-up results, among newly diagnosed HIV/AIDS patients with a CD4 cell count < 200 cells/µl and CD4 cell count ≥ 200 cells/µl. The model can predict changes in immune function and thereby calculate the prognosis of HIV/AIDS patients, allowing an appropriate selection of clinical antiviral drugs. The model has the potential for considerable cost savings for diagnosis and follow-up. The benefits to infected patients are clear.

METHODS

Ethics Approval

The research protocol used in the present study has been reviewed by the Ethics Committee of the First Affiliated Hospital of Kunming Medical University. Informed consent was obtained from all participants included in the study prior to enrollment, and all information and data were confirmed for analysis.

Sample Collection

A retrospective study was conducted on HIV/AIDS diagnosis and follow-up patients from the First Affiliated Hospital of Kunming Medical University. Of the 96 patients, the longest follow-up time was 9.9 years while the shortest was 2.6 years, with a median duration of 5.9 years. All confirmed patients were screened and the presence of HIV antibody confirmed by standard methods, Western blot analysis, and nucleic acid testing as a measure of HIV viral load as a supplementary test, if necessary. All confirmed patients were diagnosed in accordance with the national technical specifications for HIV/AIDS Testing, 2020. Among 96 patients with HIV/AIDS, according to the Chinese AIDS diagnosis and treatment guidelines, the main treatment regimen was lamivudine + zidovudine + efavirenz (3TC+AZT+EFV) and lamivudine + Zidovudine + nevirapine (3TC+AZT+NVP), and the dosage was strictly in accordance with the guidelines, and in accordance with China's AIDS diagnosis and treatment guidelines.

Of the 96 HIV/AIDS patients who began and adhered to highly active ART (HAART) during the 10 year period from September 2009 to September 2019, 54 had a CD4 cell count of < 200 cells/µl while 42 had a count of ≥200 cells/µl. In accordance with the requirements of the National information management standards for free antiviral therapy, routine blood tests, liver and kidney function, and blood lipid levels were followed up 3, 6, 9, and 12 months before and after treatment. Free lymphocyte subset counts and viral load tests were performed once per year. All test results were recorded for analysis and evaluation.

The treatment plan and medical inclusion criteria for patients were those stated in the “National Manual for Free HIV Antiviral Treatment (2nd edition)”. All patients signed the “Informed Consent for Free HIV antiviral treatment” document, allowing drugs to be provided free of charge by the state.

Laboratory Testing

A 2ml sample of venous blood was collected from each HIV/AIDS patient on an empty stomach at each time point. Blood cells were analyzed by flow cytometry (FACScan II, BD Biosciences, San Jose, CA) using a combined CD3/CD4/CD8/CD45 Multitest reagent (BD Biosciences, San Jose, CA), allowing the absolute number of lymphocyte subsets to be measured and analyzed. All tests were completed less than 4 hours after venous blood collection. White blood cell (WBC) and platelet counts and hemoglobin (Hb) concentration were measured by routine blood testing using a Nisen Meikang automatic hematocyte counter (Japan). Total cholesterol (TC), total triglyceride (TG), alanine transaminase (ALT), aspartate aminotransferase (AST), and creatinine levels, aspects of blood lipids, and liver and renal function tests were performed using a Roche Cobas 8000 analyzer. Samples were prepared using a High Pure System Viral Acid kit, while a COBAS TaqMan 48 analyzer was used for automatic amplification and measurement. Samples were tested after routine daily indoor quality control testing.

Data Preprocessing

For each sample, we deleted records with missing CD4/CD8 ratio or anti-HIV treatment, all variables (including gender, age, marital status, route of infection, liver and kidney function, blood lipid levels, routine blood tests, lymphoid subsets, and HIV viral load data at diagnosis and at each follow-up) were first normalized then processed in accordance with the following formulae:

$$x_i = v_0 + \left[ \left( \frac{v_1 - v_0}{d_1} \right)^2, \frac{v_2 - v_1}{d_2 - d_1}, \ldots, \frac{v_i - v_{i-1}}{d_i - d_{i-1}} \right]^2$$

$$y_i = \frac{d_{i+1}}{d_i} \cdot v_{i+1}$$

$$X = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}, Y = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix}$$

where $v_i$ indicates the tested value of a clinicopathological variable at day $i$, $v_0$ represents a baseline measurement value, $x_i$ is a generated variable and $y_i$ is the corresponding score at day $i$. 


The generated matrix $X$ and vector $Y$ were then used for construction of the various machine learning models.

**Support Vector Machine Model**

A support vector machine (SVM) is a category of generalized linear classifier that performs binary classification on data by supervised learning, having as a decision boundary the maximum margin hyperplane that solves the learning samples (Huang et al., 2018). SVM uses a hinge loss function to calculate the empirical risk and adds a regularization term to the solution system to optimize the structural risk. The model operates as a robust classifier using sparse data. SVM can perform nonlinear classification through kernel methods and represents a common kernel learning method. Using such methods, the robustness and sparsity of an SVM reduces the computational and memory overhead of the kernel matrices while ensuring that a reliable solution is obtained. The present study used the “SVR” function in the “scikit-learn” Python package to build the model. The parameter settings are: kernel=‘rbf’, degree=3, gamma=‘scale’, and $C=1.0$.

**Random Forest Model**

A random forest (RF) refers to a classifier using multiple decision trees to train and predict samples. Output categories are determined by the mode of the output category of the individual trees (Blanchet et al., 2020). Random forests have the advantages of generating highly accurate classifiers while dealing with a large number of input variables and balancing errors. The present study used the “RandomForestRegressor” function in the “scikit-learn” Python package with all models constructed using the following parameters: n_estimators=100, criterion=‘squared_error’, min_samples_split=2, and min_samples_leaf=1.

**Multi-Layer Perceptron Model**

A multi-layer perceptron (MLP) is a class of feedforward artificial neural network. Neural networks are operational models that consist of interconnections between a large number of nodes (or neurons). Each node represents a specific output function, described as an excitation function. The connection between every two nodes represents a weighted value for the signal passing through the connection, known as the weight, equivalent to the memory of the artificial neural network (Renganathan, 2019). The output of the network varies according to the method by which the network is connected, the weight value, and the excitation function. The network itself is generally an approximation of a certain algorithm or function in nature, and may also be an expression of a logic strategy. An MLP consists of at least three layers of nodes: an input layer, a hidden layer, and an output layer. Except for the input nodes, each node represents a neuron using a nonlinear activation function. The present study used the “MLPRegressor” function in the “scikit-learn” Python package, with parameter settings: solver=“lbfgs’, alpha=1e-5, hidden_layer_sizes= (Li et al., 2010; Chen et al., 2015), and random_state=1.

**Statistical Analysis**

Continuous variables (including age and baseline CD4 level) are presented as means ± standard deviation while the means of these variables in the 2 groups were compared using a student’s $t$-test. Categorical variables (sex, marital status, and HIV transmission route) are presented as numbers (or percentages of cases) while the prevalence of these variables was compared using a Pearson’s Chi-squared test. Due to the small sample sizes for a number of variables, comparisons were conducted using a Pearson’s Chi-squared test with Yates’ continuity correction. Pearson correlation analysis was used to calculate the pairwise correlation coefficients among all clinicopathological variables in the whole cohort. Pearson correlation analysis and univariate linear regression were used to explore the correlation between the original score and the predicted score generated by the three machine learning models. P-values < 0.05 were considered statistically significant. An independent sample t-test was used for statistical analysis of the biochemical indices and the viral load in each time group, for which 0.05 was the significance level.

**RESULTS**

**Population Characteristics**

The population of patients was divided into two groups based on baseline CD4 concentrations (CD4 cell count < 200 cells/µl or ≥ 200 cells/µl). There were no differences in sex, age, or route of HIV transmission between the two groups. However, there was a significant difference in marital status between the two groups. The data indicated that the majority of patients with a CD4 cell count < 200 cells/µl were married, while a higher proportion of unmarried, divorced, and widowed patients were observed in the CD4 cell count ≥200 cells/µl (Table 1).

**Comparison of Clinical Indicators and Viral Load in Each Group**

As described above, all patients were categorized into a baseline CD4 cell concentration greater than or equal to 200, or less than 200 µl/ml. Depending on the follow-up period (from 0 to 9.8 years), the data were divided into 10 follow-up period groups (including the baseline group). All test indicators were compared between the two CD4 cell count groups at intervals of one year. It was found that there were significant differences in WBCs (P = 0.018), platelets (P = 0.001), ALT (P = 0.022), AST (P = 0.002), and hemoglobin (P = 0.002) between the two groups for follow-up periods of up to 4 years. There were significant differences in hemoglobin (P = 0.002) and AST (P = 0.002) between the two groups in the first year after diagnosis. For the follow-up period of 5 years, there was a significant difference in TC (P = 0.04) between the two groups, and a significant difference in creatinine (P = 0.014) for the 8 year follow-up (Tables 2A, 2B).

**Correlation of Clinical Indices**

The results indicate that the ALT and AST levels demonstrated a strong positive correlation ($r = 0.587$) and the CD4 level was also strongly positively correlated with the CD4/CD8 ratio ($r = 0.541$), whereas the CD8 level was strongly negatively correlated with the CD4/CD8 ratio ($r = -0.543$, Figure 1). However, the CD4 level was only weakly positively correlated
with the CD8 level \((r = 0.166)\). Furthermore, CD4 was positively correlated with WBCs \((r = 0.261)\) and platelets \((r = 0.347)\) while CD8 was positively correlated with WBCs \((r = 0.317)\). Platelets were negatively correlated with ALT \((r = -0.229)\) and AST \((r = -0.251)\), and positively correlated with WBCs \((r = 0.280)\).

### Predictive Performance of the Three Machine Learning Models

In patients with a CD4 cell count of \(< 200 \text{ cells/µl}\), there were significant correlations between the predicted results of the SVM model and the patient data for CD4 \((r = 0.390, P = 0.045)\), CD4/CD8 ratio \((r = 0.721, P < 0.001)\), platelets \((r = 0.435, P = 0.022)\), TG \((r = 0.614, P = 0.005)\), and AST \((r = 0.569, P = 0.012)\). Additionally, the predicted results of the RF model were significantly correlated with the patient data for CD8 \((r = 0.368, P = 0.028)\), CD4/CD8 ratio \((r = 0.662, P = 0.002)\), platelets \((r = 0.563, P = 0.013)\), and TG \((r = 0.536, P = 0.008)\). Finally, the predicted results of the MLP model were significantly correlated with the patient data for CD8 \((r = 0.412, P = 0.008)\), CD4/CD8 ratio \((r = 0.554, P = 0.015)\), and platelets \((r = 0.451, P = 0.016)\). For the CD4 cell count \(\geq 200 \text{ cells/µl}\) group, a significant correlation was observed for data predicted by the SVM model and the patient data for CD4 \((r = 0.365, P = 0.036)\), CD4/CD8 ratio \((r = 0.807, P < 0.001)\), WBCs \((r = 0.577, P = 0.005)\), TC \((r = 0.482, P = 0.011)\), and ALT \((r = 0.362, P = 0.035)\). The results predicted by the RF model were significantly correlated with the patient data for CD4 \((r = 0.513, P = 0.002)\), CD8 \((r = 0.634, P = 0.003)\), CD4/CD8 ratio \((r = 0.898, P < 0.001)\), WBCs \((r = 0.452, P = 0.008)\), and platelets \((r = 0.484, P = 0.004)\), while there were significant correlations for the MLP model for CD4 \((r = 0.356, P = 0.028)\), CD8 \((r = 0.315, P = 0.032)\), and CD4/CD8 ratio \((r = 0.837, P < 0.001)\). The results above demonstrate that the three machine learning models exhibited a superior predictive performance in patients with a CD4 cell count \(\geq 200 \text{ cells/µl}\) than in those with a CD4 cell count \(< 200 \text{ cells/µl}\) (Table 3).

### Predictions of the CD4/CD8 Ratio

Based on the results above, we found that the best predictive performance for CD4/CD8 ratio was achieved by the machine learning model. All three models demonstrated highly consistent predictions (Figure 2). In patients with a CD4 cell count of \(< 200 \text{ cells/µl}\), the SVM model displayed the best predictive performance \((r^2 = 0.519)\), followed by the RF model \((r^2 = 0.438)\), with the MLP model \((r^2 = 0.307)\) found to be worst. In patients with a CD4 cell count of \(\geq 200 \text{ cells/µl}\), the RF model exhibited the best predictive performance \((r^2 = 0.806)\), followed by the MLP model \((R^2 = 0.700)\), with the SVM model found to be worst \((r^2 = 0.651)\). The results indicate that it may be appropriate to utilize the SVM model to predict the CD4/CD8 ratio for patients with a CD4 cell count \(< 200 \text{ cells/µl}\), and the RF model for those with a CD4 cell count of \(\geq 200 \text{ cells/µl}\) (Figure 3).

### DISCUSSION

The number and function of lymphocytes are directly related to immune system function. The CD4 cell count is among the most critical indicators of immune function, lower counts indicative of weaker immune system function (Frontiers in Cellular and Infection Microbiology, 2018). However, not all T-cell subsets become attenuated. CD4\(^+\) T cells are helper lymphocytes that secrete cytokines that activate other immune cells. CD8\(^+\) T lymphocytes, also known as cytotoxic T cells, directly destroy virus-infected and tumor cells. Following HIV infection, the synthesis of CD4\(^+\) T cells is reduced and their destruction increases causing their number to progressively decrease, although the number of CD8\(^+\) T cells increases significantly and they become functionally activated (Masia et al., 2016). Therefore, while observing the destruction of immune function in patients following HIV infection, or its reconstruction, in addition to an intuitive index of plasma viral load, attention should also be paid to the number of CD4\(^+\) T cells, the absolute number of CD8\(^+\) T cells, the CD4/CD8 ratio, and other immune activation parameters (Cohen Stuart et al., 2000). The CD4/CD8 ratio is often described as a marker of immune status in the general population and is of interest in...
### TABLE 2A

| CD4 Baseline | 0-1 year | 1-2 years | 3-4 years | 4-5 years |
|--------------|----------|-----------|-----------|-----------|
| N            | Mean ± SD | p value   | N            | Mean ± SD | p value   | N            | Mean ± SD | p value   | N            | Mean ± SD | p value   |
| Viral load   |          |           | ≧200        | 20 14      | 3.62 ± 20.22 | 0.782 | 18 1683.06 ± 7134.36 | 0.491 | 25 845.36 ± 4195.21 | 0.321 | 23 34.04 ± 140.38 | 0.938 |
|             |          |           | <200        | 19 58156.68 ± 156034.31 | 0.018 | 32 633.63 ± 3584.32 | 0.291 | 29 6739.34 ± 29093.25 | 0.349 | 37 30.51 ± 185.61 | 0.491 |
| WBC         |          |           | ≧200        | 41 6.47 ± 6.69 | 0.121 | 68 5.36 ± 1.54 | 0.018 | 62 5.40 ± 1.68 | 0.259 | 54 5.67 ± 1.47 | 0.118 |
|             |          |           | <200        | 49 147.04 ± 81.44 | 0.416 | 114 177.41 ± 67.02 | 0.018 | 97 178.04 ± 59.48 | 0.721 | 78 186.60 ± 56.71 | 0.007 |
| Creatinine  |          |           | ≧200        | 40 70.64 ± 24.19 | 0.734 | 63 65.65 ± 23.43 | 0.781 | 63 69.12 ± 17.66 | 0.678 | 53 69.06 ± 19.03 | 0.211 |
|             |          |           | <200        | 48 72.31 ± 21.72 | 0.018 | 93 66.50 ± 14.55 | 0.291 | 93 70.43 ± 20.25 | 0.491 | 74 85.82 ± 95.50 | 0.056 |
| ALT         |          |           | ≧200        | 40 2.46 ± 4.33 | 0.639 | 61 1.87 ± 1.41 | 0.937 | 60 1.76 ± 0.91 | 0.389 | 53 3.17 ± 8.44 | 0.481 |
|             |          |           | <200        | 49 41.94 ± 44.42 | 0.281 | 109 43.22 ± 64.64 | 0.018 | 95 38.03 ± 33.06 | 0.491 | 74 35.07 ± 24.09 | 0.007 |
| Hb          |          |           | ≧200        | 41 146.98 ± 17.58 | 0.109 | 66 146.82 ± 19.06 | 0.109 | 62 150.11 ± 16.48 | 0.723 | 54 155.54 ± 18.39 | 0.381 |
|             |          |           | <200        | 46 130.24 ± 29.37 | 0.018 | 104 141.31 ± 23.26 | 0.109 | 89 151.15 ± 18.52 | 0.723 | 74 152.47 ± 20.29 | 0.381 |
| TC          |          |           | ≧200        | 35 5.05 ± 6.11 | 0.299 | 61 4.41 ± 1.22 | 0.223 | 60 4.73 ± 0.81 | 0.534 | 45 4.44 ± 1.01 | 0.139 |
|             |          |           | <200        | 39 3.98 ± 1.78 | 0.018 | 90 6.97 ± 16.29 | 0.018 | 84 5.21 ± 5.91 | 0.491 | 55 5.58 ± 5.87 | 0.139 |

The difference in bold is statistically significant. Symbol* represents a statistically different value.
| Table 2B | Statistical analysis of clinicopathological variables in each time group. |
|----------|---------------------------------------------------------------------|
| **CD4** | **Mean ± SD** | **p value** | **CD4 cell counts** | **Mean ± SD** | **p value** | **CD8 ratio** | **Mean ± SD** | **p value** |
| <200 cells/μl | 23 | 4.24 ± 0.37 | 0.039 | 12 | 2.18 ± 0.42 | 0.519 | 9 | 1.59 ± 0.36 | 0.014* |
| 200-299 cells/μl | 44 | 4.64 ± 0.37 | 0.007 | 34 | 2.48 ± 0.42 | 0.063 | 17 | 2.08 ± 0.36 | 0.001* |
| 300-399 cells/μl | 63 | 5.04 ± 0.37 | 0.001 | 54 | 2.63 ± 0.42 | 0.001 | 21 | 2.31 ± 0.36 | 0.001 |
| 400-499 cells/μl | 83 | 5.43 ± 0.37 | 0.001 | 74 | 2.83 ± 0.42 | 0.001 | 28 | 2.51 ± 0.36 | 0.001 |
| ≥500 cells/μl | 103 | 5.83 ± 0.37 | 0.001 | 94 | 3.03 ± 0.42 | 0.001 | 35 | 2.71 ± 0.36 | 0.001 |

**Note:** All variables in the two groups were compared using the Student’s t-test. The differences in bold are statistically significant.
FIGURE 1 | Correlations for all clinicopathological variables. The orange color indicates a positive correlation while cyan indicates a negative correlation. ALT, alanine aminotransferase; AST, aspartate aminotransferase; CD4, CD4+ T cell; CD8, CD8+ T cell; TC, total cholesterol; TG, total cholesterol; Hb, hemoglobin; WBC, white blood cell.

TABLE 3 | Predictive performance of the three machine learning models for clinicopathological variables.

| Variables                  | SVM model |            | RF model |            | MLP model |            |
|----------------------------|-----------|------------|----------|------------|-----------|------------|
|                            |           | R         | P        | R         | P         | R         | P         |
| CD4 cell count < 200 cells/μl |           |           |          |           |           |           |           |
| CD4                        | 0.390     | 0.045     | 0.219    | 0.185     | 0.257     | 0.207     |
| CD8                        | 0.025     | 0.241     | 0.368    | 0.028     | 0.412     | 0.008     |
| CD4/CD8 ratio              | 0.721     | < 0.001   | 0.662    | 0.002     | 0.554     | 0.015     |
| WBC                        | 0.293     | 0.200     | 0.293    | 0.243     | 0.297     | 0.116     |
| Platelets                  | 0.435     | 0.022     | 0.563    | 0.013     | 0.451     | 0.016     |
| Creatinine                 | 0.136     | 0.205     | 0.069    | 0.225     | 0.206     | 0.177     |
| TC                         | 0.219     | 0.105     | 0.112    | 0.403     | 0.126     | 0.431     |
| TG                         | 0.614     | 0.005     | 0.536    | 0.008     | 0.098     | 0.599     |
| ALT                        | 0.237     | 0.101     | 0.259    | 0.057     | 0.105     | 0.859     |
| AST                        | 0.569     | 0.012     | 0.082    | 0.442     | 0.271     | 0.064     |
| Hb                         | 0.031     | 0.728     | 0.042    | 0.543     | 0.073     | 0.652     |
| CD4 cell count ≥ 200 cells/μl |           |           |          |           |           |           |           |
| CD4                        | 0.365     | 0.006     | 0.513    | 0.002     | 0.356     | 0.028     |
| CD8                        | 0.190     | 0.108     | 0.634    | 0.003     | 0.315     | 0.032     |
| CD4/CD8 ratio              | 0.807     | < 0.001   | 0.898    | < 0.001   | 0.837     | < 0.001   |
| WBC                        | 0.577     | 0.005     | 0.452    | 0.008     | 0.383     | 0.068     |
| Platelets                  | 0.290     | 0.231     | 0.484    | 0.004     | 0.213     | 0.185     |
| Creatinine                 | 0.091     | 0.381     | 0.170    | 0.386     | 0.068     | 0.827     |
| TC                         | 0.482     | 0.011     | 0.241    | 0.052     | 0.261     | 0.127     |
| TG                         | 0.191     | 0.279     | 0.078    | 0.598     | 0.058     | 0.691     |
| ALT                        | 0.362     | 0.035     | 0.320    | 0.065     | 0.082     | 0.541     |
| AST                        | 0.298     | 0.161     | 0.277    | 0.191     | 0.183     | 0.241     |
| Hb                         | 0.120     | 0.381     | 0.254    | 0.232     | 0.081     | 0.391     |

Pearson correlation statistics are displayed as correlation coefficients (R) and P values.
CONCLUSION

In the present study, three machine learning models, namely SVM, RF and MLP, were constructed by including clinical monitoring indicators such as routine blood examination, lymphocyte subset counts, viral load, liver and renal function, and blood lipid levels in HIV/AIDS patients at baseline and the follow-up period. Baseline and follow-up results were used to analyze and predict the outcome of these measures after treatment and follow-up testing. The results demonstrated that the predictive capability of the three models was better for the group with a CD4 cell count ≥200 cells/μl than for patients with < 200 cells/μl. For both groups, the three models yielded the best predictive performance for the CD4/CD8 ratio, for which

FIGURE 2 | Comparisons of CD4/CD8 ratio predictions for the three machine learning models. Each point represents a sample. SVM, support vector machine; RF, random forest; MLP, multi-layer perceptron.
the results were highly consistent. In patients with a CD4 cell count of < 200 cells/μl, the SVM model exhibited the best performance for predicting the CD4/CD8 ratio, while in patients with a CD4 cell count of ≥200 cells/μl, the RF model was best.

DATA AVAILABILITY STATEMENT

The original contributions presented in the study are included in the article/supplementary material. Further inquiries can be directed to the corresponding authors.

ETHICS STATEMENT

The research protocol used in the present study has been reviewed by the Ethics Committee of the First Affiliated Hospital of Kunming Medical University. Informed consent was obtained from all participants included in the study prior to enrollment, and all information and data were confirmed for analysis.

AUTHOR CONTRIBUTIONS

BL, MS, Y-QK, YL, and DL conceived and designed the study. BL, ML, Y-QK, XW, and RZ collected the reagents and study materials. BL, ML, YS, XL, XW, and RZ performed the laboratory experiments. BL, RZ, MS, YL, and DL analyzed the data. BL, ML, MS, YL, and DL wrote and revised the manuscript. All authors approved the final manuscript.

FUNDING

The study was supported by the Special Funds for Yunnan Province Medical academic leader (D-2018025), a Scientific
REFERENCES

Abbass, A. K., and Lichtman, A. H. (2008). “Congenital and Acquired Immuno Deficiency”, in Basic Immunology. Data Status, 209–223. (2018). HIV Monitoring. Available at: https://www.hiv-monitoring.nl/english/patients-and-public/about-hiv/ (Accessed on October 12).

Blanchet, L., Vitale, R., van Vorstenbosch, R., Stavropoulos, G., Pender, J., Jonkers, D., et al. (2020). Constructing Bi-Plots for Random Forest: Tutorial. Anal. Chim. Acta 1131, 146–155. doi: 10.1016/j.aca.2020.06.043

Carpenter, C. C., Cooper, D. A., Fischl, M. A., Gatell, J. M, Gazzard, B. G, Hamner, S. M, et al. (2000). Antiretroviral Therapy in Adults: Updated Recommendations of the International AIDS Society-USA Panel[J]. JAMA 283 (3), 381–390. doi: 10.1001/jama.2000.17025

Chen, M., Ma, Y. M., Luo, H. B., Chen, H. C., Yang, C. J., et al. (2018). The Changing HIV-1 Genetic Characteristics and Transmitted Drug Resistance Among Recently Infected Population in Yunnan, China. Epidemiol. Infect. 146 (6), 775–781. doi: 10.1017/S0950268818000389

Chen, Y. H., Chen, Y. X., Luo, H. M., Dai, H., Song, L., et al. (2015). Multiple Introduction and Naturally Occurring Drug Resistance of HCV Among HIV-Infected Intravenous Drug Users in Yunnan: An Origin of China’s HIV/HCV Epidemics. PloS One 10 (11), e0142543. doi: 10.1371/journal.pone.0142543

Chen, R. Y., Westfall, A. O., Hardin, J. M., Miller-Hardwick, C., Stringer, J. S., et al. (2015). Multiple Applications of Support Vector Machine (SVM) Learning in Cancer Genomics. Cancer Genomics Proteomics 15 (1), 41–51. doi: 10.21873/cgp.20063

Jia, Y., Sun, J., Fan, L., Song, D., Tian, S., Yang, Y., et al. (2008). Estimates of HIV Prevalence in a Highly Endemic Area of China: Dehong Prefecture, Yunnan Province. Int. J. Epidemiol. 37 (6), 1287–1296. doi: 10.1093/ije/dyn196

Jia, Z., Wang, L., Chen, R. Y., Li, D., Wang, L., Qin, Q., et al. (2011). Tracking the Evolution of HIV/AIDS in China From 1989–2009 to Inform Future Prevention and Control Efforts. PloS One 6 (10), e25671. doi: 10.1371/journal.pone.0025671

Li, L., Chen, L., Yang, S., Liu, Y., Li, H., Bao, Z., et al. (2010). Near Full-Length Genomic Characterization of a Novel HIV Type 1 Subtype B/C Recombinant Strain From Yunnan, China. AIDS Res. Hum. Retroviruses 26 (6), 711–716. doi: 10.1089/aid.2010.0001

Li, Y. Y., Jin, Y. M., He, L. P., Bai, J. S., Liu, J., Yu, M., et al. (2016). Clinical Analysis of HIV/AIDS Patients With Drug Eruption in Yunnan, China. Sci. Rep. 6, 35938. doi: 10.1038/srep35938

Li, J., Li, Y., Yang, S., Li, J., Zhang, M., Yang, C., et al. (2016). The Identification of a Novel HIV-1 CRF01_AE/B Recombinant Based on Near Full-Genome Genomic Analysis in Yunnan Province, China. AIDS Res. Hum. Retroviruses 32 (5), 467–470. doi: 10.1089/aid.2015.0245

Li, Y., Miao, J., Miao, Z., Song, Y., Wen, M., Zhang, Y., et al. (2017). Identification of a Novel HIV Type 1 Circulating Recombinant Form (CRF86_BCO) Among Heterosexuals in Yunnan, China. AIDS Res. Hum. Retroviruses 33 (3), 279–283. doi: 10.1089/aid.2016.0188

Li, M., and Ye, H. (2010). A Fractional Differential Equation Model of CD4+ T Cells Infected With HIV With Cure Rate [J]. J. Donghua. Univ. (Nat. Sci. Ed.) 36 (01), 103–108+114.

Lu, X., Su, B., Xia, H., Zhang, X., Liu, Z., Ji, Y., et al. (2016). Low Double-Negative CD4+CD8- T Cells Are Associated With Incomplete Restoration of CD4 + T Cells and Higher Immune Activation in HIV-1 Immunological non-Responders. Front. Immunol. 7, 579. doi: 10.3389/fimmu.2016.00579

Masia, M., Padilla, S., Barber, X., Sanchez, M., Terol, G., Lidon, F., et al. (2016). Comparative Impact of Suppressive Anti Retro - Viral Regimens on the CD4/CD8 T-Cell Ratio: A Cohort Study. Med. (Baltimore) 95 (11), e3108. doi: 10.1097/MID.0000000000003108

Mussini, C., Lorenzini, P., Cozzi-Lepri, A., Lapadula, G., Marchetti, G., Nicastri, E., et al. (2015). CD4/CD8 Ratio Normalisation and non-AIDS-Related Events in Individuals With HIV Who Achieve Viral Load Suppression With Antiretroviral Therapy: An Observational Cohort Study. Lancet HIV 2, e98–e106. doi: 10.1016/S2352-3018(15)00006-5

Pereison, A. S., and Nelson, P. W. (1999). Mathematical Analysis of HIV-Idnynamics In Vivo[J]. SIAM. Rev. 41 (1), 3–4. doi: 10.1137/S0036144598335107

Rajasuriar, R., Wright, E., and Lewin, S. R. (2015). Impact of Antiretroviral Therapy (ART) Timing on Chronic Immune Activation/in - Flammation and End-Organ Damage. Curr. Opin. HIV AIDS. 10 (1), 35–42. doi: 10.1097/ COH.0000000000000118

Renganathan, V. (2019). Overview of Artificial Neural Network Models in the Biomedical Domain. Britisil. Lek. Lysty. 120 (7), 536–540. doi: 10.4149/ BLL.2019_087

Ren, L., Li, J., Zhou, S., Xia, X., Xie, Z., Liu, P., et al. (2017). Prognosis of HIV Patients Receiving Antiretroviral Therapy According to CD4 Counts: A Long-
Term Follow-Up Study in Yunnan, China. Sci. Rep. 7 (1), 9595. doi: 10.1038/s41598-017-10105-7.
Rowley, C. F. (2014). Developments in CD4 and Viral Load Monitoring in Resource-Limited Settings. Clin. Infect. Dis. 58 (3), 407–412. doi: 10.1093/cid/cit733
Serrano-Villar, S., Gutierrez, C., Vallejo, A., Hernandez-Novoa, B., Diaz, L., Abad Fernandez, M., et al. (2013). The CD4/CD8 Ratio in HIV-Infected Subjects is Independently Associated With T-Cell Activation Despite Long-Term Viral Suppression. J. Infection. 66 (1), 57–66. doi: 10.1016/j.jinf.2012.09.013
Singh, Y., and Mars, M. (2010). Support Vector Machines to Forecast Changes in CD4 Count of HIV-1 Positive Patients[J]. Sci. Res. Essay. 5 (17), 2384–2390.
Su, Y., Ding, G., Reilly, K. H., Norris, J. L., Liu, H., Li, Z., et al. (2016). Loss to Follow-Up and HIV Incidence in Female Sex Workers in Kaiyuan, Yunnan Province China: A Nine Year Longitudinal Study. BMC Infect. Dis. 16 (1), 526. doi: 10.1186/s12879-016-1854-y
Wang, Y., Liang, Y., Feng, Y., Wang, B., Li, Y., Wu, Z., et al. (2015). HIV-1 Prevalence and Subtype/Recombinant Distribution Among Travelers Entering China From Vietnam at the HeKou Port in the Yunnan Province, China, Between 2003 and 2012. J. Med. Virol. 87 (9), 1500–1509. doi: 10.1002/jmv.24202
Wang, J., Shen, C., Guo, C., Liang, L., and Wu, H. (2006). Relationship Between CD4+ T Lymphocytes and Total Number of Lymphocytes in Adult HIV/AIDS Patients and its Significance[J]. Chin. J. Gen. Med. 22, 1859–1861.
Wang, X., and Song, X. (2009). Characterization of a htlV-i Infection Model With Cure Rate and ATL Response[J]. J. Appl. Math. 22 (03), 589–595.
Xiao, Y., Kristensen, S., Sun, J., Lu, L., and Vermund, S. H. (2007). Expansion of HIV/AIDS in China: Lessons From Yunnan Province. Soc. Sci. Med. 64 (3), 665–675. doi: 10.1016/j.socscimed.2006.09.019
Xu, Y., Fu, L. R., Jia, M., Dai, G., Wang, Q., Huang, P., et al. (2014). HIV Prevalence and Associated Factors Among Foreign Brides From Burma in Yunnan Province, China. PloS One 9 (12), e115599. doi: 10.1371/journal.pone.0115599
Zhu, J., Yuan, R., Hu, D., Zhu, Z., Wang, N., and Wang, B. (2018). HIV Prevalence and Correlated Factors of Female Sex Workers and Male Clients in a Border Region of Yunnan Province, China. Int. J. STD. AIDS 29 (5), 424–434. doi: 10.1177/0956462417730258

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

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

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