



## OPEN ACCESS

## EDITED BY

Wenle Li,  
Xiamen University, China

## REVIEWED BY

Shengtao Dong,  
Second Affiliated Hospital of Dalian Medical  
University, China  
Cheng Xu,  
Zhejiang Taizhou Hospital, China  
Haosheng Wang,  
Nanjing Drum Tower Hospital, China

## \*CORRESPONDENCE

Xinli Zhan  
zhanxinli@stu.gxmu.edu.cn

<sup>†</sup>These authors have contributed equally to this work

## SPECIALTY SECTION

This article was submitted to Orthopedic Surgery, a section of the journal Frontiers in Surgery

RECEIVED 29 August 2022

ACCEPTED 21 October 2022

PUBLISHED 06 January 2023

## CITATION

Ye Z, Zhu J, Liu C, Lu Q, Wu S, Zhou C, Liang T, Jiang J, Li H, Chen T, Chen J, Deng G, Yao Y, Liao S, Yu C, Sun X, Chen L, Guo H, Chen W, Jiang W, Fan B, Tao X, Yang Z, Gu W, Wang Y and Zhan X (2023) Difference between the blood samples of patients with bone and joint tuberculosis and patients with tuberculosis studied using machine learning. *Front. Surg.* 9:1031105. doi: 10.3389/fsurg.2022.1031105

## COPYRIGHT

© 2023 Ye, Zhu, Liu, Lu, Wu, Zhou, Liang, Jiang, Li, Chen, Chen, Deng, Yao, Liao, Yu, Sun, Chen, Guo, Chen, Jiang, Fan, Tao, Yang, Gu, Wang and Zhan. This is an open-access article distributed under the terms of the [Creative Commons Attribution License \(CC BY\)](https://creativecommons.org/licenses/by/4.0/). 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.

# Difference between the blood samples of patients with bone and joint tuberculosis and patients with tuberculosis studied using machine learning

Zhen Ye<sup>†</sup>, Jichong Zhu<sup>†</sup>, Chong Liu, Qing Lu, Shaofeng Wu, Chenxing Zhou, Tuo Liang, Jie Jiang, Hao Li, Tianyou Chen, Jiarui Chen, Guobing Deng, Yuanlin Yao, Shian Liao, Chaojie Yu, Xuhua Sun, Liyi Chen, Hao Guo, Wuhua Chen, Wenyong Jiang, Binguang Fan, Xiang Tao, Zhenwei Yang, Wenfei Gu, Yihan Wang and Xinli Zhan\*

First Affiliated Hospital of Guangxi Medical University, Nanning, China

**Background:** Tuberculosis (TB) is a chronic infectious disease. Bone and joint TB is a common type of extrapulmonary TB and often occurs secondary to TB infection. In this study, we aimed to find the difference in the blood examination results of patients with bone and joint TB and patients with TB by using machine learning (ML) and establish a diagnostic model to help clinicians better diagnose the disease and allow patients to receive timely treatment.

**Methods:** A total of 1,667 patients were finally enrolled in the study. Patients were randomly assigned to the training and validation cohorts. The training cohort included 1,268 patients: 158 patients with bone and joint TB and 1,110 patients with TB. The validation cohort included 399 patients: 48 patients with bone and joint TB and 351 patients with TB. We used three ML methods, namely logistic regression, LASSO regression, and random forest, to screen the differential variables, obtained the most representative variables by intersection to construct the prediction model, and verified the performance of the proposed prediction model in the validation group.

**Results:** The results revealed a great difference in the blood examination results of patients with bone and joint TB and those with TB. Infectious markers such as hs-CRP, ESR, WBC, and NEUT were increased in patients with bone and joint TB. Patients with bone and joint TB were found to have higher liver function burden and poorer nutritional status. The factors screened using ML were PDW, LYM, AST/ALT, BUN, and Na, and the nomogram diagnostic model was constructed using these five factors. In the training cohort, the area under the curve (AUC) value of the model was 0.71182, and the C value was 0.712. In the validation cohort, the AUC value of the model was 0.6435779, and the C value was 0.644. **Conclusion:** We used ML methods to screen out the blood-specific factors—PDW, LYM, AST/ALT, BUN, and Na<sup>+</sup>—of bone and joint TB and constructed a diagnostic model to help clinicians better diagnose the disease in the future.

## KEYWORDS

tuberculosis, bone tuberculosis, joint tuberculosis, machine learning algorithms, prediction model, nomogram

## Introduction

Tuberculosis (TB) is a chronic infectious disease mainly caused by *Mycobacterium tuberculosis* (1). The main transmission route of TB is respiratory tract transmission. The prevalence of TB in China is approximately 66/100,000 in sputum smear-positive patients and approximately 119/100,000 in sputum smear-negative patients (2). TB tends to occur in people with low immunity, such as patients with diabetes, infants, patients with AIDS, and immunosuppressant users (3). Common symptoms include cough, sputum, hemoptysis, and dyspnea (4). Severe complications can lead to cardiopulmonary failure and even death.

Extrapulmonary TB is reported in 14% of TB cases (5). Bone and joint TB is an infectious disease caused by tuberculosis bacilli through blood, lymph, or direct spread to bone tissue, most of which are secondary to TB. Bone and joint TB tends to occur in the spine, hip joint, and knee joint, often presenting with pain in the joints (6). Bone and joint TB is diagnosed by blood examination, imaging, and pathological biopsy. Commonly used treatment methods include drug therapy and surgery (7).

Machine learning (ML) is the science of how computers learn from data. ML has its roots in computer science and is a subfield of artificial intelligence. It is a cross between statistics, which deals with learning relationships from data, and computer science, which focuses on efficient algorithms. ML is extensively used in clinical data processing (8, 9).

Numerous studies have been conducted on TB, but few studies have focused on the differences between patients with bone and joint TB and patients with TB. The gold standard for the diagnosis of bone and joint TB is pathological results but many primary hospitals lack imaging tools and pathological examination and can thus judge bone and joint TB only by symptoms and blood examination results. In addition, bone and joint TB is often difficult to get TB infected tissues in minimally invasive puncture diagnosis, which easily affects the clinician's judgment of the condition. In this study, we collected the clinical data of patients with bone and joint TB and common TB to determine the differences by using ML methods to help clinicians better diagnose and treat bone and joint TB.

## Patients and methods

### Patients

This study was voluntary, and all patients signed an informed consent form. This study was approved by the Ethics Committee of the First Affiliated Hospital of Guangxi Medical University ([Supplementary Materials 2](#)) [No. 2022-KY-E-(287)].

From 2012 to 2022, we collected a total of 12,770 hospitalized patients diagnosed with TB in the First Affiliated Hospital of Guangxi Medical University. The inclusion criteria for patients with pulmonary TB were as follows: (1) *Mycobacterium tuberculosis* detected in sputum smears or lesions and diagnosed as TB by two or more specialists. (2) Complete erythrocyte sedimentation rate (ESR) examination, blood routine examination, liver function examination, renal function examination, high sensitivity C-reactive protein (hs-CRP), and plasma electrolyte examination data were available for the patient. (3) Patients voluntarily participated in the study. In the case of minors, their guardians signed on their behalf. (4) The patient exhibited good compliance and no serious cardiovascular and cerebrovascular diseases. The inclusion criteria for patients with bone and joint TB were as follows: (1) Patients were hospitalized in the Department of Spinal Bone Disease, the First Affiliated Hospital of Guangxi Medical University, and diagnosed as bone and joint TB by surgery or biopsy. (2) Complete erythrocyte sedimentation rate examination, blood routine examination, liver function examination, renal function examination, and hs-CRP and plasma electrolyte examination data were available for the patient. (3) Patients voluntarily participated in the study. In the case of minors, their guardians signed on their behalf. (4) The patient had good compliance and no serious cardiovascular and cerebrovascular diseases.

The exclusion criteria were as follows: (1) Patients who did not have *Mycobacterium tuberculosis* in sputum smears or lesions. (2) Patients with missing clinical data. (3) Patients with serious cardiovascular and cerebrovascular diseases. (4) Patients who did not agree to participate in the study.

A total of 1,667 patients were enrolled in the study. Among them, 206 patients were diagnosed with bone and joint TB, and 1,461 patients were diagnosed with common TB. Among the patients diagnosed with bone and joint TB, eight were hip TB, 14 were knee TB, eight were cervical TB, 63 were lumbar TB, two were thoracolumbar TB, and 111 were thoracic TB. We randomly assigned patients to the training and validation cohorts. There were 1,268 patients in the training cohort: 158 patients with bone and joint TB and 1,110 patients with TB. The validation cohort included 399 patients: 48 patients with bone and joint TB and 351 patients with TB.

All data were collected from the information system of the First Affiliated Hospital of Guangxi Medical University, and the patients' ID numbers were used for the search. Blood routine examination included white blood cell (WBC) count, red blood cell (RBC) count, hematocrit value (HCT), mean corpuscular volume (MCV), mean corpuscular hemoglobin (MCH), mean corpuscular hemoglobin concentration (MCHC), Platelet hematocrit (PLT), platelet distribution width (PDW), neutrophil percentage (NEUT), lymphocytes percentage (LYM), monocyte percentage (MO), eosinophils percentage (EO), RBC volume distributing width (CV), and thrombocytocrit (PCT).



TABLE 1 The differences ESR and blood routine examination.

Type	Training cohort			Validation cohort		
	Bone and joint tuberculosis (N = 158)	Tuberculosis (N = 1110)	P-value	Bone and joint tuberculosis (N = 48)	Tuberculosis (N = 351)	P-value
<b>Sex</b>						
Male	108	753	0.896	33	232	0.715
Female	50	357		15	119	
<b>Age</b>						
Mean (SD)	56.92 (17.2)	55.73 (18.6)	0.450	60.42 (17.44)	56.08 (17.8)	0.113
<b>ESR</b>						
Mean (SD)	51.67 (31.5)	44.83 (30.8)	0.009	56.91 (31.29)	46.9 (30.9)	0.036
<b>WBC</b>						
Mean (SD)	8.99 (3.8)	8.37 (4.85)	0.121	9.69 (9.25)	8.22 (3.56)	0.051
<b>RBC</b>						
Mean (SD)	3.8 (0.83)	4.08 (0.78)	<0.001	3.68 (4.10)	4.10 (0.78)	0.001
<b>HCT</b>						
Mean (SD)	0.31 (0.056)	0.34 (0.06)	<0.001	0.31 (0.054)	0.34 (0.06)	<0.001
<b>MCV</b>						
Mean (SD)	83.48 (8.49)	84.56 (9.79)	0.189	84.16 (8.75)	84.85 (9.91)	0.645
<b>MCH</b>						
Mean (SD)	27.3 (3.43)	27.69 (4.1)	0.213	27.6 (3.28)	27.80 (3.78)	0.709
<b>MCHC</b>						
Mean (SD)	325.87 (12.6)	326.53 (18.3)	0.657	327.34 (10.42)	326.89 (12.04)	0.805
<b>PLT</b>						
Mean (SD)	329.7 (143.2)	309.1 (135.1)	0.077	307.5 (121.2)	322.9 (138.5)	0.464
<b>PDW</b>						
Mean (SD)	0.153 (0.03)	0.165 (0.007)	<0.001	0.154 (0.03)	0.1423 (0.03)	0.017
<b>NEUT</b>						
Mean (SD)	0.71 (0.13)	0.65 (0.13)	<0.001	0.70 (0.12)	0.66 (0.13)	0.029
<b>LYM</b>						
Mean (SD)	0.162 (0.09)	0.21 (0.10)	<0.001	0.165 (0.09)	0.204 (0.1)	0.010
<b>MO</b>						
Mean (SD)	0.086 (0.035)	0.092 (0.040)	0.059	0.090 (0.036)	0.093 (0.037)	0.594
<b>EO</b>						
Mean (SD)	0.037 (0.05)	0.04 (0.04)	0.455	0.038 (0.039)	0.038 (0.04)	0.965
<b>CV</b>						
Mean (SD)	0.166 (0.089)	0.158 (0.031)	0.004	0.170 (0.04)	0.15 (0.03)	<0.001
<b>PCT</b>						
Mean (SD)	0.26 (0.104)	0.24 (0.095)	0.031	0.25 (0.09)	0.27 (0.11)	0.093

The red text means that the *p* value was <0.05. SD, standard deviation. ESR, erythrocyte sedimentation rate; WBC, White blood cell count; RBC, Red blood cell count; HCT, hematocrit value; MCV, Mean corpuscular volume; MCH, Mean corpuscular hemoglobin; MCHC, Mean corpuscular hemoglobin concentration; PLT, Platelet hematocrit; PDW, Platelet distribution width; NEUT, neutrophil percentage; LYM, lymphocytes percentage; MO, Monocyte percentage; EO, eosinophils percentage; CV, RBC volume distributing width; PCT, thrombocytocrit.

Liver and kidney function examination included total bilirubin (TBil), direct bilirubin (DBil), indirect bilirubin (IBil), total protein (TP), albumin (ALB), globulin (GLB), gamma-glutamyl transpeptidase (GGT), total bile acid (TBA), aspartate aminotransferase (AST), alanine aminotransferase (ALT), A alkaline phosphatase (ALP), prealbumin (PAB), cholinesterase

(ChE), blood urea nitrogen (BUN), creatinine (Cr), uric acid (UA), bicarbonate radical (HCO), creatinine clearance rate (Ccr), and Cysteine C (Cys-C). Plasma electrolyte examination included K<sup>+</sup>, Na<sup>+</sup>, Cl<sup>-</sup>, Ca<sup>2+</sup>, and Mg<sup>2+</sup>. The data of all screened patients were complete, but hs-CRP was not tested in some patients, and hs-CRP was statistically analyzed separately.

TABLE 2 The differences liver function examination, renal function examination and plasma electrolyte examination.

Type	Training cohort			Validation cohort		
	Bone and joint tuberculosis (N = 158)	Tuberculosis (N = 1110)	P-value	Bone and joint tuberculosis (N = 48)	Tuberculosis (N = 351)	P-value
<b>TBil</b>						
Mean (SD)	14.56 (25.9)	9.78 (11.6)	<0.001	11.2 (7.10)	7.99 (5.51)	<0.001
<b>DBil</b>						
Mean (SD)	8.06 (18.7)	4.69 (8.1)	<0.001	5.79 (4.45)	3.96 (3.94)	0.003
<b>IBil</b>						
Mean (SD)	6.59 (8.22)	5.09 (4.27)	<0.001	5.32 (3.73)	3.96 (3.95)	0.004
<b>DBil/TBil</b>						
Mean (SD)	0.48 (0.15)	0.44 (0.15)	0.001	0.51 (0.15)	4.03 (2.77)	0.612
<b>TP</b>						
Mean (SD)	64.9 (7.42)	65.9 (8.15)	0.148	8.27 (1.19)	7.48 (0.40)	0.006
<b>ALB</b>						
Mean (SD)	32.8 (4.74)	34.96 (5.8)	<0.001	32.8 (4.03)	35.87 (5.31)	<0.001
<b>GLB</b>						
Mean (SD)	32.23 (7.47)	30.96 (7.44)	0.044	30.2 (7.92)	30.23 (6.10)	0.964
<b>ALB/GLB</b>						
Mean (SD)	1.08 (0.34)	1.20 (0.37)	<0.001	1.13 (0.30)	1.24 (0.32)	0.03
<b>GGT</b>						
Mean (SD)	66.24 (52.8)	68.43 (73.6)	0.718	56.03 (37.65)	67.02 (79.5)	0.347
<b>TBA</b>						
Mean (SD)	13.75 (33)	11.07 (22.62)	0.192	7.58 (9.38)	8.67 (12.5)	0.563
<b>AST</b>						
Mean (SD)	29.62 (18.32)	29.5 (44.1)	0.974	37.17 (64.24)	28.45 (27.16)	0.094
<b>ALT</b>						
Mean (SD)	21.77 (17.42)	25.62 (42.9)	0.265	24.0 (40.32)	24.1 (21.87)	0.979
<b>AST/ALT</b>						
Mean (SD)	1.75 (1.09)	1.42 (0.84)	<0.001	1.65 (1.12)	1.50 (1.43)	0.476
<b>ALP</b>						
Mean (SD)	107.8 (49.75)	102.34 (66.65)	0.321	94.15 (33.11)	102.43 (81.0)	0.484
<b>PAB</b>						
Mean (SD)	139.7 (69.2)	171.1 (73.5)	<0.001	0.170 (0.04)	0.15 (0.03)	0.002
<b>BUN</b>						
Mean (SD)	5.31 (4.83)	4.57 (3.45)	0.018	5.34 (4.48)	4.75 (2.83)	0.22
<b>Cr</b>						
Mean (SD)	83.44 (105.66)	76.04 (62.45)	0.209	74.27 (49.8)	82.54 (83.9)	0.505
<b>UA</b>						
Mean (SD)	308.08 (167.81)	350.36 (194.84)	0.010	293.65 (167.5)	345.74 (180.66)	0.06
<b>HCO</b>						
Mean (SD)	25.1 (4.21)	25.1 (3.79)	0.976	25.83 (4.82)	25.78 (4.23)	0.943
<b>Ccr</b>						
Mean (SD)	83.19 (30.16)	88.07 (30.58)	0.061	85.0 (33.27)	97.8 (35.45)	0.018
<b>Cys-C</b>						
Mean (SD)	1.10 (0.64)	1.02 (0.7)	0.200	1.12 (0.77)	0.94 (0.61)	0.057
<b>K<sup>+</sup></b>						
Mean (SD)	4.02 (0.47)	4.0 (0.49)	0.721	3.98 (0.36)	3.95 (0.52)	0.693
<b>Na<sup>+</sup></b>						

(continued)

TABLE 2 Continued

Type	Training cohort			Validation cohort		
	Bone and joint tuberculosis (N = 158)	Tuberculosis (N = 1110)	P-value	Bone and joint tuberculosis (N = 48)	Tuberculosis (N = 351)	P-value
Mean (SD) Cl <sup>-</sup>	137.25 (3.79)	138.67 (3.94)	<0.001	138.0 (3.94)	138.5 (0.365)	0.462
Mean (SD) Ca <sup>2+</sup>	101.86 (4.45)	102.87 (4.58)	0.009	101.9 (4.56)	100.65 (4.20)	0.052
Mean (SD) Mg <sup>2+</sup>	2.16 (0.16)	2.18 (0.16)	0.101	2.14 (0.17)	2.18 (0.17)	0.05
Mean (SD)	0.80 (0.12)	0.86 (0.14)	<0.001	0.797 (0.105)	0.872 (0.121)	<0.001

The red text means that the *p* value was <0.05. SD, standard deviation. TBil, total bilirubin; DBil, direct bilirubin; IBil, indirect bilirubin; TP, total protein; ALB, albumin; GLB, globulin; GGT, gamma-glutamyl transpeptidase; TBA, Total bile acid; AST, aspartate aminotransferase; ALT, alanine aminotransferase; ALP, A alkaline phosphatase; PAB, prealbumin; ChE, cholinesterase; BUN, Blood urea nitrogen; Cr, creatinine; UA, Uric acid; HCO, bicarbonate radical; Ccr, Creatinine clearance rate; Cys-C, Cysteine C.

TABLE 3 The differences hs-CRP examination.

Type	Training cohort			Validation cohort		
	Bone and joint tuberculosis (N = 125)	Tuberculosis (N = 1110)	P-value	Bone and joint tuberculosis (N = 38)	Tuberculosis (N = 351)	P-value
<10	29 (23.2%)	442 (39.8%)	<0.001	5 (13.2%)	148 (42.2%)	0.001
≥10	96 (76.8%)	668 (60.2%)		33 (86.8%)	203 (57.8%)	

The red text means that the *P* value was <0.05. hs-CRP, High sensitivity C-reactive protein.

## Statistical analysis

IBM SPSS Statistics 23 and R software (version 4.1.3; <https://www.R-project.org>) were used for data analysis and visualization. Student's *t*-test was used to compare the means of continuous variables between the two cohorts (patients with bone and joint TB and patients with TB) (10). The *t*-test data were normally distributed with uniform variance (11). Hs-CRP and gender differences were tested by chi-square test (12). Multiple checks were performed on the calculations to ensure that they were correct. A two-sided probability of less than 0.05 was considered statistically significant for all analyzed data (13, 14).

Univariate and multivariate logistic regression were analyzed and visualized using the “rms,” “glmnet,” and “plyr” R packages. LASSO regression was analyzed and visualized using the “glmnet” package (15). The random forest was analyzed and visualized using the “randomForest” package. The nomogram construction and C value calculation were performed using the “rms” package (16). The area under the curve (AUC) of the ROC curve and Harrell's concordance index were used to evaluate the performance of nomogram predictions (17). Harrell's concordance index (C-index) was calculated to assess nomogram discrimination by using a bootstrap method with 1,000 samples (18). The “corrplot” package was used to analyze

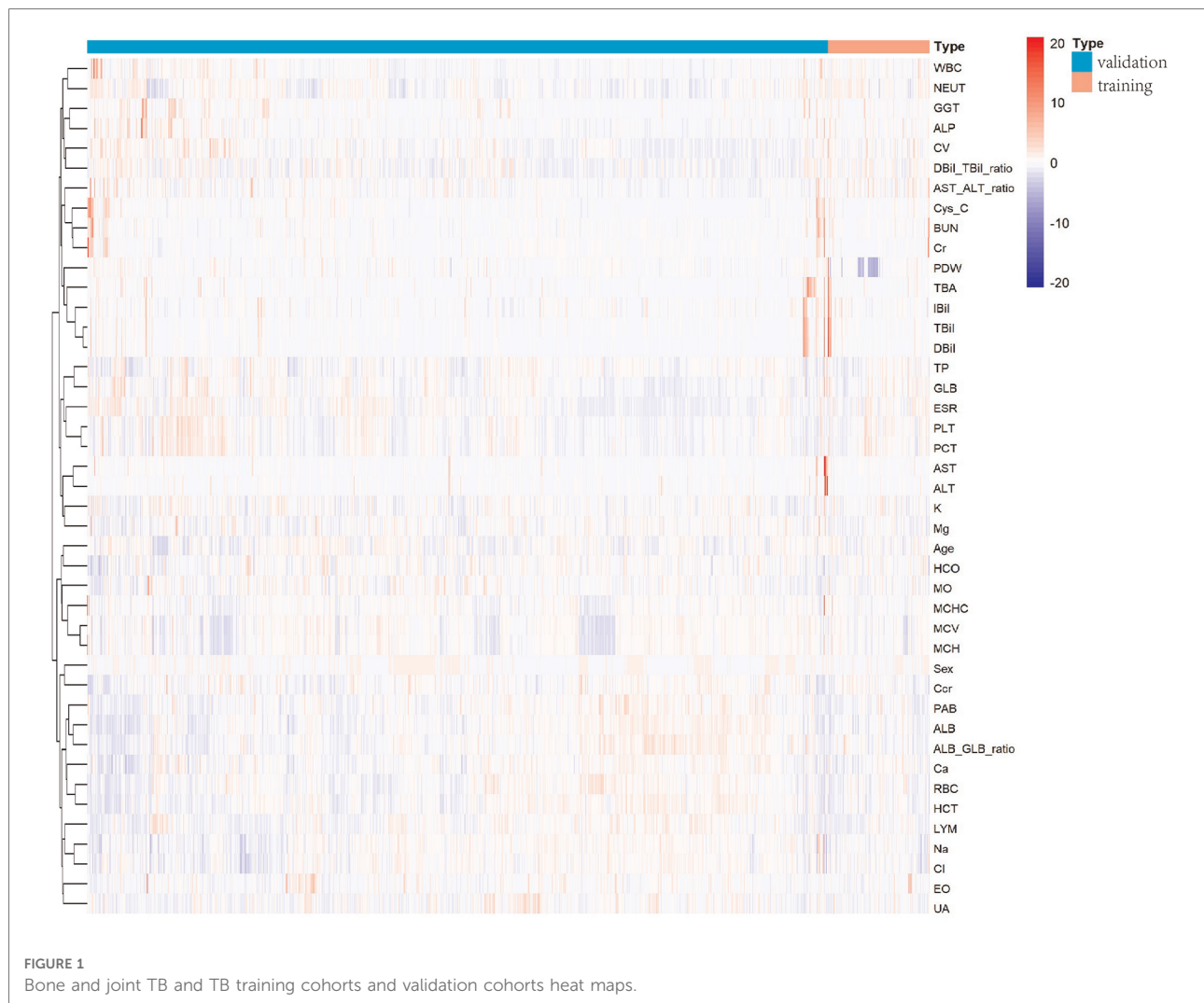
the correlation of the independent variables. The “rma” and “rmda” packages were used to calculate the net benefits and visualize the decision curve (19).

## Random forest

Random forest is a learning machine based on decision trees and can be used to predict continuous variables and improve prediction accuracy (20, 21). An increase in mean squared (% IncMSE) and an increase in node purity (IncNodePurity) are two methods of random forest for prediction. %IncMSE is a random assignment of each predictor variable. If this variable is more important, the prediction error will be larger if it is randomly replaced. IncNodePurity is measured by the sum of squared residuals. In both cases, the larger the value, the more important the prediction variable. Both can be regarded as important indicators of predictive variables (22).

## Results

Tables 1–3 list the differences in ESR, blood routine examination, liver function examination, renal function examination, and plasma electrolytes between patients with



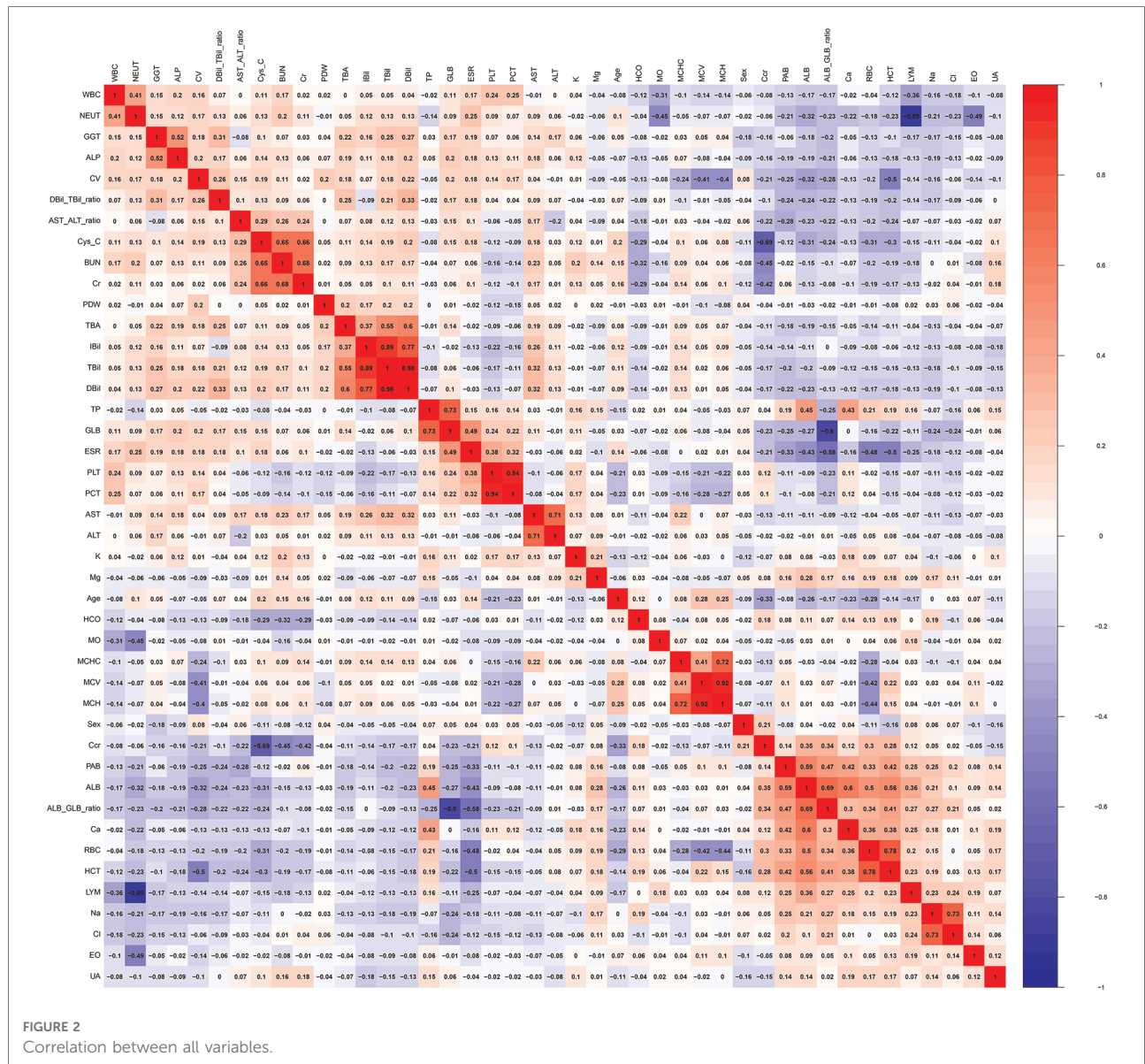
bone and joint TB and patients with TB. In the ESR and blood routine examination (**Table 1**), ESR was significantly increased in patients with TB, but patients with bone and joint TB had higher values than patients with TB. The RBC, HCT, PDW, and LYM of patients with bone and joint TB were lower than those of patients with TB. NEUT and CV of patients with bone and joint TB were higher than those of patients with pulmonary TB.

In the liver function examination (**Table 2**), TBil, DBil, IBil, DBil/IBil, GLB, and AST/ALT in patients with bone and joint TB were higher than those in patients with TB, and ALB, ALB/GLB, and PAB in patients with bone and joint TB were lower than those in patients with TB. In the renal function examination (**Table 2**), UA was lower, and BUN was higher in patients with bone and joint TB than in those with TB. Ccr was lower in patients with bone and joint TB in both training and validation cohorts, but there was no significant difference in the training cohort. In the examination of

plasma electrolytes (**Table 2**),  $\text{Na}^+$ , Cl, and Mg in patients with bone and joint TB were lower than those in patients with TB, and the differences were statistically significant. As can be seen in **Table 3**, the proportion of  $\text{hs-CRP} \geq 10$  was significantly higher in patients with bone and joint TB than in patients with TB.

The heat map (**Figure 1**) shows the distribution of all the variables in the training and validation cohorts. **Figure 2** illustrates the relationship between all the variables. Positive correlations were noted between BUN and Cys-C, Cr and BUN, TBil and IBil, DBil and TBil, DBil and IBil, GLB and TB, PCT and PLT, ALT and AST, MCH and MCHC, MCH and MCV, ALB/GLB and ALB, HCT and RBC, and  $\text{Cl}^-$  and  $\text{Na}^+$ . a clear negative correlation was observed between LYM and NEUT, Ccr and Cys-C, and ALB/GLB and GLB.

We screened the variables by using univariate and multivariate logistic regression, Lasso regression, and random forest to obtain the factors with the most obvious differences



between the two cohorts. **Table 4** presents the results of univariate regression and multivariate regression. **Figure 3A** shows the results of Lasso regression for all the variables. As can be seen from **Figure 3B**, the best effect was achieved when 19 factors were selected to be included in the Lasso regression model; **Table 5** lists the 19 variables included in the Lasso regression results.

**Figure 4A** depicts the 30 most important variables screened by the two calculation methods of %IncMSE and IncNodePurity. The effect was better when 9–18 factors were screened out in the random forest and included in the model (**Figure 4B**). Finally, we used %IncMSE and IncNodePurity to screen out the first 18 important variables and combined them with the factors screened by univariate logistic regression and the 19 factors screened by Lasso

regression to screen out five factors as the final model (**Figure 5**). The final five factors screened were PDW, LYM, AST/ALT, BUN, and Na<sup>+</sup>.

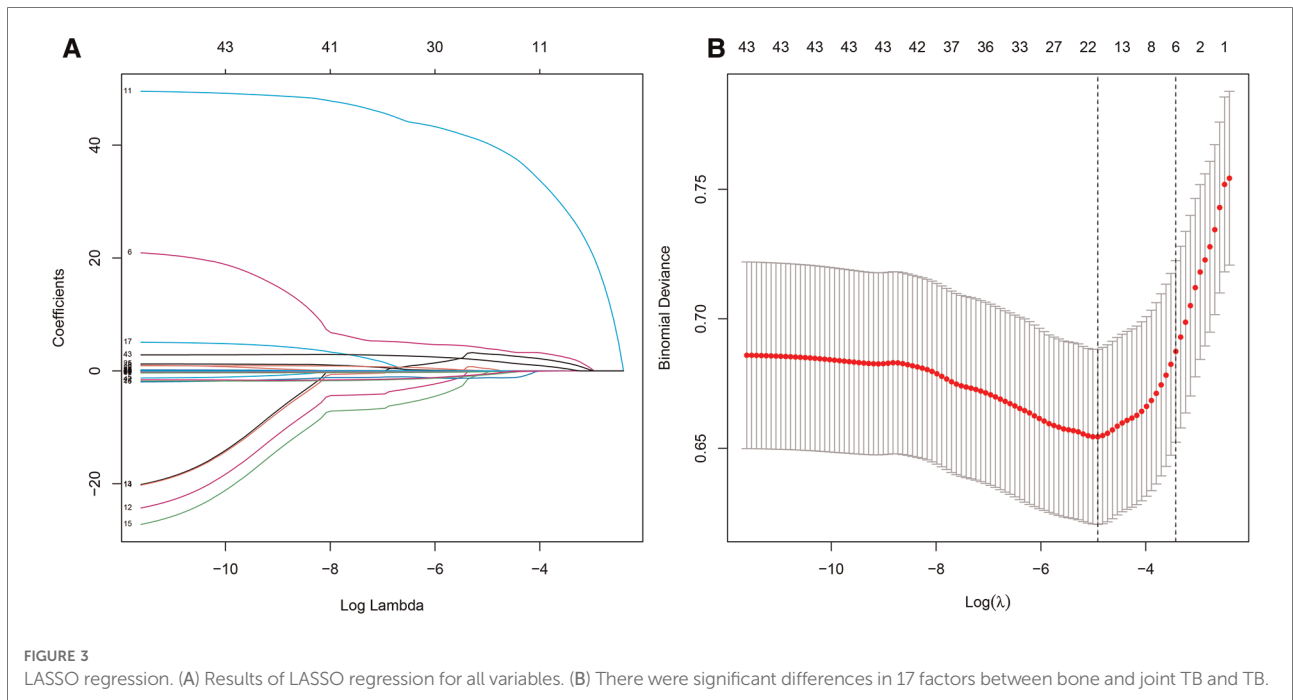
We applied PDW, LYM, AST/ALT, BUN, and Na<sup>+</sup> to construct the nomogram diagnostic model (**Figure 6A**). The nomogram exhibited a good fitting degree (**Figure 6B**). The AUC value of the diagnostic model was 0.71182 (**Figure 6C**). A decision curve was constructed to analyze the clinical utility of the model; the model exhibited clinical utility when the threshold of the model was in the range of 3%–100%, and the NONE line of the decision curve was above the ALL line (**Figure 6E**). The AUC value obtained by placing the diagnostic model in the validation cohort was 0.6435779 (**Figure 7A**), and a good fit was obtained in the validation cohort (**Figure 7B**).



TABLE 4 Univariate logistic regression and multivariate logistic regression predicted bone tuberculosis.

Type	Univariate OR (95% CI)	P-value	Multivariate OR (95% CI)	P-value
Age	0.9965 (0.9873–1.0055)	0.450	/	/
ALB	1.0641 (1.034–1.0947)	<0.001	0.9210 (0.8575–0.9893)	0.0241
ALB/GLB	2.5574 (1.5724–4.2222)	0.0002	2.8507 (0.5749–14.1353)	0.1997
ALP	0.9988 (0.9967–1.0013)	0.3232	/	/
ALT	1.0050 (0.9987–1.0148)	0.2437	/	/
AST	0.9999 (0.9967–1.0052)	0.9736	/	/
AST/ALT	0.7277 (0.6246–0.8488)	<0.001	0.8341 (0.6887–1.0103)	0.0636
BUN	0.9595 (0.9269–0.9964)	0.0227	0.9698 (0.9182–1.0243)	0.2719
Ca	2.3550 (0.8407–6.5216)	0.1009	/	/
Ccr	1.0054 (0.9998–1.0858)	0.0607	0.9999 (0.9925–1.0074)	0.9789
Cl	1.0482 (1.0115–1.0858)	0.0092	0.97789 (0.9209–1.0384)	0.4656
Cr	0.9988 (0.9970–1.0010)	0.2234	/	/
CV	0.0010 (0.00001–0.1244)	0.0040	0.2172 (0.0002–282.726)	0.6765
Cys_C	0.8789 (0.7251–0.9917)	0.2074	/	/
DBil	0.9793 (0.9662–0.9917)	0.0015	0.8154 (0.5165–1.2874)	0.3811
DBil/TBil	0.1789 (0.0630–0.5114)	0.0013	0.3233 (0.0610–0.1847)	0.1847
EO	5.0283 (0.1035–483.49)	0.4545	/	/
ESR	0.9932 (0.9880–0.9984)	0.0097	1.0036 (0.9952–1.0120)	0.4051
GGT	1.0004 (0.9982–1.0031)	0.7181	/	/
GLB	0.9788 (0.9588–0.9999)	0.0450	1.0207 (0.9666–1.0779)	0.4603
HCO	0.9993 (0.9565–1.0433)	0.9756	/	/
HCT	1875.70 (124.71 –29,129.04)	<0.001	5447.49 (10.1422–292,592.7)	0.0073
IBil	0.9581 (0.9327–0.9838)	0.0014	0.7664 (0.4870–0.2502)	0.2502
K	0.9397 (0.6709–1.3260)	0.7206	/	/
LYM	237.97 (35.73–1725.94)	<0.001	164.9533 (1.2679–0.0398)	0.0398
MCH	1.0275 (0.9852–1.0719)	0.2102	/	/
MCHC	1.0024 (0.9933–1.0140)	0.6562	/	/
MCV	1.0113 (0.9943–1.0283)	0.1886	/	/
Mg	28.9297 (7.4546–116.75)	<0.001	13.7496 (2.8109–67.256)	0.0012
MO	90.2245 (0.9762–10,355.38)	0.0574	52.8651 (0.0973–28,721.44)	0.2169
Na <sup>+</sup>	1.0896 (1.0467–1.1340)	<0.001	1.0232 (0.9545–1.0968)	0.5176
NEUT	0.0255 (0.0063–0.0993)	<0.001	4.6243 (0.0962–222.31)	0.4383
PAB	1.0066 (1.0040–1.0093)	<0.001	1.0025 (0.9989–1.0061)	0.1715
PCT	0.16798 (0.0337–0.8800)	0.0318	145.246 (0.2563–82,313.954)	0.1238
PDW	1.054 × 10 <sup>19</sup> (2.7267 × 10 <sup>14</sup> –1.4052 × 10 <sup>24</sup> )	<0.001	4.3985 × 10 <sup>21</sup> (5.4315 × 10 <sup>15</sup> –3.5619 × 10 <sup>27</sup> )	<0.001
PLT	0.9990 (0.9978–1.0001)	0.0771	0.9961 (0.9917–1.005)	0.0855
RBC	1.5783 (1.2730–1.9625)	<0.001	0.7185 (0.4649–1.1102)	0.1364
Sex	1.0241 (0.7196–1.4749)	0.8965	/	/
TBA	0.9963 (0.9910–1.0026)	0.2010	/	/
TBil	0.9851 (0.9758–0.9938)	0.0011	1.2349 (0.7852–1.9421)	0.36108
TP	1.0155 (0.9947–1.0369)	0.1475	/	/
UA	1.0013 (1.0003–1.0023)	0.0099	1.0012 (1.0001–1.0023)	0.0397
WBC	0.9774 (0.9495–1.0089)	0.1292	/	/

The red text means that the *p* value was <0.05. ESR, erythrocyte sedimentation rate; WBC, White blood cell count; RBC, Red blood cell count; HCT, hematocrit value; MCV, Mean corpuscular volume; MCH, Mean corpuscular hemoglobin; MCHC, Mean corpuscular hemoglobin concentration; PLT, Platelet hematocrit; PDW, Platelet distribution width; NEUT, neutrophil percentage; LYM, lymphocytes percentage; MO, Monocyte percentage; EO, eosinophils percentage; CV, RBC volume distributing width; PCT, thrombocytocrit; TBil, total bilirubin; DBil, direct bilirubin; IBil, indirect bilirubin; TP, total protein; ALB, albumin; GLB, globulin; GGT, gamma-glutamyl transpeptidase; TBA, Total bile acid; AST, aspartate aminotransferase; ALT, alanine aminotransferase; ALP, A alkaline phosphatase; PAB, prealbumin; ChE, cholinesterase; BUN, Blood urea nitrogen; Cr, creatinine; UA, Uric acid; HCO, bicarbonate radical; Ccr, Creatinine clearance rate; Cys-C, Cysteine C.

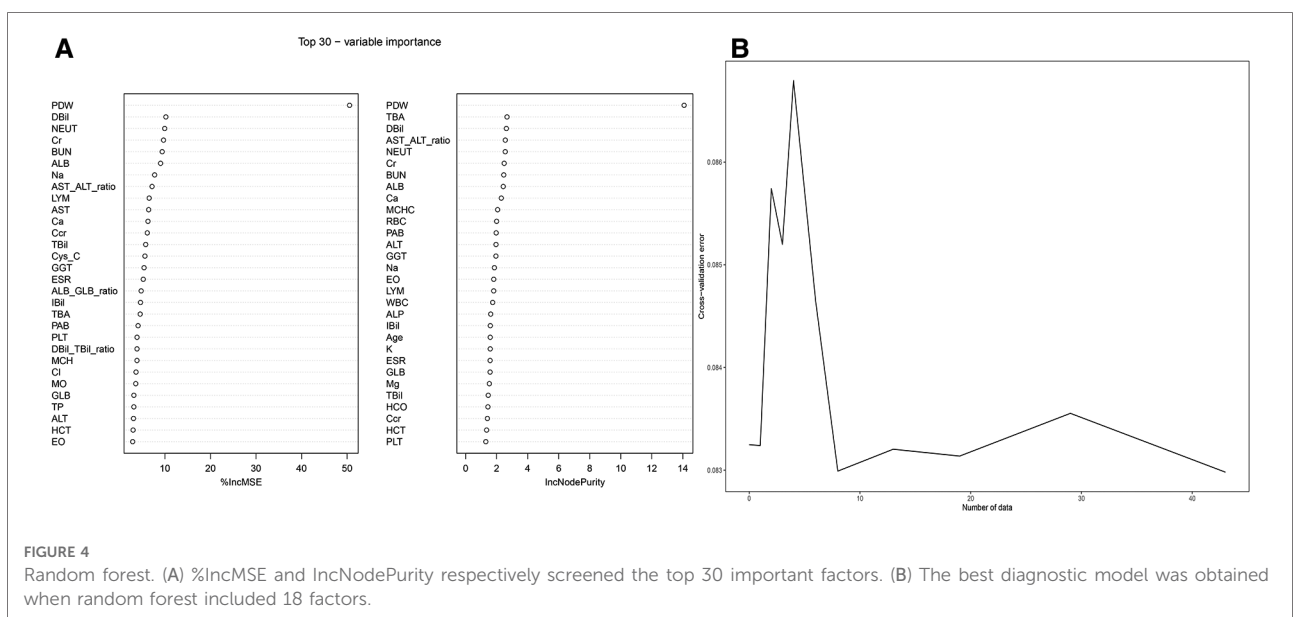


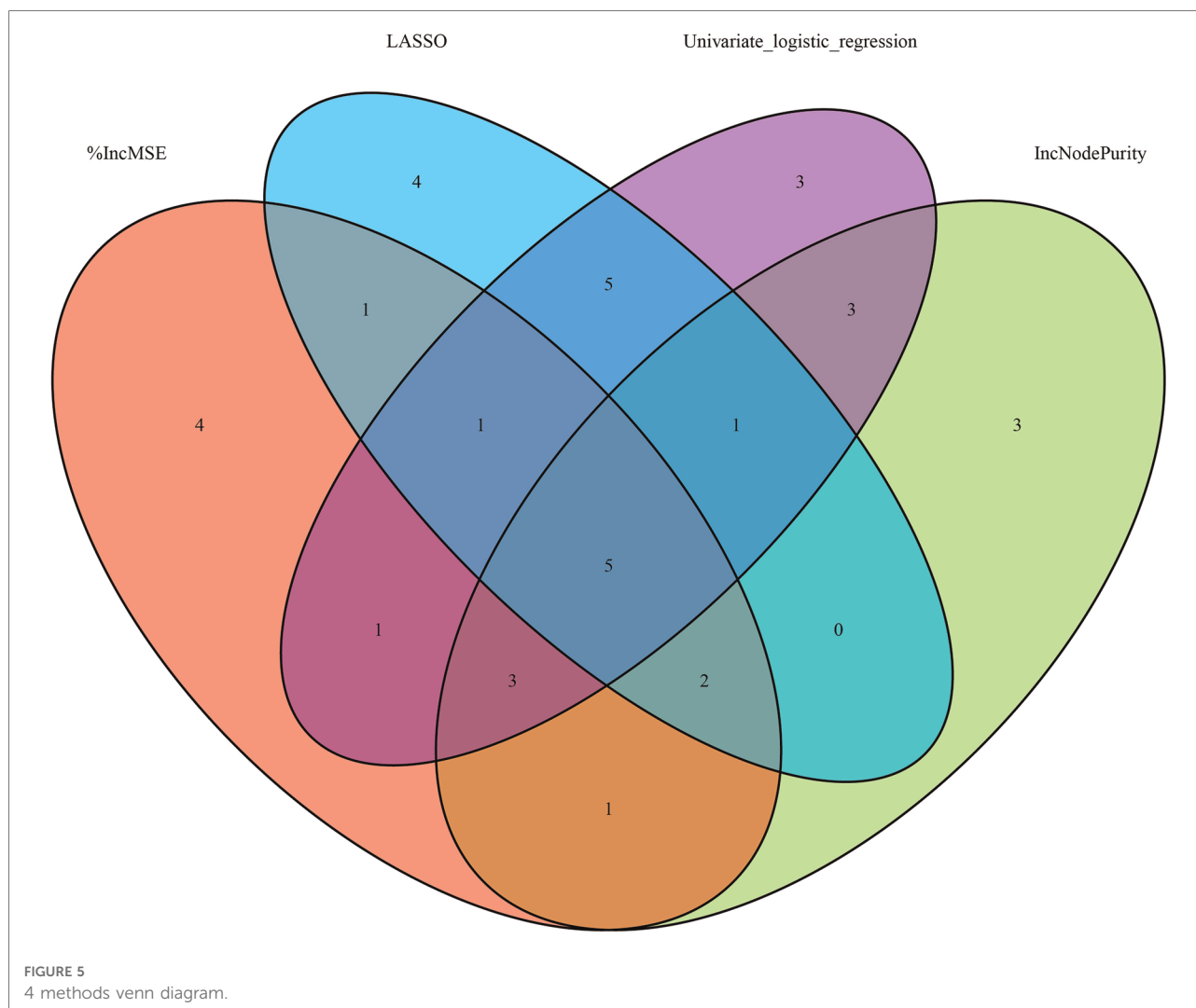
## Discussion

The aim of this study was to explore the differences between patients with bone and joint TB and patients with TB to screen out the most representative factors through the application of ML in the clinical data of the two groups. It is hoped that in the future, it will help clinicians to diagnose bone and joint TB earlier and more easily, and help patients get earlier treatment (23).

**TABLE 5** Nineteen variables screened by lasso regression.

PDW	LYM	MO	EO	CV
IBil	DBil/TBil	GGT	AST/ALT	PAB
BUN	UA	HCO	K	Na <sup>+</sup>
Ca	Mg			

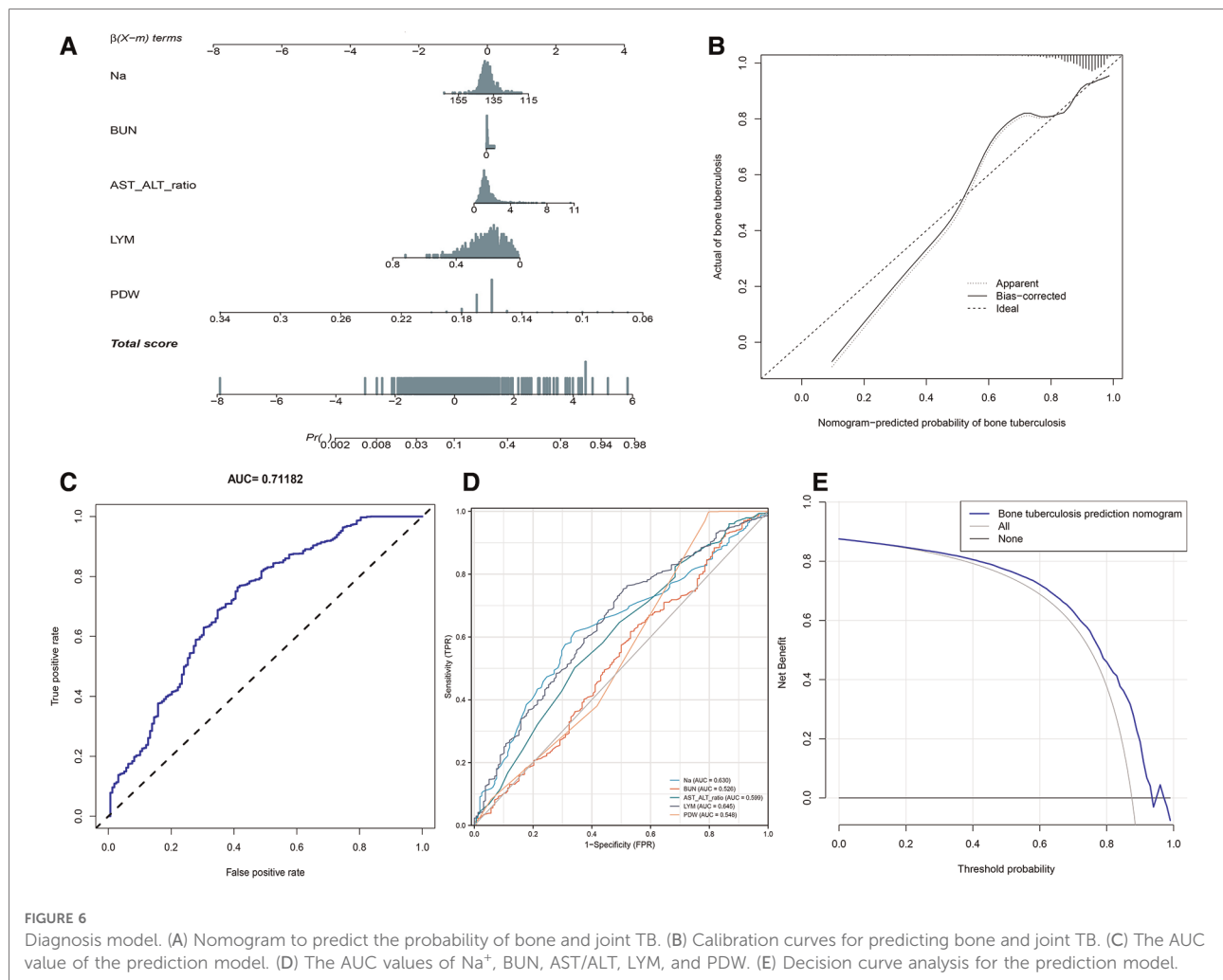




TB is a chronic inflammatory reaction caused by *Mycobacterium tuberculosis*. The specific response of humans to *Mycobacterium tuberculosis* is called “tuberculosis granuloma.” It is a nodular mass that mixes inflammatory cells and immune cells in the local collection (5). “Granulomatous inflammation” is used to describe a mononuclear infiltration of macrophages and lymphocytes (24). Tissue injury and inflammatory stimulation cause elevated hs-CRP and ESR levels. hs-CRP and ESR can be used as activity indicators to monitor diseases (25, 26). As can be seen in **Table 1**, the average values of these two factors are much greater in patients with TB than in normal controls. This is the same as observed in the study by Li et al. (27). WBC and NEUT are commonly used as indicators of bacterial infection in clinical practice (28). The mean values of hs-CRP, ESR, WBC, and NEUT in patients with TB were lower than those in patients with bone and joint TB, indicating that patients with bone and joint TB were more seriously infected.

PDW is a common indicator in routine blood examination; hemodilution or some blood-related diseases can lead to a decrease in PDW. In the PDW control study of TB and normal subjects by Xu et al., the mean PDW of TB subjects was lower than that of normal subjects (29). LYM is a type of WBC, originally derived from bone marrow hematopoietic stem cells, which differentiate into cells with different functions (30). Lymphocytes play an important role in adaptive immunity, enabling the immune system to recognize and remember antigens. LYM reduction is common in infections, TB, immune system diseases, and the use of hormones (31). LYM in patients with TB is lower than that in normal subjects (32).

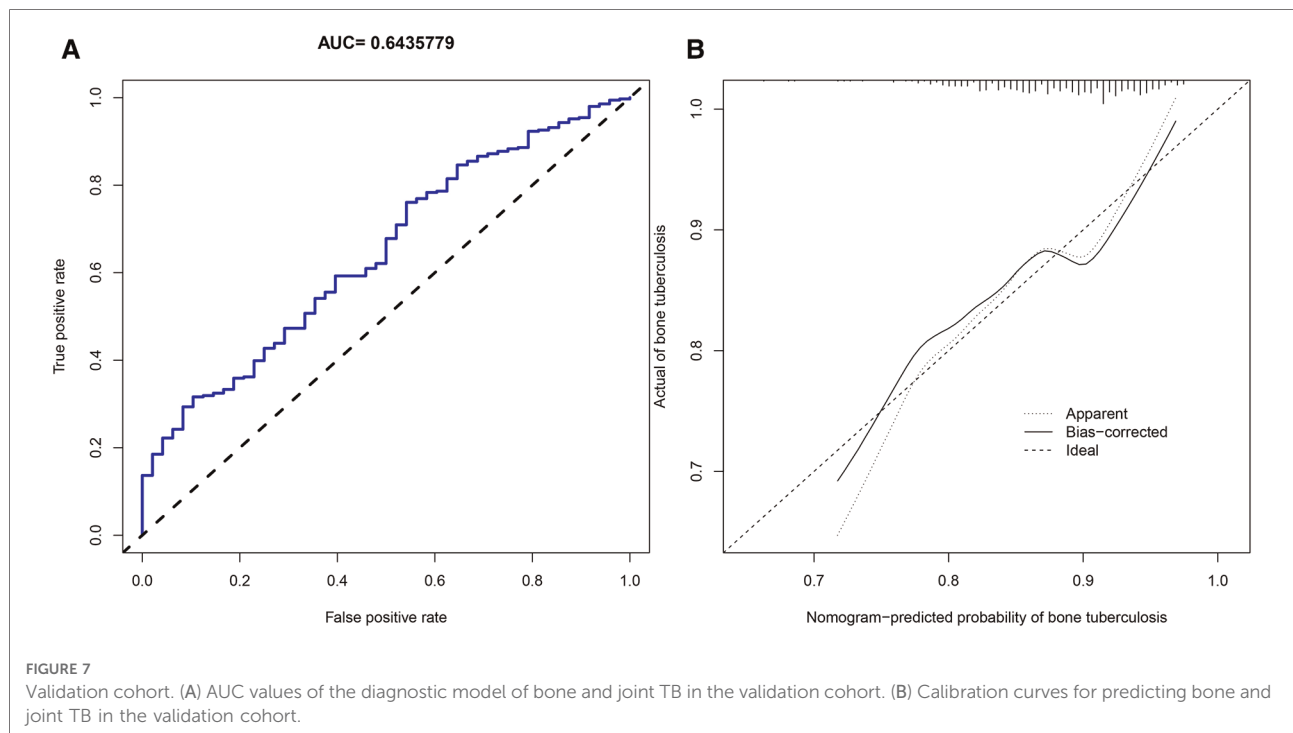
We found that the mean PDW and LYM values of patients with bone and joint TB on routine blood tests were significantly lower than those of patients with TB. The lower mean value of LYM in patients with bone and joint TB indicates that such patients have lower immune capacity and



are at higher levels of inflammation (30). This issue is also illustrated by the inverse relationship between LYM and NEUT inflammatory markers in Figure 2. Lymphocytes are divided into T lymphocytes, B lymphocytes, and NK cells. Different lymphocytes have different characteristics. Lymphocytopenia in patients with bone and joint TB is more obvious and needs further study. PLT was elevated in patients with bone and joint TB (Table 1), and the average value was higher than normal. PLT is an inflammatory marker and is elevated in patients with TB (33). PLT and PDW exhibited a negative correlation (Figure 2). We believe that the relative increase in PLT in patients with bone and joint TB leads to a decrease in PDW.

Bilirubin is a pigment made from hemoglobin in red blood cells. When hemoglobin dies, it breaks down to biliverdin, which in response to NADPH and H ions generates bilirubin (34). In the liver function test, TBil, DBil, and IBil of patients with bone and joint TB were higher than those of patients with TB, indicating a correlation between aging erythrocytes in patients with bone and joint TB. The decrease in RBC in

patients with bone and joint TB proved our view (Table 1). However, AST/ALT was increased in patients with bone and joint TB, mainly due to the decrease in ALT. ALT is a sensitive indicator of liver damage (35). The higher average value of ALT in patients with TB indicates that patients with pulmonary TB may have a greater liver function burden. Antituberculous drugs also affect liver function, and the reason for the increase in ALT needs to be further investigated. ALB is commonly used as a measure of nutritional status, and TB is a wasting disease, which is confirmed by the fact that ALB was lower than normal in both groups (36). The average ALB of patients with bone and joint TB was lower than that in patients with TB, which may be related to the longer and more severe infection, leading to more decline in nutritional status. The increase in GLB is usually due to the invasion of viruses and bacteria. The immune system produces increased GLB against viruses and bacteria, which also indicates that patients with bone and joint TB are more severely attacked by *Mycobacterium tuberculosis* (37).



Furthermore, BUN increased, and UA decreased in patients with bone and joint TB.

BUN and Cr are both waste products of human metabolism. The kidney is also the only organ that clears Cys-C. In clinical practice, the increase of BUN, Cys-C and Cr is often used to indicate abnormal renal function, so there is a positive correlation between them (38). BUN is the product of the catabolism of human protein; 90% of BUN needs to be excreted by the kidney (39). Protein breakdown in patients with bone and joint TB is greater than that in patients with TB, and the nutrient consumption is greater. UA content is usually proportional to the intake of purine content, and there was no difference in the diet between the two groups; Thus, only the endogenous reduction of purine production in UA can be considered. Currently, it is not known how UA works, and further research is required. In the examination of plasma electrolytes,  $\text{Na}^+$ ,  $\text{Cl}^-$ , and  $\text{Mg}^{2+}$  were slightly decreased in patients with bone and joint TB, but there is no clear implication in clinical practice.

We used univariate logistic regression, LASSO regression, and random forest to screen the clinical data of 206 patients with bone and joint TB and 1,461 patients with TB to identify the five most variable factors. Finally, we constructed the nomogram diagnostic model by using five factors, namely PDW, LYM, AST/ALT, BUN, and  $\text{Na}^+$ , and achieved relatively good efficiency. The proposed diagnostic model will help clinicians more easily diagnose patients with bone and joint TB, prevent the progression of TB in time, and give patients more timely treatment. Especially in the face of patients with atypical bone and joint TB, many

doctors are afraid to use anti-TB drugs without clear pathological results. Our diagnostic model can help doctors to a certain extent. Of course, a more accurate diagnosis of bone and joint TB still requires the imaging and pathological results of TB and the experience of clinicians.

Few blood investigation studies on patients with bone and joint TB have been conducted. We used ML methods to screen patients' ESR, blood routine, liver and kidney function, and plasma electrolytes, which is a novel feature of our study. We used ML to process big data, screen out specific factors, and construct diagnostic models. However, our study also has some limitations. For example, the data used in the study belonged to a single hospital. The use of multicenter data will provide more convincing results. In addition, cases of bone and joint TB are relatively rare. The number gap between bone and joint TB and TB patients is also large, which can also cause errors in the accuracy of the model. The diagnostic performance of the nomogram is not good enough. This study is only a preliminary exploration of the blood characteristics of patients with bone and joint TB, and there are many directions for further research.

## Conclusion

We used ML to screen out the blood-specific factors PDW, LYM, AST/ALT, BUN, and  $\text{Na}^+$  of bone and joint TB and constructed a diagnostic model to help clinicians better diagnose the disease in the future.



## 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 author/s.

## Ethics statement

All the patients in the study volunteered and signed informed consent. Patients younger than 18 years old shall be signed by the patient's guardian. All subjects had their names and ID numbers hidden to protect data confidentiality. This study was approved by the Ethics Committee of the First Affiliated Hospital of Guangxi Medical University.

## Author contributions

All the authors contributed to the research. ZY, JZ, CL, XZ and SW contributed to the design of the study and the writing and revision of the paper. QL, XS, TL, JJ, HL, TC and JC participated in data collection for the study. GD, SL, CY, LC, CZ and HG were mainly responsible for data collation and data analysis. YY, WC, WJ and BF were responsible for data visualization. XT, ZY, WG and YW were responsible for data verification. All authors contributed to the article and approved the submitted version.

## Funding

This work was sponsored by the National Natural Science Foundation of China (81560359 and 81860393). Funding

## References

1. Sia JK, Rengarajan J. Immunology of Mycobacterium tuberculosis infections. *Microbiol Spectr.* (2019) 7(4). doi: 10.1128/microbiolspec.GPP3-0022-2018
2. Li XX, Wang LX, Zhang H, Jiang SW, Fang Q, Chen JX, et al. Spatial variations of pulmonary tuberculosis prevalence co-impacted by socio-economic and geographic factors in People's Republic of China, 2010. *BMC Public Health.* (2014) 14:257. doi: 10.1186/1471-2458-14-257
3. MacNeil A, Glaziou P, Sismanidis C, Date A, Maloney S, Floyd K. Global epidemiology of Tuberculosis and progress toward meeting global targets—worldwide, 2018. *MMWR Morb Mortal Wkly Rep.* (2020) 69(11):281–5. doi: 10.15585/mmwr.mm6911a2
4. Suárez I, Fünßer SM, Kröger S, Rademacher J, Fätkenheuer G, Rybniker J. The diagnosis and treatment of Tuberculosis. *Dtsch Arztebl Int.* (2019) 116(43):729–35. doi: 10.3238/arztebl.2019.0729
5. Basaraba RJ, Hunter RL. Pathology of tuberculosis: how the pathology of human tuberculosis informs and directs animal models. *Microbiol Spectr.* (2017) 5(3). doi: 10.1128/microbiolspec.TBTB2-0029-2016
6. Johansen IS, Nielsen SL, Hove M, Kehrer M, Shakar S, Wøyen AV, et al. Characteristics and clinical outcome of bone and joint Tuberculosis from 1994

bodies had no role in the study design, collection, analysis, and interpretation of the data or in writing the manuscript.

## Acknowledgments

We are grateful to Dr. Xinli Zhan and Dr. Chong Liu (Spine and Osteopathy Ward, The First Affiliated Hospital of Guangxi Medical University) for his kindly assistance in all stages of the present study.

## 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.

## Supplementary material

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fsurg.2022.1031105/full#supplementary-material>.

7. Tuli SM. General principles of osteoarticular tuberculosis. *Clin Orthop Relat Res.* (2002) 398:11–9. doi: 10.1097/00003086-200205000-00003
8. Deo RC. Machine learning in medicine. *Circulation.* (2015) 132(20):1920–30. doi: 10.1161/CIRCULATIONAHA.115.001593
9. Li W, Liu Y, Liu W, Tang ZR, Dong S, Li W, et al. Machine learning-based prediction of lymph node metastasis among osteosarcoma patients. *Front Oncol.* (2022) 12:797103. doi: 10.3389/fonc.2022.797103
10. Kim TK. T test as a parametric statistic. *Korean J Anesthesiol.* (2015) 68(6):540–6. doi: 10.4097/kjae.2015.68.6.540
11. Rochon J, Gondan M, Kieser M. To test or not to test: preliminary assessment of normality when comparing two independent samples. *BMC Med Res Methodol.* (2012) 12:81. doi: 10.1186/1471-2288-12-81
12. McHugh ML. The chi-square test of independence. *Biochem Med (Zagreb).* (2013) 23(2):143–9. doi: 10.11613/BM.2013.018
13. Guyatt G, Jaeschke R, Heddle N, Cook D, Shannon H, Walter S. Basic statistics for clinicians: 1. Hypothesis testing. *CMAJ.* (1995) 152(1):27–32.

14. Li W, Wang H, Dong S, Tang ZR, Chen L, Cai X, et al. Establishment and validation of a nomogram and web calculator for the risk of new vertebral compression fractures and cement leakage after percutaneous vertebroplasty in patients with osteoporotic vertebral compression fractures. *Eur Spine J.* (2022) 31(5):1108–21. doi: 10.1007/s00586-021-07064-z
15. Guo C, Gao YY, Ju QQ, Zhang CX, Gong M, Li ZL. The landscape of gene co-expression modules correlating with prognostic genetic abnormalities in AML. *J Transl Med.* (2021) 19(1):228. doi: 10.1186/s12967-021-02914-z
16. Pavey TG, Gilson ND, Gomersall SR, Clark B, Trost SG. Field evaluation of a random forest activity classifier for wrist-worn accelerometer data. *J Sci Med Sport.* (2017) 20(1):75–80. doi: 10.1016/j.jsams.2016.06.003
17. Janssens A, Martens FK. Reflection on modern methods: revisiting the area under the ROC curve. *Int J Epidemiol.* (2020) 49(4):1397–403. doi: 10.1093/ije/dyz274
18. Zhang M, Wang X, Chen X, Zhang Q, Hong J. Novel immune-related gene signature for risk stratification and prognosis of survival in lower-grade glioma. *Front Genet.* (2020) 11:363. doi: 10.3389/fgene.2020.00363
19. Li W, Dong S, Wang B, Wang H, Xu C, Zhang K, et al. The construction and development of a clinical prediction model to assess lymph node metastases in osteosarcoma. *Front Public Health.* (2021) 9:813625. doi: 10.3389/fpubh.2021.813625
20. Li L, Prato CG, Wang Y. Ranking contributors to traffic crashes on mountainous freeways from an incomplete dataset: a sequential approach of multivariate imputation by chained equations and random forest classifier. *Accid Anal Prev.* (2020) 146:105744. doi: 10.1016/j.aap.2020.105744
21. Li W, Wang J, Liu W, Xu C, Li W, Zhang K, et al. Machine learning applications for the prediction of bone cement leakage in percutaneous vertebroplasty. *Front Public Health.* (2021) 9:812023. doi: 10.3389/fpubh.2021.812023
22. Zhang H, Zhang X, Hua L, Li L, Tian L, Zhang X, et al. An exploratory analysis of forme fruste keratoconus sensitivity diagnostic parameters. *Int Ophthalmol.* (2022) 42(8):2473–81. doi: 10.1007/s10792-022-02246-0
23. Li W, Liu W, Hussain Memon F, Wang B, Xu C, Dong S, et al. An external-validated prediction model to predict lung metastasis among osteosarcoma: a multicenter analysis based on machine learning. *Comput Intell Neurosci.* (2022) 2022:2220527. doi: 10.1155/2022/2220527
24. Wang X, Liu Y. Offense and defense in granulomatous inflammation disease. *Front Cell Infect Microbiol.* (2022) 12:797749. doi: 10.3389/fcimb.2022.797749
25. Stikbakke E, Richardsen E, Knutsen T, Wilsgaard T, Giovannucci EL, McTiernan A, et al. Inflammatory serum markers and risk and severity of prostate cancer: the PROCA-life study. *Int J Cancer.* (2020) 147(1):84–92. doi: 10.1002/ijc.32718
26. Lapić I, Padoan A, Bozzato D, Plebani M. Erythrocyte sedimentation rate and C-reactive protein in acute inflammation. *Am J Clin Pathol.* (2020) 153(1):14–29. doi: 10.1093/ajcp/qaq142
27. Li W, Liu Z, Xiao X, Xu Z, Sun Z, Zhang Z, et al. Early surgical intervention for active thoracic spinal tuberculosis patients with paraparesis and paraplegia. *BMC Musculoskelet Disord.* (2021) 22(1):213. doi: 10.1186/s12891-021-04078-y
28. Zheng S, Wang Z, Qin S, Chen JT. Usefulness of inflammatory markers and clinical manifestation for an earlier method to diagnosis surgical site infection after spinal surgery. *Int Orthop.* (2020) 44(11):2211–9. doi: 10.1007/s00264-020-04567-0
29. Xu F, Qu S, Wang L, Qin Y. Mean platelet volume (MPV): new diagnostic indices for co-morbidity of tuberculosis and diabetes mellitus. *BMC Infect Dis.* (2021) 21(1):461. doi: 10.1186/s12879-021-06152-1
30. Almezghwi K, Serte S. Improved classification of white blood cells with the generative adversarial network and deep convolutional neural network. *Comput Intell Neurosci.* (2020) 2020:6490479. doi: 10.1155/2020/6490479
31. Adamstein NH, Ridker PM. The neutrophil-lymphocyte ratio: considerations for clinical application. *Eur Heart J.* (2021) 42(22):2216–7. doi: 10.1093/eurheartj/ehab166
32. Iliaz S, Iliaz R, Ortakoylu G, Bahadir A, Bagci BA, Caglar E. Value of neutrophil/lymphocyte ratio in the differential diagnosis of sarcoidosis and tuberculosis. *Ann Thorac Med.* (2014) 9(4):232–5. doi: 10.4103/1817-1737.140135
33. Zhao Y, Yin L, Patel J, Tang L, Huang Y. The inflammatory markers of multisystem inflammatory syndrome in children (MIS-C) and adolescents associated with COVID-19: a meta-analysis. *J Med Virol.* (2021) 93(7):4358–69. doi: 10.1002/jmv.26951
34. Shibahara S, Kitamura T, Takahashi K. Heme degradation and human disease: diversity is the soul of life. *Antioxid Redox Signaling.* (2002) 4(4):593–602. doi: 10.1089/15230860260220094
35. Ladue JS, Wroblewski F. Serum glutamic pyruvic transaminase SGP-T in hepatic disease: a preliminary report. *Ann Intern Med.* (1956) 45(5):801–11. doi: 10.7326/0003-4819-45-5-801
36. Mendez CM, McClain CJ, Marsano LS. Albumin therapy in clinical practice. *Nutr Clin Pract.* (2005) 20(3):314–20. doi: 10.1177/0115426505020003314
37. Ballow M. Mechanisms of action of intravenous immune serum globulin in autoimmune and inflammatory diseases. *J Allergy Clin Immunol.* (1997) 100(2):151–7. doi: 10.1016/S0091-6749(97)70217-3
38. Gao F, Qian M, Liu G, Ao W, Dai D, Yin C. USP10 Alleviates sepsis-induced acute kidney injury by regulating Sirt6-mediated Nrf2/ARE signaling pathway. *J Inflamm (Lond).* (2021) 18(1):25. doi: 10.1186/s12950-021-00291-7
39. Weiner ID, Mitch WE, Sands JM. Urea and ammonia metabolism and the control of renal nitrogen excretion. *Clin J Am Soc Nephrol.* (2015) 10(8):1444–58. doi: 10.2215/CJN.10311013