Heliyon 10 (2024) e24189

Contents lists available at ScienceDirect

Heliyon

journal homepage: www.cell.com/heliyon

Research article

Construction of a novel lower-extremity peripheral artery disease subtype prediction model using unsupervised machine learning and neutrophil-related biomarkers

Lin Zhang [a], Yuanliang Ma [a], Que Li [a], Zhen Long [a], Jiangfeng Zhang [a], Zhanman Zhang [a], Xiao Qin [a,*]

[a] *The First Affiliated Hospital of Guangxi Medical University, Nanning, Guangxi, PR China*

ARTICLE INFO

Keywords:
Machine learning
Predictive model
Lower-extremity peripheral artery disease
Immune cell infiltration
Neutrophils

ABSTRACT

Lower-extremity peripheral artery disease (LE-PAD) is a prevalent circulatory disorder with risks of critical limb ischemia and amputation. This study aimed to develop a prediction model for a novel LE-PAD subtype to predict the severity of the disease and guide personalized interventions. Additionally, LE-PAD pathogenesis involves altered immune microenvironment, we examined the immune differences to elucidate LE-PAD pathogenesis. A total of 460 patients with LE-PAD were enrolled and clustered using unsupervised machine learning algorithms (UMLAs). Logistic regression analyses were performed to screen and identify predictive factors for the novel subtype of LE-PAD and a prediction model was built. We performed a comparative analysis regarding neutrophil levels in different subgroups of patients and an immune cell infiltration analysis to explore the associations between neutrophil levels and LE-PAD. Through hematoxylin and eosin (H&E) staining of lower-extremity arteries, neutrophil infiltration in patients with and without LE-PAD was compared. We found that UMLAs can helped in constructing a prediction model for patients with novel LE-PAD subtypes which enabled risk stratification for patients with LE-PAD using routinely available clinical data to assist clinical decision-making and improve personalized management for patients with LE-PAD. Additionally, the results indicated the critical role of neutrophil infiltration in LE-PAD pathogenesis.

1. Introduction

Lower-extremity peripheral artery disease (LE-PAD) is a prevalent vascular condition characterized by the narrowing of arteries in the lower extremities. The progressive narrowing of the arteries in the legs reduces the blood flow and oxygen delivery to the lower limbs, which impairs muscle function and overall quality of life [1]. The prevalence of LE-PAD among the population aged $\geq$70 years is approximately 15%–20 % in Western countries, whereas, in China, LE-PAD shows a high incidence rate of 15.91 % [2,3]. In certain instances, individuals afflicted with LE-PAD manifest pronounced limb ischemia and tissue necrosis attributable to the absence of systematic follow-up care. This deficiency culminates in the imperative requirement for limb amputation and, in certain scenarios, leads to mortality, primarily caused by sepsis, multiple organ failure, and assorted complications [4].

* Corresponding author. Department of Vascular Surgery Ward, The First Affiliated Hospital of Guangxi Medical University, No.6 of Shuangyong Road, Nanning, Guangxi, 530021, PR China

E-mail address: dr_qinxiao@hotmail.com (X. Qin).

https://doi.org/10.1016/j.heliyon.2024.e24189
Received 31 August 2023; Received in revised form 20 November 2023; Accepted 4 January 2024
Available online 6 January 2024

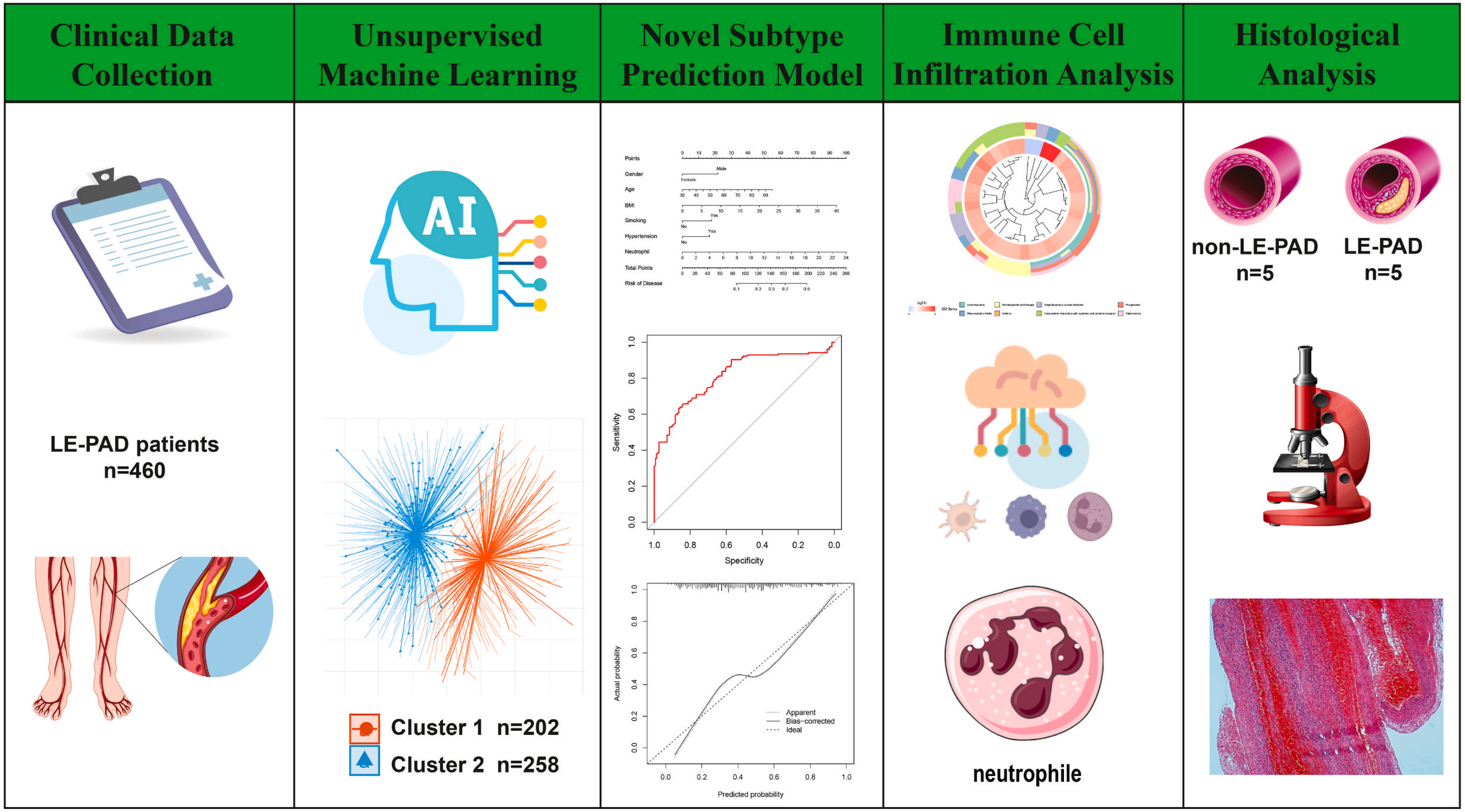


Fig. 1. Graphical abstract of this study. LE-PAD: lower-extremity peripheral artery disease.

With the advancements in artificial intelligence (AI), unsupervised machine learning is being applied in disease diagnosis, classification, and treatment. Unsupervised machine learning algorithms (UMLAs) are employed to analyze and cluster unlabeled datasets, which facilitates discovering unidentified patterns or data groupings without human intervention. UMLAs can be used to cluster patients based on their disease characteristics and effectively classify a heterogeneous cohort with accuracy and rationality [5]. Kobayashi et al. identified three echocardiographic phenotypes in the STANISLAS cohort using k-means clustering, demonstrating that echocardiographic data-based data-driven classification can help identify profiles with varying long-term heart failure risks in asymptomatic middle-aged individuals [6]. However, the application of UMLAs in clinical studies on LE-PAD is lacking.

In recent years, some studies highlighted the value of inflammation as an important element in the development, progression, and prognosis of atherosclerosis [7,8]. A study revealed an extreme diversity of human neutrophils *in vivo* and novel organized cellular functions, such as neutrophil extracellular trap (NET) generation and inflammasome activation [9]. These functions are associated with LE-PAD [10], and these inflammatory markers are useful in clinical studies on the risk stratification and prognosis of patients with atherosclerosis [11,12].

Herein, we used UMLAs to investigate heterogeneity among patients with LE-PAD. The clinical data were used to construct a novel LE-PAD subtype predictive model based on two clusters identified by UMLAs. Furthermore, an examination of the correlation between LE-PAD and neutrophil levels was conducted by scrutinizing disparities in immune cell infiltration between samples sourced from individuals with LE-PAD and those from healthy controls. This analysis was corroborated through histological examination. The outcomes demonstrated the precise classification of patients with LE-PAD into severe and mild categories using UMLAs, grounded in their clinical data. The developed LE-PAD subtype prediction model showed good accuracy. The results of this study suggested an association between LE-PAD occurrence/progression and neutrophil levels. The graphical abstract of this study is illustrated in Fig. 1.

2. Materials and methods

2.1. Patients

We retrospectively analyzed the clinical data of patients with lower-extremity arteriosclerosis admitted to Guangxi Medical University between January 2014 and January 2023. All individuals satisfying the diagnostic criteria for LE-PAD as per the directives outlined by the Vascular Surgery Branch of the Surgery Society of Chinese Medical Association [13] were included. Exclusion criteria were as follows: patients with (a) acute limb ischemia; (b) acute infections, autoimmune diseases, or other inflammatory conditions; (c) malignant tumors or congenital vascular diseases; and (d) incomplete case data, insufficient follow-up time, or lost to follow-up. Based on the inclusion and exclusion criteria, 460 eligible patients were enrolled in the study.

The clinical data collected included predictor variables for UMLAs clustering and outcome variables for assessing clustering accuracy. The predictor variables included patients' baseline data, comorbidities, and immune cell data from routine blood tests. The outcome variables included the follow-up data, ankle–brachial index (ABI), and Rutherford classification of patients. ABI is a simple, noninvasive measurement used to assess LE-PAD by comparing the blood pressure in the ankles to that in the arms. A normal ABI ranges between 0.9 and 1.3, and an ABI of <0.9 indicates LE-PAD, with lower values indicating more severe obstruction [14]. The Rutherford classification is a system used to categorize the severity of LE-PAD by correlating clinical symptoms with the severity of LE-PAD, with higher categories indicating more severe obstruction and ischemia [15]. All predictor variables were derived from clinical data obtained during the patients' initial visit. Subsequently, all outcome variables were extracted from the clinical records of patients who revisited the hospital for either re-examination or telephone follow-up within a period ranging from six months to one year. Additionally, between January 2022 and January 2023, lower-limb artery tissue samples were collected from five patients with LE-PAD undergoing amputation because of lower-limb ischemic necrosis. The lower-limb artery tissue samples obtained from five trauma patients without LE-PAD undergoing amputation procedures were denoted as the control group.

This study was conducted following the Declaration of Helsinki and was approved by the Ethics Committee of The First Affiliated Hospital of Guangxi Medical University.

2.2. Data normalization and UMLAs

The development of UMLAs was accomplished using R software version 4.2.1. Normalization of data for patients with LE-PAD was executed employing the "scale" function within the "factoextra" package [16]. The "Fpc" package was used to determine the optimal number of clusters (K-value) by calculating the silhouette coefficient (SC) [17]. Subsequently, patients were classified into clusters using the K-means cluster algorithm. K-means clustering, a popular UMLAs, can effectively group patients on the basis of disease characteristics and accurately classify a heterogeneous cohort [18,19]. Based on the predictor variables, UMLAs categorized patients into two clusters, following which, the differences in the outcome variables between the two clusters were analyzed to verify the accuracy of UMLA clustering.

2.3. Construction of a predictive model for the novel LE-PAD subtype

Univariate and multivariate binary logistic regression analyses were performed using statistically significant predictor variables. The results of logistic regression analyses were compared, and a nomogram model [20] was developed. The performance of the nomogram was assessed via receiver operating characteristic (ROC) curves and C-index calibration [20], and a p-value of <0.05 was considered statistically significant.

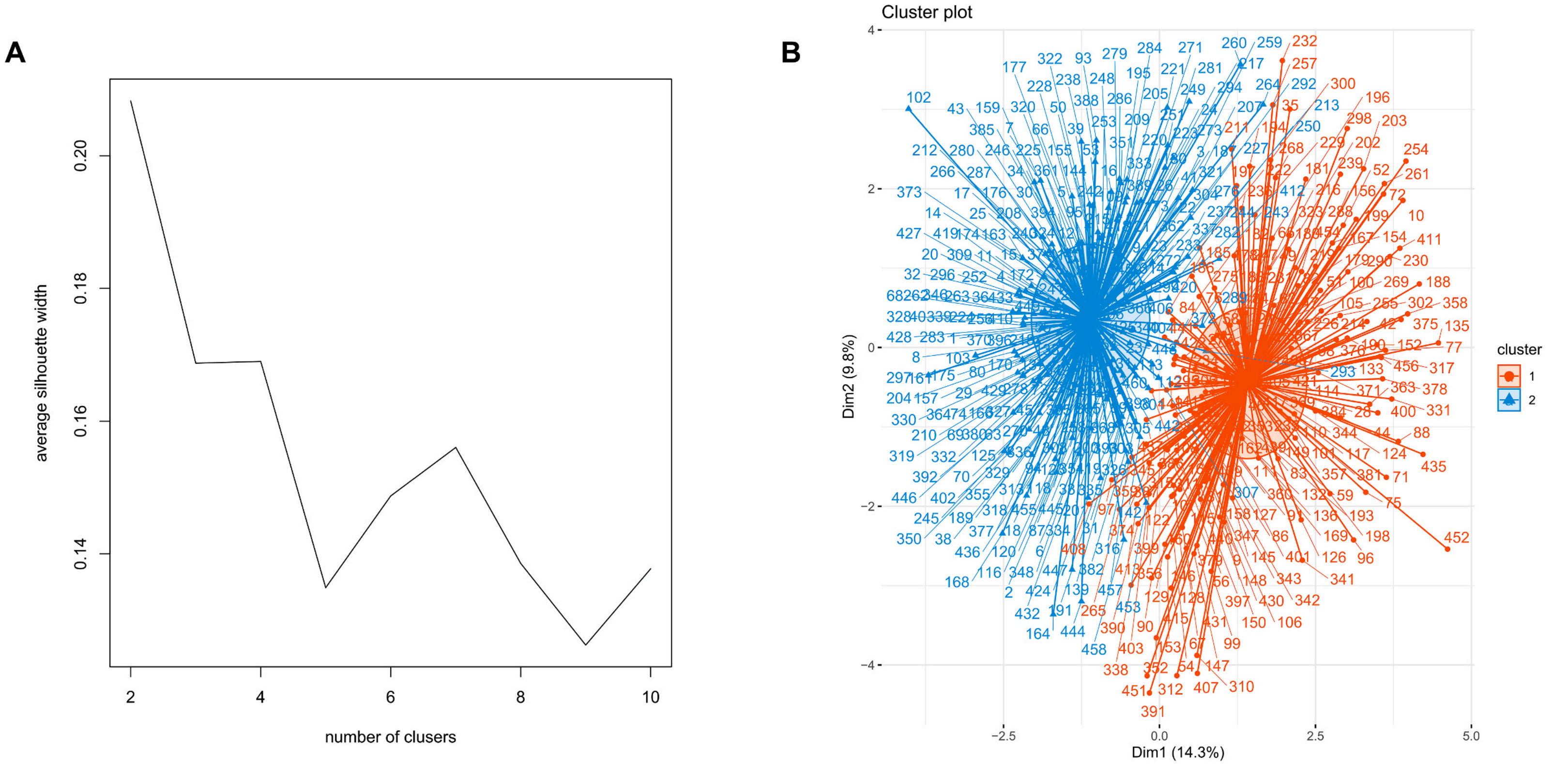


Fig. 2. Result of unsupervised machine learning. (A) Optimal clustering number of the K-means clustering algorithm was determined by Silhouette coefficient (SC). The peak of the curve is the best value for the Silhouette coefficient (Y-axis); the best number of clusters was equal to 2 (X–axis). (B) Scatter plots of patients' clinical data. Scatter points on the graph represent each patient. The K-means algorithm divides patients into two clusters. The red scatter represents cluster 1 and the blue scatter represents cluster 2. (For interpretation of the references to colour in this figure legend, the reader is referred to the Web version of this article.)

2.4. Data downloading and processing

We used “lower extremity peripheral artery disease” as a keyword to search related gene expression datasets from the Gene Expression Omnibus (GEO) database. The datasets GSE100927 [21] and GSE113873 [22] were analyzed. GSE100927 data were normalized, and differentially expressed genes (DEGs) were obtained using the “limma” package in the R software [23]. The fold changes (FCs) were calculated for individual gene expression. Genes meeting the specific cutoff criteria of $p < 0.05$ and $|\log FC| > 1$ were defined as DEGs. Following this, Gene Ontology (GO) annotation [24] and Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway enrichment analysis [25] of DEGs were performed using the R packages, including “clusterProfifiler,” “org,” “Hs.eg.db,” “enrichplot,” and “ggplot2,” to determine the functions of common DEGs [26].

2.5. Identification of hub genes

For the identification of hub genes, least absolute shrinkage and selection operator (LASSO) regression and support vector machine recursive feature elimination (SVM-RFE) analyses were performed to rank individual features on the basis of importance using the “glmnet,” “rms,” “e1071,” “kernlabt,” and “caret” R packages [27–29]. DEGs screened through these two machine learning methods for feature selection were intersected using the R package “veen” to obtain the hub genes [30]. Thereafter, the expression of the identified hub genes was validated in GSE113873 [22] using the Wilcoxon test, and a p-value of <0.05 was considered statistically significant [31]. Additionally, to assess the predictive accuracy of the hub genes, ROC curves were generated using the pROC package

Table 1
Predictor variable’ s baseline characteristics.

Characteristic	cluster			p-value
	Overall, N = 460	1, N = 202	2, N = 258	
Gender				<0.001
Female	173 (38 %)	50 (25 %)	123 (48 %)	
Male	287 (62 %)	152 (75 %)	135 (52 %)	
Age				<0.001
Median (IQR)	59 (50, 69)	63 (54, 71)	54 (47, 66)	
BMI				<0.001
Median (IQR)	22.5 (20.0, 25.0)	23.4 (22.0, 25.8)	21.3 (19.5, 24.5)	
Smoking				<0.001
No	162 (35 %)	44 (22 %)	118 (46 %)	
Yes	298 (65 %)	158 (78 %)	140 (54 %)	
Hypertension				<0.001
No	294 (64 %)	107 (53 %)	187 (72 %)	
Yes	166 (36 %)	95 (47 %)	71 (28 %)	
Diabetes				<0.001
No	350 (76 %)	138 (68 %)	212 (82 %)	
Yes	110 (24 %)	64 (32 %)	46 (18 %)	
MD				0.908
No	432 (94 %)	190 (94 %)	242 (94 %)	
Yes	28 (6.1 %)	12 (5.9 %)	16 (6.2 %)	
CD				0.123
No	433 (94 %)	194 (96 %)	239 (93 %)	
Yes	27 (5.9 %)	8 (4.0 %)	19 (7.4 %)	
RD				0.404
No	418 (91 %)	181 (90 %)	237 (92 %)	
Yes	42 (9.1 %)	21 (10 %)	21 (8.1 %)	
White blood cell (10ˆ9/L)				0.119
Median (IQR)	7.0 (5.4, 9.7)	7.3 (5.6, 10.0)	6.7 (5.3, 9.2)	
Erythrocyte (10ˆ12/L)				0.075
Median (IQR)	4.25 (3.76, 4.68)	4.36 (3.62, 5.02)	4.18 (3.80, 4.63)	
Hemoglobin (g/L)				0.322
Median (IQR)	116 (98, 131)	119 (93, 136)	116 (103, 128)	
Platelet (10ˆ9/L)				0.285
Median (IQR)	220 (156, 275)	207 (154, 274)	222 (163, 274)	
Neutrophil (10ˆ9/L)				<0.001
Median (IQR)	5.09 (3.62, 7.60)	6.23 (4.21, 8.69)	4.52 (3.18, 6.88)	
Lymphocyte (10ˆ9/L)				0.236
Median (IQR)	1.53 (1.05, 1.89)	1.46 (1.04, 1.85)	1.56 (1.06, 1.89)	
Monocyte (10ˆ9/L)				0.410
Median (IQR)	0.49 (0.38, 0.66)	0.52 (0.38, 0.67)	0.49 (0.38, 0.65)	
Eosinophil (10ˆ9/L)				0.215
Median (IQR)	0.15 (0.07, 0.21)	0.15 (0.08, 0.23)	0.14 (0.07, 0.20)	
Basophil (10ˆ9/L)				0.908
Median (IQR)	0.030 (0.020, 0.040)	0.030 (0.020, 0.048)	0.030 (0.020, 0.040)	

BMI: body mass index; CD: cerebrovascular disease; MD: myocardial disease; RD: renal disease.

in the R language [32].

2.6. *Estimation of immune infiltration-related cells and genes*

The immune cell composition of the GEO datasets was analyzed using Cell-type Identification by Estimating Relative Subsets of RNA Transcripts (CIBERSORT) to assess immunological infiltration in patients with LE-PAD. CIBERSORT is a bioinformatics software that characterizes immune cell composition and expresses the components as a matrix [33]. The relationships of the hub genes with 22 immune cells were analyzed to identify hub gene-associated immune cells.

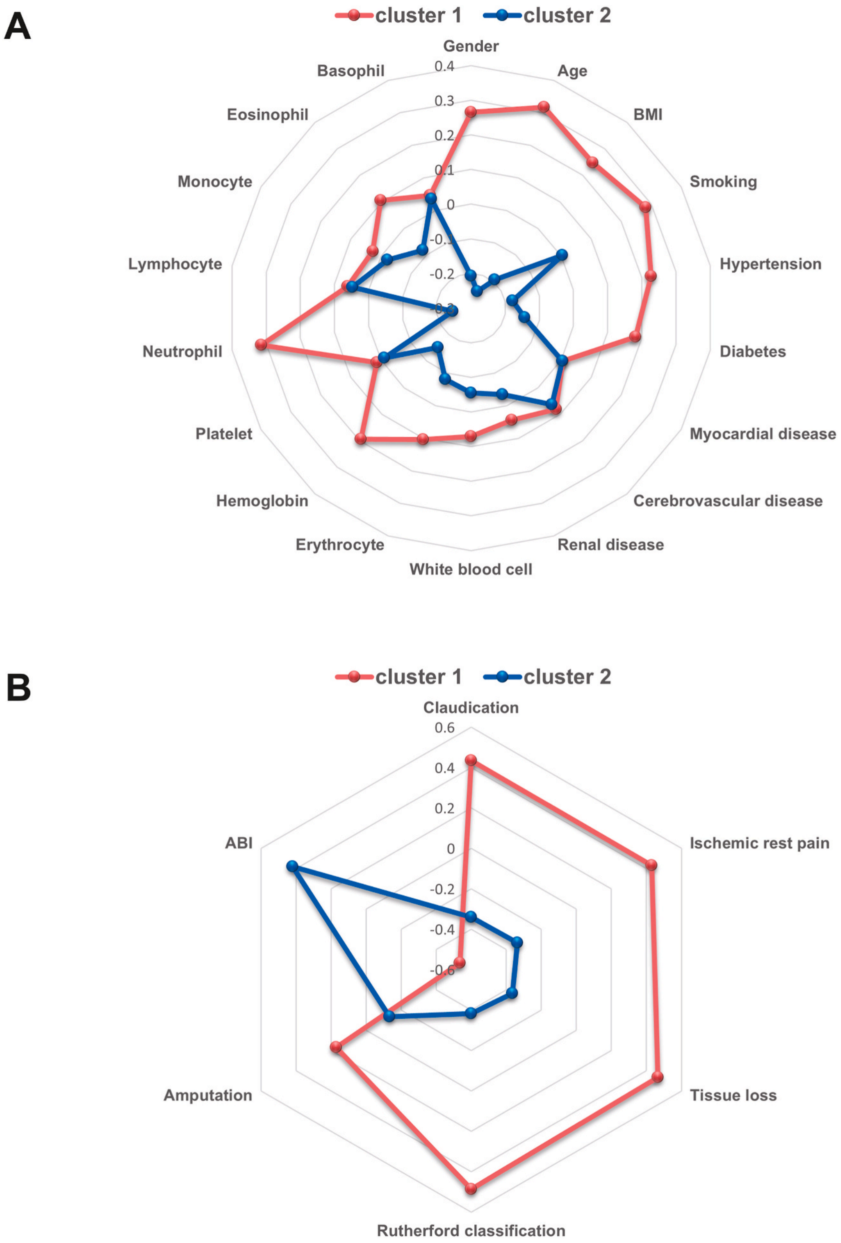


Fig. 3. Radargram of LE-PAD patients' clustering and validation variables in two clusters. (A)Radargram of LE-PAD patients' clustering variables in two clusters. (B) Radargram of LE-PAD patients' validation variables in two clusters. ABI: ankle-brachial index; BMI: body mass index; CD: cerebrovascular disease; MD: myocardial disease; RD: renal disease.

2.7. Validating immune cell infiltration by histological analysis

Formalin-fixed, paraffin-embedded sections (5-μm thickness) of lower-limb artery tissues were obtained from patients with LE-PAD and trauma patients without LE-PAD. The sections underwent deparaffinization, rehydration, and hematoxylin and eosin (H&E) staining. Neutrophil infiltration among the samples from the patients was compared by light microscopy.

2.8. Statistical analysis

We performed statistical analysis using the International Business Machines Corporation Statistical Package for the Social Sciences 26.0 and R 4.2.1 software. The clinical data are presented as the median (P25 and P75). Depending on the data type, Student's t-test, Mann–Whitney *U* test, or chi-square test were performed. The p-value of <0.05 was considered statistically significant.

3. Results

3.1. UMLAs results

We used UMLAs to cluster patients with LE-PAD. Fig. 2A shows the optimal clustering number from the K-means algorithm, and the curve peak indicates the best value for the SC (Y-axis) [17], suggesting two as the optimal cluster number. This algorithm sufficiently clustered the present clinical data into clusters 1 and 2 (Fig. 2B). Table 1 presents the K-means clustering results for the predictor variables. Gender distribution was significantly different among the two clusters ($p < 0.001$), with cluster 1 having a higher proportion of males (75 %) than cluster 2 (52 %). Cluster 1 also exhibited significantly higher values for age, body mass index (BMI), neutrophil levels, smoking status, hypertension, and diabetes prevalence than cluster 2 ($p < 0.001$). Furthermore, no significant differences were observed for other predictor variables. Fig. 3A shows the radargram of the predictor variables.

4. Comparison of outcome variables between the two clusters

Table 2 shows differences in outcome variables between the two clusters. Cluster 1 had a significantly lower ABI than cluster 2 ($p < 0.001$). Furthermore, cluster 1 showed significantly higher Rutherford classification, claudication and tissue loss severity, and ischemic rest pain prevalence than did cluster 2 ($p < 0.001$). Additionally, cluster 1 showed a higher amputation rate than cluster 2 ($p = 0.001$). These findings indicated worse conditions and prognoses for patients in cluster 1 than for those in cluster 2 (Fig. 3B). The disparities observed in the outcome variables between the two clusters substantiate the precision of the UMLAs clustering in this investigation. Consequently, UMLAs effectively segregated patients with LE-PAD into distinct severe and mild groups based on the acquired clinical data.

Table 2
Comparison of outcome variables between two clusters.

Characteristic	cluster			p-value
	Overall, N = 460	1, N = 202	2, N = 258	
Claudication				<0.001
Asymptomatic	12 (2.6 %)	4 (2.0 %)	8 (3.1 %)	
Mild	41 (8.9 %)	6 (3.0 %)	35 (14 %)	
Moderate	213 (46 %)	57 (28 %)	156 (60 %)	
Severe	194 (42 %)	135 (67 %)	59 (23 %)	
Ischemic rest pain				<0.001
No	279 (61 %)	80 (40 %)	199 (77 %)	
Yes	181 (39 %)	122 (60 %)	59 (23 %)	
Tissue loss				<0.001
No	304 (66 %)	86 (43 %)	218 (84 %)	
Minor	103 (22 %)	75 (37 %)	28 (11 %)	
Major	53 (12 %)	41 (20 %)	12 (4.7 %)	
Rutherford classification				<0.001
0	12 (2.6 %)	4 (2.0 %)	8 (3.1 %)	
1	40 (8.7 %)	6 (3.0 %)	34 (13 %)	
2	213 (46 %)	57 (28 %)	156 (60 %)	
3	14 (3.0 %)	13 (6.4 %)	1 (0.4 %)	
4	25 (5.4 %)	6 (3.0 %)	19 (7.4 %)	
5	103 (22 %)	75 (37 %)	28 (11 %)	
6	53 (12 %)	41 (20 %)	12 (4.7 %)	
Amputation				0.001
No	433 (94 %)	182 (90 %)	251 (97 %)	
Yes	27 (5.9 %)	20 (9.9 %)	7 (2.7 %)	
ABI				<0.001
Median (IQR)	0.55 (0.39, 0.69)	0.39 (0.32, 0.60)	0.65 (0.51, 0.74)	

ABI: ankle-brachial index.

4.1. Construction of a predictive model for the novel LE-PAD subtype

Variables with significant differences between the two clusters ($p < 0.05$) were gender, age, BMI, smoking status, hypertension and diabetes prevalence, and neutrophil levels. Univariate and multivariate logistic regression analyses were performed on these variables. The results showed the following six independent clustering prediction factors: gender, age, BMI, smoking status, hypertension prevalence, and neutrophil levels ($p < 0.05$) (Table 3). A forest plot was constructed based on the binary logistic regression results (Fig. 4A). Further, a nomogram was established based on the six independent risk factors (Fig. 4B). The ROC area under the curve (AUC) of the nomogram was 0.759 (95 % confidence interval: 0.717–0.802) (Fig. 4C). Furthermore, calibration curves validated the actual and predicted nomogram probabilities (Fig. 4D) with a C-index of 0.761. These results showed the good accuracy of the predictive model for the novel LE-PAD subtype.

Through binary logistic regression analysis, we identified a heightened neutrophil count as a significant risk factor for the unfavorable prognosis of Lower Extremity Peripheral Artery Disease (LE-PAD). Consequently, we employed immunoinfiltration analysis to delve deeper into understanding the nuanced role of neutrophils in the onset and progression of LE-PAD.

4.2. Immune cell infiltration analysis verified the heterogeneity of neutrophils in LE-PAD

According to the aforementioned analyses, neutrophil levels were significantly higher in the severe LE-PAD cluster than in the mild cluster. Therefore, we further investigated the relationship between neutrophil levels and LE-PAD.

The volcano plot showed that 338 DEGs were identified from GSE100927 (Fig. 5A). The heatmaps of these DEGs are shown in Fig. 5B. GO and KEGG pathway enrichment analyses were performed to further examine biological data associated with these DEGs. Fig. 6A shows that the top three terms associated with biological processes in the GO enrichment analysis were “leukocyte cell−cell adhesion,” “leukocyte-mediated immunity,” and “positive regulation of leukocyte activation.” The top three terms associated with cellular components were “major histocompatibility complex (MHC) class II protein complex,” “endocytic vesicle,” and “external side of the plasma membrane.” The top three terms associated with molecular functions were “MHC protein complex binding,” “MHC class II protein complex binding,” and “immune receptor activity.” The KEGG pathway enrichment analysis revealed that the DEGs were mainly enriched in leishmaniasis, hematopoietic cell lineage, and *Staphylococcus aureus* infection (Fig. 6B).

The DEGs were subjected to LASSO regression and SVM-RFE analyses to identify candidate genes. LASSO regression screened 12 genes (Fig. 7A), whereas SVM-RFE screened eight genes (Fig. 7B). The intersection of the Venn diagrams showed the following three genes: CSF3 (colony-stimulating factor 3), SAA2 (serum amyloid A-2 protein), and CXCR2 (C-X-C motif chemokine receptor 2) (Fig. 7C). These observations were further validated using the GSE113873 dataset. Within the trio of candidate genes, CXCR2 exhibited a noteworthy upregulation in patients with LE-PAD compared to normal individuals (Fig. 8A–C). The determination of CXCR2 expression and the generation of ROC curves, facilitated by the pROC package in R software, were executed to assess its diagnostic accuracy (Fig. 8D–E). The AUC values for CXCR2 surpassed 0.9 across all datasets, underscoring its robust discriminative capability. Consequently, CXCR2 emerged as the identified hub gene.

The relationship between the hub gene CXCR2 and immune cells was investigated using CIBERSORT. The violin plot revealed patients with LE-PAD exhibited significantly higher neutrophil levels than did the controls (Fig. 9A). A strong positive correlation between CXCR2 expression and neutrophil activation was observed in the LE-PAD datasets ($p < 0.001$) (Fig. 9B). Lollipop plots showed correlation coefficients for CXCR2 and each immune cell type (Fig. 9C). Overall, the results suggested a significant correlation of neutrophil activation with LE-PAD pathology and CXCR2 expression.

Table 3
Univariate and multivariate analysis of influencing factors.

			Univariable			Multivariable		
Characteristic	N	Event N	OR	95 % CI	p-value	OR	95 % CI	p-value
Gender								
Female	173	50	–	–		–	–	
Male	287	152	2.77	1.86, 4.17	<0.001	2.41	1.30, 4.51	0.005
Age	460	202	1.05	1.03, 1.07	<0.001	1.03	1.01, 1.05	<0.001
BMI	460	202	1.14	1.08, 1.20	<0.001	1.10	1.03, 1.17	0.004
Smoking								
No	162	44	–	–		–	–	
Yes	298	158	3.03	2.01, 4.61	<0.001	1.99	1.07, 3.73	0.031
Hypertension								
No	294	107	–	–		–	–	
Yes	166	95	2.34	1.59, 3.46	<0.001	1.84	1.16, 2.91	0.009
Diabetes								
No	350	138	–	–		–	–	
Yes	110	64	2.14	1.39, 3.32	<0.001	1.20	0.70, 2.05	0.507
Neutrophils	460	202	1.21	1.14, 1.30	<0.001	1.17	1.09, 1.27	<0.001

BMI: body mass index; CI: confidence interval.

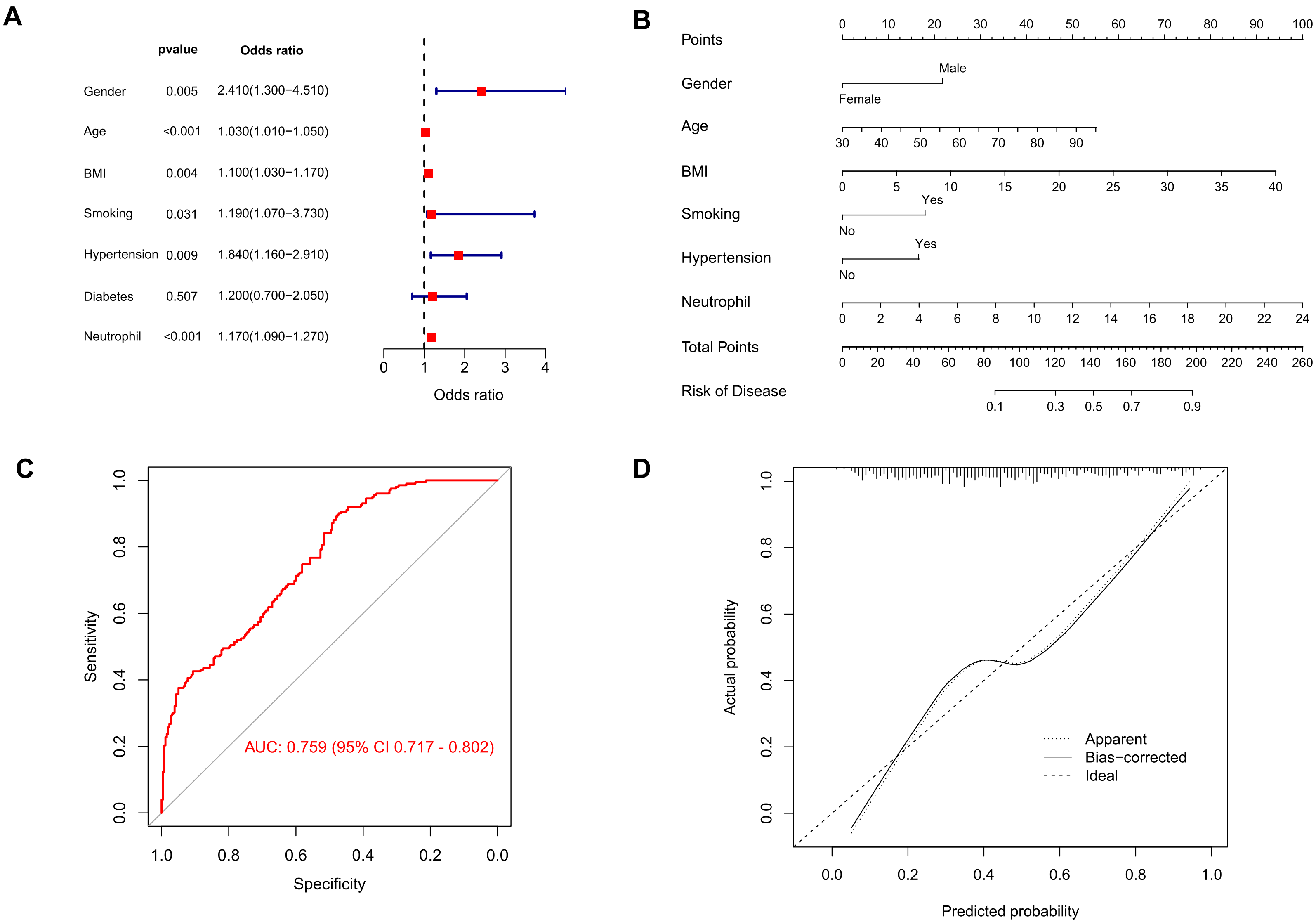
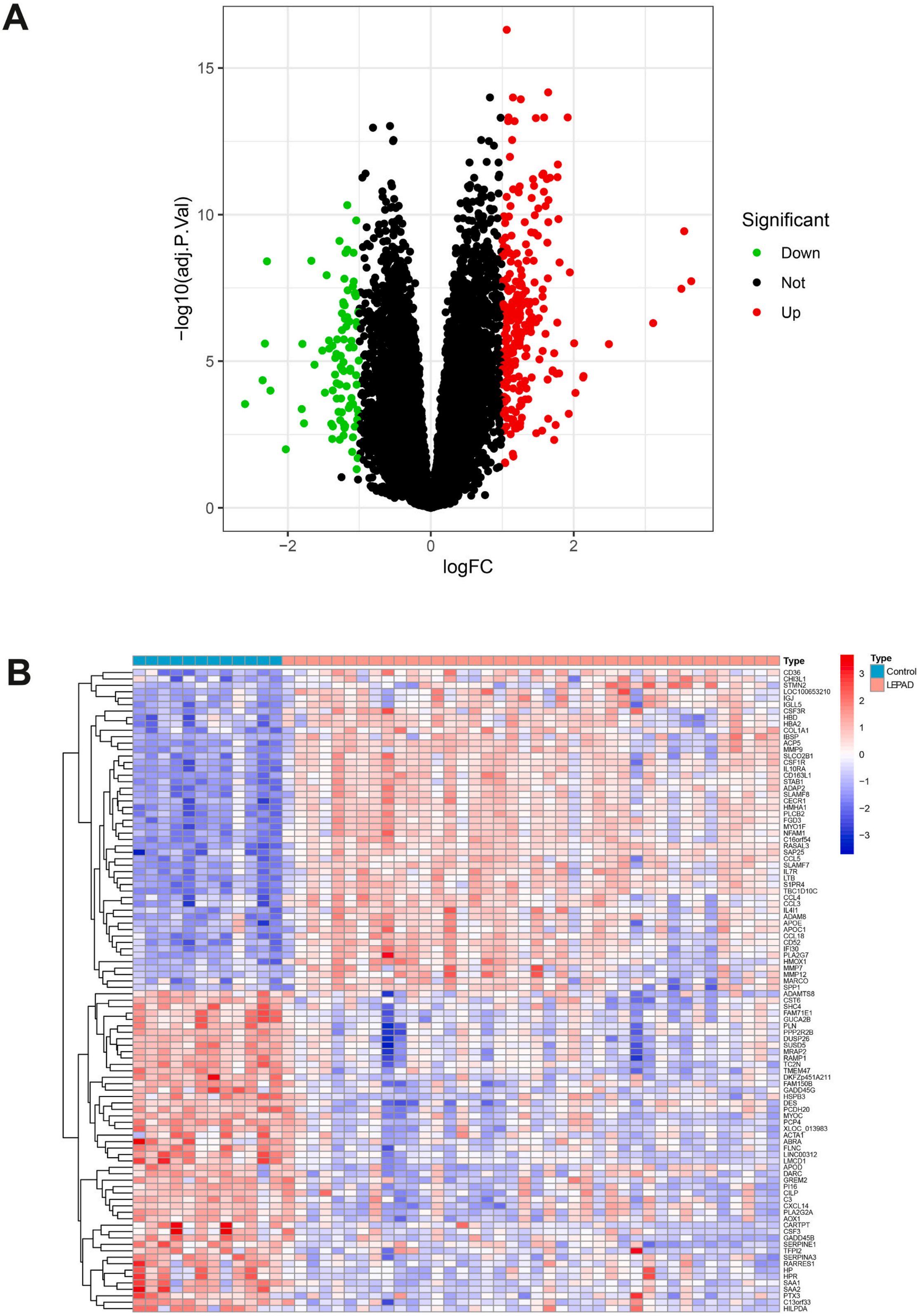


Fig. 4. Construction and validation of the predictive model. (A)Forest plot of the predictor variables. (B) Nomogram for prediction model. (C) AUC of the nomogram. (D) Calibration curves for predictive model. AUC: area under the curve; BMI: body mass index; CI: confidence interval.



(caption on next page)

Fig. 5. Identification of DEGs. (A)Volcano plot revealing 338 DEGs between the LE-PAD patients and healthy controls. Red points are represented up-regulated genes and green points are represented down-regulated genes. (B) Heat map of DEGs between LE-PAD patients and healthy controls. Red areas are represented up-regulated genes and blue areas are represented down-regulated genes. (For interpretation of the references to colour in this figure legend, the reader is referred to the Web version of this article.)

4.3. Validating immune cell infiltration by histological analysis

To verify neutrophil infiltration in LE-PAD, we collected lower limb artery samples from five patients with LE-PAD who were undergoing amputation and five controls without LE-PAD (Fig. 10A). The collected artery samples were subjected to H&E staining. Specimens from patients with LE-PAD manifested heterogeneous neutrophil infiltration. Specifically, two samples exhibited mild infiltration, whereas three samples displayed pronounced neutrophil infiltration within the intima and media layers (Fig. 10B). Conversely, negligible neutrophil infiltration was observed within the control samples (Fig. 10C).

Overall, the results showed a critical role for neutrophil infiltration in LE-PAD pathogenesis. Furthermore, we observed a significant positive correlation between neutrophil activation and CXCR2 expression.

5. Discussion

5.1. Clinical significance of the AI-based modeling of the novel LE-PAD subtype

With the advancement of AI, UMLAs are increasingly used in clinical diagnosis and treatment. These algorithms use clustering methods specifically designed to integrate heterogeneous clinical data in an unsupervised manner, thus, grouping patients on the basis of similar characteristics rather than relying on prior knowledge (5). Algorithms of this nature possess the capacity to discern intricate patterns within complex data, thereby aiding physicians in clinical decision-making. Zheyu Wang et al. employed UMLAs to formulate a risk score for the progression of Alzheimer's disease [34]. In a similar vein, Maddali et al. assessed clinical classifier models rooted in ML for acute respiratory distress syndrome, incorporating diverse clinical, biological, and treatment response features [35]. Their results indicated that these AI models could provide valuable prognostic information and help devise treatment strategies against the syndrome.

In the present study, UMLAs classified 18 predictor variables obtained based on the information of patients with LE-PAD into two distinct clusters (Table 1). Fig. 3A shows substantial heterogeneity in the clinical data between the two clusters. Further analysis of follow-up data, ABI, and Rutherford classifications in these two clusters revealed significantly worse conditions and prognoses for patients in cluster 1 than in cluster 2 (Fig. 3B). Hence, we defined these two clusters as novel LE-PAD machine learning subtypes, namely NLEPADML1 (cluster 1) and NLEPADML2 (cluster 2). Univariate and multivariate logistic regression analyses were performed to develop a predictive model for identifying patients in NLEPADML1 who were at a high risk of disease progression and poor outcomes. Our study demonstrates the potential utility of utilizing UMLAs to identify novel subtypes and construct prediction models in LE-PAD patients undergoing endovascular therapy. The resultant prediction model could effectively stratify LE-PAD risk, facilitating personalized care through medications, lifestyle changes, close monitoring, and timely procedures. The present study highlights the importance of using AI-based methods in advancing precision medicine for this heterogeneous disease.

In this study, gender, age, BMI, smoking status, hypertension prevalence, and neutrophil levels were identified as risk factors contributing to an unfavorable prognosis in patients with LE-PAD. Some studies reveal an escalating prevalence of LE-PAD with advancing age, notably with individuals aged 70 and above exhibiting a substantially higher incidence compared to those under 70. The likelihood of an adverse prognosis proportionally increases with age [36]. Research indicates that male patients with LE-PAD have a heightened probability of experiencing an unfavorable prognosis, potentially linked to their increased propensity for smoking and the presence of risk factors like diabetes and coronary heart disease. Smoking, in contrast to non-smoking, amplifies the risk of LE-PAD [37]. Despite the risk reduction associated with smoking cessation, contemporary studies suggest that it may take up to three decades for the LE-PAD risk to diminish to the level observed in non-smokers [38]. Numerous epidemiological and longitudinal studies identify hypertension as a major risk factor for the development of LE-PAD [39]. Even mild to moderate hypertension and dyslipidemia have a multiplicative detrimental effect on atherosclerosis risk [40]. Additionally, more than 65 % of adults with LE-PAD are reported to be overweight or obese [41]. High fat mass is associated with a decline in ambulatory status and vascular health in patients with LE-PAD and claudication [42]. In conclusion, the majority of patients with LE-PAD exhibit multiple risk factors, and the cumulative effects of these factors can accelerate atherosclerotic processes.

The ability to discern novel patient subgroups from complex, multifaceted data that are not apparent with traditional techniques is the key feature of UMLAs applications [43]. By clustering patients on the basis of intrinsic data patterns, we identified a high-risk phenotype that would likely not have been identified by the existing standard statistical methods. The model developed from this data-driven subgroup exhibited good predictive performance for poor outcomes (Fig. 4). Furthermore, this study exemplifies how UMLAs approaches can help develop a personalized medicine model by accurately predicting risks for individual patients. The present prediction model could stratify LE-PAD risk, enabling personalized care through medications, lifestyle changes, close monitoring, and timely procedures. The model help optimize disease management to mitigate adverse events in patients at high risk. With further validation, the implementation of such AI models can assist clinicians in applying personalized therapies and surveillance strategies. Jun Ma et al. used UMLAs to develop a prediction model with high accuracy that could meet doctors' needs for individualized preoperative and surgical safety evaluations in laparoendoscopic single-site surgery [44].

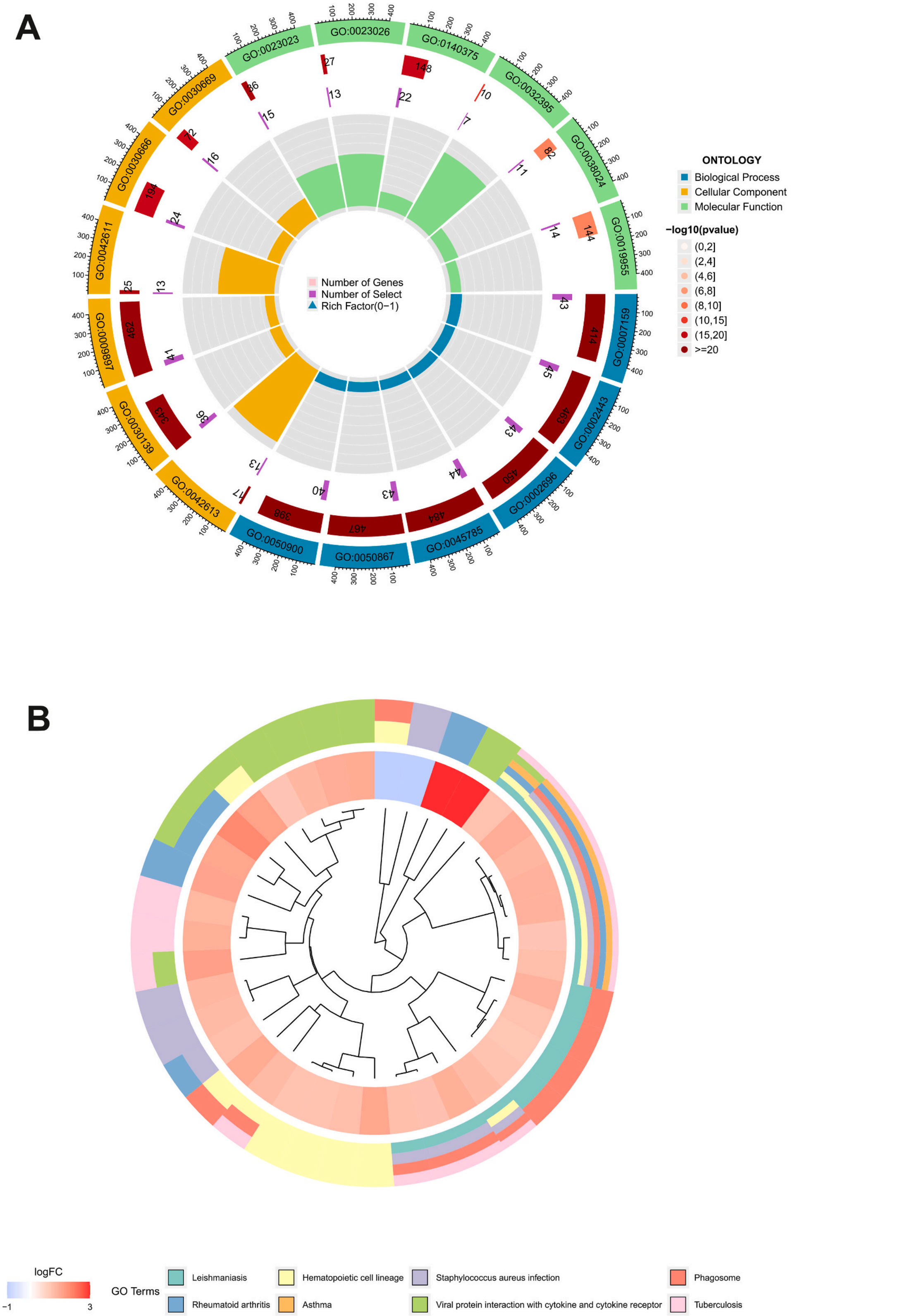


Fig. 6. GO and KEGG pathway enrichment analysis of the DEGs. (A)GO terms in biological process, cellular component, and molecular function were used for functional enrichment clustering analysis on the DEGs. (B) KEGG pathway analysis was performed on the DEGs.

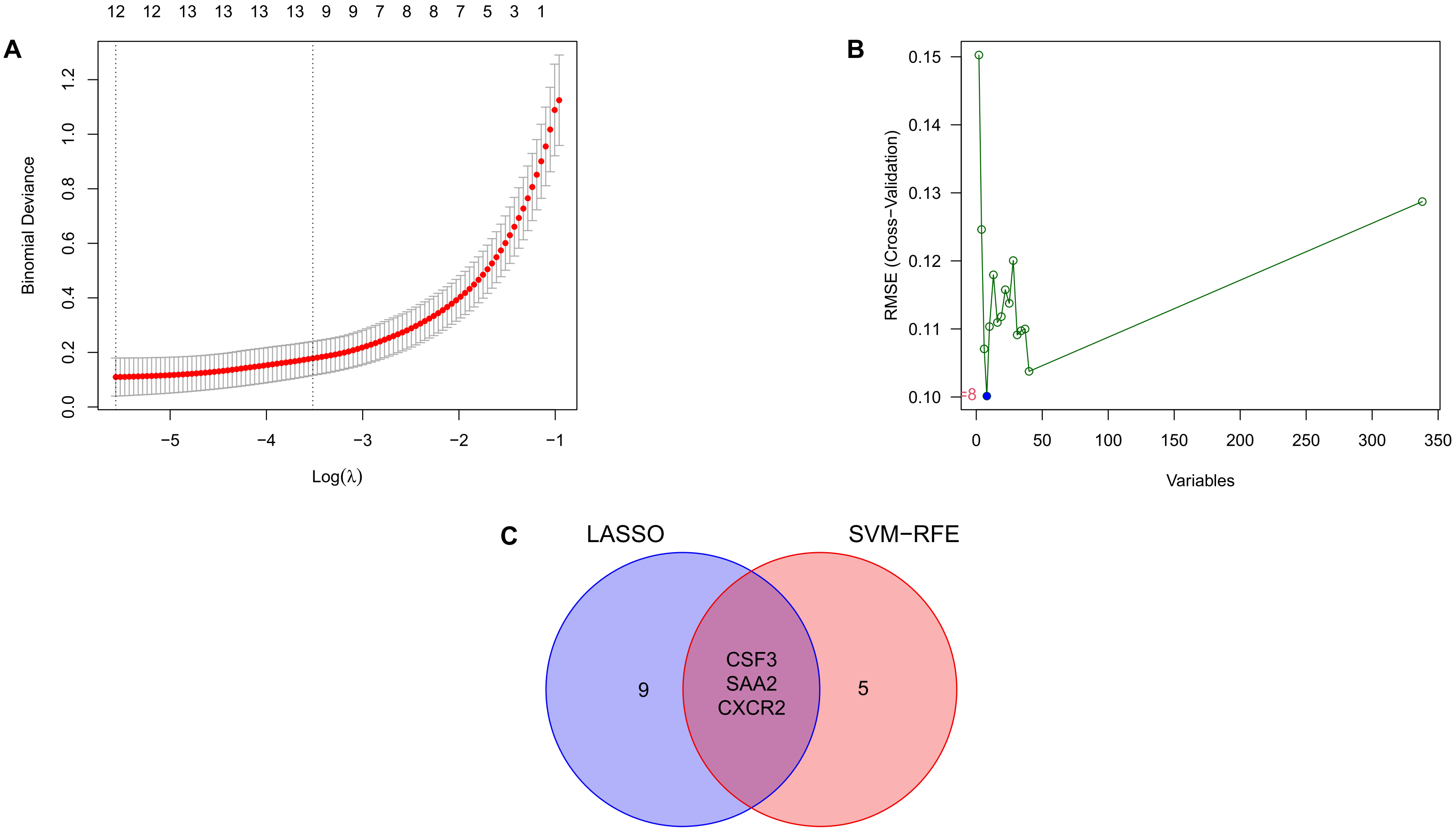


Fig. 7. Screening of candidate genes. (A)The result of LASSO regression. (B) The result of SVM-RFE analyses. (C) Three overlapping genes were filtered using venn diagram. CSF3: Colony Stimulating Factor 3; CXCR2: C-X-C Motif Chemokine Receptor 2 SAA2: Serum Amyloid A-2 Protein; SVM-RFE: support vector machine recursive feature elimination; LASSO: least absolute shrinkage and selection operator.

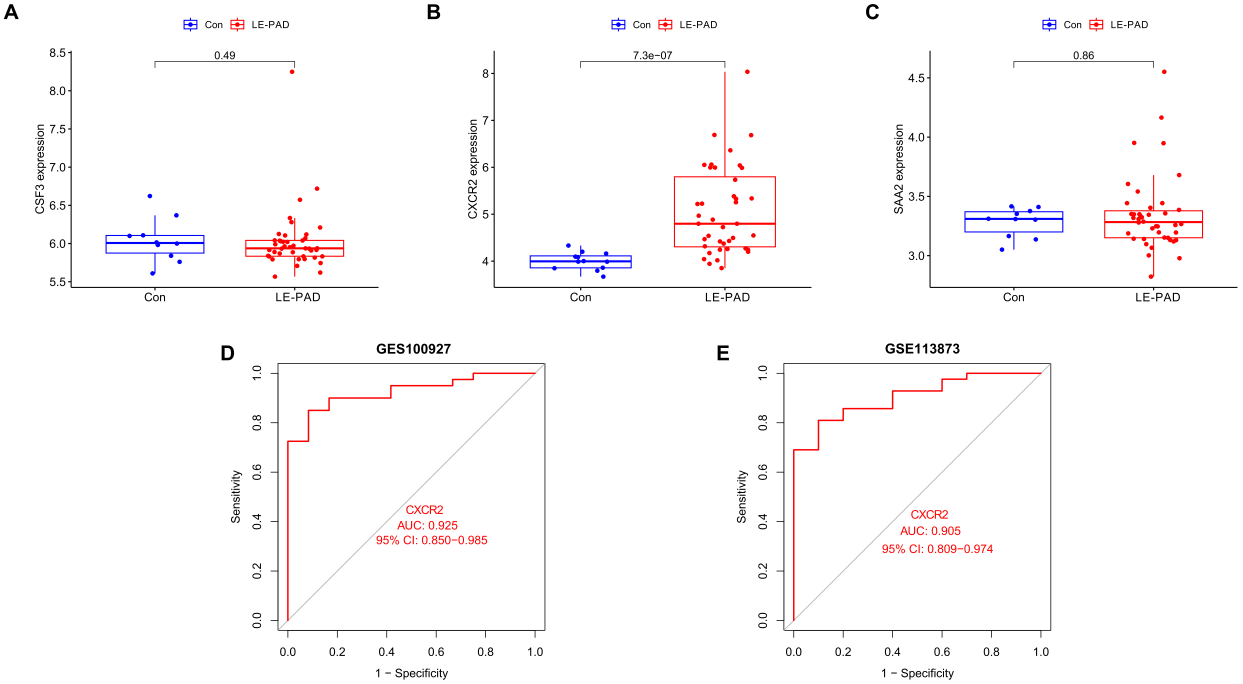


Fig. 8. Validation of hub gene (A–C) CSF3, CXCR2 and SAA2 were validated in GSE47472. The boxplots show that the expression level of the CXCR2 is higher in LE-PAD samples. (D–E) ROC curves were drawn to evaluate the accuracy of the CXCR2 in diagnosing LE-PAD. CSF3: Colony Stimulating Factor 3; CXCR2: C-X-C Motif Chemokine Receptor 2 SAA2: Serum Amyloid A-2 Protein; AUC: area under the curve; CI: confidence interval; LE-PAD: lower-extremity peripheral artery disease.

Moreover, there is substantial scope for expanding UMLAs applications in medicine. These approaches can be applied to large integrated biomedical datasets with deeper phenotyping. Models incorporating multi-modal data such as imaging, genetics, and biomarkers may reveal further novel patient subgroups and refined predictive abilities [45].

5.2. Heterogeneity of neutrophils in LE-PAD is associated with CXCR2 dysregulation

Herein, we identified heterogeneity in neutrophils among patients with LE-PAD in NLEPADML1 and NLEPADML2. Furthermore, neutrophil levels were one of the independent predictors for patients in NLEPADML1 (Fig. 4A), an ML subtype of LE-PAD with a severer disease condition. Based on the results of immune cell infiltration analysis, the heterogeneity of neutrophils was validated.

The present findings showed a significant positive correlation between neutrophil activation and CXCR2 expression and a G-protein coupled receptor responsible for cellular signal transduction in this activation process [46]. In neutrophils, CXCR2 is activated by the chemokine C-X-C motif chemokine ligand 8 (interleukin-8, IL-8) to regulate neutrophil recruitment from blood to inflammation sites [47]. CXCR2 is associated with atherosclerosis progression [48]. IL-8 interacts with CXCR2 in neutrophils, thereby forming NETs via Src/extracellular signal-regulated kinase and p38/mitogen-activated protein kinase signaling pathways, thus, aggravating atherosclerosis progression *in vivo* [49].

The existing studies on LE-PAD have focused on the diagnostic and prognostic utility of biomarkers to indicate neutrophil activity. Selvaggio et al. found a significantly higher neutrophil-to-lymphocyte ratio (NLR) in patients with LE-PAD showing an ABI of ≤0.9 compared with those showing an ABI of 0.9–1.4 [50]. Celebi et al. demonstrated via angiography that patients with LE-PAD showed a significantly higher NLR than did patients without LE-PAD, suggesting NLR as a simple and reproducible LE-PAD marker [51]. Myeloperoxidase, one of the extracellular proteases stored in neutrophil granules and released upon activation, probably plays a role in atherogenesis. It is a heme-containing peroxidase stored in azurophilic granules and released extracellularly during degranulation [52]. In a study of patients with LE-PAD treated with endovascular therapy, immunohistochemical and immunofluorescence analyses revealed the presence of proteases, including myeloperoxidase, in inflammatory cells and nearly half of the debris particles [53]. Matijevic et al. found higher myeloperoxidase levels in patients with LE-PAD than in healthy individuals [54].

Accumulating evidence supports the central role of neutrophils and their mediators in LE-PAD, highlighting the need for further studies on their etiopathogenic mechanisms and utility as diagnostic and prognostic markers across the PAD severity spectrum. The complementary assessment of neutrophil activity and novel mechanisms may improve risk stratification and help develop personalized treatments for patients with LE-PAD.

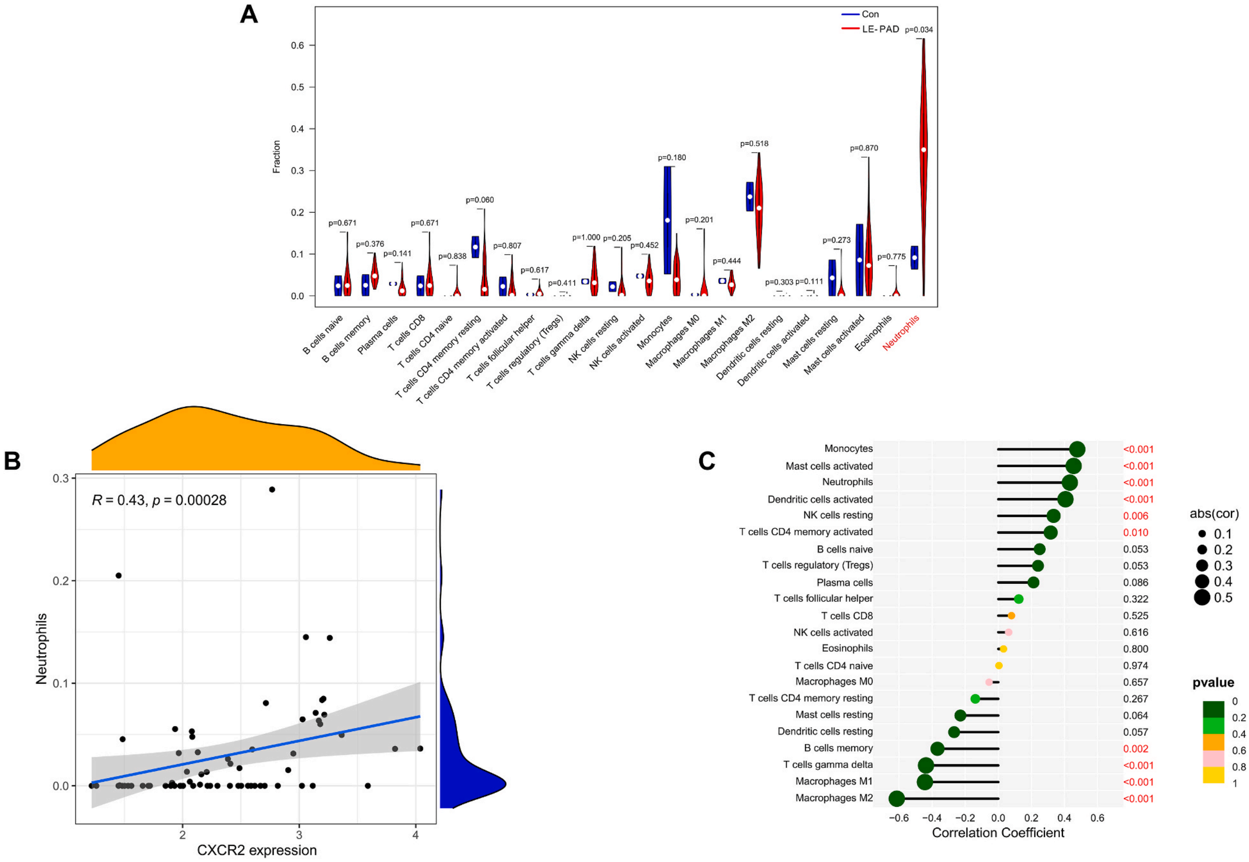


Fig. 9. Association between the hub gene and immune infiltration (A)Violin plot of the difference in immune cells between LE-PAD and controls. (B) Correlation plot of CXCR2 with neutrophils activation. (C) Correlation coefficient of CXCR2 with 22 immune cells in AAA datasets. CXCR2: C-X-C Motif Chemokine Receptor 2; LE-PAD: lower-extremity peripheral artery disease.

5.3. Limitations of the study

This present study has some limitations. First, owing to its retrospective design, selection bias cannot be excluded. Second, the data were collected for a single center; thus, external validation in a larger and multicenter setting is needed to optimize the predictive model. Finally, more *in vitro* and *in vivo* studies are required to explore neutrophil-associated LE-PAD pathogenesis.

6. Conclusions

UMLAs furnish prediction models for a novel subtype among patients with LE-PAD. This model facilitates risk stratification in LE-PAD patients utilizing readily available clinical data, thereby assisting in clinical decision-making and augmenting personalized management for individuals with LE-PAD. Furthermore, the findings underscore the pivotal role of neutrophil infiltration in the pathogenesis of LE-PAD.

Ethics statement

The studies involving human participants were reviewed and approved by the Ethics Committee of The First Affiliated Hospital of Guangxi Medical University. The approval number: 2023-E340-01.

Data availability statement

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

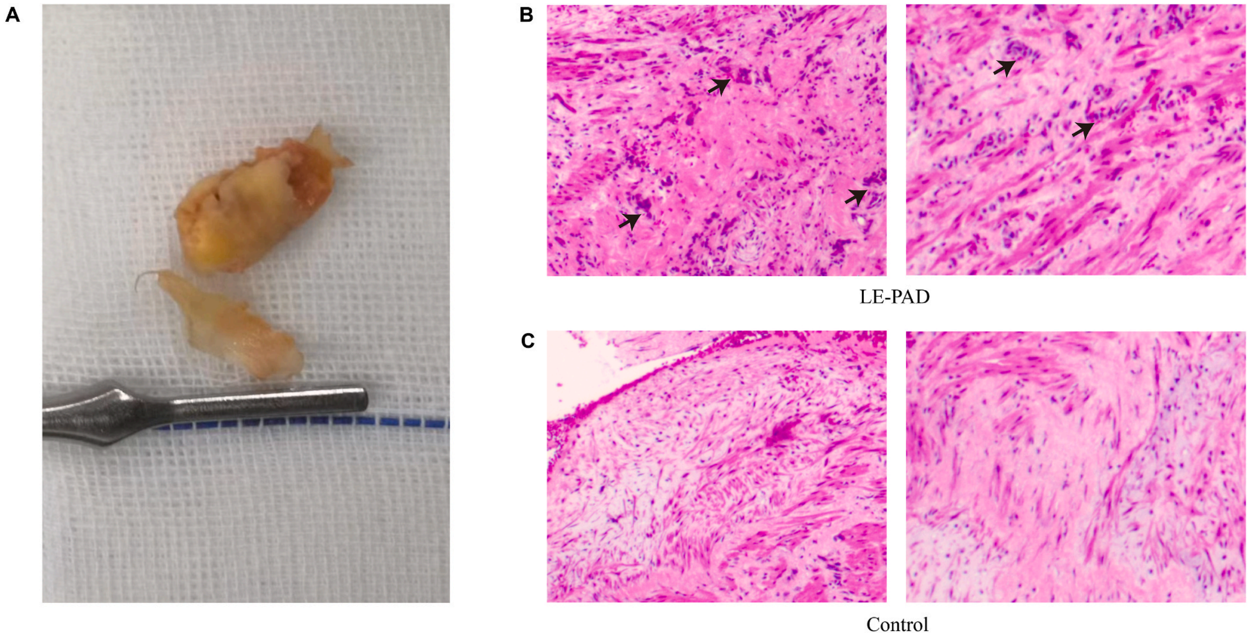


Fig. 10. Neutrophil infiltration in lower limb artery samples. (A) An illustration of lower limb artery tissues collection. (B) H&E staining of lower limb artery samples from patients with LE-PAD (black arrow, neutrophil infiltration). (C) H&E staining of lower limb artery samples from control.

CRediT authorship contribution statement

Lin Zhang: Conceptualization, Data curation, Visualization, Writing - original draft, Writing - review & editing. **Yuanliang Ma:** Data curation, Formal analysis, Methodology, Writing - review & editing. **Que Li:** Conceptualization, Data curation, Methodology, Validation, Writing - review & editing. **Zhen Long:** Data curation, Resources, Validation, Visualization, Writing - review & editing. **Jiangfeng Zhang:** Formal analysis, Methodology, Resources, Supervision, Writing - review & editing. **Zhanman Zhang:** Formal analysis, Methodology, Resources, Supervision, Writing - review & editing. **Xiao Qin:** Conceptualization, Formal analysis, Funding acquisition, Project administration, Supervision, Writing - review & editing.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

The authors are very grateful to the National Natural Science Foundation of China (81960091) for support in all stages of this study.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.heliyon.2024.e24189.

References

[1] W. Lian, et al., Clinical significance of endothelin-1 and C reaction protein in restenosis after the intervention of lower extremity arteriosclerosis obliterans, J. Invest. Surg. 34 (7) (2021) 765–770.

[2] W. Yao, et al., Effects of valsartan on restenosis in patients with arteriosclerosis obliterans of the lower extremities undergoing interventional therapy: a prospective, randomized, single-blind trial, Med Sci Monit 26 (2020) e919977.

[3] M. Ye, et al., Neutrophil-lymphocyte ratio and platelet-lymphocyte ratio predict severity and prognosis of lower limb arteriosclerosis obliterans, Ann. Vasc. Surg. 64 (2020) 221–227.

[4] V.A. Badtieva, D.N. Voroshilova, N.V. Sichinava, [Use of enhanced external counterpulsation in the treatment and rehabilitation of patients with atherosclerosis obliterans of the lower extremity], Vopr. Kurortol. Fizioter. Lech. Fiz. Kul't. 96 (4) (2019) 5–11.

[5] J. Dong, et al., Machine learning model for early prediction of acute kidney injury (AKI) in pediatric critical care, Crit. Care 25 (1) (2021) 288.

[6] M. Kobayashi, et al., Machine learning-derived echocardiographic phenotypes predict heart failure incidence in asymptomatic individuals, JACC Cardiovasc Imaging 15 (2) (2022) 193–208.

[7] G.K. Hansson, Inflammation, atherosclerosis, and coronary artery disease, N. Engl. J. Med. 352 (16) (2005) 1685–1695.

[8] G.R. Geovanini, P. Libby, Atherosclerosis and inflammation: overview and updates, Clin. Sci. (Lond.) 132 (12) (2018) 1243–1252.

[9] E. Montaldo, et al., Cellular and transcriptional dynamics of human neutrophils at steady state and upon stress, Nat. Immunol. 23 (10) (2022) 1470–1483.

[10] C. Silvestre-Roig, et al., Neutrophils as regulators of cardiovascular inflammation, Nat. Rev. Cardiol. 17 (6) (2020) 327–340.

[11] M.H. Criqui, et al., Biomarkers in peripheral arterial disease patients and near- and longer-term mortality, J. Vasc. Surg. 52 (1) (2010) 85–90.

[12] H. Vidula, et al., Comparison of effects of statin use on mortality in patients with peripheral arterial disease with versus without elevated C-reactive protein and d-dimer levels, Am. J. Cardiol. 105 (9) (2010) 1348–1352.

[13] D. Shen, L. Fan, J. Li, Analysis of the effect of color Doppler ultrasonography in the diagnosis of arteriosclerotic occlusive disease of lower extremities, Minerva Surg 77 (2) (2022) 188–191.

[14] V. Aboyans, et al., Measurement and interpretation of the ankle-brachial index: a scientific statement from the American Heart Association, Circulation 126 (24) (2012) 2890–2909.

[15] H.C. Rieß, et al., Indicators of outcome quality in peripheral arterial disease revascularisations - a Delphi expert consensus, Vasa 47 (6) (2018) 491–497.

[16] S. Wu, et al., Genome-wide identification of immune-related alternative splicing and splicing regulators involved in abdominal aortic aneurysm, Front. Genet. 13 (2022) 816035.

[17] H. Kim, et al., Pathological gait clustering in post-stroke patients using motion capture data, Gait Posture 94 (2022) 210–216.

[18] E. Ahlqvist, et al., Novel subgroups of adult-onset diabetes and their association with outcomes: a data-driven cluster analysis of six variables, Lancet Diabetes Endocrinol. 6 (5) (2018) 361–369.

[19] M.J. Brusco, E. Shireman, D. Steinley, A comparison of latent class, K-means, and K-median methods for clustering dichotomous data, Psychol. Methods 22 (3) (2017) 563–580.

[20] Y.Q. Huang, et al., Development and validation of a radiomics nomogram for preoperative prediction of lymph node metastasis in colorectal cancer, J. Clin. Oncol. 34 (18) (2016) 2157–2164.

[21] M. Steenman, et al., Identification of genomic differences among peripheral arterial beds in atherosclerotic and healthy arteries, Sci. Rep. 8 (1) (2018) 3940.

[22] C. Parmer, et al., Skeletal muscle expression of adipose-specific phospholipase in peripheral artery disease, Vasc. Med. 25 (5) (2020) 401–410.

[23] M.E. Ritchie, et al., Limma powers differential expression analyses for RNA-sequencing and microarray studies, Nucleic Acids Res. 43 (7) (2015) e47.

[24] P. Gaudet, C. Dessimoz, Gene Ontology: pitfalls, biases, and remedies, Methods Mol. Biol. 1446 (2017) 189–205.

[25] M. Kanehisa, et al., KEGG: new perspectives on genomes, pathways, diseases and drugs, Nucleic Acids Res. 45 (D1) (2017) D353–D361.

[26] G. Yu, et al., clusterProfiler: an R package for comparing biological themes among gene clusters, OMICS 16 (5) (2012) 284–287.

[27] S. Narala, et al., Application of least absolute shrinkage and selection operator logistic regression for the histopathological comparison of chondrodermatitis nodularis helicis and hyperplastic actinic keratosis, J. Cutan. Pathol. 48 (6) (2021) 739–744.

[28] S. Kim, Margin-maximised redundancy-minimised SVM-RFE for diagnostic classification of mammograms, Int J Data Min Bioinform 10 (4) (2014) 374–390.

[29] P. Golpour, et al., Comparison of support vector machine, naïve bayes and logistic regression for assessing the necessity for coronary angiography, Int J Environ Res Public Health 17 (18) (2020).

[30] N. Wang, et al., Identification of SMIM1 and SEZ6L2 as potential biomarkers for genes associated with intervertebral disc degeneration in pyroptosis, Dis. Markers 2022 (2022) 9515571.

[31] F. Dexter, Wilcoxon-Mann-Whitney test used for data that are not normally distributed, Anesth. Analg. 117 (3) (2013) 537–538.

[32] X. Robin, et al., pROC: an open-source package for R and S+ to analyze and compare ROC curves, BMC Bioinf. 12 (2011) 77.

[33] J.I. Kawada, et al., Immune cell infiltration landscapes in pediatric acute myocarditis analyzed by CIBERSORT, J. Cardiol. 77 (2) (2021) 174–178.

[34] Z. Wang, et al., AD risk score for the early phases of disease based on unsupervised machine learning, Alzheimers Dement 16 (11) (2020) 1524–1533.

[35] M.V. Maddali, et al., Validation and utility of ARDS subphenotypes identified by machine-learning models using clinical data: an observational, multicohort, retrospective analysis, Lancet Respir. Med. 10 (4) (2022) 367–377.

[36] Y.B. Yang, et al., A risk predictor of restenosis after superficial femoral artery stent implantation: relevance of mean platelet volume, BMC Cardiovasc. Disord. 20 (1) (2020) 361.

[37] E.M. Willigendael, et al., Influence of smoking on incidence and prevalence of peripheral arterial disease, J. Vasc. Surg. 40 (6) (2004) 1158–1165.

[38] N. Ding, et al., Cigarette smoking, smoking cessation, and long-term risk of 3 major atherosclerotic diseases, J. Am. Coll. Cardiol. 74 (4) (2019) 498–507.

[39] M. Ito, Y. Mishima, [Risk factor, natural history and prognosis of the patients with arteriosclerosis obliterans], Nihon Geka Gakkai Zasshi 97 (7) (1996) 476–480.

[40] Multiple risk factor intervention trial. Risk factor changes and mortality results, Multiple Risk Factor Intervention Trial Research Group. 1982. Jama 277 (7) (1997) 582–594.

[41] J. Thomas, et al., Nutritional status of patients admitted to a metropolitan tertiary care vascular surgery unit, Asia Pac. J. Clin. Nutr. 28 (1) (2019) 64–71.

[42] A.W. Gardner, et al., Dietary intake of participants with peripheral artery disease and claudication, Angiology 62 (3) (2011) 270–275.

[43] D. Demanse, et al., Unsupervised machine-learning algorithms for the identification of clinical phenotypes in the osteoarthritis initiative database, Semin. Arthritis Rheum. 58 (2023) 152140.

[44] J. Ma, et al., Prediction model of laparoendoscopic single-site surgery in gynecology using machine learning algorithm, Wideochir Inne Tech Maloinwazyjne 16 (3) (2021) 587–596.

[45] C.J. Haug, J.M. Drazen, Artificial intelligence and machine learning in clinical medicine, 2023, N. Engl. J. Med. 388 (13) (2023) 1201–1208.

[46] N.B. Boppana, et al., Blockade of CXCR2 signalling: a potential therapeutic target for preventing neutrophil-mediated inflammatory diseases, Exp Biol Med (Maywood) 239 (5) (2014) 509–518.

[47] D. Muthas, et al., Neutrophils in ulcerative colitis: a review of selected biomarkers and their potential therapeutic implications, Scand. J. Gastroenterol. 52 (2) (2017) 125–135.

[48] W.A. Boisvert, L.K. Curtiss, R.A. Terkeltaub, Interleukin-8 and its receptor CXCR2 in atherosclerosis, Immunol. Res. 21 (2–3) (2000) 129–137.

[49] Z. An, et al., Neutrophil extracellular traps induced by IL-8 aggravate atherosclerosis via activation NF-κB signaling in macrophages, Cell Cycle 18 (21) (2019) 2928–2938.

[50] S. Selvaggio, et al., Platelet-to-lymphocyte ratio, neutrophil-to-lymphocyte ratio and monocyte-to-HDL cholesterol ratio as markers of peripheral artery disease in elderly patients, Int. J. Mol. Med. 46 (3) (2020) 1210–1216.

[51] S. Celebi, B. Berkalp, B. Amasyali, The association between thrombotic and inflammatory biomarkers and lower-extremity peripheral artery disease, Int. Wound J. 17 (5) (2020) 1346–1355.

[52] J.M. Kinkade Jr., et al., Differential distribution of distinct forms of myeloperoxidase in different azurophilic granule subpopulations from human neutrophils, Biochem. Biophys. Res. Commun. 114 (1) (1983) 296–303.

[53] H. Maezawa, et al., The histological characteristics and virtual histology findings of the tissues obtained by a distal protection device during endovascular therapy for peripheral artery disease, J. Cardiol. 69 (1) (2017) 125–130.

[54] N. Matijevic, et al., The ARIC carotid MRI study of blood cellular markers: an inverse association of monocyte myeloperoxidase content with peripheral arterial disease, Angiology 62 (3) (2011) 237–244.