scientific reports

Check for updates

OPEN

Establishing a clinical prediction model for diabetic foot ulcers in type 2 diabetic patients with lower extremity arteriosclerotic occlusion using machine learning

Yubo Wang[1,8], Chunyu Jiang[2,8], Yi qi Xing[1,8], Linxuan Zou[1], Mingzhi Song[1], Xueling Qu[3], Zhuqiang Jia[4,5], Lin Zhao[6], Xin Han[4,7✉], Junwei Zong[1✉] & Shouyu Wang[1✉]

The burden of diabetic foot ulcers (DFU) is exacerbated in diabetic patients with concomitant arteriosclerotic occlusion disease (ASO) in the lower extremities, who experience more severe symptoms and poorer prognoses. The study aims to develop a predictive model grounded in machine learning (ML) algorithms, specifically tailored to forecast the occurrence of DFU in diabetic patients with lower extremity ASO. The study involves the data from diabetic patients diagnosed with lower extremity ASO from January 1, 2011 to August 31, 2023. We conducted quality control on the data. Subsequently, the dataset was divided into a training set comprising patients before 2020 and a validation set comprising patients in 2020 and onwards. Patients were stratified into the DFU group or the non-DFU group based on the occurrence of DFU. Intergroup comparisons were conducted to analyze the differences between these two groups. Logistic regression analyses, 3 kinds of machine learning algorithms, a predictive model and nomogram was formulated to estimate the risk of DFU occurrence among diabetic patients with lower extremity ASO. Internal validation of the model was undertaken using the bootstrap method, combing with external temporal validation, with the results visually presented through the Receiver Operating Characteristic (ROC) curve and the Calibration curve. To evaluate the clinical practicality of the model, Decision Curve Analysis (DCA) and Clinical Impact Curve (CIC) were employed. Body Mass Index (BMI), hypertension, coronary heart disease, diabetic nephropathy, the number of lower leg artery occlusions, controlling glucose by insulin injection, age, number of cigarettes smoked per day, diastolic blood pressure, and C-reactive protein (CRP) were utilized to construct a clinical prediction model. This model exhibited a high predictive performance (AUC = 0.962), and the results from both internal validation and external temporal validation further confirmed its high accuracy and reproducibility (AUC = 0.968 and AUC = 0.977, respectively). Additionally, DCA and CIC demonstrated the high clinical practicality of this model. The clinical prediction model exhibited excellent accuracy and reproducibility, along with broad clinical practicality. It provides a good reference for the diagnosis and treatment of DFU.

Keywords Diabetic foot ulcer, Arteriosclerotic occlusion, Risk factors, Clinical prediction model, Nomogram

Diabetic foot ulcer (DFU) refers to the occurrence of infection, ulceration or destruction of tissue in the foot of diabetes patients, usually accompanied by neuropathy of the lower extremities and (or) vascular lesions of

[1]Department of Orthopaedic Surgery, The First Affiliated Hospital of Dalian Medical University, Dalian, China. [2]Department of Trauma Surgery, The Second Affiliated Hospital of Harbin Medical University, Dalian, China. [3]Pelvic Floor Repair Center, Dalian Women and Children Medical Center (Group), Dalian, China. [4]Naqu People's Hospital, Tibet, China. [5]The First Affiliated Hospital of Dalian Medical University, Dalian, China. [6]Department of Quality Management, Dalian Municipal Central Hospital, Dalian, China. [7]Department of Orthopaedic Surgery, The Second Affiliated Hospital of Dalian Medical University, Dalian, China. [8]Yubo Wang, Chunyu Jiang and Yi qi Xing contributed equally to this work and share first authorship. ✉email: hanxin1973@163.com; aweizone@163.com; wangshouyu666@126.com

nature portfolio

the lower extremities, a serious diabetic complication. Since 1980, people have been paying attention gradually, and the incidence rate in our country is increasing year by year. According to the latest statistics released by the International Diabetes Federation (IDF), there are currently about 537 million people living with diabetes worldwide[1], 15 to 25 percent of people with diabetes are at risk of developing DFU[2]. DFU wounds are often prolonged, wound enlargement, infection and even lower limb amputation. As the leading cause of non-traumatic amputation of the lower extremity, approximately one in five DFU patients will need to undergo amputation[3]. About 10 percent of DFU patients die within 1 year and 40 percent within 5 years, and the extremely high mortality rate of this disease even exceeds that of certain malignancies[4]. DFU not only brings suffering to patients and their families, but also puts a heavy burden on social economy and medical resources[5]. Therefore, in clinical work, active prevention, screening of high-risk patients, treatment of underlying diseases and foot wounds are the key tasks to improve the efficacy of DFU and reduce the disability rate and mortality.

Large vessel artery disease dysfunction caused by type 2 diabetes (T2DM) are considered to be important reasons for the formation and treatment of DFU[6]. Arteriosclerotic occlusion (ASO) in the lower extremity is a common complication of diabetes[7]. ASO in diabetic patients is often manifested as multiplanar occlusion of lower limb vessels, in which femoral artery and lower leg artery are dominant. The incidence of ASO in diabetic patients is as high as 50%[8]. Due to the age, obesity and weakness of diabetic patients leading to less activity, coupled with limb paresthesia and other characteristics, some typical symptoms of lower limb ischemia, such as lameness and resting pain, are often covered up in the early stage of ASO[9]. In addition, DFU progresses rapidly in diabetic patients combined with ASO of the lower extremities, so that patients often have a large area of foot ulcers and gangrene when they seek treatment. Even after successful reconstruction of arterial blood flow of the lower extremities, some patients still face the problem of foot wounds that are difficult to heal. Studies have shown that 20% to 70% of lower extremity ASO patients will be complicated by foot ulcers, of which 35% of ulcers progress and lead to amputation due to the lack of active and effective intervention[8]. Therefore, predicting the risk of DFU in diabetic patients with lower extremity ASO is helpful for early diagnosis and treatment and prevention of DFU.

The clinical prediction model related to DFU is currently in its initial stage. Boyko et al.[10] conducted a prospective cohort study on 1285 patients with diabetes, and the results showed that HBA1c, visual impairment, neuropathy, foot infection history, foot ulcer history and amputation history were related factors for the occurrence of DFU, and established a prediction model. This model can predict the occurrence of DFU within 1 and 5 years. However, this study lacked external validation and included only older men. Heald et al.[11] followed up 16,852 patients with diabetes in primary health clinics through logistic multivariate regression, and finally obtained five risk factors, namely HBA1c, age, neuropathy, creatinine, and cerebral infarction history. However, the prediction ability of the model was only moderate. Although there is a strong association between ASO and DFU, this link is often underestimated in medical practice. At present, there are few studies on the risk factors of DFU in diabetic patients with lower extremity ASO and the establishment of related prediction models. Therefore, this study will focus on diabetic patients with ASO of lower limb arteries, and combine machine learning (ML) algorithm to screen factors related to DFU occurrence of ASO of lower limb arteries, establish a clinical prediction model and nomogram, and verify its accuracy and clinical practicality. It can be used for the early screening of DFU and help clinicians make decisions for the next intervention, which is conducive to the prevention of further DFU formation in diabetic patients with ASO of lower limb arteries and the reasonable allocation of medical resources.

Methods

Participants

The participants of this study were diabetic patients with lower extremity ASO who were admitted to the First Affiliated Hospital of Dalian Medical University from January 1, 2011 to August 31, 2023. This study is a retrospective study. All methods were performed in accordance with the relevant guidelines, and has been approved by the Clinical Ethics Committee of the hospital (approval No. PJ-KS-KY-2024-32) and due to the retrospective nature of the study. Clinical Ethics Committee of the hospital waived the need of obtaining informed consent. The composition and review procedures of the Ethics Committee of the First Affiliated Hospital of Dalian Medical University are in line with the Helsinki Recommendations, the International Ethical Guidelines for Human Biomedical Research, the Measures for Ethical Review of Biomedical Research Involving Humans, the Measures for Ethical Review of Life Science and Medical Research Involving Humans, the Quality Management Standards for Drug Clinical Trials, and the Quality of Medical device Clinical Trials International ethical principles such as Management Code and Guiding Principles for the Ethical Review of Drug Clinical Trials and relevant regulations and guiding principles in China are required. The ethical review and work process shall not be influenced by any organization or individual other than the Ethics Committee, and shall exercise its duties and rights in accordance with the above norms. Inclusion criteria were: (1) Diabetic patients diagnosed with lower extremity ASO; (2) hospitalized for more than 72 h; (3) Complete clinical data, including demographic characteristics, medical history information, lower extremity vascular CTA and vascular Doppler ultrasonography, and laboratory results. The exclusion criteria were: (1) Gestational diabetes mellitus; (2) Type 1 diabetes; (3) Duplicate medical records of patients due to multiple hospitalizations; (4) Patients with incomplete clinical data; (5) Critical death cases.

Study design

According to the definition of DFU, the diagnostic criteria are "infection, ulceration, or tissue destruction in the foot of a person who has been diagnosed with diabetes for the first time or has a history of diabetes, usually accompanied by lower extremity neuropathy and/or lower extremity vascular disease." Taking January 1, 2020 as the boundary, the study subjects were divided into training set and validation set according to the time of visit,

and were divided into DFU group and non-DFU group according to whether they had DFU. First, the differences in indicators between the DFU group and the non-DFU group were analyzed. Then Univariate and multivariate logistic regression analyzed the risk factors for DFU in diabetic patients with lower limb arteriosclerotic occlusion. The machine learning algorithm was used to further screen the risk factors and establish a clinical prediction model, which was visualized by nomogram to verify and evaluate the prediction ability and clinical practicability of the model. The research flow chart is shown in Fig. 1.

Data quality control

All data comes from the hospital's clinical data platform, which has concealed patients' personally identifiable information, and the collection and collation process of all clinical data is carried out by trained clinical reviewers. This study is a retrospective study, combined with the actual clinical situation, previous diagnosis and treatment experience, and the results of previous literature studies[10–12], we collected demographic characteristics (height, weight, BMI, age, sex, education, smoking and drinking history), medical history information (chief complaint, present history, past medical history, comorbidities, signs and symptoms), lower extremity vascular CTA and vascular Doppler ultrasonography, and laboratory results. In this study, the results of the first examination after admission were selected for statistical analysis, due to the different ways of conducting laboratory tests and examinations, the results of laboratory tests are the first within 24 h, and the results of imaging examinations are the first results of the entire hospital stay because some examinations require an appointment. Due to the long time span of this study, there are different units of the same index, such as the biochemical index of liver and kidney function, which is converted into a unified unit. We perform an outlier test to remove outliers that fall outside the 1%–99% range. Indicators whose data is missing by more than 20% were deleted, and those with less than 20% missing values were processed by multiple imputation (MI) method. Whether continuous variables were converted into categorical variables was determined by the actual clinical significance. For example, when
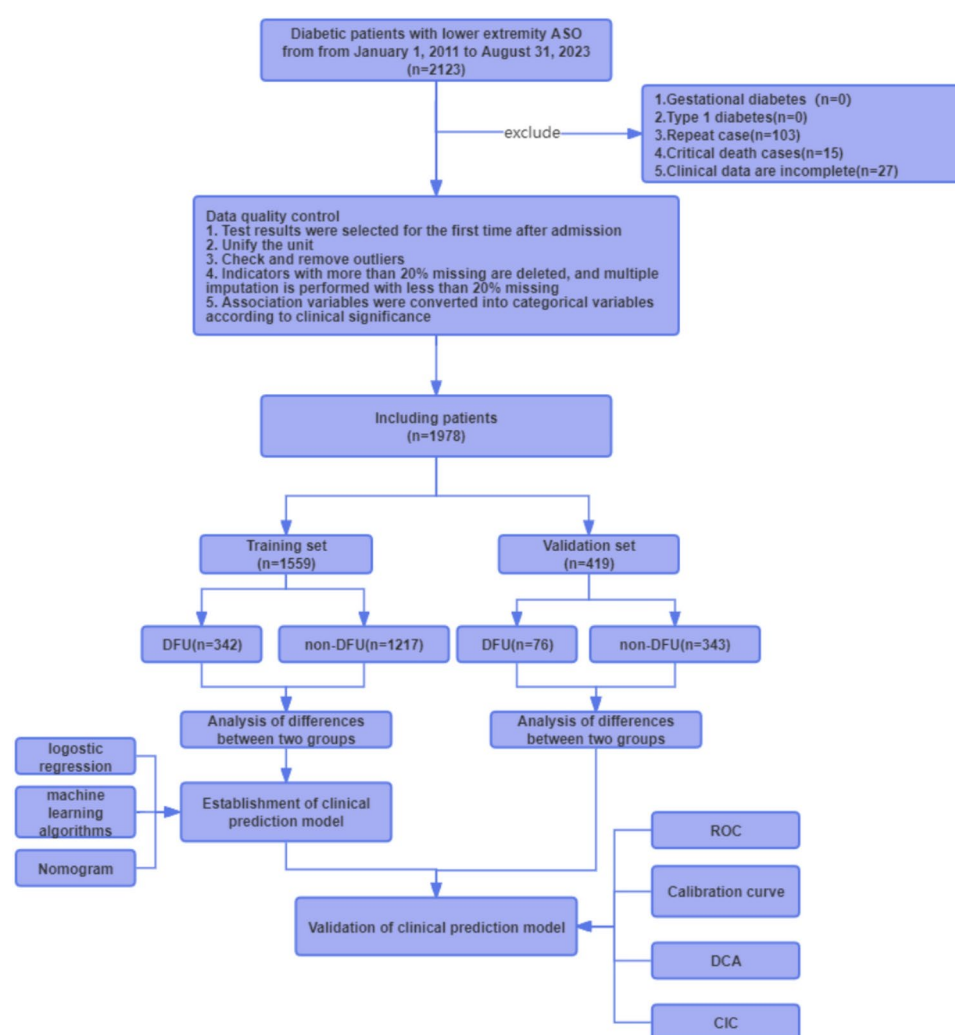
Fig. 1. Flowchart shows the process of data analysis and establishment of clinical prediction model in this study.

a continuous variable increased or decreased, indicating different clinical significance, it was converted into categorical variables, such as blood glucose.

Data analysis

Continuous variables with normal distribution were presented in the form of mean and standard deviation (SD). Those that do not conform to the normal distribution are represented by median [M(P25,P75)], and categorical variables are represented by "percentages". The t test (continuous variables conforming to normal distribution), Kruscal-Whallis H test (continuous variables not conforming to normal distribution), χ^2 test, or Fisher exact test (categorical variables) were used. Univariate and multivariate logistic regression, least absolute selection and shrinkage operator (LASSO), support vector machines (SVM), random forest (RF) machine learning algorithm were used to identify DFU independent risk factors and establishes a nomogram for prediction. The internal validation Bootstrap method (1000 resamples) is combined with the external period validation to ensure the accuracy of the validation. The validation results are presented by Receiver operating characteristic curve (ROC) and calibration curve. Decision curve analysis (DCA) and Clinical impact curve (CIC) were used to evaluate the clinical utility of the model[13–15]. R software (version 4.2.1R) was used for all statistical analyses. "naniar" package for checking missing values, with "mice" package for MI, "boot" package for significance difference, "MASS" package for predictor screening, "glmnet" package for LASSO analysis, "randomForest" package for RF analysis, "msvmRFE" package for SVM analysis, "glmnet" package for model establishing, "pROC" package for ROC, "rms" package for calibration plot, "regplot" package for nomogram diagram, "rmda" package for DCA and CIC. P value < 0.05 was considered statistically significant.

Results

Demographic and clinical data differences analysis

A total of 2123 diabetic patients with lower extremity ASO were collected in this study. According to the medical identification in the case information, 103 repeated cases were removed, 15 cases of critical death and 27 cases of missing examination results. There were 1978 eligible cases, of which 418 were DFU patients, accounting for 21.1% of the total. According to the treatment time, the subjects were divided into training set (n = 1559) and validation set (n = 419). In the training set, 342 cases were in the DFU group, 1217 cases were in the non-DFU group, 76 cases were in the validation set, and 343 cases were in the non-DFU group. A total of 54 indicators were included in this study, indicators with statistical deletion greater than 20% (D-Dimer, ALP, PCT) were deleted and finally 51 indicators were included for statistical analysis (supplementary information 1, supplementary information 2 Figure S1). Smoking, BMI, age, smoking time, number of cigarettes smoked per day, course of diabetes, history of hypertension, history of coronary heart disease, proportion of diabetic nephropathy, systolic blood pressure, diastolic blood pressure, proportion of swelling of lower limb, proportion of iliac artery occlusion, and proportion of femoral artery occlusions, number of lower leg artery occlusion, proportion of controlling glucose by insulin injection, proportion of controlling glucose by oral medicine, albumin (ALB), uric acid (UA), C-reactive protein (CRP), erythrocyte sedimentation rate (ESR), lymphocyte percent (Lymph p), hemoglobin (Hb), globulin (GLO) and neutrophil percent (Neut p) showed statistically signifcant differences in both Training set and Validation set (p < 0.05). Among them, the average age of DFU patients in the Training set was 78.7 years, and that of Non-DFU patients was 70.4 years. The mean age of DFU patients in the Validation set was 74.5 years and that of Non-DFU patients was 70.1 years. The proportion of smoking history was 62.9% in DFU patients and 38.5% in Non-DFU patients in Training set. 67.1% of DFU patients in Validation set and 40.8% of Non-DFU patients had a history of smoking. The mean BMI of DFU patients in Training set was 28.1 and 23.4, respectively. The average BMI of DFU patients in Validation set was 28.5, and that of Non-DFU patients was 23.5. The proportion of Hb reduction in DFU patients in Training set was 64.3%, and that in Non-DFU patients was 40.9%. In Validation set, 67.1% of DFU patients and 42.3% of Non-DFU patients had reduced hemoglobin. In Training set, 88.3% of DFU patients and 58.9% of Non-DFU patients had a history of hypertension. 89.5% of DFU patients in Validation set and 61.2% of Non-DFU patients had decreased hemoglobin. 52.6% of DFU patients in Training set and 7.7% of Non-DFU patients had Diabetic nephropathy. The proportion of DFU patients with a history of Diabetic nephropathy in the Validation set was 52.6%. The percentage of Lower leg artery occlusion was 96.2% in DFU patients and 70.5% in Non-DFU patients. The percentage of DFU patients with Lower leg artery occlusion in the Validation set was 97.4%. In the Training set, 67.3% of DFU patients and 34.9% of Non-DFU patients used insulin to control blood glucose. In the Validation set, 63.2% of DFU patients and 37.3% of Non-DFU patients used insulin for glycemic control. The demographic and clinical data differences between DFU and non-DFU are shown in supplementary information 2 Table S1.

Univariate logistic regression and multivariate logistic analysis of DFU independent correlation factors

Univariate logistic regression and Multivariate logistic regression were used to screen the predictors. Results showed age, hypertension, diastolic blood pressure, coronary heart disease, cerebral infarction, diabetic nephropathy, lower limb swelling and numbness, number of cigarettes smoked per day, number of lower limb artery occlusion segments, iliac artery occlusion, femoral artery occlusion, lower leg artery occlusion, number of lower leg artery occlusion, controlling glucose by insulin injection, BMI, diabetes course, hemoglobin (Hb), C-reactive protein (CRP) were independent correlation factors of DFU (P < 0.05), and the analysis results were shown in supplementary information 2 Table S2, supplementary information 2 Table S3 and Fig. 2.
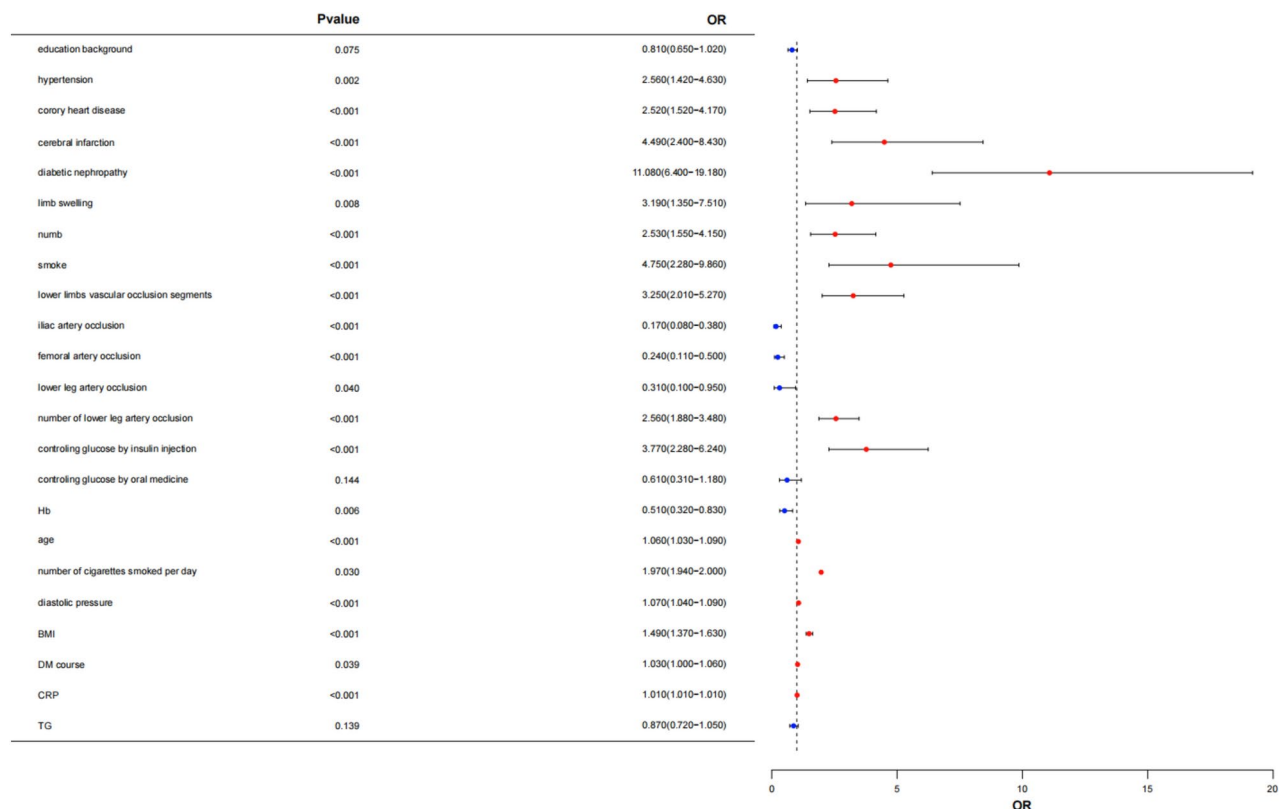
Fig. 2. Multivariate logistic regression analysis of DFU-related risk factors in diabetic patients with ASO of the lower extremities.

Clinical prediction model and nomogram were established by combining machine learning algorithm

We further analyzed independent correlation factors using LASSO, SVM and RF ML algorithms, and the results were intersections to obtain 10 correlation factors, which were BMI, hypertension, coronary heart disease, diabetic nephropathy, number of lower leg artery occlusion, controlling glucose by insulin injection, age, number of cigarettes smoked per day, diastolic blood pressure and CRP, all were independent risk factors, as shown in Fig. 3. We constructed a clinical prediction model consisting of these 10 characteristic independent risk factors, the linear combination of models is expressed as Eq. (1):

$$
\begin{aligned}
Z = &-24.79763 + 0.69377 * (\text{hypertension}) \\
&+ 1.16253 * (\text{conory heart disease}) \\
&+ 2.38991 * (\text{diabetic nephropathy}) \\
&+ 0.87378 * (\text{number of lower leg artery occlusion equals 1}) \\
&+ 2.15340 * (\text{number of lower leg artery occlusion equals 2}) \\
&+ 2.74218 * (\text{number of lower leg artery occlusion equals 3}) \\
&+ 1.43573 * (\text{controlling glucose by insulin injection}) \\
&+ 0.06383 * (\text{diastolic pressure}) \\
&+ 0.36660 * (\text{BMI}) + 0.01107 * (\text{CRP})
\end{aligned}
\tag{1}
$$

DFU risk probability (P) is expressed as Eq. (2):

$$
(\text{P}) = \frac{1}{1 + e^{-Z}}
\tag{2}
$$

e is the base of the natural logarithm (about 2.718). The composition and coefficients of the model are shown in supplementary information 2 Table S4. And we presented it as a nomogram, as shown in Fig. 4, the score of the first action item, the total score of the second-to-last action, the probability of the last action occurring DFU, and each remaining line represents a related factor. According to the value of a single factor of a certain research object, the corresponding score in the first row is found at the position of the line segment, and the total score
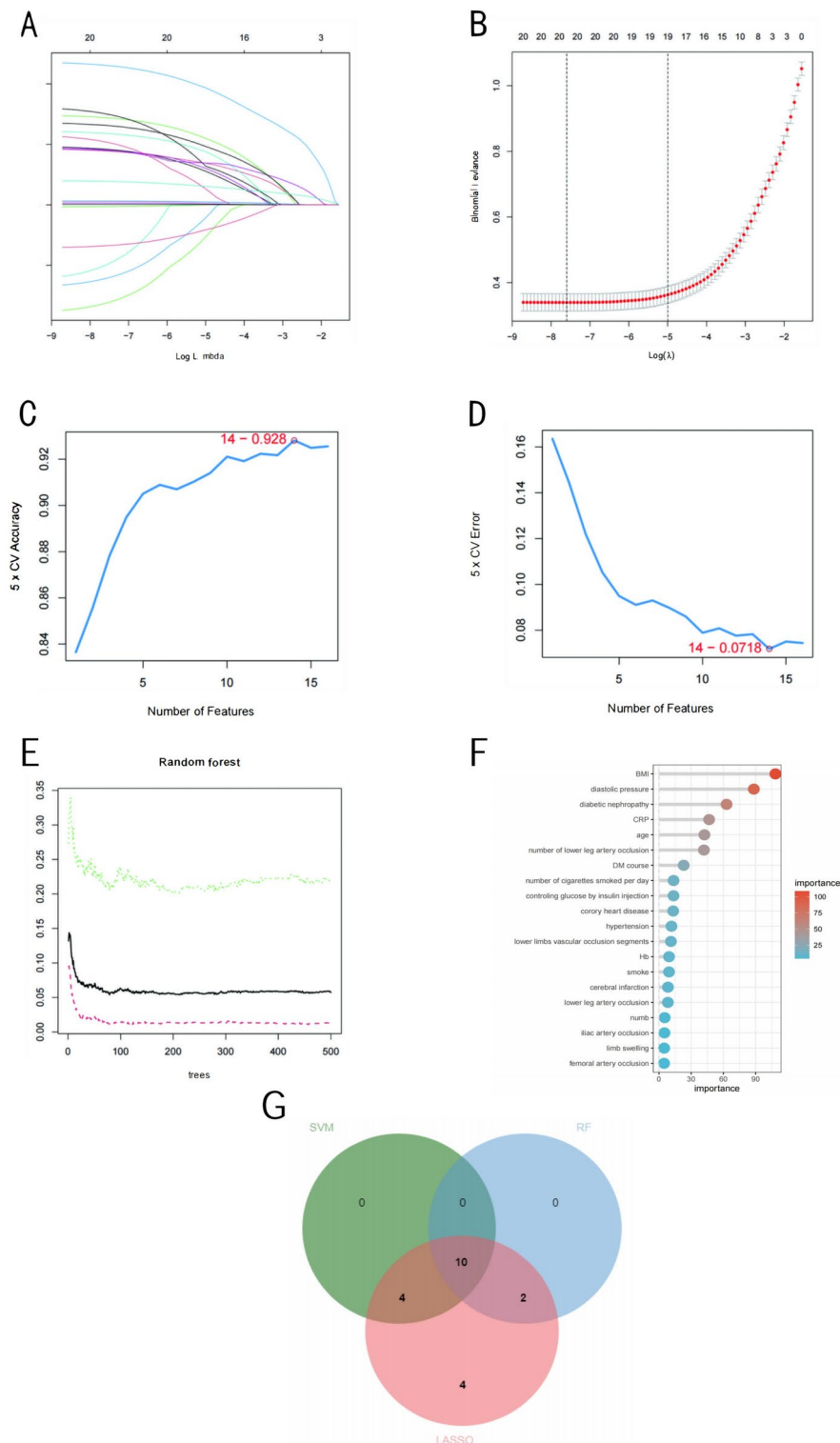
Fig. 3. Machine learning screening DFU characteristic correlation factors. LASSO algorithm (**A**, **B**) obtained 20 characteristic correlation factors. The dashed line on the left is λ min, and the corresponding λ of this dashed line is 20, which means that under this λ value, the model deviation is minimum and the fitting effect is highest. Results of SVM algorithm (**C**, **D**) obtained 14 characteristic correlation factors. The accuracy was highest when 14 factors were used, and the error rate was lowest when 14 factors were used. RF algorithm (**E**, **F**) ranks the importance of all correlation factors included. The horizontal coordinate represents importance. Combined with clinical significance, this study selected a total of 12 relevant factors with an importance greater than 10. Intersection of the results of three machine learning algorithms to obtain 10 related factors (**G**).
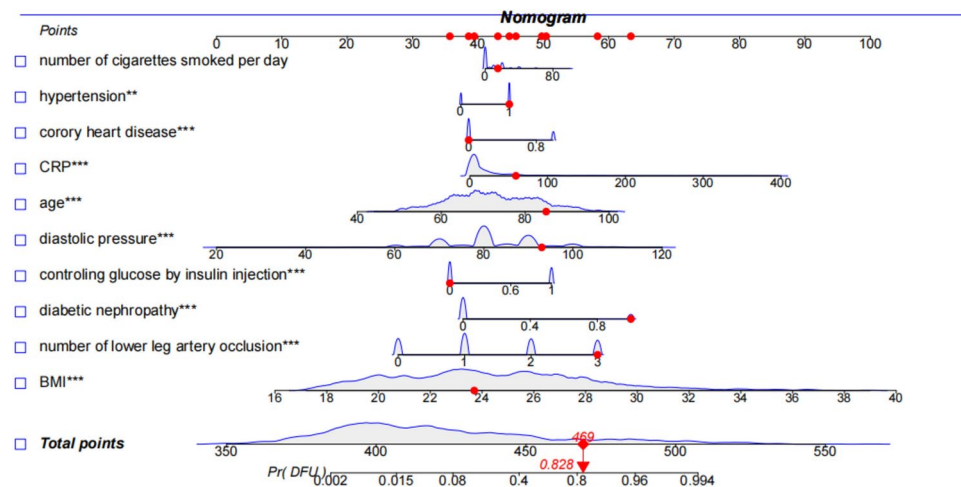
Fig. 4. Nomogram of DFU risk in diabetic patients with ASO of lower extremities. The red points in the figure represent the actual values of each index of a patient. The patient's CRP was 50 mg/L, BMI was 24, diastolic blood pressure was 95 mmHg, he smoked about 10 cigarettes a day, was 84 years old, did not use insulin, had three lower leg arteries occlusion, had diabetic nephropathy, had a history of hypertension, and had no history of coronary heart disease. The final score was 469 points. The probability of DFU occurrence is 0.828. * represents the significance of the factor's contribution to the model.
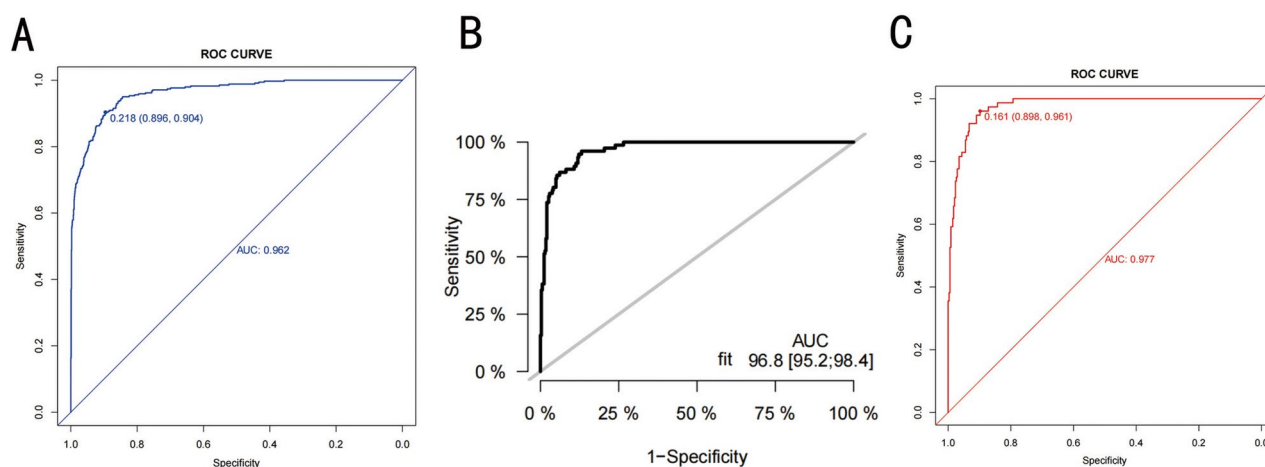
Fig. 5. ROC curve with AUC of training set (**A**), internal validation using the bootstrap method (resampling = 1000) (**B**), and validation set (**C**).

is added by adding the single score, and the probability of occurrence of DFU is obtained by corresponding to the last row.

Validation of predictive model efficacy and evaluation of clinical practicability

ROC curve results of the prediction model established in this study showed that the area under the curve (AUC) was 0.962, the specificity was 0.896, and the sensitivity was 0.904 (Fig. 5A). All three indexes were high, and the calibration curve was highly coincident with the ideal result (Fig. 6A), indicating that the model had excellent prediction ability. In this study, the effectiveness of the model was verified through the combination of internal bootstrap validation and external temporal validation of the validation set. The internal validation results showed that the area under the curve (AUC) reached 0.968, and the calibration curve was highly coincident with the ideal result (Figs. 5B, 6B). The results of external temporal validation showed that the ROC curve of the prediction model showed that the AUC reached 0.977, the specificity was 0.898, and the sensitivity was 0.961 (Fig. 5C). All three indexes were high, and the calibration curve was highly coincident with the ideal result (Fig. 6C). The results of internal validation and external validation confirm that the model has excellent prediction efficiency and strong repeatability. The DCA curve and CIC curve were used to demonstrate the clinical practical value of the prediction model. According to the DCA curve of the prediction model and the DCA curve verified by external time periods (Fig. 7A and B), the net benefit corresponding to the curve was above 0 in a wide range of decision thresholds (0.2–1.0). And it is far from the two extreme curves of "None" and "All". The CIC curves
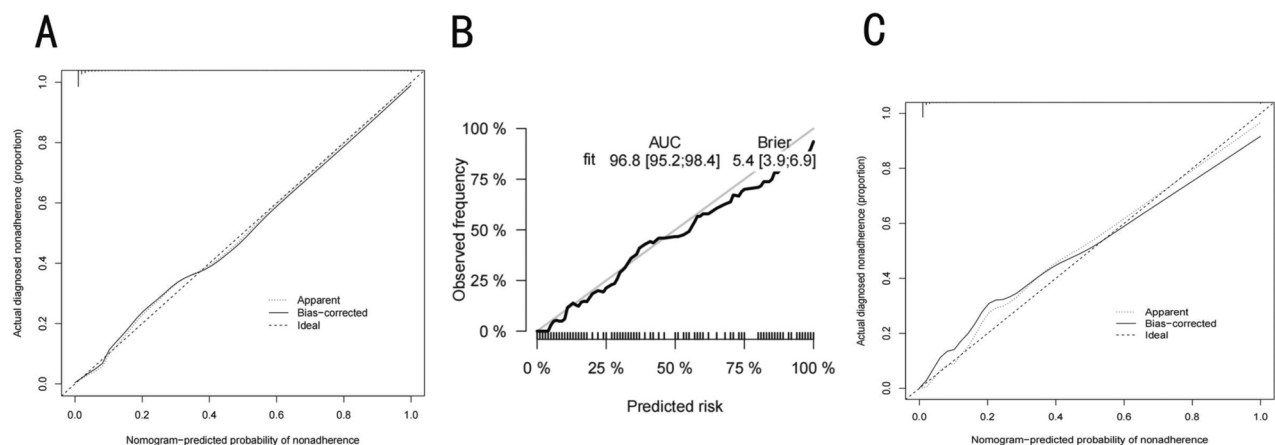
Fig. 6. Calibration plots of training set (**A**), internal validation using the bootstrap method (resampling = 1000) (**B**), and validation set (**C**). The black dashed diagonal line of calibration plots indicates the perfect prediction of the ideal model. The black solid line represents the performance of the nomogram.
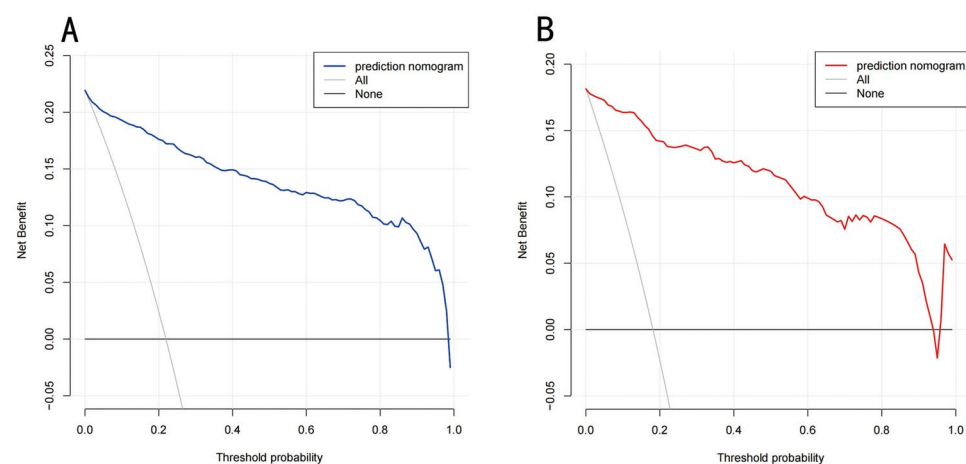
Fig. 7. DCA of the training set (**A**) and validation set (**B**). The horizontal coordinate represents the risk threshold, and the vertical coordinate corresponding to the curve is the net benefit of taking treatment measures. When the risk threshold is in a certain range, the greater the corresponding net benefit, and the farther away from the two extreme oblique lines of "None" and "All", the better the applicability of the model in making clinical decisions.

of the prediction model and external time period validation (Fig. 8A and B) show that the results predicted by the model are close to the actual results in a wide range of risk thresholds (0.2–1.0 and 0.4–1.0), which indicate practical application value in a wider range of clinical situations.

Discussion

According to the latest epidemiological statistics, the number of people with diabetes worldwide has climbed to a staggering 537 million, a figure that underscores the importance of diabetes as a global public health challenge[1]. The total number of diabetics in China has reached 116 million, accounting for about a quarter of the global total[16]. Between 9.1 million and 26.1 million people with diabetes progress to DFU each year[8], this complication is marked by a poor prognosis and a heavy socio-economic burden. According to statistics, up to 17% of newly diagnosed DFU patients require minor amputations, and 5% of patients have to face major amputations within one year of the onset of the disease, and the risk of amputation is significantly increased in patients with lower limb ischemia[17]. As the overall incidence of diabetes has increased and the life expectancy of patients with diabetes has increased, the proportion of hospitalized patients with DFU has also shown an increasing trend. According to the study, about 20 percent of DFU patients require hospitalization[18]. The significant increase in hospitalization costs for DFU patients is an issue that cannot be ignored for both the healthcare system and the social economy[19].

ASO in diabetic patients was independently associated with DFU occurrence and risk of amputation[20]. ASO significantly accelerated the progression of diabetic patients to DFU. Due to the lack of clear specific symptoms
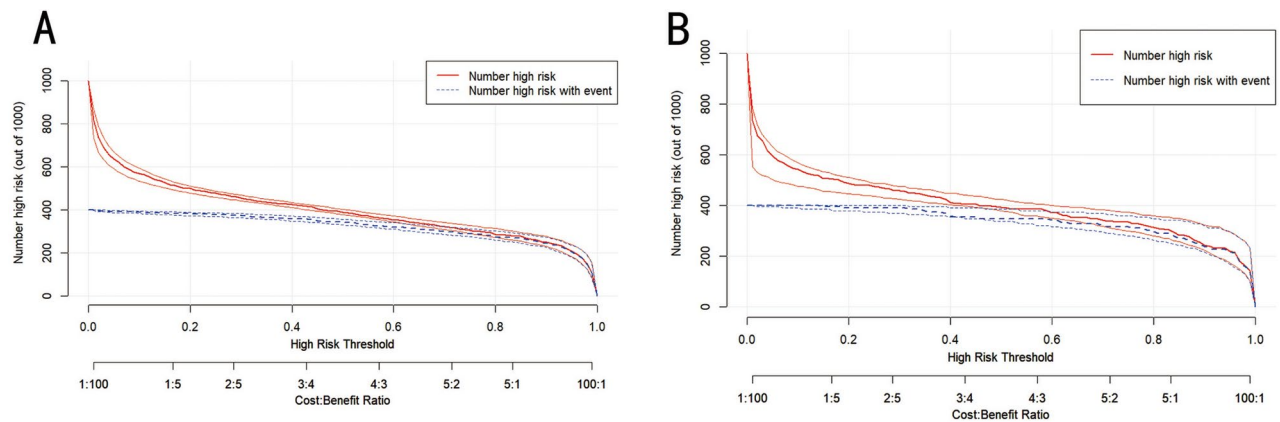
Fig. 8. CIC of the training set (**A**) and validation set (**B**). The red curve (Number high risk) represents the number of people classified as positive by the model under each risk threshold (horizontal coordinate). The blue curve (Number high risk with event) shows the number of true positives under each threshold. When the risk threshold is within a certain range, the corresponding prediction model determines that the positive population is highly close to the true positive population, which confirms that the prediction model has high clinical prediction ability and practical value within this range.

of DFU at the early stage and the lack of awareness of early screening, DFU patients with ASO of lower extremities tend to present more severe pain, infection and osteomyelitis than other DFU patients, and have a poor prognosis[3]. Treatment strategies for DFU patients with lower extremity ASO often include distal vascular reconstructive surgery or angioplasty. However, even after revascularization, patients still need to undergo long-term treatment and follow-up to manage the ulcerated wounds that have developed. For those patients who are not suitable for vascular surgery due to poor vascular conditions and other reasons, the wound treatment of DFU faces greater challenges. Therefore, early screening of DFU is particularly important for diabetic patients with lower limb ASO. By collecting a large number of clinical data and combining univariate and multivariate Logistic regression, 10 characteristic DFU-related independent risk factors including BMI, hypertension, coronary heart disease, diabetic nephropathy, number of lower leg artery occlusion, controlling glucose by insulin injection, age, daily smoking, diastolic blood pressure and CRP were finally screened by combining various ML algorithms, and a prediction model was established.

BMI is widely used in the public health field as an important indicator to assess the degree of obesity and health status of the population in a country or region. A recent Mendelian randomized analysis involving risk factors related to DFU, diabetic polyneuropathy (DPN) and diabetic peripheral arterial disease (DPAD) evaluated the value of BMI in the public health field. The results showed that BMI was a common independent risk factor for diabetes peripheral vascular disease, diabetic neuropathy and DFU[21], a meta-analysis showed that BMI > 24.5 was an independent risk factor for new DFU[22]. Other studies have shown that the risk of DFU increases with the increase of BMI in the 25–34.9 range[23], for every 20 kg increase in body weight, the risk of foot ulcers increases by 20%[24]. As shown in this study, the increase of BMI is an independent risk factor for DFU, and the score increases gradually with the increase of BMI in the column chart. From the perspective of biomechanics, the increase of BMI increases the pressure of the human body on the foot. Meanwhile, obesity is thought to increase the risk of foot deformities[25], from pathophysiological perspective, obesity leads to changes in skin barrier function. Lipid deposition in obese patients leads to the occurrence and aggravation of peripheral vascular occlusion. At the same time, the increase of BMI and obesity may lead to the occurrence of diabetic microvascular complications, skin barrier function and microcirculation disorders[26,27]. Foot ischemia increases tissue loss, increases foot pressure, changes in anatomical structure, and decreases skin barrier function, leading to an increased risk of foot trauma forming ulcers. However, there are differences between the results of this study and those of some studies regarding people with too low BMI[28], current literature has not fully explained the relationship between such people and DFU. By comparing this study with previous studies in terms of patient population characteristics and research methods, we believe that the reason for this difference may be that the proportion of such people in existing studies is too small, and small samples cannot truly reflect the actual situation. Further large-scale studies are needed to clarify the relationship between low BMI and DFU. Smoking is a risk factor for the occurrence and progression of ASO in the lower limbs, and a large amount of smoking will accelerate the progression of ischemia in the lower limbs, resulting in tissue necrosis. Studies have shown that smoking is closely related to the degree of DFU lesions and the risk of amputation[29,30]. Age is also one of the risk factors for the occurrence of DFU. In elderly patients, abnormal metabolism, decline of immune function, aging of blood vessels, nerves and immune system are often combined with a variety of underlying diseases, such as hypertension and coronary heart disease, leading to ischemia, infection and ulcers. At the same time, in elderly diabetic patients, long-term hyperglycemia damages blood vessels and nerves, leading to lower extremity artery stenosis, occlusion and severe neuropathy, and ultimately leads to the occurrence of DFU. Hypertension is one of the main complications of diabetes. In this study, 65.57% of the patients complicated with hypertension accounted for all the subjects. Arterial wall thickening and elasticity gradually decreased in

patients with hypertension, thus reducing the compliance of blood vessels[31]. In addition, high blood pressure also causes damage to the ductal skin cells, which not only weakens their barrier function, but also inhibits the ability and bioavailability of NO, an important molecule that maintains vasomotor balance and inhibits arteriosclerosis[32]. The combination of these factors leads to further destruction of blood vessel wall and reduced blood supply in ASO patients of the lower extremities. In addition, we also found that the increase of diastolic blood pressure is also an independent risk factor for DFU. The increase of diastolic blood pressure can reflect the increase of peripheral vascular resistance, which is more detrimental to the blood supply of lower limbs, and aggravates foot ischemia and tissue necrosis. This study also found that other diabetic comorbidities such as coronary heart disease and diabetic nephropathy may lead to the development of DFU. In patients with coronary heart disease, the heart pumping function is weakened, the lower limb perfusion is insufficient, and the tissue hypoxia is aggravated. Kidney disease can lead to edema of the lower extremities, resulting in reduced local perfusion of the skin, resulting in ischemia[33], patients with diabetic kidney disease who receive dialysis for a long time will have peripheral neuropathy, the atrophy of nerve fibers leading to changes in the structure of the foot, the most typical of which is the removal of the protective fat pad from under the metatarsal head, the loss of the so-called "protective pad", which increases the risk of ulcer formation[34]. Despite similar results from previous studies[35]. The association between insulin and DFU has not been widely reported or thoroughly studied. In clinical work, we found that patients who used insulin to control blood sugar generally had poor blood sugar control in the past, ineffective oral drugs for sugar control, or delayed diabetes treatment. Therefore, we suggest that the potential reasons why insulin use is a risk factor for the development of DFU are the higher severity of diabetes and the later timing of diabetes intervention. The anatomical structure of lower leg artery is characterized by small diameter, so ASO of lower extremity in diabetic patients is more likely to lead to lower leg artery occlusion than in non-diabetic patients[36]. Therefore, the occlusion of lower leg arteries can reflect the lower limb blood circulation of diabetic patients with lower limb ASO. Our study showed that the more lower leg lesions involved, the higher the severity of ischemia, and the greater the possibility of DFU. CRP can reflect the level of inflammation in the body, and chronic inflammation is also the manifestation of DFU[37], vascular endothelial injury in diabetic patients with ASO of lower extremity and long-term exposure of foot tissue to high glucose environment will lead to the increase of the level of inflammation in the body. On the other hand, the occurrence of diabetic foot in such patients is often accompanied by severe osteomyelitis and tissue necrosis of the foot, which usually causes bacterial infection[38]. Some studies have shown that serum CRP levels are increased in patients with DFU compared with those without DFU. CRP levels were determined to be the most reliable indicator of DFU infection and the extent of infection[39,40], this study also found that CRP level in the DFU group was significantly higher than that in the non-DFU group, and after univariate and multi-factor analysis and screening by machine learning, only CRP level among the indicators related to inflammation and infection was an independent risk factor for the occurrence of DFU.

In this study, a variety of machine learning algorithms were combined to further screen the results of multi-factor Logistic regression. Logistic regression has the problem of weak feature selection ability and poor generalization ability due to underfitting[41]. Therefore, on the basis of Logistic regression, this study selected three widely used machine learning algorithms, namely LASSO, SVM and RF. By setting λ, Lasso gradually reduced the variables to 0, selected the most characteristic variables, and simplified the model to improve the generalization ability[42]. The number of variables corresponding to λ-min has the best fitting effect, with more variables and complex model, while the number of variables corresponding to λ-se is less. The model is brief and convenient for clinical use, but the fitting effect is weaker than lambda-min[43]. In this study, the LASSO results λ-min and λ-se were 20 and 19 respectively, and the difference was small, so we chose λ-min to ensure the fitting effect. RF can reduce the complexity of the model while preventing overfitting and improving the ability to generalize[44]. However, this algorithm is highly dependent on the correlation of input factors. If the correlation of each factor is low, the authenticity and interpretation of the result will be affected[45]. Therefore, this study combined clinical practical significance with the importance of RF results of each factor to comprehensively evaluate. In order to improve the authenticity and interpretability of the model, we further carry out SVM algorithm, this algorithm can get a result with the minimum cumulative error between the real and the result[46]. Another advantage of this algorithm is that it has better prediction performance for external data. This algorithm also has drawbacks, such as taking too long to compute when the amount of data is large. The prediction model established in this study has been verified by a combination of internal and external validation methods, and its good prediction ability and repeatability have been determined. Its clinical practical value has been determined by the DCA and CIC curves. Meanwhile, the presentation of the nomogram also makes the model more intuitive and easy for clinicians and patients to understand and use. For diabetic patients with ASO of the lower extremities, besides daily foot care and active treatment of underlying diseases, this study provides clinicians with new ideas for DFU prevention and screening, which is helpful to supplement and improve existing prevention and treatment measures.

It should not be ignored that this study still has some limitations: First, this study is a retrospective study, and the problem of information bias should not be ignored; Secondly, the results of this study come from three large hospitals, and the accuracy and practicability of the model need to be tested in primary medical institutions. Finally, most of the data included in this study came from medical records, and information not recorded in the cases may also be a potential factor affecting the accuracy of the results.

Conclusion

BMI, hypertension, coronary heart disease, diabetic nephropathy, number of lower leg artery occlusion, controlling glucose by insulin injection, age, daily smoking amount, diastolic blood pressure and CRP were the most characteristic factors associated with DFU, and all were independent risk factors. The clinical prediction model based on these risk factors has perfect efficacy and repeatability, and has a wide range of clinical

practicability. Visualization of the nomogram makes the prediction model convenient for clinical application, and the establishment of the prediction model is conducive to the implementation of early screening and personalized treatment plans.

Data availability

All data generated or analyzed during the study are available from the manuscript and supplementary information 1 and 2.

35   Crawford, F. et al. The risk of foot ulceration in people with diabetes screened in community settings: Findings from a cohort study. *QJM Int. J. Med.* **104**(5), 403–410 (2011).

36. Zhang, S. et al. Application opportunity of Doppler ultrasound combined with CT angiography in diabetic lower extremity arterial disease and the analysis of the risk factors. *Front. Endocrinol.* **14**, 1257241 (2024).

37. Demirdal, T. & Sen, P. The significance of neutrophil-lymphocyte ratio, platelet-lymphocyte ratio and lymphocyte-monocyte ratio in predicting peripheral arterial disease, peripheral neuropathy, osteomyelitis and amputation in diabetic foot infection. *Diabetes Res. Clin. Pract.* **144**, 118–125 (2018).

38. Kang, H., Choi, S., Park, Y.-G., Choi, J. & Lim, C. Risk factors for major lower limb amputation and effect of endovascular revascularization in patients with diabetic foot wound. *Indian J. Orthop.* **58**(4), 379–386 (2024).

39. Park, J. H. et al. Role of procalcitonin in infected diabetic foot ulcer. *Diabetes Res. Clin. Pract.* **128**, 51–57 (2017).

40. Korkmaz, P. et al. The role of serum procalcitonin, interleukin-6, and fibrinogen levels in differential diagnosis of diabetic foot ulcer infection. *J. Diabetes Res.* **2018**(1), 7104352 (2018).

41. Wang, Y., Dou, L., Wang, N., Zhao, Y. & Nie, Y. An analysis of factors influencing cognitive dysfunction among older adults in Northwest China based on logistic regression and decision tree modelling. *BMC Geriatr.* **24**(1), 405 (2024).

42. Feng, Y. et al. LASSO-derived prognostic model predicts cancer-specific survival in advanced pancreatic ductal adenocarcinoma over 50 years of age: A retrospective study of SEER database research. *Front. Oncol.* **13**, 1336251 (2024).

43. Chu, F., Liang, T., Chen, C. P., Wang, X. & Ma, X. Compact broad learning system based on fused lasso and smooth lasso. *IEEE Trans. Cybern.* **54**(1), 435–448 (2023).

44. Kim, K.-M., Kim, J.-H., Rhee, H.-S. & Youn, B.-Y. Development of a prediction model for the depression level of the elderly in low-income households: Using decision trees, logistic regression, neural networks, and random forest. *Sci. Rep.* **13**(1), 11473 (2023).

45   Hu, J. & Szymczak, S. A review on longitudinal data analysis with random forest. *Brief. Bioinform.* **24**(2), bbad002 (2023).

46. Jiang, H., Zou, B., Xu, C., Xu, J. & Tang, Y. Y. SVM-boosting based on Markov resampling: Theory and algorithm. *Neural Netw.* **131**, 276–290 (2020).

Acknowledgements

Author contributions

Yubo Wang, Shouyu Wang, Xin Han and Junwei Zong participated in the study design; Yubo Wang and Chunyu Jiang retrieved and analyzed the data; Yubo Wang and Yi qi Xing wrote the main manuscript; Yubo Wang, Linxuan Zou, Mingzhi Song, Xueling Qu Zhao , Zhuqiang Jia were responsible for drawing diagrams; Xin Han ,Junwei Zong Shouyu Wang were involved in the supervision, review, and revision of the manuscript. All the authors have read and approved the final manuscript.

Funding

Declarations

Competing interests

The authors declare no competing interests.

Ethical approval

The study protocol has been approved by the Clinical Ethics Committee of the hospital (approval No. PJ-KS-KY-2024-32).

Additional information

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1038/s41598-025-96502-9.

Correspondence and requests for materials should be addressed to X.H., J.Z. or S.W.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.