BMC Endocrine Disorders

RESEARCH

Open Access

Establishment and external validation of an early warning model of diabetic peripheral neuropathy based on random forest and logistic regression

Lujie Wang[1†], Jiajie Li[4†], Yixuan Lin[2], Huilun Yuan[1], Zhaohui Fang[2], Aihua Fei[3], Guoming Shen[1*] and Aijuan Jiang[1*]

Abstract

Objective The primary objective of this study was to investigate the risk factors for diabetic peripheral neuropathy (DPN) and to establish an early diagnostic prediction model for its onset, based on clinical data and biochemical indices.

Methods Retrospective data were collected from 1,446 diabetic patients at the First Affiliated Hospital of Anhui University of Chinese Medicine and were split into training and internal validation sets in a 7:3 ratio. Additionally, 360 diabetic patients from the Second Affiliated Hospital were used as an external validation cohort. Feature selection was conducted within the training set, where univariate logistic regression identified variables with a p-value < 0.05, followed by backward elimination to construct the logistic regression model. Concurrently, the random forest algorithm was applied to the training set to identify the top 10 most important features, with hyperparameter optimization performed via grid search combined with cross-validation. Model performance was evaluated using ROC curves, decision curve analysis, and calibration curves. Model fit was assessed using the Hosmer-Lemeshow test, followed by Brier Score evaluation for the random forest model. Ten-fold cross-validation was employed for further validation, and SHAP analysis was conducted to enhance model interpretability.

Results A nomogram model was developed using logistic regression with key features: limb numbness, limb pain, diabetic retinopathy, diabetic kidney disease, urinary protein, diastolic blood pressure, white blood cell count, HbA1c, and high-density lipoprotein cholesterol. The model achieved AUCs of 0.91, 0.88, and 0.88 for the training, validation, and test sets, respectively, with a mean AUC of 0.902 across 10-fold cross-validation. Hosmer-Lemeshow test results showed p-values of 0.595, 0.418, and 0.126 for the training, validation, and test sets, respectively. The random forest model demonstrated AUCs of 0.95, 0.88, and 0.88 for the training, validation, and test sets, respectively, with a mean

[†]Lujie Wang and Jiajie Li contributed equally to this work.

*Correspondence:
Guoming Shen
shengm_66@163.com
Aijuan Jiang
jiangaijuan@ahtcm.edu.cn

Full list of author information is available at the end of the article

AUC of 0.886 across 10-fold cross-validation. The Brier score indicates a good calibration level, with values of 0.104, 0.143, and 0.142 for the training, validation, and test sets, respectively.

Conclusion The developed nomogram exhibits promise as an effective tool for the diagnosis of diabetic peripheral neuropathy in clinical settings.

Keywords Diabetic peripheral neuropathy, Nomogram model, Risk factors, Predictive model

Introduction

The prevalence of diabetes mellitus and its associated complications is rising rapidly worldwide [1]. Diabetic peripheral neuropathy (DPN) is a common complication, and is primarily characterized by neuropathic pain and acral paresthesia, with more advanced cases exhibiting motor neuron involvement and a potential need for amputation [2, 3]. Affected patients often experience dysfunctional nerve conduction that can contribute to diabetic foot ulcers [4]. Early stages of DPN can often undetected, with up to 50% of cases being asymptomatic [5]. By the time symptoms such as pain and numbness arise in the affected limb, the associated nerve may have already sustained irreversible damage. Sensory testing and comprehensive clinical scores used to diagnose DPN are highly dependent on subjective patient responses such that instances of small nerve fiber involvement may be overlooked [6], complicating the early diagnosis of this condition. Current treatments for DPN primarily focus on pain relief, glycemic control, and restoring metabolic homeostasis; however, these approaches frequently result in suboptimal outcomes [7]. There is thus a clear need to define novel approaches to preventing DPN or diagnosing it in its early stages in order to lower associated rates of disability and mortality, contributing to better patient quality of life. Nomograms are effective, easy-to-use tools that integrate a range of risk factors while enabling the personalized calculation of a given individual's risk of a particular condition in light of their clinical status. Sample collection for this study was carried out across two medical centers, enhancing the diversity of the sample set and thereby bolstering the generalizability of the model. Based on the data from 1806 diabetic patients, this research explores the etiological factors of DPN and establishes a nomogram for individualized risk prediction.

Materials and methods

Study population

From January 2018 to June 2022, the medical records of 1500 diabetic patients from the First Affiliated Hospital of Anhui University of Chinese Medicine were retrospectively reviewed, and concurrently, the records of 400 diabetic patients were collated from the Second Affiliated Hospital of Anhui University of Chinese Medicine. Patients aged 18 years or older with type 1 or type 2 diabetes mellitus were eligible for inclusion. Those with gestational diabetes, acute infectious diseases, severe conditions such as heart failure, severe liver or kidney disease, malignant tumors, or missing baseline data were excluded. Upon patient admission, the clinical diagnosis of DPN was made by physicians in accordance with the 2016 Chinese Medicine Clinical Diagnosis and Treatment Guidelines for DPN [8]. The Michigan Diabetic Neuropathy Score (MDNS) [9] and Toronto Clinical Scoring System (TCSS) [10] were used for scoring, and grading was performed according to the patient's score.

Data collection and preprocessing

In this study, all case data were double-entered by two researchers and subsequently reviewed by a third person, ensuring strict adherence to Good Clinical Practice (GCP) guidelines throughout the entire process. Collected patient data included age, gender, height, weight, BMI [weight (kg)/height^2(m^2)], systolic blood pressure (SBP), diastolic blood pressure (DBP), history of drinking, history of smoking, family history of diabetes, duration of diabetes, limb pain and limb numbness. Evaluated comorbid conditions included diabetic retinopathy (DR), Diabetic kidney disease (DKD), hyperlipidemia and hypertension. Laboratory test results included measures of fasting blood glucose (FBG), hemoglobin A1c (HbA1c), triglycerides (TG), total cholesterol (TC), high-density lipoprotein cholesterol (HDL-C), low-density lipoprotein cholesterol (LDL-C), hemoglobin (HGB), platelets (PLT), red blood cells (RBC), white blood cells (WBC), urine glucose (UGLU), urine protein (UPRO), urine red blood cells (URBC), urine white blood cells (UWBC), blood urea nitrogen (BUN), aspartate aminotransferase (AST) and alanine aminotransferase (ALT).

Statistical analyses

Statistical analyses were conducted using Python (version 3.12) and R studio (version 4.2.3). In the training set, logistic regression variables were selected using backward elimination with a 0.05 p-value threshold, ensuring model parsimony and preventing overfitting by retaining only statistically significant predictors [11]. Simultaneously, the random forest algorithm identified the top ten factors based on importance scores, offering a nonlinear and interaction-aware complement to the linear logistic regression analysis. The RF algorithm's ensemble

approach, which constructs a multitude of decision trees and aggregates their results, provides a robust mechanism against overfitting, making it suitable for both internal and external validation datasets [12]. The predictive performance of the selected factors was evaluated using statistical analyses, including receiver operating characteristic (ROC) curves, decision curve analysis (DCA), and clinical calibration curves. These analyses aimed to assess the true-positive rate against the false-positive rate and the concordance between predicted probabilities and observed outcomes. Ten-fold cross-validation was employed to evaluate the model's generalization on the training set. Calculate the Hosmer-Lemeshow test to validate the model's goodness-of-fit. The Brier score was computed to evaluate the overall accuracy of probabilistic predictions made by the random forest model, reflecting the degree of calibration between predicted probabilities and actual outcomes.

SHAP (Shapley Additive Explanations) methods are used to interpret machine learning model predictions by determining each feature's contribution to the final prediction [13]. Shapley values, derived from cooperative game theory, measure feature importance by averaging a feature's marginal contributions across all possible combinations of features [14]. After validating the predictive value and clinical relevance of the features, factors identified through backward elimination were selected to construct the nomogram. A nomogram, by representing intricate mathematical relationships and facilitating the estimation of dependent variables via multiple independent variables, functions as a visual tool in statistics, providing an intuitive interface that converts multivariate regression models into interpretable charts, thereby bridging the gap between sophisticated statistical analysis and practical application in fields such as medicine and risk assessment [15]. The above process is detailed in the flowchart (Fig. 1).

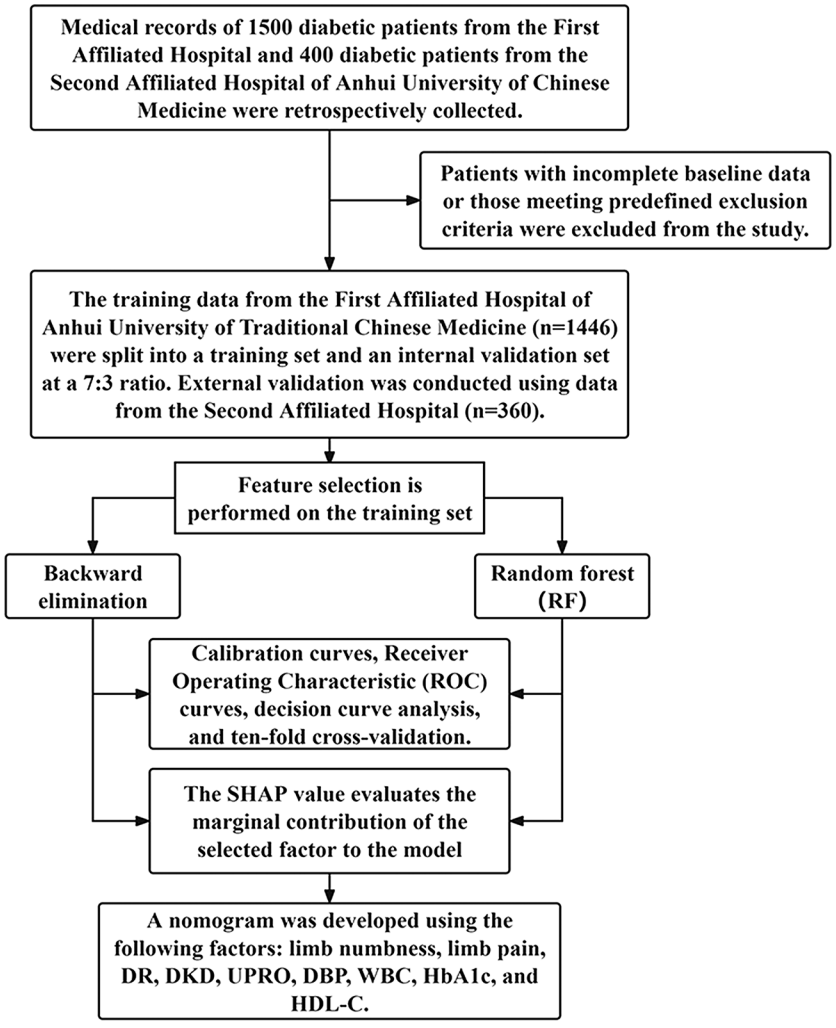

Fig. 1 Development and validation flowchart for diabetic neuropathy prediction models. Flowchart of feature selection and nomogram building

Results

Comprehensive analysis of baseline data and key feature selection

Data from 1806 cases were finally included in this study for analysis. The data from 1,446 patients at the First Affiliated Hospital of Anhui University of Traditional Chinese Medicine were split into a training set ($n=1,012$) and an internal validation set ($n=434$) in a 7:3 ratio, with data from 360 patients at the Second Affiliated Hospital used as the external validation set. Categorical variables were described using frequencies and percentages, while continuous variables were summarized as mean ± standard deviation for normally distributed data, and as medians with interquartile ranges for non-normally distributed data (Table 1). Comparison of baseline data between DPN patients and non-DPN patients (Table 2). Univariate logistic regression analysis (Table 3) identified several factors significantly associated with DPN ($p<0.05$). These factors included sex, limb numbness, limb pain, diabetic retinopathy, diabetic kidney disease, urinary glucose levels, urinary protein levels, age, diabetes duration, diastolic blood pressure, hemoglobin, white blood cell count, urinary white blood cell count, HbA1c, and low-density lipoprotein cholesterol.

The factors selected by the logistic regression model through backward elimination include limb numbness, limb pain, DR, DKD, URPO, DBP, WBC, HbA1c, and HDL-C; the factors identified by the random forest model based on feature importance include limb numbness, limb pain, WBC, CREA, diabetes duration, PLT, HbA1c, TG, UWBC, and FPG.

Development and validation of LR and RF models

Receiver Operating Characteristic curves were plotted to assess model discrimination. The logistic regression model showed an AUC of 0.91 (95% CI: 0.89–0.92) on the training set, 0.88 (95% CI: 0.85–0.91) on the validation set, and 0.88 (95% CI: 0.85–0.92) on the test set (Fig. 2A). The random forest model demonstrated an AUC of 0.95 (95% CI: 0.94–0.96) on the training set, 0.88 (95% CI: 0.85–0.91) on the validation set, and 0.88 (95% CI: 0.84–0.91) on the test set (Fig. 2B). Decision Curve Analysis was conducted to evaluate clinical usefulness, with both models showing consistent net benefit across varying threshold probabilities (Fig. 3). Calibration curves demonstrated the alignment between predicted and observed probabilities for both models, with the logistic regression model showing better calibration across all sets, while the random forest model showed some deviation, particularly on the training set (Fig. 4). The Hosmer-Lemeshow test results indicated that the logistic regression model had p-values of 0.595 for the training set, 0.418 for the validation set, and 0.126 for the test set. The Brier scores for the random forest model were 0.104 for the training set, 0.143 for the validation set, and 0.142 for the test set, indicating reasonable calibration. During 10-fold cross-validation, the logistic regression model achieved a mean AUC of 0.902, while the random forest model had a mean AUC of 0.886.

SHAP analysis: evaluating feature impact and predictive contributions

The SHAP summary plots for both models illustrate the contributions of features to the predictions. In Fig. 5A, the logistic regression model shows the linear impact of features, with limb numbness and limb pain as notable contributors. In contrast, Fig. 5B presents the random forest model' s SHAP summary plot, which captures the non-linear interactions among features. Unlike the logistic regression model, the random forest model underscores the role of variables such as creatinine, diabetes duration, and urinary white blood cell count, reflecting the model' s ability to account for complex, non-linear relationships within the data. The color gradient from red to blue indicates feature values, with higher values corresponding to a greater influence on the prediction of diabetic peripheral neuropathy. The SHAP bar plots further distinguish the models by ranking features based on their mean impact on the model' s output. SHAP bar plots for both the logistic regression and random forest models, ranking features by their mean impact on the model's predictions (Fig. 6).

Nomogram for predictive modeling of DPN

Based on model evaluation and SHAP analysis, the random forest model exhibited slightly lower generalization ability compared to the logistic regression model and showed a tendency towards underfitting. Therefore, the factors selected by the logistic regression model were used to construct the nomogram (Fig. 7). The nomogram visually represents the contributions of various clinical factors to the risk of developing DPN. Key predictors include limb numbness, limb pain, DR, DKD, URPO, DBP, WBC, HbA1c, and HDL-C. Each factor is assigned a point value based on its presence or severity, and the total score is calculated by summing these points. This total score corresponds to the diagnostic probability of DPN, providing clinicians with a tool to estimate a patient' s risk and support personalized clinical decision-making.

Discussion

To address the limitations in medical resource availability, a nomogram model was herein developed to predict the risk of DPN in patients with diabetes based on limb numbness, limb pain, DR, DKD, URPO, DBP, WBC, HbA1c, and HDL-C. Employing a 10 g monofilament examination approach [16] and corresponding

Table 1 Comparison of baseline characteristics between training and validation sets

Characteristics		All subjects($n=1806$)	Training set($n=1012$)	Validation set($n=434$)	External validation set($n=360$)
Sex	Female	1062(58.80%)	617(60.97%)	240(55.30%)	205(56.94%)
	Male	744(41.20%)	395(39.03%)	194(44.70%)	155(43.06%)
Limb numbness	Yes	812(44.96%)	452(44.66%)	199(45.85%)	161(44.72%)
	No	994(55.04%)	560(55.34%)	235(54.15%)	199(55.28%)
Limb pain	Yes	1356(75.08%)	764(75.49%)	319(73.50%)	273(75.83%)
	No	450(24.92%)	248(24.51%)	115(26.50%)	87(24.17%)
History of smoking	Yes	1246(68.99%)	677(66.90%)	313(72.12%)	256(71.11%)
	No	560(31.01%)	335(33.10%)	121(27.88%)	104(28.89%)
History of drinking	Yes	1319(73.03%)	714(70.55%)	332(76.50%)	273(75.83%)
	No	487(26.97%)	298(29.45%)	102(23.50%)	87(24.17%)
Family history of diabetes	Yes	1334(73.86%)	723(71.44%)	332(76.50%)	279(77.50%)
	No	472(26.14%)	289(28.56%)	102(23.50%)	81(22.50%)
DPN	Yes	944(52.27%)	533(52.67%)	224(51.61%)	187(51.94%)
	No	862(47.73%)	479(47.33%)	210(48.39%)	173(48.06%)
DR	Yes	1445(80.01%)	821(81.13%)	342(78.80%)	282(78.33%)
	No	361(19.99%)	191(18.87%)	92(21.20%)	78(21.67%)
DKD	Yes	1586(87.82%)	890(87.94%)	385(88.71%)	311(86.39%)
	No	220(12.18%)	122(12.06%)	49(11.29%)	49(13.61%)
Hypertension	Yes	851(47.12%)	481(47.53%)	205(47.24%)	165(45.83%)
	No	955(52.88%)	531(52.47%)	229(52.76%)	195(54.17%)
Hyperlipidemia	Yes	1031(57.09%)	580(57.31%)	248(57.14%)	203(56.39%)
	No	775(42.91%)	432(42.69%)	186(42.86%)	157(43.61%)
UGLU	Negtive	1059(58.64%)	592(58.50%)	260(59.91%)	207(57.50%)
	1+	102(5.65%)	62(6.13%)	27(6.22%)	13(3.61%)
	2+	92(5.09%)	55(5.43%)	18(4.15%)	19(5.28%)
	3+	386(21.37%)	211(20.85%)	85(19.59%)	90(25.00%)
	4+	167(9.25%)	92(9.09%)	44(10.14%)	31(8.61%)
URPO	Negtive	1647(91.20%)	924(91.30%)	398(91.71%)	325(90.28%)
	Weakly positive	41(2.27%)	18(1.78%)	9(2.07%)	14(3.89%)
	1+	48(2.66%)	31(3.06%)	11(2.53%)	6(1.67%)
	2+	41(2.27%)	20(1.98%)	12(2.76%)	9(2.50%)
	3+	29(1.61%)	19(1.88%)	4(0.92%)	6(1.67%)
Age		58.00 (67.00, 51.00)	58.00 (67.00, 50.00)	59.00 (68.00, 52.00)	59.00 (69.00, 52.00)
Duration of Diabetes		7.00 (13.00, 2.00)	7.00 (13.00, 2.00)	8.00 (14.00, 2.00)	7.00 (12.00, 2.00)
SBP		132.00 (144.00, 120.00)	132.00 (144.00, 120.00)	133.48 ± 16.22	131.50 (144.00, 120.00)
DBP		80.00 (88.00, 74.00)	80.00 (88.00, 75.00)	80.00 (89.75, 74.25)	80.00 (88.00, 74.00)
HGB		135.00 (148.00, 123.00)	136.00 (149.00, 124.00)	133.85 ± 18.89	135.00 (147.00, 122.00)
RBC		4.53 (4.95, 4.13)	4.56 (5.00, 4.14)	4.47 (4.89, 4.13)	4.49 (4.88, 4.06)
WBC		5.92 (7.09, 4.93)	5.91 (7.07, 4.94)	5.89 (7.07, 4.91)	5.98 (7.21, 4.92)
PLT		190.00 (233.00, 153.00)	187.00 (230.00, 152.00)	193.00 (230.75, 156.00)	196.00 (237.00, 154.00)
UWBC		4.60 (16.90, 1.90)	4.45 (17.45, 1.90)	4.20 (14.15, 1.80)	5.10 (18.32, 1.88)
URBC		4.90 (11.38, 2.50)	5.05 (11.30, 2.50)	4.50 (11.30, 2.10)	5.00 (11.40, 2.60)
FPG		7.39 (10.01, 5.91)	7.36 (9.93, 5.92)	7.46 (10.06, 5.94)	7.42 (10.18, 5.88)
HbA1c		7.80 (9.70, 6.68)	7.78 (9.51, 6.65)	7.87 (9.90, 6.70)	7.88 (9.62, 6.62)
TG		1.54 (2.30, 1.06)	1.52 (2.35, 1.06)	1.56 (2.25, 1.02)	1.58 (2.28, 1.05)
TC		4.52 (5.23, 3.83)	4.55 (5.29, 3.85)	4.47 (5.12, 3.81)	4.50 (5.15, 3.78)
LDL-C		2.69 (3.30, 2.11)	2.71 (3.31, 2.14)	2.70 (3.26, 2.09)	2.64 (3.39, 2.06)
HDL-C		1.13 (1.38, 0.94)	1.12 (1.39, 0.94)	1.17 (1.43, 0.95)	1.13 (1.34, 0.94)
CREA		61.35 (73.60, 51.00)	61.05 (74.33, 51.50)	61.95 (73.97, 50.20)	61.35 (72.53, 50.90)
BUN		5.74 (6.95, 4.80)	5.76 (6.95, 4.81)	5.77 (7.09, 4.82)	5.64 (6.78, 4.72)
ALT		19.00 (28.00, 13.83)	19.00 (29.00, 14.00)	18.00 (27.00, 13.00)	19.00 (26.25, 13.00)
AST		18.00 (23.00, 15.00)	18.00 (23.00, 15.00)	18.00 (22.08, 15.00)	18.00 (23.05, 15.00)

Table 2 Comparison of baseline characteristics between patients with and without diabetic peripheral neuropathy

Characteristics		non-DPN($n=944$)	DPN($n=862$)
Sex	Female	336 (35.59%)	408 (47.33%)
	Male	608 (64.41%)	454 (52.67%)
Limb numbness	Yes	241 (25.53%)	753 (87.35%)
	No	703 (74.47%)	109 (12.65%)
Limb pain	Yes	136 (14.41%)	314 (36.43%)
	No	808 (85.59%)	548 (63.57%)
History of smoking	Yes	313 (33.16%)	247 (28.65%)
	No	631 (66.84%)	615 (71.35%)
History of drinking	Yes	280 (29.66%)	207 (24.01%)
	No	664 (70.34%)	655 (75.99%)
Family history of diabetes	Yes	236 (25.00%)	236 (27.38%)
	No	708 (75.00%)	626 (72.62%)
DR	Yes	108 (11.44%)	253 (29.35%)
	No	836 (88.56%)	609 (70.65%)
DKD	Yes	77 (8.16%)	143 (16.59%)
	No	867 (91.84%)	719 (83.41%)
Hypertension	Yes	517 (54.77%)	438 (50.81%)
	No	427 (45.23%)	424 (49.19%)
Hyperlipidemia	Yes	408 (43.22%)	367 (42.58%)
	No	536 (56.78%)	495 (57.42%)
UGLU	Negtive	594 (62.92%)	465 (53.94%)
	1+	41 (4.34%)	61 (7.08%)
	2+	42 (4.45%)	50 (5.80%)
	3+	184 (19.49%)	202 (23.43%)
	4+	83 (8.79%)	84 (9.74%)
URPO	Negtive	888 (94.07%)	759 (88.05%)
	Weakly positive	16 (1.69%)	25 (2.90%)
	1+	14 (1.48%)	34 (3.94%)
	2+	18 (1.91%)	23 (2.67%)
	3+	8 (0.85%)	21 (2.44%)
Age		58.00 (49.00, 67.00)	59.00 (52.00, 67.00)
Duration of Diabetes		6.00 (2.00, 10.25)	10.00 (4.00, 14.00)
SBP		132.00 (120.00, 144.00)	132.00 (121.00, 144.00)
DBP		82.00 (75.00, 90.00)	80.00 (74.00, 88.00)
HGB		137.00 (126.00, 149.00)	133.00 (121.00, 146.00)
RBC		4.56 (4.19, 4.99)	4.48 (4.06, 4.91)
WBC		6.09 (5.07, 7.23)	5.74 (4.84, 6.95)
PLT		192.00 (156.00, 235.25)	188.00 (151.00, 229.00)
UWBC		4.10 (1.80, 13.72)	4.95 (1.90, 21.77)
URBC		4.70 (2.30, 10.50)	5.10 (2.60, 12.20)
FPG		7.29 (5.92, 9.91)	7.55 (5.88, 10.20)
HbA1c		7.60 (6.59, 9.48)	8.00 (6.80, 9.90)
TG		1.58 (1.09, 2.39)	1.50 (1.03, 2.25)
TC		4.54 (3.84, 5.24)	4.50 (3.82, 5.20)
LDL-C		2.76 (2.18, 3.41)	2.61 (2.02, 3.22)
HDL-C		1.11 (0.93, 1.34)	1.17 (0.95, 1.42)
CREA		62.80 (52.40, 74.60)	60.00 (49.02, 72.28)
BUN		5.63 (4.78, 6.84)	5.84 (4.82, 7.11)
ALT		20.00 (14.00, 30.00)	18.00 (13.00, 26.08)
AST		18.85 (15.00, 24.00)	17.70 (15.00, 22.00)

Table 3 Univariate logistic regression analysis of the training cohort

Characteristics	β	SE	OR	95%CI	Z	*P* Value
Sex	-0.367	0.129	0.690	0.54–0.89	-2.840	0.005
Limb numbness	3.207	0.178	24.700	17.43–34.99	18.041	< 0.001
Limb pain	1.239	0.157	3.450	2.54–4.69	7.906	< 0.001
History of smoking	-0.118	0.134	0.890	0.68–1.16	-0.878	0.380
History of drinking	-0.192	0.139	0.830	0.63–1.08	-1.387	0.165
Family history of diabetes	-0.015	0.139	0.980	0.75–1.29	-0.110	0.912
DR	1.001	0.169	2.720	1.95–3.79	5.912	< 0.001
DKD	0.934	0.204	2.550	1.71–3.80	4.574	< 0.001
Hypertension	-0.117	0.126	0.890	0.70–1.14	-0.924	0.355
Hyperlipidemia	-0.170	0.127	0.840	0.66–1.08	-1.333	0.183
UGLU	0.091	0.042	1.090	1.01–1.19	2.163	0.031
URPO	0.210	0.086	1.230	1.04–1.46	2.434	0.015
Age	0.010	0.005	1.010	1.00-1.02	1.981	0.048
Duration of Diabetes	0.049	0.009	1.050	1.03–1.07	5.381	< 0.001
SBP	0.001	0.003	1.000	0.99–1.01	0.279	0.780
DBP	-0.014	0.006	0.990	0.97-1.00	-2.496	0.013
HGB	-0.011	0.003	0.990	0.98-1.00	-3.244	0.001
RBC	-0.093	0.070	0.910	0.80–1.04	-1.334	0.182
WBC	-0.119	0.035	0.890	0.83–0.95	-3.356	0.001
PLT	-0.001	0.001	1.000	1.00–1.00	-1.132	0.258
UWBC	0.003	0.001	1.000	1.00–1.00	2.950	0.003
URBC	0.002	0.002	1.000	1.00-1.01	1.168	0.243
FPG	0.002	0.018	1.000	0.97–1.04	0.135	0.893
HbA1c	0.057	0.028	1.060	1.00-1.12	2.032	0.042
TG	-0.006	0.023	0.990	0.95–1.04	-0.283	0.777
TC	-0.001	0.052	1.000	0.90–1.11	-0.012	0.990
LDL-C	-0.161	0.068	0.850	0.75–0.97	-2.359	0.018
HDL-C	0.144	0.127	1.150	0.90–1.48	1.130	0.258
CREA	-0.001	0.001	1.000	1.00–1.00	-0.574	0.566
BUN	0.005	0.014	1.010	0.98–1.03	0.374	0.709
ALT	-0.003	0.003	1.000	0.99-1.00	-1.190	0.234
AST	-0.001	0.004	1.000	0.99–1.01	-0.243	0.808

clinical assessments can detect neuropathy in its more advanced stages [17], but some reports have found that nerve conduction velocity testing cannot effectively monitor lesions affecting smaller nerve fibers [18]. In contrast to traditional scoring strategies, the developed nomogram incorporates a wider range of predictive factors and can be applied more flexibly in clinical settings. The Nomogram' s utility is grounded in its ability to integrate diverse clinical variables into a unified predictive framework. This approach underscores the Nomogram's potential to enhance prognostic evaluations and optimize patient-centered care in diabetes management.

The hyperglycemic-induced stress state is accompanied by a crosstalk between metabolic, immune, genetic, and epigenetic factors. Metabolic syndrome leads to an insufficient uptake of glucose by peripheral nerves, and this bioenergetic imbalance, in turn, disrupts the coupling between neurons and glial cells [7]. Reduced mitochondrial plasticity in peripheral nerves, along with DNA damage, triggers significant impediments in biosynthetic and metabolic functions. Under the high metabolic stress, the inadequacy of energy supply to distal axons diminishes neural conduction efficiency, impedes sensory signal transmission, and ultimately leads to degenerative changes in nerve fibers and axonal death [19]. This study focused on the individualized assessment of risk in patients with diabetes, as all patients are at potential risk of DPN after developing hyperglycemia although their progression may vary [20]. Diabetes duration was positively correlated with the incidence of DPN in this analysis, consistent with prior evidence linking poor glycemic control and diabetes duration to microvascular complications [21]. As the disease course grows longer, dysregulated glucose metabolism can adversely affect nerve fibers and compensatory functions can be disrupted, leading to reductions in nerve cell numbers more extensive nerve damage, and an elevated DPN risk [22, 23]. DPN is not solely a consequence of hyperglycemia; rather, it can emerge from a combination of cumulative and additive factors.

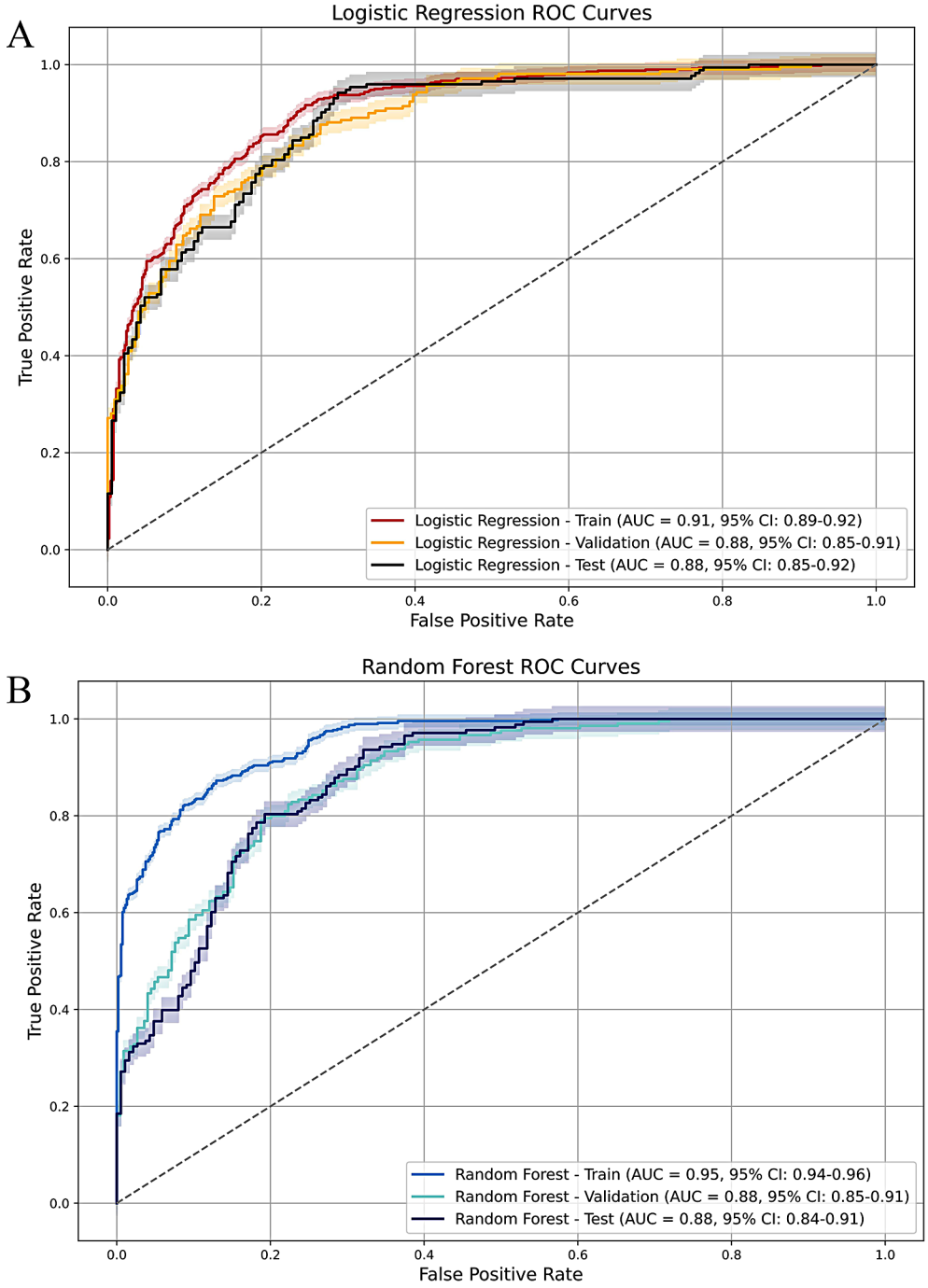

Fig. 2 ROC curves for logistic regression and random forest models. Figure 2A illustrates the ROC curves for the logistic regression model, highlighting its ability to distinguish between positive and negative cases across the training, validation, and test datasets, with AUC values and confidence intervals provided. Figure 2B shows the ROC curves for the random forest model, which, while performing strongly on the training set, suggests potential overfitting when compared to the validation and test sets. This comparison underscores the importance of evaluating model performance across different datasets to ensure generalizability and avoid overfitting

In this study, blood lipid levels were higher in early-stage diabetes patients but declined as the disease progressed. Altered energy metabolism may underlie these changes, and dysregulated lipid signaling could play a role in the onset or progression of DPN. In the present study, higher levels of blood lipids were detected in early-stage diabetes patients whereas these levels declined with the prolongation of the disease course. Altered energy metabolism may contribute to these differences in pathological status in diabetes patients. Given the complexity of

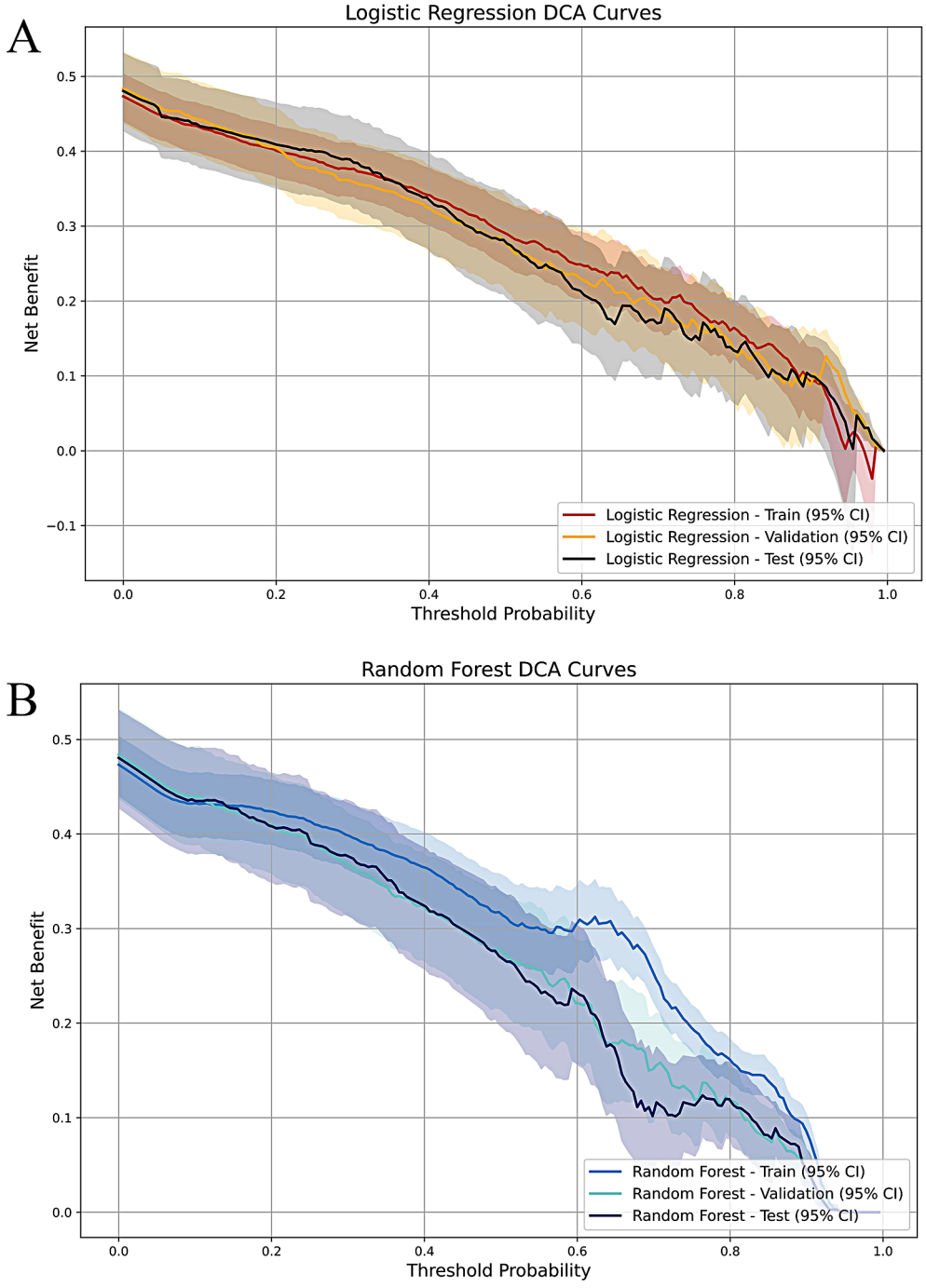

Fig. 3 Decision curve analysis for logistic regression and random forest. Figure 3A shows the DCA for the logistic regression model, evaluating net benefits at various threshold probabilities. Figure 3B illustrates the DCA for the random forest model, where a notable rise in net benefit is observed at certain thresholds. This suggests that the random forest model may provide greater clinical value in specific decision-making scenarios

the processes that link nerve damage to clinically detectable neuropathy, it is feasible that dysregulated lipid-related signaling is involved in DPN onset or progression. The study indicates that reduced SBP and elevated LDL levels are significant risk factors for DPN, particularly among European patients, underscoring the crucial role of lipid profiles in its pathogenesis [24]. The more pronounced disruption of glycolipid metabolism evident in early-stage diabetes can irreversibly damage nerve cells through glycolipid toxicity. High lipid levels can induce Schwann cell apoptosis in experimental settings, particularly in the context of elevated glucose levels [25, 26]. While the precise role of dyslipidemia in the pathogenesis of DPN remains to be defined, lipid-lowering

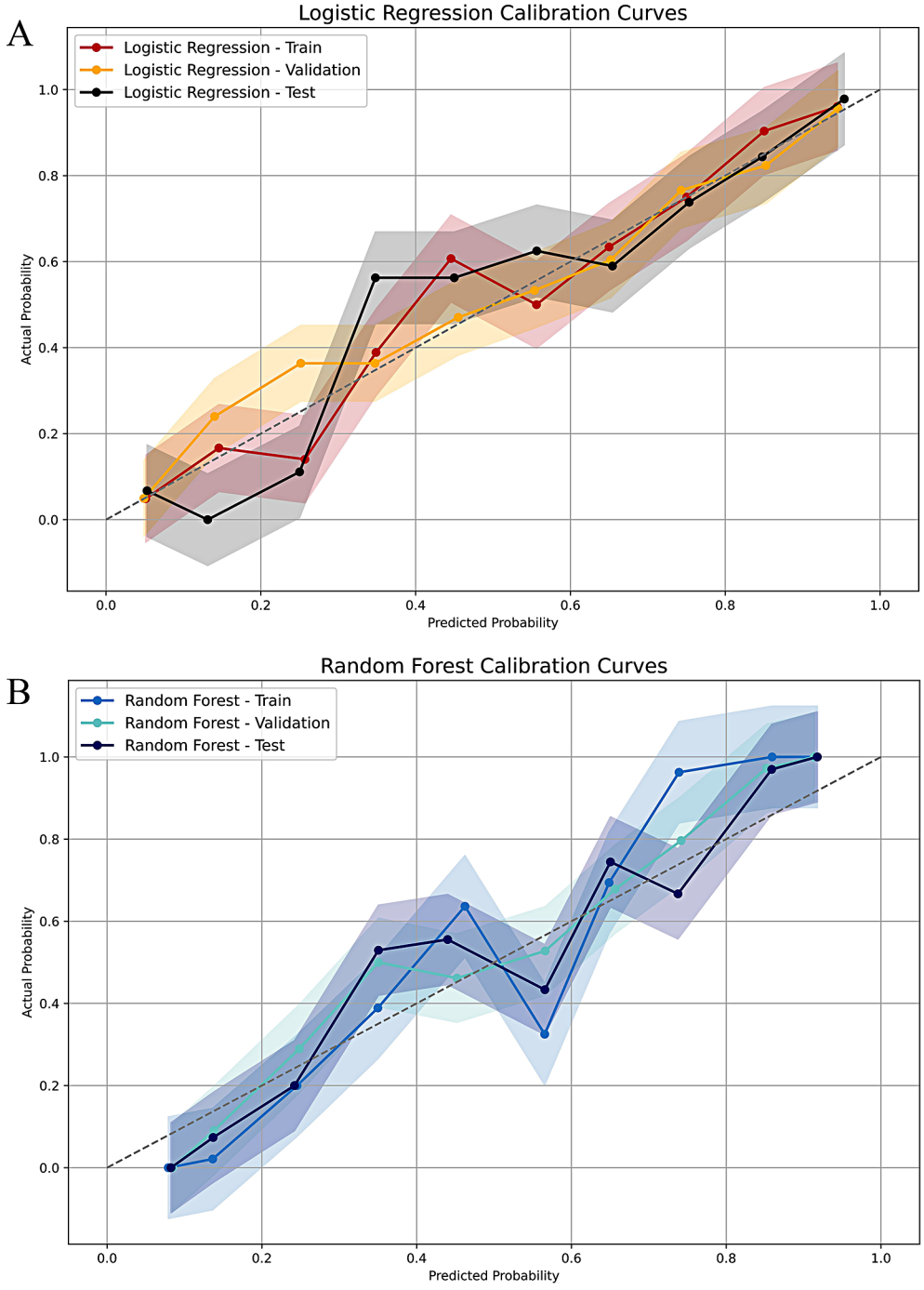

Fig. 4 Calibration curves for logistic regression and random forest. Figure 4A shows the calibration curves for the logistic regression model, depicting the alignment between predicted probabilities and observed outcomes in the training, validation, and test datasets, and evaluating the model's calibration accuracy. Figure 4B presents the calibration curves for the random forest model, indicating how well the predicted probabilities correspond with actual results across the datasets, thereby assessing the model's reliability

therapy can support a better neuronal blood supply conducive to greater neuroprotective efficacy [27]. As diabetes becomes more advanced, however, affected patients can present with chronic wasting characterized by low levels of fat and persistent hyperglycemia. Insulin imbalances alter peripheral glucose metabolism while also more broadly adversely impacting energy metabolism throughout the body. Both urinary glucose excretion and insulin imbalances can prevent excessive tissue nutrient accumulation while also impacting myelinated nerve fiber repair [28, 29]. Patients with diabetes reportedly exhibit impaired mitochondrial function [30], higher

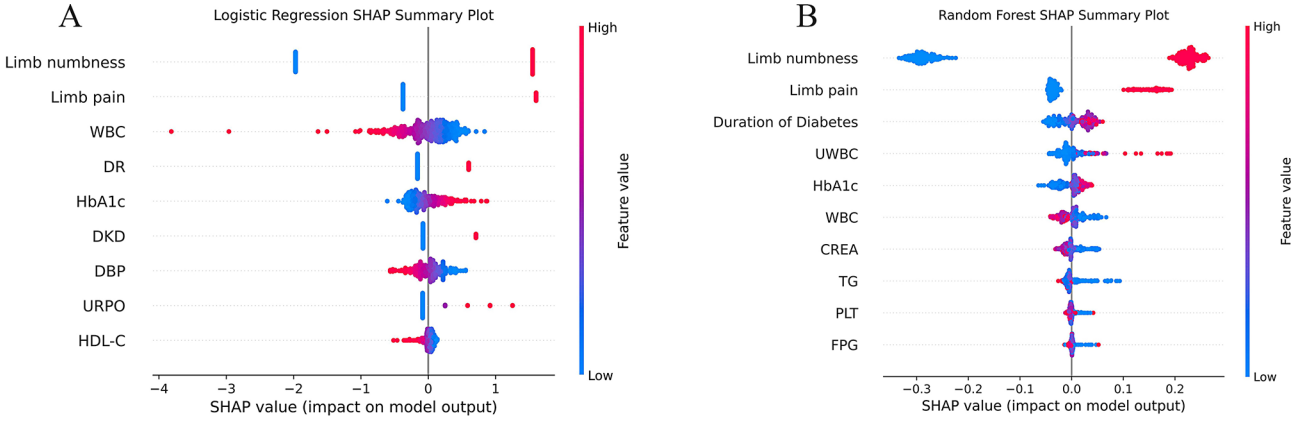

Fig. 5 SHAP summary plots for logistic regression and random forest models. Figure 5A demonstrates the impact of key features in a linear model, suggesting that certain symptoms like limb numbness and limb pain have a predictable, direct relationship with DPN risk. Figure 5B reveals the more complex, non-linear interactions present in the random forest model, where factors such as diabetes duration and creatinine levels influence risk in a multifaceted manner. These plots highlight the importance of considering both direct and complex clinical interactions when assessing DPN risk, suggesting the need for an integrated approach to patient management that takes multiple aspects of health data into account

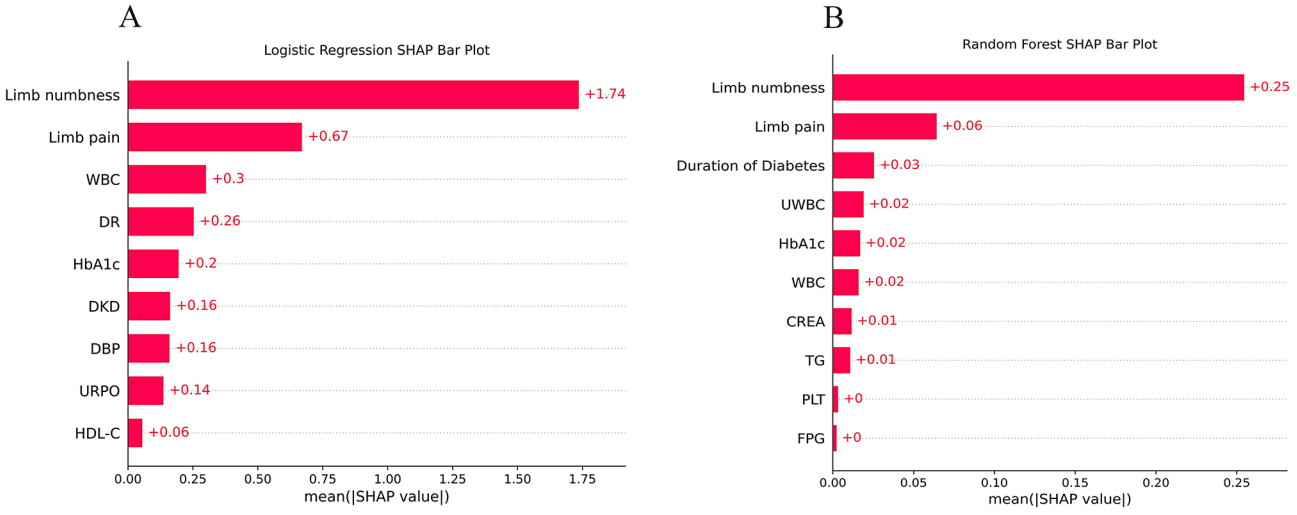

Fig. 6 SHAP bar plots for logistic regression and random forest models. Figure 6A shows the SHAP bar plot for the logistic regression model, where limb numbness (+1.74) and limb pain (+0.67) have the highest linear impact. Figure 6B illustrates the SHAP bar plot for the random forest model, with limb numbness (+0.25), limb pain (+0.06), and diabetes duration (+0.03) contributing non-linearly. The plots indicate that the significance of each feature differs between the two models, highlighting the distinct ways in which these features influence their respective predictions

levels of oxygen utilization, and enhanced lipid availability [23]. These findings suggest that the need for persistent compensatory adjustments in diabetes patients ultimately leads to the onset of chronic wasting disease with concomitant shifts in lipid levels. Pathological metabolic perturbations and gene reprogramming are often the key to DPN [31]. A number of studies have shown that Schwann cells play an important role in the synthesis of myelin sheath in peripheral nerves, and the metabolic coupling phenomenon between axons and Schwann cells can affect the occurrence and development of peripheral neuropathy.

The present analyses revealed that three forms of diabetic microangiopathy, including retinopathy, nephropathy, and neuropathy, were often comorbid with one another. This suggests that these microvascular conditions are associated with similar risk factors that can exacerbate their development or progression. Consistently, a prior cross-sectional analysis identified strong clustering among these three conditions such that they co-occurred more often than would be expected by chance [32]. Vascular endothelial cell damage can result from impaired microvascular circulation [33], and this process can contribute to nerve damage and associated declines in peripheral nerve function. Previous studies have not established a clear correlation between DPN and diabetic nephropathy [34], whereas DPN has been found to be more closely associated with diabetic

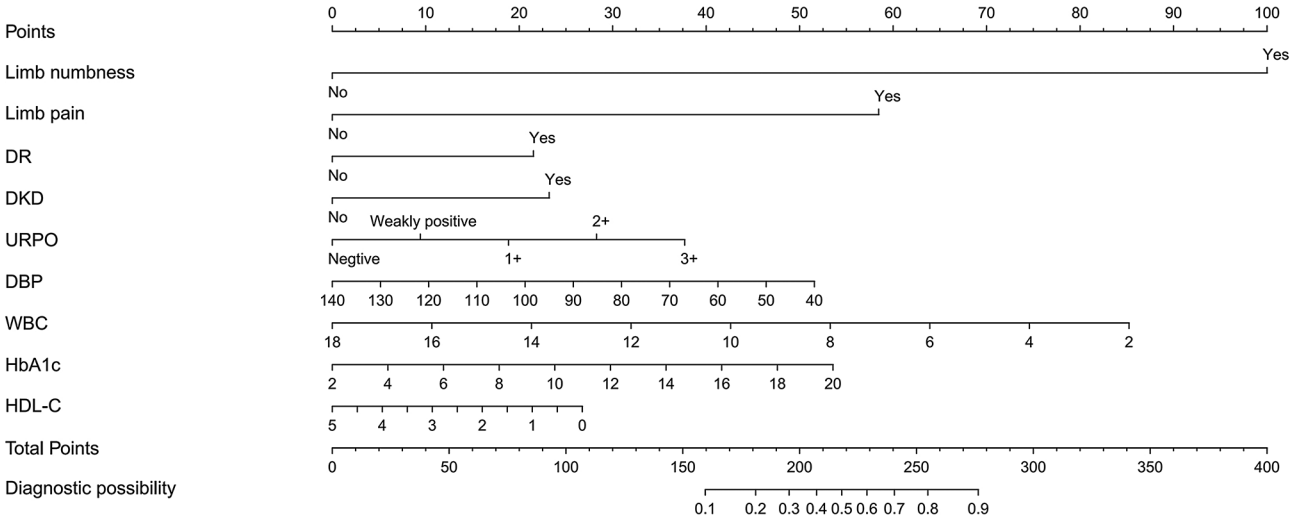

Fig. 7 Nomogram for predicting condition probability based on clinical parameters. Each clinical factor, including limb numbness, limb pain, DR, DKD, URPO, DBP, WBC, HbA1c, and HDL-C, is represented by a horizontal line segment. The length of each line reflects the weight or importance of that factor in predicting DPN. By aligning a patient's data with the respective values on each line, a corresponding point value is determined. The total score, obtained by summing these points, maps onto a scale at the bottom of the nomogram, which indicates the probability of developing DPN

retinopathy [20]. There remains ongoing scientific debate as to whether vascular insufficiency can trigger the incidence of peripheral neuropathy directly or indirectly reduces the ability of nerve cells to tolerate damage such as that resulting from glucotoxicity [35]. Various overlapping phenotypes contribute to disease development [36], potentially explaining significant interactions among these three distinct forms of microangiopathy. Limited testing methods and variations in timing limit efforts to detect microvascular interactions, but these findings nonetheless emphasize the importance of considering that these disease origins may be linked at earlier time points than otherwise noted.

This predictive model offers advantages over more invasive or exhaustive testing, as all of the included variables can be readily assessed, thereby lowering patient medical costs and discomfort. Some reports have suggested that the underlying basis for peripheral neuropathy can be discerned in ~73% of cases based on the results of laboratory tests, physical examination, and a medical history [37]. However, the study has certain limitations. For instance, incorporating sensitivity analyses could enhance the model's stability and reliability. Moreover, increasing the sample size and integrating data from a broader range of medical centers could improve the generalizability of the findings, thereby further validating and strengthening the model's applicability. The lack of data on patients' medication use is a limitation that could affect the interpretation of the results. Addressing these limitations in future research could provide a more comprehensive understanding of the factors influencing diabetic peripheral neuropathy.

Summary

In conclusion, the predictive nomogram established in this study can determine the odds of a given diabetic individual having DPN based on their physical condition and test results, thus offering a valuable tool for individualized screening that can better guide the management of DPN in clinical practice.

Abbreviations

DPN	Diabetic peripheral neuropathy
DKD	Diabetic kidney disease
DR	Diabetic retinopathy
HbA1c	Hemoglobin A1c
BMI	Body Mass Index
DBP	Diastolic blood pressure
SBP	Systolic blood pressure
TG	Triglycerides
UGLU	Urinary glucose
URBC	Urinary red blood cells
UWBC	Urinary white blood cells
UPRO	Urinary protein
HGB	Hemoglobin
PLT	Platelets
RBC	Red blood cells
WBC	White blood cells
FBG	Fasting blood glucose
TC	Total cholesterol
LDL-C	Low-density lipoprotein cholesterol
HDL-C	High-density lipoprotein cholesterol
CREA	Creatinine
BUN	Blood urea nitrogen
ALT	Alanine aminotransferase
AST	Aspartate aminotransferase
ROC	Receiver operating characteristic curve
DCA	Decision curve analysis
AUC	Area under the curve

Acknowledgements

Not applicable.

Author contributions
Lujie Wang: Conceptualization, Software, Writing- Original draft preparation. Jiajie Li: Validation, Investigation. Yixuan Lin: Methodology. Huilun Yuan: Data Curation. Zhaohui Fang: Resources. Aihua Fei: Resources. Guoming Shen (Corresponding Author): Supervision. Aijuan Jiang (Corresponding Author): Writing- Reviewing and Editing.

Funding
Scientific Research Project of higher education in Anhui Province (No.2022AH050480), National Natural Science Foundation of China (No.81874457), Key Research and Development Program of Anhui Province (No.202104j07020006).

Data availability
Due to the sensitive nature of the data involved in this study, including personal health information protected under privacy laws and regulations, it is not possible to make the data publicly available. The data is subject to ethical restrictions imposed by the Ethics Committee of the First Affiliated Hospital of Anhui University of Chinese Medicine. However, access to the data may be granted upon a reasonable and justified request to the aforementioned Ethics Committee and the researchers of this study. Data access will be considered for qualified researchers who meet the necessary ethical and legal criteria and agree to adhere to the confidentiality requirements. Such access will be granted under appropriate conditions designed to ensure the privacy and confidentiality of the data.

Declarations

Ethics approval and consent to participate
This study was reviewed by the Ethics Committee of the First Affiliated Hospital of Anhui University of Chinese Medicine and was exempt from informed consent, with the exemption approval number (2021MCZQ11). The external validation part was also exempt from informed consent by the Ethics Committee of the Second Affiliated Hospital of Anhui University of Chinese Medicine, with the exemption approval number (No: 2023-zjsk-06). This study did not disclose any personal privacy of patients and will not violate local data protection laws. This study is registered with the Chinese Clinical Trial Registry under the registration number ChiCTR2300076081. Registration date: September 25, 2023.

Consent for publication
Not applicable.

Competing interests
The authors declare no competing interests.

Author details
[1]School of Integrated Traditional Chinese and Western Medicine, Anhui University of Chinese Medicine, 350 Longzihu Road, Xinzhan District, Hefei City, Anhui Province 230012, China
[2]The First Affiliated Hospital of Anhui University of Chinese Medicine, Hefei 230031, China
[3]The Second Affiliated Hospital of Anhui University of Chinese Medicine, Hefei 230031, China
[4]Yunnan University of Chinese Medicine, Kunming 650500, China

Received: 5 January 2024 / Accepted: 11 September 2024
Published online: 20 September 2024

References

1. Kautzky-Willer A, Harreiter J, Pacini G. Sex and gender differences in risk, pathophysiology and complications of type 2 diabetes Mellitus. Endocr Rev. 2016;37(3):278–316.
2. Forbes JM, Cooper ME. Mechanisms of diabetic complications. Physiol Rev. 2013;93(1):137–88.
3. Sloan G, Shillo P, Selvarajah D, et al. A new look at painful diabetic neuropathy. Diabetes Res Clin Pract. 2018;144:177–91.
4. Sloan G, Selvarajah D, Tesfaye S. Pathogenesis, diagnosis and clinical management of diabetic sensorimotor peripheral neuropathy. Nat Rev Endocrinol. 2021;17(7):400–20.
5. American Diabetes Association. Classification and diagnosis of diabetes: standards of Medical Care in Diabetes-2020. Diabetes Care. 2020;43(Suppl 1):S14–31.
6. Malik RA. Diabetic neuropathy: a focus on small fibres. Diabetes Metab Res Rev. 2020;36(Suppl 1):e3255.
7. Elafros MA, Andersen H, Bennett DL, et al. Towards prevention of diabetic peripheral neuropathy: clinical presentation, pathogenesis, and new treatments. Lancet Neurol. 2022;21(10):922–36.
8. Fang CH, Wu YL, Zhao JD. Guidelines for clinical diagnosis and treatment of diabetic peripheral neuropathy (2016 edition). J Tradit Chin Med. 2017;58(07):625–30.
9. Herman WH, Pop-Busui R, Braffett BH, et al. Use of the Michigan Neuropathy Screening Instrument as a measure of distal symmetrical peripheral neuropathy in type 1 diabetes: results from the Diabetes Control and Complications Trial/ Epidemiology of Diabetes Interventions and complications. Diabet Med. 2012;29(7):937–44.
10. Bril V, Tomioka S, Buchanan RA, et al. mTCNS Study Group. Reliability and validity of the modified Toronto clinical neuropathy score in diabetic. Diabet Med. 2009;26(3):240–6.
11. Chan KY, Kwong CK, Dillon TS, et al. Reducing overfitting in manufacturing process modeling using a backward elimination based genetic programming[J]. Appl Soft Comput. 2011;11(2):1648–56.
12. Lee S, Zhou J, Wong WT, et al. Glycemic and lipid variability for predicting complications and mortality in diabetes mellitus using machine learning[J]. BMC Endocr Disorders. 2021;21:1–15.
13. Lundberg S. A unified approach to interpreting model predictions[J]. arXiv preprint arXiv:1705.07874, 2017.
14. Ponomartseva DA, Derevitskii IV, Kovalchuk SV, et al. Prediction model for thyrotoxic atrial fibrillation: a retrospective study[J]. BMC Endocr Disorders. 2021;21:1–14.
15. Balachandran VP, Gonen M, Smith JJ, et al. Nomograms in oncology: more than meets the eye[J]. Lancet Oncol. 2015;16(4):e173–80.
16. Feng Y, Schlösser FJ, Sumpio BE. The Semmes Weinstein monofilament examination as a screening tool for diabetic peripheral neuropathy. J Vasc Surg. 2009;50(3):675–82.
17. Javed S, Hayat T, Menon L, et al. Diabetic peripheral neuropathy in people with type 2 diabetes: too little too late. Diabet Med. 2020;37(4):573–9.
18. Al-Bazz DY, Nelson AJ, Burgess J, et al. Is nerve Electrophysiology a Robust primary endpoint in clinical trials of treatments for Diabetic Peripheral Neuropathy? Diagnostics (Basel). 2022;12(3):731.
19. Chowdhury SK, Smith DR, Fernyhough P. The role of aberrant mitochondrial bioenergetics in diabetic neuropathy. Neurobiol Dis. 2013;51:56–65.
20. Bjerg L, Hulman A, Carstensen B, et al. Development of Microvascular complications and Effect of concurrent risk factors in type 1 diabetes: a Multistate Model from an Observational Clinical Cohort Study. Diabetes Care. 2018;41(11):2297–305.
21. Tesfaye S, Stevens LK, Stephenson JM, et al. Prevalence of diabetic peripheral neuropathy and its relation to glycaemic control and potential risk factors: the EURODIAB IDDM complications Study. Diabetologia. 1996;39(11):1377–84.
22. Liu YP, Shao SJ, Guo HD. Schwann cells apoptosis is induced by high glucose in diabetic peripheral neuropathy. Life Sci. 2020;248:117459.
23. Selvarajah D, Kar D, Khunti K, et al. Diabetic peripheral neuropathy: advances in diagnosis and strategies for screening and early intervention. Lancet Diabetes Endocrinol. 2019;7(12):938–48.
24. Naqvi SS, Z H, Imani S, Hosseinifard H, et al. Associations of serum low-density lipoprotein and systolic blood pressure levels with type 2 diabetic patients with and without peripheral neuropathy: systemic review, meta-analysis and meta-regression analysis of observational studies[J]. BMC Endocr Disorders. 2019;19:1–16.
25. Vincent AM, Hayes JM, McLean LL, et al. Dyslipidemia-induced neuropathy in mice: the role of oxLDL/LOX-1. Diabetes. 2009;58(10):2376–85.
26. Padilla A, Descorbeth M, Almeyda AL, et al. Hyperglycemia magnifies Schwann cell dysfunction and cell death triggered by PA-induced. Brain Res. 2011;1370:64–79.
27. Davis TM, Yeap BB, Davis WA, et al. Lipid-lowering therapy and peripheral sensory neuropathy in type 2 diabetes: the Fremantle Diabetes Study. Diabetologia. 2008;51(4):562–6.

28. Ruegsegger GN, Creo AL, Cortes TM, et al. Altered mitochondrial function in insulin-deficient and insulin-resistant states. J Clin Invest. 2018;128(9):3671–81.
29. Dhatariya KK, Glaser NS, Codner E, et al. Diabetic ketoacidosis. Nat Rev Dis Primers. 2020;6(1):40.
30. Donath MY, Dinarello CA, Mandrup-Poulsen T. Targeting innate immune mediators in type 1 and type 2 diabetes. Nat Rev Immunol. 2019;19(12):734–46.
31. Tzvetanova ID, Nave KA. Axons hooked to Schwann cell metabolism. Nat Neurosci. 2014;17(10):1293–5.
32. Bjerg L, Hulman A, Charles M, et al. Clustering of microvascular complications in type 1 diabetes mellitus. J Diabetes Complications. 2018;32(4):393–9.
33. Churdchomjan W, Kheolamai P, Manochantr S, et al. Comparison of endothelial progenitor cell function in type 2 diabetes with good and poor glycemic control[J]. BMC Endocr Disorders. 2010;10:1–10.
34. Li J, Cao Y, Liu W, et al. Correlations among Diabetic Microvascular complications: a systematic review and Meta-analysis. Sci Rep. 2019;9(1):3137.
35. Calcutt NA. Diabetic neuropathy and neuropathic pain: a (con)fusion of pathogenic mechanisms? Pain. 2020;161(Suppl 1):S65–86.
36. Rosenbloom AL. Distinguishing type 1 and type 2 diabetes at diagnosis. What is the problem? Pediatr Diabetes. 2007;8(2):51–2.
37. Lau KHV. Laboratory evaluation of Peripheral Neuropathy. Semin Neurol. 2019;39(5):531–41.

Publisher's note