



Risk assessment of coronary artery occlusion based on integrated Chinese and western medicine data

ZHANG Jiyu, XU Jiatusuo*, TU Liping, WANG Yu

School of Traditional Chinese Medicine, Shanghai University of Traditional Chinese Medicine, Shanghai 200120, China

ARTICLE INFO

Article history

Received 30 August 2024

Accepted 29 November 2024

Available online 25 December 2024

Keywords

Machine learning

Coronary heart disease (CHD)

Traditional Chinese medicine (TCM)

Diagnosis

Risk prediction

Comprehensive risk model

ABSTRACT

Objective To develop an integrated risk model for coronary artery occlusion based on data of both traditional Chinese medicine (TCM) and western medicine data, and to evaluate the contribution of TCM-specific indicators to conventional coronary heart disease (CHD) risk prediction.

Methods Data of TCM indicators (tongue, facial, and pulse diagnostics) and clinical parameters from patients diagnosed with CHD at the Cardiology Department of Shanghai Baoshan Hospital of Integrated Traditional Chinese and Western Medicine, from October 3, 2023 to March 15, 2024, were collected. Important variables were identified using importance screening and correlation analysis with CHD risk factors and laboratory markers. Six machine learning models including logistic regression (LR), decision tree (DT), support vector machine (SVM), k-nearest neighbors (KNN), random forest (RF), and gradient boosting (GB), were applied to evaluate the risk of coronary artery obstruction by combining clinical and TCM data of CHD. Model performance was assessed using metrics such as accuracy, precision, and recall, with reliability validated through ten-fold cross-validation.

Results A total of 288 patients were included in the study. Fifteen clinical risk factors, including body mass index (BMI), myoglobin, and alcohol consumption history, were incorporated into the diagnostic models. The KNN model showed good performance when combining clinical data with tongue and facial data. The SVM model performed well when clinical data was combined with pulse data. Among all the models, the KNN model with 10-fold cross-validation, which integrates the three types of TCM diagnostic data (tongue, face, and pulse) with clinical data, performs the best (accuracy: 0.837, precision: 0.814, and recall: 0.809).

Conclusion Incorporating TCM diagnostic data can enhance the accuracy of coronary artery obstruction risk assessment. The KNN prediction model that integrate tongue, facial, and pulse data performs the best and can be recommended as a clinical decision support tool.

1 Introduction

The incidence of coronary heart disease (CHD) has been rising steadily in recent years. Given the invasiveness of

coronary angiography, which is considered the “gold standard” for diagnosing CHD, non-invasive diagnostic approaches have garnered significant attention. Traditional Chinese medicine (TCM) diagnostics, which

*Corresponding author: XU Jiatusuo, E-mail: xjt@fudan.edu.cn.

Peer review under the responsibility of Hunan University of Chinese Medicine.

DOI: 10.1016/j.dcmcd.2025.01.009

Citation: ZHANG JY, XU JT, TU LP, et al. Risk assessment of coronary artery occlusion based on integrated Chinese and western medicine data. Digital Chinese Medicine, 2024, 7(4): 419-428.

include inspection, auscultation and olfaction, inquiry, and palpation, offer a non-invasive alternative that avoids the risks associated with invasive procedures. Moreover, diagnostic methods based on the objectification of TCM four diagnostic techniques have demonstrated a certain degree of specificity [1, 2]. With the rapid development of artificial intelligence in the medical field, research on CHD risk assessment using deep learning has also advanced significantly [3-5]. Feature-based imaging can be a reference for combined CHD prediction and may provide an alternative for CHD screening in clinical practice [6].

In recent years, machine learning and deep learning have become focal points in CHD research. These techniques enable the prediction and evaluation of CHD occurrence and progression through the selection and combined analysis of relevant indicators [7-14]. Machine learning-based assessment methods can swiftly predict the onset of CHD in patients, facilitating early screening and reducing the discomfort associated with invasive diagnostics. TCM has notable advantages in diagnosing non-invasive CHD. Practitioners can assess and treat heart-related issues based on syndrome differentiation by utilizing TCM diagnostic methods such as inspection, auscultation and olfaction, inquiry, and palpation, collectively known as “four diagnostic methods”. However, due to the inherent subjectivity of TCM diagnostics, variations in diagnosis may occur between practitioners. Recent research into the objectification of TCM diagnostic information has shown that objectifying these methods benefits CHD diagnosis [15]. TCM diagnostic approaches, including facial diagnosis [16], tongue diagnosis [17], and pulse diagnosis [18], offer specific advantages in diagnosing CHD. TCM emphasizes a “combined four-diagnostic approach”, which integrates various diagnostic techniques to provide a comprehensive judgment of disease, aligning in specific ways with modern deep learning’s multi-modal data integration. Integrating data from multiple sources enable a comprehensive evaluation of outputs. In medical multi-modal integration, data from different sources can complement each other, enhancing the model’s predictive accuracy and robustness [19].

Therefore, this study aims to develop a non-invasive diagnostic model by combining tongue, facial, and pulse diagnosis data from TCM with laboratory tests from western medicine, creating an integrated TCM-western model. Through the integration of TCM and western medical data, this research aspires to develop a non-invasive diagnostic tool for assessing coronary artery obstruction, ultimately assisting in clinical decision-making for CHD.

2 Data and methods

2.1 Data collection

The cases were sourced from inpatients in the Cardiology Department of Shanghai Baoshan Hospital of Integrated

Traditional Chinese and Western Medicine between October 3, 2023 to March 15, 2024. This study has been reviewed and approved by the Ethics Committee of Shuguang Hospital Affiliated to Shanghai University of Traditional Chinese Medicine (2020-916-125), with clinical registration number ChiCTR2100043546. It complies with the Helsinki Declaration of the World Medical Association Guidelines. All participants in this study provided written informed consent prior to their inclusion. The objectives, procedures, potential risks, and anticipated benefits of the study were comprehensively communicated to each participant. The research was conducted in full compliance with the principles of the Declaration of Helsinki, ensuring that the rights, safety, and welfare of all participants were safeguarded throughout the study.

2.2 Diagnostic criteria

The diagnostic criteria for acute and chronic CHD were based on the 9th edition of *Internal Medicine*, published by the People’s Medical Publishing House [20]. The diagnosis includes patients with coronary heart disease, specifically classified as follows. (i) Chronic CHD: including stable angina pectoris (SAP), myocardial ischemia, and post-myocardial infarction conditions. (ii) Acute coronary syndrome (ACS): including acute myocardial infarction (AMI) and unstable angina pectoris (UAP). Patients were included if they met the diagnostic criteria for CHD, exhibited typical symptoms of chest pain (such as paroxysmal angina or compressive pain), had diminished heart sounds upon auscultation, showed abnormal S-T segment changes on electrocardiograms, and had coronary angiography results showing at least one vessel stenosis, as outlined in the 2019 ESC Guidelines for the Diagnosis and Management of Chronic Coronary Syndromes [21].

2.3 Inclusion criteria

Patients enrolled in this study were required to meet the diagnostic criteria for CHD, be aged between 20 and 85 years, and voluntarily sign an informed consent form after being fully informed of the study’s objectives, procedures, potential risks, and benefits. In addition, participants were required to provide complete and reliable clinical data and possess the physical and mental capacity to comply with the study requirements. The inclusion criteria were designed to ensure the standardization of study participants and the reliability of the research findings, while adhering to ethical principles.

2.4 Exclusion criteria

Patients were excluded if they did not meet the inclusion criteria for CHD, were younger than 20 years or older than 85 years, had malignant tumors, were pregnant or breastfeeding, or if their data were incomplete. Patients

were also excluded if they were in a “critical condition,” defined as those requiring mechanical ventilation, having hemodynamic instability, or experiencing life-threatening organ failure. Furthermore, patients with other severe cardiovascular diseases, such as advanced heart failure (New York Heart Association class III – IV) or severe arrhythmias, as well as those with significant hepatic or renal dysfunction, such as decompensated liver cirrhosis or end-stage renal disease requiring dialysis, were excluded.

2.5 Measurements and indicator analysis

The patients were classified into two groups according to whether the degree of coronary artery occlusion was greater than 75%. At the same time, the clinical indicators of patients were differentiated according to the results of two categories.

In the context of CHD diagnosis, facial feature images of patients can serve as one of the diagnostic references [22-25]. TCM pulse indicators are also associated with cardiovascular and cerebrovascular diseases [26]. The study utilized the TFDA-1 digital tongue and facial diagnosis instrument, developed by the Shanghai University of Traditional Chinese Medicine, to collect tongue and facial images of the patients. The images were analyzed using the university’s proprietary TCM Tongue Diagnosis Analysis System (TDAS) V2.0.

Tongue body indicators in the evaluation include colour space values from different colour domains. Texture indicators include contrast (CON), angular second moment (ASM), entropy (ENT), and mean (MEAN). Texture indicators reflect the fineness and depth of texture in the image. An enormous ASM value corresponds to smaller CON, ENT, and MEAN values, indicating finer texture. PerAll and PerPart represent tongue coating indices, where PerAll is the ratio of tongue coating area to the total tongue area, and PerPart is the ratio of tongue coating area to the area without coating. “TB” represents the tongue body, and “TC” represents the tongue coating.

In the evaluation indicators of the pulse wave, h1 represents the compliance of large arteries and the left ventricular ejection function, while h3 indicates arterial elasticity and peripheral resistance, h4 pertains to peripheral vascular resistance and the aortic valve function, and h5 is associated with extensive artery compliance and aortic valve function. The t1 aligns with the left ventricular rapid ejection phase, t4 corresponds to the left ventricular systolic period, and t5 relates to the left ventricular diastolic period. t denotes the cardiac cycle of the left ventricle. The h3/h1 ratio provides insights into peripheral resistance and vascular wall compliance, while the h1/t1 ratio characterizes the strength of cardiovascular function. The h4/h1 ratio is indicative of peripheral vascular resistance. In the w1/t ratio, w1 measures the primary wave’s upper one-third, associated with the duration of elevated aortic pressure. Similarly, in the w2/t ratio, w2 measures the upper one-fifth of the primary wave, reflecting the duration

of elevated aortic pressure. The t1/t ratio evaluates the strength of cardiac ejection function, and the t4/t5 ratio captures heart rate variability.

The analysis system integrates four core categories of indicators. The first is tongue indicators, which include texture parameters (TB_Con, TB_ASM, TB_ENT, and TB_MEAN) and multidimensional color space data. The second is tongue coating indicators, encompassing coating coverage ratios (perPart and perAll), texture features (TC_Con, TC_ASM, TC_ENT, and TC_MEAN), and their color information. The third category is facial indicators, primarily collecting color space parameters from the face and lips. All color data is based on four standard color spaces: HSV [hue (H), saturation (S), value (V)], RGB [red (R), green (G), blue (B)], Lab [lightness (L), green-red chromaticity (a), blue-yellow chromaticity (b)], and YCr-Cb [luminance (Y), red chrominance (Cr), blue chrominance (Cb)]. The final category is pulse diagnosis indicators, which include pulse strength (h1, h3, h4, h5), temporal characteristics (t, t1, t4, t5), and several crucial ratio parameters (h3/h1, h1/t1, h4/h1, t1/t, t4/t5, w1/t, w2/t).

2.6 Indicator importance screening

CHD risk factors, including total cholesterol, triglycerides, high-density lipoprotein, low-density lipoprotein, glycated hemoglobin, creatine kinase-myocardial band (CK-MB), myoglobin, history of diabetes, and history of hypertension, were integrated with thromboelastography indicators (clotting time, clot formation rate, clot aggregation rate, clot lysis rate, percentage of clot lysis, coagulation index, platelet AA inhibition rate, and platelet ADP inhibition rate), hemorheological indicators [whole blood reduction at 200 s^{-1} , 30 s^{-1} , 5 s^{-1} , and 1 s^{-1} ; plasma viscosity; whole blood high-shear relative index; whole blood low-shear relative index; erythrocyte aggregation index; and Casson viscosity], and coagulation indicators [prothrombin time, international normalized ratio, partial thromboplastin time, fibrinogen concentration, thrombin time, D-dimer assay, fibrin (ogen) degradation products, and antithrombin III activity assay] for model selection. The optimal model for predicting individual clinical indicators was used as the basis for feature importance selection, employing the Permutation Importance method. This method’s principle involves randomly shuffling each feature’s values and observing the degree of decline in model performance metrics (e.g., accuracy and F1 score). Features causing greater declines are deemed to have higher contributions to the model. To mitigate overfitting due to excessive data dimensionality, a threshold of 0.5 was applied for feature importance scores, retaining only features with importance scores above 0.5. This selection method was independently applied to tongue, facial, and pulse diagnoses. For clinical data, the top 15 features ranked by importance were included in the selection process.

2.7 Indicator correlation screening

As identified by the deep learning model, TCM indicators with an importance score greater than 0.5 were further analyzed for their correlation with CHD risk factors and laboratory indicators. TCM indicators with a correlation coefficient less than 0.3 were filtered out.

2.8 Risk assessment model

The machine learning framework was developed using the Python (V3.12.3). Six commonly used deep learning multi-classification models were employed for model prediction and evaluation, including logistic regression (LR), decision tree (DT), support vector machine (SVM), k-nearest neighbors (KNN), random forest (RF), and gradient boosting (GB).

A total of 30 parameter combinations were tested, including baseline factors alone, as well as combinations of baseline factors with tongue image data, pulse data, facial data, and the integration of tongue, facial, and pulse data. The optimal model was selected based on these combinations. To ensure robustness, the mean result from ten-fold cross-validation was used as the final model performance metric, minimizing the impact of data selection variability.

2.9 Statistical analysis

The data were analyzed using SPSS 27.0. Measurement data conforming to a normal distribution were expressed as the mean \pm standard deviation (SD); while non-normally distributed data were presented as the median and interquartile range median (Q_1 , Q_3). For inter-group

comparisons, an independent samples t test was employed for data meeting the assumptions of normal distribution and homogeneity of variance, whereas the Mann-Whitney U test was used for data not satisfying these assumptions. $P < 0.05$ was considered statistically significant.

3 Results

3.1 Comparison of clinical data

A total of 288 CHD patients were included, comprising 140 males and 148 females. As shown in Table 1, when coronary artery occlusion greater than 75% was used as binary data in patients with CHD, statistical differences were observed in height, clotting time, regulation index, total cholesterol, and high-density lipoprotein ($P < 0.05$).

3.2 Indicator parameter screening results

3.2.1 Indicator importance screening (i) Baseline indicator importance. The top 15 indicators contributing to the model were selected as the basis for model construction. These included BMI, myoglobin, family history, drinking history, antithrombin III activity, triglycerides, CK-MB, fibrin (ogen) degradation products, clot aggregation rate (deg), clotting time, fibrinogen, D-dimer, whole blood high shear relative index, clot lysis rate, and clot dissolution percentage (Figure 1A).

(ii) Importance of lip colour indicators. These indicators include H_Lips, a_Lips, B_Lips, and S_Lips values (Figure 1B).

(iii) Importance of facial colour indicators. These

Table 1 Baseline data of included CHD patients

Classification	Age (year)	Height (m)	Body mass index (kg/m ²)	Presence of hypertension	Presence of diabetes	Current smoking status	Clotting time (min)	Clot formation rate (min)
Vascular obstruction < 75% (n = 172)	68.00 (62.00, 75.00)	1.63 (1.58, 1.70)	25.24 (22.73, 27.23)	116	57	56	5.00 (4.38, 5.80)	1.75 (1.48, 2.12)
Vascular obstruction \geq 75% (n = 116)	70.00 (61.00, 75.25)	1.65 (1.60, 1.72)	24.46 (22.84, 26.54)	76	48	56	4.95 (4.00, 5.80)	1.70 (1.30, 2.12)
<i>t/U</i> value	9 339	8 533.5	10 621	10 621	10 168	9 154	8 408	5 541
<i>P</i> value	0.358 2	0.037 1	0.352 5	0.352 5	0.735 1	0.155 2	0.007 4	0.542 9
Classification	Clot aggregation rate (%)	Clot dissolution rate (%)	Clot lysis percentage (%)	Coagulation index	Platelet aggregation (mm)	Inhibition rate (%)	Platelet aggregation (mm)	Inhibition rate (%)
Vascular obstruction < 75% (n = 172)	66.30 (61.80, 69.53)	0.00 (0.00, 0.33)	0.00 (0.00, 0.00)	0.61 \pm 2.55	14.70 (10.30, 23.40)	90.60 (69.70, 98.21)	33.90 (23.93, 45.25)	51.95 \pm 24.21
Vascular obstruction \geq 75% (n = 116)	66.60 (61.80, 71.35)	0.00 (0.00, 0.33)	0.00 (0.00, 0.10)	1.21 \pm 2.58	15.00 (10.10, 24.80)	92.60 (72.90, 98.90)	32.10 (23.57, 46.33)	53.20 \pm 23.96
<i>t/U</i> value	4 915	10 399	4 918.5	- 1.67	2 810	2 667.5	2 302	- 0.3
<i>P</i> value	0.395 3	0.501 0	0.270 4	0.095 7	0.858 5	0.718 5	0.973 9	0.764 2

Table 1 Continued

Classification	Whole blood viscosity at 200 s ⁻¹ (mPas)	Whole blood viscosity at 30 s ⁻¹ (mPas)	Whole blood viscosity at 5 s ⁻¹ (mPas)	Whole blood viscosity at 1 s ⁻¹ (mPas)	Plasma viscosity (mPas)	High shear relative index	Low shear relative index	Erythrocyte aggregation index
Vascular obstruction < 75% (n = 172)	4.19 ± 0.89	5.18 ± 1.05	7.77 (6.80, 8.70)	15.22 (12.97, 17.89)	1.55 (1.44, 1.65)	2.76 (2.38, 3.07)	10.20 ± 2.15	3.79 ± 0.76
Vascular obstruction ≥ 75% (n = 116)	4.30 ± 0.87	5.30 ± 1.09	8.08 (6.66, 9.19)	15.09 (13.01, 18.74)	1.59 (1.45, 1.65)	2.81 (2.50, 3.08)	10.34 ± 2.42	3.75 ± 0.68
t/U value	- 0.61	- 0.57	1529	1566.5	1565.5	1509.5	- 0.34	0.28
P value	0.544 9	0.571 6	0.700 0	0.862 8	0.857 9	0.620 0	0.736 4	0.779 8

Classification	Casson viscosity (mPas)	Prothrombin time	International normalized ratio	Activated partial thromboplastin time (s)	Fibrinogen (g/L)	D-Dimer assay (µg/mL)	Fibrin (ogen) degradation products (µg/mL)	Antithrombin III activity (%)
Vascular obstruction < 75% (n = 172)	3.54 (3.04, 4.03)	11.40 (10.90, 11.90)	0.99 (0.94, 1.04)	25.30 (21.90, 29.20)	2.96 (2.57, 3.41)	0.28 (0.18, 0.55)	2.50 (1.85, 2.50)	90.35 ± 11.84
Vascular obstruction ≥ 75% (n = 116)	3.28 (3.05, 4.01)	11.20 (10.80, 11.90)	0.97 (0.94, 1.04)	26.20 (22.50, 28.10)	3.12 (2.63, 3.62)	0.29 (0.20, 0.47)	2.50 (1.70, 2.50)	90.97 ± 12.63
t/U value	1 653	10 287.5	10 273	9 580.5	8 950	9 618.5	10 157	- 0.42
P value	0.755 3	0.564 2	0.578 3	0.713 8	0.172 6	0.755 5	0.69	0.674 6

Classification	Total cholesterol (mmol/L)	Triglycerides (mmol/L)	High-density lipoprotein (mmol/L)	Low-density lipoprotein (mmol/L)	Glycated hemoglobin (%)	Creatine kinase (U/L)	Creatine kinase isoenzyme (U/L)	Myoglobin (ng/mL)
Vascular obstruction < 75% (n = 172)	4.21 (3.41, 5.10)	1.32 (0.94, 1.84)	1.09 (0.93, 1.28)	2.51 (1.88, 3.12)	6.15 (5.70, 7.00)	77.00 (56.00, 98.00)	1.12 (0.70, 1.62)	39.15 (30.72, 53.83)
Vascular obstruction ≥ 75% (n = 116)	3.80 (3.15, 4.53)	1.44 (1.05, 1.94)	0.96 (0.85, 1.13)	2.19 (1.82, 2.84)	6.30 (5.80, 7.30)	80.00 (58.75, 103.00)	1.18 (0.77, 1.73)	43.39 (32.81, 57.86)
t/U value	9 788	7 735.5	10 508	9 400	7 299.5	9 252.5	9 148	8 886.5
P value	0.020 1	0.294 1	0.000 5	0.091 7	0.262 5	0.296 9	0.281 8	0.116 2

indicators include B_Face, H_Face, S_Face, L_Face, a_Face, b_Face, and Cb_face values (Figure 1C).

(iv) Importance of tongue indicators. These include the TB_H, TB_S, TB_Lb, TB_B, and TB_Cb values, as well as tongue texture indicators perPart_Tougue and perAll_Tougue (Figure 1D).

(v) Importance of pulse indicators. These selected indicators include pulse wave time indicators (t5, t4, and t1), pulse wave amplitude indicators (h4, h3, h1, and h5), and pulse wave ratio indicators (t4/t5, h1/t1, h3/h1, and h4/h1) (Figure 1E).

3.2.2 Indicator correlation screening (i) Correlation analysis of facial colour indicators and risk factors. As shown in Figure 2A, the B_Face, H_Face, S_Face, L_Face, b_Face, Cb_Face, and Cr_Face values were found to have significant correlations with various hemodynamic, coagulation, and metabolic parameters.

(ii) Correlation analysis of tongue body indicators and risk factors. As shown in Figure 2B, the perAll, perPart, b, and Cb indicators showed significant correlations with various coagulation and blood parameters.

(iii) Correlation analysis of pulse indicators and risk factors. Figure 2C demonstrates that the pulse indicators (h3/h1, t1, t4, t5, h1/t1, h4/h1, t1/t, t4/t5, and h5) show significant correlations with various hemorheological, coagulation, and metabolic parameters.

(iv) Correlation analysis of lip colour indicators and risk factors. Figure 2D illustrates that the L_Lips, Y_Lips, and G_Lips values show significant correlations with hemorheological parameters, fibrin degradation products, age, and hypertension.

3.3 Risk assessment model results

Coronary heart disease risk factors and the top 15 laboratory indicators ranked by importance were incorporated into the basic model (BMI, myoglobin, family history, history of alcohol consumption, antithrombin III activity, triglycerides, creatine kinase-MB, fibrinogen degradation products, clot aggregation rate deg, clotting time, fibrinogen, D-dimer assay, whole blood high-shear relative index, clot lysis rate, and percentage of clot lysis).

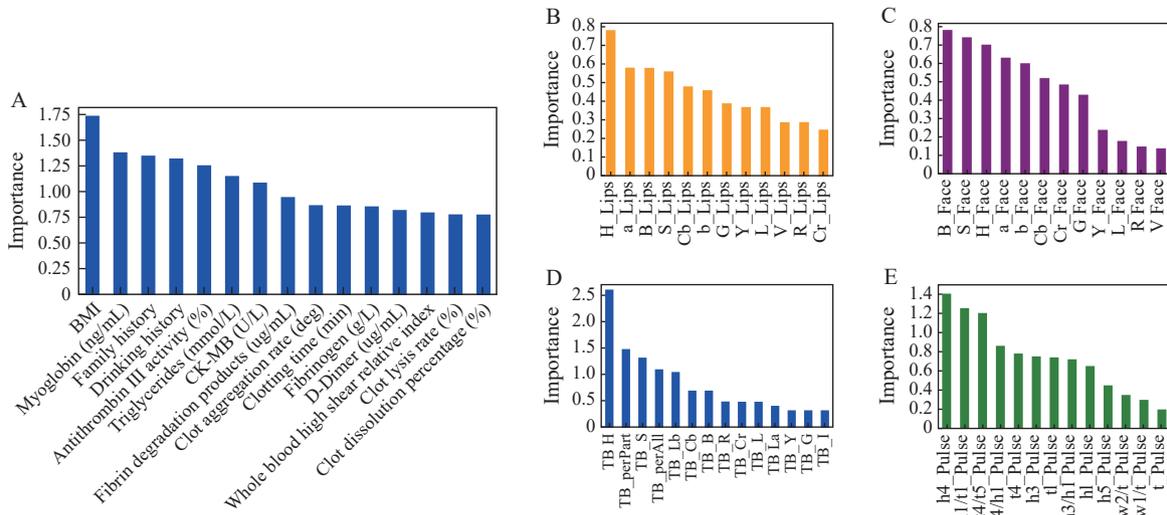


Figure 1 Parameter importance screening results of laboratory and TCM diagnostic indicators

A, the top 15 CHD clinical risk factors and laboratory indicators. B, importance of lips indicators. C, importance of facial indicators. D, importance of tongue indicators. E, importance of pulse indicators.

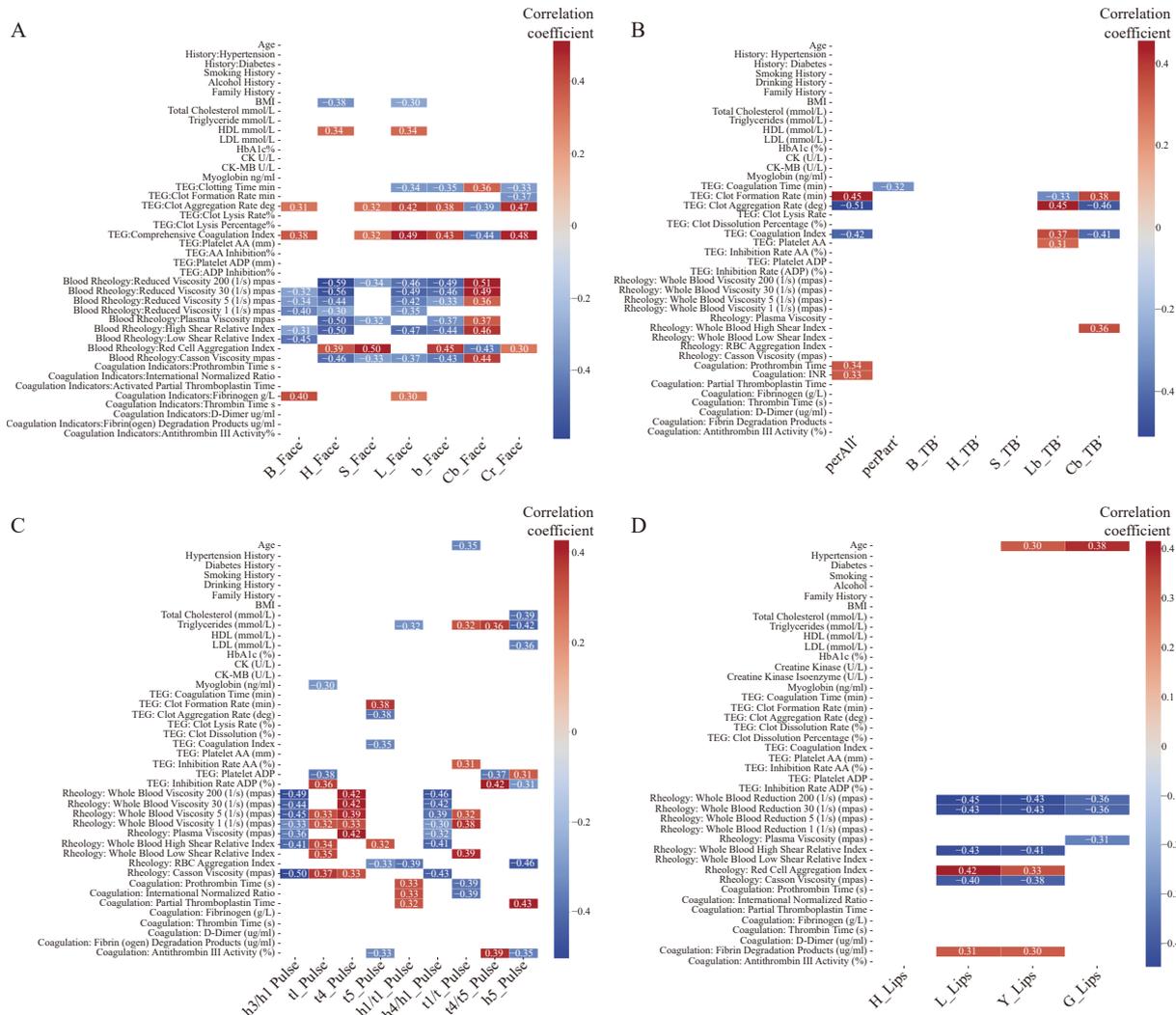


Figure 2 Correlation analysis between TCM diagnostic indicators, laboratory indicators, and risk factors

A, correlation between facial complexion indicators, laboratory indicators, and risk factors (absolute value greater than 0.3). B, correlation between tongue image indicators, laboratory indicators, and risk factors (absolute value greater than 0.3). C, correlation between pulse indicators, laboratory indicators, and risk factors (absolute value greater than 0.3). D, correlation between lip indicators, laboratory indicators, and risk factors (absolute value greater than 0.3).

Additionally, based on the importance of TCM indicators and their correlations, tongue indicators (perAll, perPart, TB-Lb, and TB-Cb) and pulse indicators (t5, t1, h4, h5, t4/t5, h1/t1, h3/h1, t1/t, and h4/h1) were included in the model. Facial colour indicators (Facial B, H, S, L, b, Cb, Cr, and Lips L, G, Y) were also integrated into the selection parameters.

3.3.1 Modelling based on clinical risk Six commonly used deep learning multi-classification models, including LR, DT, RF, GB, SVM, and KNN, were employed for model prediction and evaluation. Among them, the KNN model performed the best, while the SVM model exhibited comparable performance (Table 2).

Table 2 Performance comparison of different deep learning models

Model	Accuracy	Precision	Recall
LR	0.545	0.588	0.545
DT	0.545	0.588	0.545
RF	0.636	0.636	0.636
GB	0.455	0.539	0.455
SVM	0.727	0.529	0.612
KNN	0.727	0.702	0.708

3.3.2 Modeling based on clinical risk and tongue indicators As shown in Table 3, the predictive model incorporating tongue indicators achieved the best performance with the KNN model, followed by the SVM model.

Table 3 Performance comparison of tongue image indicator integration across models

Model	Accuracy	Precision	Recall
LR	0.545	0.588	0.545
DT	0.636	0.509	0.636
RF	0.545	0.485	0.545
GB	0.364	0.482	0.364
SVM	0.727	0.529	0.727
KNN	0.751	0.762	0.734

3.3.3 Modeling based on clinical risk and facial indicators As shown in Table 4, the KNN model achieved the highest performance, followed by the SVM model.

3.3.4 Modeling based on clinical risk and pulse indicators As shown in Table 5, the predictive model incorporating pulse indicators performed best with the SVM model, followed by the KNN model.

3.3.5 Modeling based on clinical risk, tongue, facial, and pulse indicators As shown in Table 6, the predictive model integrating tongue, facial, and pulse indicators achieved the best performance with the KNN model, followed by the SVM model. As illustrated in Figure 3, we conducted ten-fold cross-validation on the models to minimize data fluctuations during model training, with

the mean of ten-fold cross-validation results as the final output.

Table 4 Performance comparison of facial indicator integration across models

Model	Accuracy	Precision	Recall
LR	0.545	0.588	0.545
DT	0.545	0.588	0.545
RF	0.455	0.455	0.455
GB	0.455	0.539	0.455
SVM	0.727	0.529	0.727
KNN	0.812	0.796	0.762

Table 5 Performance comparison of pulse indicator integration across models

Model	Accuracy	Precision	Recall
LR	0.636	0.715	0.636
DT	0.455	0.539	0.455
RF	0.636	0.715	0.636
GB	0.455	0.539	0.455
SVM	0.818	0.855	0.818
KNN	0.771	0.778	0.758

Table 6 Performance comparison of tongue, facial, and pulse indicator integration across models

Model	Accuracy	Precision	Recall
LR	0.545	0.673	0.545
DT	0.545	0.485	0.545
RF	0.727	0.760	0.727
GB	0.455	0.539	0.455
SVM	0.818	0.855	0.818
KNN	0.837	0.814	0.809

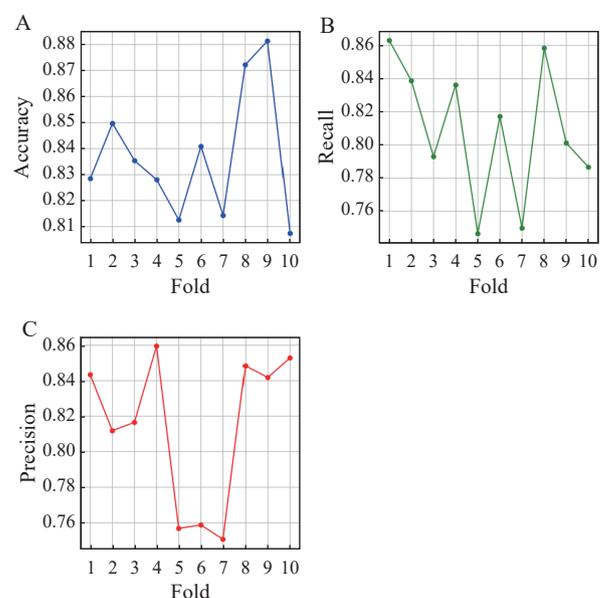


Figure 3 Ten-fold cross-validation of the KNN model using tongue, facial, pulse, and clinical risk data

A – C, ten-fold cross-validation of accuracy, precision, and recall, respectively.

4 Discussion

In recent years, there has been a growing body of research focused on the characteristics and information-based application of TCM diagnosis for CHD [26-30]. Specifically, TCM diagnostic methods, such as facial, pulse, and tongue diagnosis, have been shown to provide additional diagnostic insights. By integrating these methods with western medicine laboratory tests and risk factor analysis, a more comprehensive and accurate evaluation system is established. This approach has enhanced diagnostic accuracy and opened new avenues for non-invasive diagnosis of CHD. From a diagnostic perspective the combination of these TCM diagnostic methods makes risk assessment for CHD more comprehensive and holistic, overcoming the limitations of solely relying on western medical diagnostic methods. Recent study have also explored this integration [31]. Therefore, this study incorporates TCM indicators such as tongue, pulse, face, and lip colour data, aiming to provide a multidimensional, non-invasive diagnostic model that leverages as much data as possible.

4.1 Analysis of the results

In this study, the machine learning model based on the KNN model algorithm performed the best, with contributions from facial diagnosis, pulse, and tongue diagnosis leading to improvements of 0.085, 0.044, and 0.024 in prediction accuracy, respectively. Although not all models benefited from the inclusion of tongue, face, and pulse indicators, integrating TCM data into the predictive model significantly enhanced prediction accuracy compared with only using clinical data.

We also analyzed why the KNN model outperformed the other six models. KNN is an instance-based, non-parametric model that classifies samples based on the distance between the test sample and the training samples. When the data has clear class boundaries and samples of the same class form clusters in the feature space, KNN excels in classification tasks. In this study, the selected tongue and facial data were represented by colour space parameters of images, while pulse diagnosis data were derived from the time and pressure parameters of pulse waves. Laboratory indicators primarily focused on thromboelastography, blood rheology characteristics, and coagulation parameters, all of which exhibited distinct spatial categorical features. Consequently, KNN performed well with these data sets characterized by spatial classification indicators. However, we acknowledge the limitations of these indicators, and as such, additional dimensional features will be incorporated into future studies to improve model predictions.

4.2 Clinical application and practice

Based on the diagnostic model developed in this study, which integrates multidimensional data from TCM diagnostic methods such as tongue diagnosis, facial diagnosis, pulse diagnosis, and laboratory test indicators, this approach demonstrates significant advantages and potential in practical clinical applications. The non-invasive diagnostic model, constructed through the integration of TCM and western medical data, not only improves the accuracy of CHD risk assessment but also provides a novel technological tool for rapid screening and early warning.

In specific clinical practice, for general health check-up populations and high-risk groups suspected of CHD, the non-invasive diagnostic model can deliver fast and precise risk assessments, assisting doctors in promptly identifying potential high-risk patients. For individuals who are unable to undergo invasive examinations, such as coronary angiography, due to physical conditions or psychological factors, this model, combining TCM indicators with western medicine laboratory data, serves as an effective alternative. It supports doctors in performing non-invasive diagnoses and provides reliable reference data for subsequent treatment planning.

4.3 Innovation and deficiency

Artificial intelligence has introduced new opportunities for integrating TCM diagnostic methods into modern medicine, providing innovative technical approaches for the non-invasive diagnosis of CHD. This study demonstrates that combining non-invasive diagnostic data from both Chinese and western medicine significantly enhances the accuracy of coronary artery occlusion risk assessment, offering broad potential for clinical application. However, several limitations should be acknowledged. First, the sample size in this study was relatively small, and a larger sample would yield more robust data for disease prediction. Second, external data validation would offer a more direct evaluation of the model's predictive capabilities in real-world settings. Last, future research will explore the incorporation of heatmaps for visualizing model learning gradients, which could facilitate a more intuitive understanding of data and provide deeper insights into machine learning.

5 Conclusion

In evaluating the severity of coronary artery occlusion risk in patients with CHD, the KNN model integrating clinical data with TCM indicators, such as tongue, facial, and pulse characteristics, demonstrates significant advantages. The research findings suggest that this comprehensive model enhances the accuracy of non-invasive

evaluation for coronary artery occlusion, providing a valuable reference for the non-invasive diagnosis of coronary artery disease.

Fundings

National Natural Science Foundation of China (82104738), High-Level Key Discipline Construction Project in Traditional Chinese Medicine Diagnostics by the National Administration of Traditional Chinese Medicine (ZYYZDXK-2023069), and Clinical Research Project in the Health Industry by Shanghai Municipal Health Commission (20244Y0129).

Acknowledgements

This study was supported by the School of Traditional Chinese Medicine at Shanghai University of Traditional Chinese Medicine and the Center for Traditional Chinese Medicine Information Science and Technology at Shanghai University of Traditional Chinese Medicine. Thanks for the clinical data support from the Department of Cardiology at Shanghai Baoshan Hospital of Integrated Traditional Chinese and Western Medicine.

Competing interests

The authors declare no conflict of interest.

References

- [1] DUAN MY, WANG CH, TAN YQ, et al. Objectification study of tongue features in 315 patients with coronary heart disease. *Journal of Traditional Chinese Medicine*, 2024, 65(9): 921-927.
- [2] GONG P, LI Y, YAO CZ, et al. Traditional Chinese medicine on the treatment of coronary heart disease in recent 20 years. *Journal of Alternative and Complementary Medicine*, 2017, 23(9): 659-666.
- [3] MUNIASAMY A, BEGUM A, SABAHATH A, et al. Coronary heart disease classification using deep learning approach with feature selection for improved accuracy. *Technology and Health Care*, 2024, 32(3): 1991-2007.
- [4] LI F, CHEN Y, XU HZ. Coronary heart disease prediction based on hybrid deep learning. *The Review of Scientific Instruments*, 2024, 95(1): 015115.
- [5] NISHI T, YAMASHITA R, IMURA S, et al. Deep learning-based intravascular ultrasound segmentation for the assessment of coronary artery disease. *International Journal of Cardiology*, 2021, 333: 55-59.
- [6] LIN S, LI ZG, FU BW, et al. Feasibility of using deep learning to detect coronary artery disease based on facial photo. *European Heart Journal*, 2020, 41(46): 4400-4411.
- [7] LI WQ, ZUO M, ZHAO HJ, et al. Prediction of coronary heart disease based on combined reinforcement multitask progressive time-series networks. *Methods*, 2022, 198: 96-106.
- [8] BAO CM, RAN H, JIA H, et al. Relationship between ultrasonic myocardial work, myocardial nuclear imaging parameters and left ventricular remodeling in coronary heart disease and its combined predictive efficiency. *Chinese Journal of Misdiagnostics*, 2024, 37(9): 53-58.
- [9] XU YJ, SONG XL, LIAO PC, et al. Construction of a nomogram prediction model for poor prognosis after coronary intervention in coronary heart disease patients. *Chinese Journal of Sanatorium Medicine*, 2024, 33(5): 18-22.
- [10] YUE HT, HE CC, CHENG YY, et al. Construction and comparison of coronary heart disease risk prediction models based on machine learning. *Chinese General Practice*, 2025, 28(4): 499-509.
- [11] LI WJ. Research and implementation of coronary heart disease diagnosis and prognosis algorithm based on neural networks. Chengdu: Sichuan University, 2023.
- [12] PENG CF, QIAN FD, WU KM, et al. Construction of a nomogram prediction model for one-year postoperative bleeding in patients with coronary heart disease and atrial fibrillation undergoing PCI based on preoperative hematological indicators and risk scores. *Sichuan Medical Journal*, 2024, 45(4): 362-368.
- [13] XIA GX, MA JJ, ZHAO ZH, et al. Construction, evaluation, and validation of a risk prediction model for coronary heart disease complicated with diabetes. *Journal of Mudanjiang Medical College*, 2024, 45(3): 47-54.
- [14] GAO XX, GAO YH. Predictive value of combined serum CTRP6 and FSTL1 testing for in-stent restenosis in coronary heart disease patients. *Chinese Journal of Molecular Cardiology*, 2024, 24(1): 5887-5892.
- [15] WANG YQ, GUO R, XU ZX, et al. Application of objective research on the four TCM diagnostic methods in coronary heart disease diagnosis. *Journal of Traditional Chinese Medicine*, 2016, 57(3): 199-203.
- [16] DU B, HU YH, JIANG YC, et al. Research ideas and discussions on the clinical application of facial recognition in blood stasis syndrome of coronary heart disease. *World Journal of Integrated Traditional and Western Medicine*, 2020, 15(2): 377-380.
- [17] WANG J, JIANG LH, CHANG TY, et al. Progress in the application of TCM tongue diagnosis in coronary heart disease. *Journal of Practical Cardiopulmonary and Cerebrovascular Diseases*, 2024, 32(3): 99-102.
- [18] ZHANG HF, LU XZ, YU ZF, et al. Exploration of pulse pattern changes in 289 coronary heart disease patients. *Western Journal of Traditional Chinese Medicine*, 2017, 30(6): 1-3.
- [19] SUN YY, JIA ZT, ZHU HY. Review on multimodal deep learning. *Computer Engineering and Applications*, 2020, 56(21): 1-10.
- [20] ZENG XF, CHEN HZ, ZHONG NS, et al. Internal Medicine (9th edition). Beijing: People's Medical Publishing House, 2018.
- [21] KNUUTI J, WIJNS W, SARASTE A, et al. 2019 ESC Guidelines for the Diagnosis and Management of Chronic Coronary Syndromes. *European Heart Journal*, 2020, 41(3): 407-477.
- [22] CAO YY, LI FF, WANG YQ, et al. Research on facial diagnostic color features in coronary heart disease patients based on clinical syndrome differentiation. *Chinese Journal of Traditional Chinese Medicine*, 2013, 31(9): 1867-1869.
- [23] GUAN Q, XU Y, YANG S, et al. Exploration of the relationship between TCM facial diagnosis characteristics and diseases. *Chinese Journal of Traditional Chinese Medicine*, 2022, 37(2): 902-905.
- [24] XU JT, JIANG T, LIU S. Research status and prospect of tongue

- image diagnosis analysis based on machine learning. *Digital Chinese Medicine*, 2024, 7(1): 3-12.
- [25] HUANG JL, LIANG MC, TANG Y, et al. Progress in the application of digital pulse diagnosis technology in cardiovascular diseases. *Chinese Journal of Traditional Chinese Medicine Information*, 2024, 31(7): 193-197.
- [26] PAN Y, XU SH, XU FW, et al. Research progress on the objectification of the four diagnostic methods in coronary heart disease. *Journal of Shanxi University of Traditional Chinese Medicine*, 2016, 17(4): 78-79.
- [27] OU Y, XIA SS, LI J, et al. Current status of the application of digital information technology in TCM clinical diagnosis of coronary heart disease. *China Medical Herald*, 2022, 19(7): 35-38.
- [28] XUE P, LIN L, ZHU YJ, et al. Exploration and construction of TCM deficiency syndrome diagnosis models in elderly coronary heart disease angina patients. *China Medical Herald*, 2024, 21(12): 27-31.
- [29] ZHANG SQ, SUN YH, XIAN NX, et al. Research progress on the objectification and intelligence of the four diagnostic methods in TCM. *Journal of Traditional Chinese Medicine Guide*, 2023, 29(6): 170-174.
- [30] GAO H, SONG XY, FENG L, et al. Research progress on the objectification of four diagnostic methods for Qi deficiency and blood stasis syndrome in coronary heart disease. *Chinese Journal of Traditional Chinese Medicine Information*, 2020, 27(5): 141-144.
- [31] LI HZ, YANG WW, QU H, et al. Reflections on the multimodal diagnostic model of integrated Chinese and Western medicine. *Journal of Cardiovascular Diseases Integrated Chinese and Western Medicine*, 2023, 21(21): 4020-4024.

基于中西医数据融合的冠状动脉阻塞风险评估

张冀豫, 许家佗*, 屠立平, 王瑜

上海中医药大学中医学院, 上海 200120, 中国

【摘要】目的 本研究旨在构建基于中西医数据融合的冠状动脉阻塞风险模型, 评价中医特色指标在传统冠心病风险预测上的增益。**方法** 收集 2023 年 10 月 3 日至 2024 年 3 月 15 日上海市宝山区中西医结合医院心内科被诊断为冠心病患者的中医舌、面、脉与临床指标数据, 通过参数重要度筛选与冠心病风险因素及实验室指标相关性分析对数据进行重要度筛选, 选择逻辑回归 (LR)、决策树 (DT)、支持向量机 (SVM)、K 临近算法 (KNN)、随机森林 (RF) 和梯度提升 (GB) 6 种机器学习模型进行冠心病临床数据与中医数据的融合冠心病血管阻塞风险评估, 采用准确度、精密度、召回率指标进行模型评价, 通过十折交叉验证评价可靠性。**结果** 288 名患者被纳入研究中, 身体质量指数 (BMI)、肌红蛋白、饮酒史等 15 个临床风险因素纳入模型诊断。KNN 模型在结合临床数据与舌面部数据时表现良好。当临床数据与脉搏数据相结合时, SVM 模型表现良好。在所有模型中, 将三种中医诊断数据类型 (舌、面、脉) 与临床数据整合后, 使用十折交叉验证的 KNN 模型表现最佳 (准确率: 0.837, 精密度: 0.814, 召回率: 0.809)。**结论** 整合中医诊断数据可提升冠心病血管阻塞评估的准确度。结合了舌、面、脉数据的 KNN 预测模型表现最佳, 可作为临床决策支持工具。

【关键词】 机器学习; 冠心病; 中医; 诊断; 风险预测; 综合风险模型