ORIGINAL RESEARCH

Construction and Comparison of Machine Learning-Based Risk Prediction Models for Major Adverse Cardiovascular Events in Perimenopausal Women

Anjing Chen[1],*, Xinyue Chang[2],*, Xueling Bian[1], Fangxia Zhang[3], Shasha Ma[4], Xiaolin Chen[5]

[1]College of Nursing, Binzhou Medical University, Shandong, 256600, People's Republic of China; [2]Vascular Surgery Department, Shandong Provincial Hospital, Binzhou, Shandong, 250001, People's Republic of China; [3]CCU, Binzhou Medical University Hospital, Binzhou, Shandong, 256600, People's Republic of China; [4]Neurology Department, Binzhou Medical University Hospital, Binzhou, Shandong, 256600, People's Republic of China; [5]Office of Health Care, Binzhou Medical University Hospital, Binzhou, Shandong, 256600, People's Republic of China

*These authors contributed equally to this work

Correspondence: Xiaolin Chen, Office of Health Care, Binzhou Medical University Hospital, No. 661, Huanghe 2nd Road, Bincheng District, Binzhou, Shandong, 256600, People's Republic of China, Tel +86 13854380372, Email xiaolin750210@sina.com

Background: Perimenopausal period is a period of physiological changes in women with signs of ovarian failure, including menopausal transition period and 1 year after menopause. Ovarian function declines in perimenopausal women and lower estrogen levels lead to changes in the function of various organs, which may lead to cardiovascular disease. Major adverse cardiovascular events (MACE) are the combination of clinical events including heart failure, myocardial infarction and other cardiovascular diseases. Therefore, this study explores the factors influencing the occurrence of MACE in perimenopausal women and establishes a prediction model for MACE risk factors using three algorithms, comparing their predictive performance.

Patients and Methods: A total of 411 perimenopausal women diagnosed with MACE at the Binzhou Medical University Hospital were randomly divided into a training set and a test set following a 7:3 ratio. According to the principle of 10 events per Variable, the training set sample size was sufficient. In the training set, Random Forest (RF) algorithm, backpropagation neural network (BPNN) and Logistic Regression (LR) were used to construct a MACE risk prediction model for perimenopausal women, and the test set was used to verify the model. The prediction performance of the model was evaluated in terms of accuracy, sensitivity, specificity, and area under the subject operating characteristic curve (AUC).

Results: A total of twenty-six candidate variables were included. The area under ROC curve of the RF model, BPNN model, and logistic regression model was 0.948, 0.921, and 0.866. Comparison of ROC curve AUC between logistic regression and RF model for predicting MACE risk showed a statistically significant difference (Z=2.278, P=0.023).

Conclusion: The RF model showed good performance in predicting the risk of MACE in perimenopausal women providing a reference for the early identification of high-risk patients and the development of targeted intervention strategies.

Keywords: major adverse cardiovascular events, machine learning, perimenopause, risk factor

Introduction

Perimenopause is a stage of normal physiological changes in women characterized by the depletion of viable follicles in the ovaries leading to the permanent cessation of menstruation. Currently, over 850 million women worldwide are in perimenopause, and the total perimenopausal population in China will account for approximately 14.29% of the world's population by 2030.[1,2] Perimenopause triggers multi-system physiological changes causing a decline in estrogen levels, but also increased body fat and unstable vasoconstriction, increasing the risk of cardiovascular disease.[3] Studies have shown that the prevalence of cardiovascular disease in perimenopausal women is 13%, and the incidence of

cardiovascular disease in postmenopausal women increases rapidly to 49.3%, with mortality and disability rates as high as 31% and 18%; cardiovascular disease constitutes the main cause of high mortality and morbidity in postmenopausal women.[4,5] Therefore, early identification of risk factors for MACE in perimenopausal women is critical.[6] If perimenopausal women at risk factors for MACE can be accurately and promptly identified and timely interventions implemented, this can reduce MACE incidence and improve quality of life.

Currently, although researchers have studied early predictive models of MACE, these models seem unsuitable for accurately predicting MACE risk in perimenopausal women. Xi et al conducted a retrospective study and constructed a risk prediction model for the occurrence of atherosclerotic cardiovascular disease.[7] Ahamed et al combine the Internet of things, machine learning, and ensemble techniques to predict cardiovascular disease. In this study, five machine learning algorithms compared and contrasted to predict vascular disease. The result is shown that the Random Forest (RF) model outperforms other models. In addition, Research team further evaluate the seven machine learning algorithms by using hyperparameter tuning and ensemble techniques. As a result, the ensemble model outperforms the other seven machine learning models with an accuracy of 87.7%.[8,9] Qian et al constructed a risk prediction model based on a total of 12 predictors and performed Logistic Regression (LR) on data from 690 perimenopausal women; however, Logistic Regression relies on the selection of features, and the performance of the model will be affected if the features are not properly selected. In contrast, the backpropagation neural network (BPNN) model can approximate any complex nonlinear function. The RF model has the advantages of evaluating the importance of features, high accuracy, and fast training speed.[10,11]

Thus, the purpose of this study was to establish risk prediction models for MACE in perimenopausal women using Logistic Regression, Backpropagation Neural Network, and Random Forest algorithms, to accurately identify risk factors for MACE in this population. Furthermore, we compared the predictive performance of the three models to identify the most appropriate one for predicting the risk for MACE in perimenopausal women, providing a basis for prevention and early intervention strategies.

Materials and Methods
Data Collection

Perimenopausal women who were admitted to the cardiovascular ward of a Grade-A general hospital in Shandong Province from October 2023 to June 2024 and diagnosed with major adverse cardiovascular events were included in the MACE group (case group). Perimenopausal women who had no adverse cardiovascular events during physical examination during the same period were selected as the non-MACE group (control group). The specific inclusion criteria are described next. MACE group: 1) met the diagnostic criteria for perimenopause; 2) major cardiovascular adverse events were diagnosed for the first time; 3) voluntary participation in the study; 4) no mental illness, consciousness disorder, or other conditions, showing good cooperation. Non-mace group: 1) the diagnostic criteria for perimenopause were met and MACE were excluded; 2) voluntary participation in this study 3) no mental illness, consciousness disorders, or other conditions, showing good cooperation. Exclusion criteria: 1) patients with a history of hormone drug therapy in the past six months; 2) patients with malignant tumors, autoimmune system diseases, blood system diseases, serious organ diseases, etc. The study utilized a 1:4 matched case-control design, ensuring that four controls were randomly selected based on specific matching criteria.[12] According to the literature review, the incidence of major cardiovascular adverse events in perimenopause was about 20%, with an estimated OR value of 3, $\alpha=0.05$, $\beta=0.10$.[10,13] The sample size formula for the matched case-control study was determined as follows: $n = \frac{(1+1/c)\,pq\,(Z\alpha+Z\beta)^2}{(p1-p0)^2}$. The results showed that 82 cases were required in the MACE group and 328 cases were required in the non-MACE group. Therefore, a total of 411 cases were included. The diagnostic criteria for perimenopause were age ≥ 40 years, menstrual cycles varying in length (ie, menstrual disorders), with 2 or more out of 10 menstrual cycles showing menstrual cycle length changes of ≥ 7 days. Major adverse cardiovascular events included sudden cardiac death, malignant arrhythmia, heart failure, non-fatal myocardial infarction, unstable angina pectoris, and surgical coronary revascularization (percutaneous coronary intervention and coronary bypass grafting).[14,15] This study was reviewed and approved by the hospital Ethics Committee (2022–357), and all patients provided informed consent.

Demographic and Clinical Data Collection

In this study, a total of 26 factors influencing MACE in perimenopausal women were identified by the research team through a literature review and expert advice, including sociodemographic characteristics, medical history assessment, and physiological indicators. 1) Sociodemographic characteristics included age, literacy level, nature of work, preference for salted diet, preference for sweets, and average daily sleep duration, the above were obtained by questionnaire. The body weight was measured with an electronic scale with a maximum capacity of 200 kg. BMI was calculated by dividing the participant's weight in kilograms by the square of their height in meter (kg/m^2) (BMI > 25kg/m^2 for overweight and obese). 2) Medical history assessment included family history of cardiovascular disease, history of total hysterectomy, history of gestational obesity (weight gain of more than 15 kg during pregnancy), age at menarche, age at menopause, reproductive life expectancy (the age of menopause minus the age of menarche), the above determined by patient self-report. 3) Physiological indicators included blood pressure, waist-to-hip ratio (WHR), bone mineral density (BMD), estradiol, hemoglobin (HB), blood glucose (BG), calcium (Ca), total cholesterol (TC), low-density lipoprotein cholesterol (LDL-C), high-density lipoprotein cholesterol (HDL-C), triglycerides (TG), lipoproteins (LP(a)), and plasma homocysteine (Hcy). Blood pressure was measured with an electronic sphygmomanometer and averaged over three measurements (Blood pressure above 140/90 is hypertension). Waist and hip circumference were measured with a tape measure. The WHR was calculated by dividing waist circumference (cm) by the hip circumference (cm) (central obesity was defined as waist-to-hip ratio greater than 0.9). BMD was measured by X-ray absorptiometry. Other physiological indexes were detected from fasting blood samples. All data collection tools are pre-tested and validated before use. Double-entry checks are implemented during data entry to reduce errors. Outliers of each variable were identified and properly handled.

Statistical Analysis

SPSS 25.0 and R 4.4.0 software were used for statistical analysis of the data. Quantitative information not conforming to a normal distribution was expressed as M (P25, P75), and comparisons between the two groups were made using the Wilcoxon test. Frequencies and percentages indicated qualitative information, and the χ^2 test as used for comparison between the two groups. Univariate analysis and logistic regression analysis were conducted to identify risk factors for MACE in perimenopausal women; subsequently, the 12 risk factors showing significant results in the logistic regression analysis were included in the risk prediction models. Due to the low percentage of patients in the MACE group about 19.95%, the training set was processed by Random oversampling with a ratio of 7:3 corresponding to the training set and test set, respectively. A total of 288 samples in the training set were established with the RF Algorithm, BPNN, and Logistic Regression analysis, and 123 samples in the test set were used to verify the models. SPSS 25.0 was used to construct Logistic Regression model to predict the influencing factors of major adverse cardiovascular events in perimenopausal women, and BPNN was constructed by using the "nnet" package, "randomForest" package, and "randomForest" package in R4.4.0 software. The BPNN model and RF model were constructed using the "nnet" and "randomForest" packages in R4.4.0 software. The ROC curves of the three models were plotted to predict the risk of MACE in perimenopausal women, and the predictive efficacy of each model was evaluated by comparing the accuracy, sensitivity, specificity, F1 index and AUC.

Results

Patient Characteristics

A total of 411 perimenopausal female patients were collected in this study. The average age was 51.00 (49.00, 53.00) years. Marital status: Married 397 cases (96.59%), divorced 8 cases (1.95%), widowed 6 cases (1.46%). A total of 232 cases (56.45%) had an education level below high school, while 179 cases (43.55%) had a high school education or above. Moreover, 108 participants (26.28%) engaged in physical work and 303 participants

Table I General Data and Univariate Analysis of Major Cardiovascular Adverse Events in Perimenopausal Women

Characteristics	MACE N=82	None-MACE N=329	P value
Age, years, n (%)			
45–50	25(30.49%)	134(40.73%)	0.197[a]
51–55	49(59.76%)	173(52.58%)	
56–60	8(9.76%)	22(6.69%)	
Overweight/obesity, n (%)	57(69.51%)	163(49.54%)	0.001[a]
Degree of education, High school and above, n (%)	7(8.54%)	172(52.28%)	<0.001[a]
Non-physical work, n (%)	45(54.88%)	258(78.42%)	<0.001[a]
Hypertension, n (%)	33(40.24%)	80(24.32%)	0.004[a]
Central obesity, n (%)	67(81.71%)	98(29.79%)	<0.001[a]
Family history, n (%)	27(32.93%)	45 (13.68)	<0.001[a]
Panhysterectomy, n (%)	7(8.54%)	12(3.65%)	0.059[a]
Gestational obesity, n (%)	50(60.98%)	41(12.46%)	<0.001[a]
Age of menarche,>13,n (%)	75(91.46%)	239(72.64%)	<0.001[a]
Age of menopause,>47,n (%)	65(79.27%)	297(90.27%)	0.006[a]
Reproductive life,>30,n (%)	64(78.05%)	305(92.71%)	<0.001[a]
Salty diet,n (%)	268(81.46%)	61(18.54%)	<0.001[a]
Sweet food n (%)	39(47.56%)	100(18.54%)	0.003[a]
Mean daily sleep<6h, n (%)	39(47.56%)	44(13.37%)	<0.001[a]
Osteopenia, n (%)	69(84.15%)	70(21.28%)	<0.001[a]
Decreased estrogen levels, n (%)	55(67.07%)	123(37.39%)	<0.001[a]
Hb,g/L, median,(IQR)	131.50(123.75–143.25)	134.00(127.00–42.00)	0.564[b]
BG,mmol/L, median,(IQR)	5.60(5.04–6.41)	4.97(4.66–5.40)	<0.001[b]
Ca,mmol/L, median,(IQR)	2.35(2.23–2.42)	2.34(2.25–2.43)	0.447[b]
TC,mmol/L, median,(IQR)	4.52(4.00–5.42)	5.12(4.49–5.77)	<0.001[b]
LDL-C,mmol/L, median,(IQR)	2.55(2.18–3.33)	3.18(2.69–3.64)	<0.001[b]
HDL-C,mmol/L, median,(IQR)	1.33(1.13–1.48)	1.48(1.27–1.71)	<0.001[b]
TG,mmol/L, median,(IQR)	1.14(0.84–1.54)	1.14(0.82–1.60)	0.976[b]
Lp(a),mmol/L, median,(IQR)	19.21(8.45–49.82)	12.83(9.21–18.82)	0.001[b]
Hcy,umol/L, median,(IQR)	10.80(8.68–13.03)	8.90(7.40–10.50)	<0.001[b]

Notes: [a]χ^2 test, categorical variables were indicated as proportions. [b]Wilcoxon Test, continuous variables as the median with the interquartile range. $P < 0.05$ statistically significant difference.
Abbreviations: Hb, hemoglobin; BG, blood glucose; Ca, calcium; TC, total cholesterol; LDL-C, low-density lipoprotein cholesterol; HDL-C, high-density lipoprotein cholesterol; TG, triglyceride; Lp(a), lipoprotein; Hcy, plasma homocysteine.

(73.72%) did not engage in physical work. The general information of the MACE and non-MACE groups is shown in Table 1.

Multivariate Logistic Regression Analysis of Major Adverse Cardiovascular Events in Perimenopausal Women

Based on the univariate and multivariate Logistic regression analysis results, 12 variables were found to be independent predictors of MACE in perimenopausal women. In a univariate analysis, overweight/obesity (p=0.001), education level (p<0.001), non-physical work (p<0.001), hypertension (p=0.004), central obesity (p<0.001), age of menarche (p<0.001), age of menopause (p=0.006), reproductive life (p<0.001), salting diet (p<0.001), sweet food (p= 0.003), mean daily sleep <6h (p<0.001), gestational obesity (p<0.001), BG ($p < 0.001$), TC (p<0.001), LDL-c (p=0.048), HDL-c (p<0.001), Lp(a) (p=0.001), Hcy (p<0.001), osteopenia ($p < 0.001$), and decreased estrogen levels ($p < 0.001$) were screened as predictors of MACE in perimenopausal women. Multivariate logistic regression was performed to screen the significant variables related to MACE, revealing that the education level (odds ratio (OR) = 0.22; 95% confidence interval (CI): (0.08–0.92); p = 0.036), central obesity (OR = 7.73; 95% CI: (2.40–24.94); p =0.001), age of menarche (OR = 5.54; 95% CI:

Table 2 Multivariate Logistic Regression Analysis of Influencing Factors of Major Adverse Cardiovascular Events in Perimenopausal Women

Variables	Multivariable [a]OR (95% CI)	P value
Overweight/obesity	1.57(0.58–4.30)	0.376
High school and above	0.22(0.08–0.92)	0.036*
Non-physical work	0.52(0.20–1.38)	0.190
Hypertension	1.06(0.41–2.75)	0.091
Central obesity	7.73(2.40–24.94)	0.001**
Family history	2.78(0.92–8.36)	0.070
Age of menarche,>13	5.54(1.40–21.93)	0.015*
Age of menopause,>47	4.02(0.73–22.01)	0.109
Reproductive life,>30	0.15(0.02–0.90)	0.037*
Salting diet	1.94(0.74–5.11)	0.181
Sweetmeat	2.66(1.05–6.73)	0.040*
Mean daily sleep<6h	3.68(1.39–9.78)	0.009**
Gestational obesity	4.90(1.85–13.00)	0.001**
BG	1.08(0.76–1.53)	0.671
TC	0.91(0.41–1.99)	0.804
LDL-C	2.80(1.01–7.73)	0.048*
HDL-C	2.40(0.48–11.92)	0.284
Lp(a)	0.97(0.95–0.99)	0.002**
Hcy	0.83(0.74–0.94)	0.003**
Osteopenia	6.15(2.33–16.24)	<0.001**
Decreased estrogen levels	3.40(1.31–8.82)	0.012*

Notes: [a]Variables from univariate logistic regression with P < 0.1 were included in the multivariate logistic regression model using a forward stepwise logistic regression analysis and removed with P > 0.05.*p<0.05, **p<0.01.
Abbreviations: OR, odds ratio; CI, confidence interval; Hb, hemoglobin; BG, blood glucose; Ca, calcium; TC, total cholesterol; LDL-C, low-density lipoprotein cholesterol; HDL-C, high-density lipoprotein cholesterol; TG, triglyceride; Lp(a), lipoprotein; Hcy, plasma homocysteine.

(1.40–21.93); p= 0.015), reproductive life (OR = 0.15; 95% CI: (0.02–0.90); p =0.037), sweet food(OR =2.66; 95% CI: (1.05–6.73); p= 0.040), mean daily sleep <6h (OR = 3.68; 95% CI: (1.39–9.78); p= 0.009), gestational obesity (OR = 4.90; 95% CI: (1.85–13.00); p= 0.001), LDL-C (OR = 2.80; 95% CI: (1.01–7.73); p =0.048), Lp(a) (OR = 0.97; 95% CI: (0.95–0.99); p= 0.002), Hcy (OR = 0.84; 95% CI: (0.74–0.94); p =0.003), osteopenia (OR = 6.15; 95% CI: (2.33–16.24); p < 0.001), and decreased estrogen levels (OR =3.40; 95% CI: (1.31–8.82) p=0.012) were influencing factors of MACE (Table 2).

Construction and Comparison of MACE Risk Prediction Models in Perimenopausal Women

The importance analysis of independent variables in the RF model showed that the order of importance was osteoporosis, gestational obesity, central obesity, Hcy, LDL-C, literacy level, Lp(a), age of menarche, average sleep time less than 6h per day, decreased estrogen level, reproductive life, and sweet tooth (Figure 1). RF model cross-validation indicated that the OOB error rate was lowest when mtry=3 and ntree=600. The accuracy was 88.62%, the sensitivity was 88.24%, the specificity was 92.16%, and the AUC was 0.948 (95CI:0.925~0.971). The model fit was the best when the 3-layer BP neural network was constructed with a hidden layer of 5. The accuracy, sensitivity, and specificity of the model in the verification set were 85.37%, 87.10%, 80.00%, and the AUC was 0.921 (95CI:0.889~0.956). The occurrence of MACE
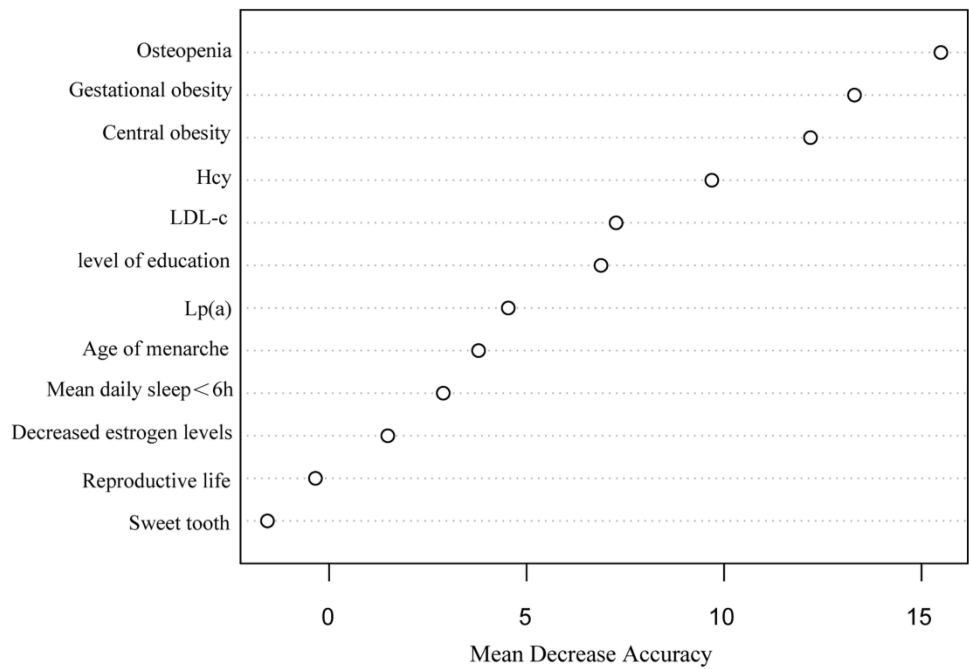
Figure 1 Risk prediction of major cardiovascular adverse events in perimenopausal women by random forest algorithm.
Abbreviations: Hcy, plasma homocysteine; LDL-C, low-density lipoprotein cholesterol; Lp(a), lipoprotein.

in perimenopausal women was taken as the dependent variable, and the influential factors with statistical significance in the multivariate logistic regression analysis were taken as the input variables of the model. The accuracy, sensitivity, and specificity of the LR in the verification set were 87.90%, 77.78%, and 89.62%, with an AUC of 0.866 (95CI:0.826~0.902). In the verification set, the accuracy (88.62%) and specificity (92.16%) of the RF model were higher than those of the BPNN model and the LR model. Moreover, the F1 index (0.727) of the BPNN model was higher than those of the RF model and LR model. When model indexes were inconsistent, the AUC value of the area under the ROC curve was used for comparison. The AUC value of the three prediction models was >0.8, with the RF model (0.948) > BPNN model (0.921) > LR model (0.866). The difference in AUC between the RF model and the LR model was statistically significant (Z=2.27, P=0.0231), while pairwise comparison of other models showed no statistical difference (Table 3 and Figure 2).

Discussion

Risk factors for MACE in perimenopausal women include the level of education, central obesity, gestational obesity, age at menarche, reproductive life span, sweet tooth, average sleep time less than 6 h per day, osteoporosis or bone mass loss,

Table 3 Comparison of Evaluation Performance of RF Model, BPNN Model, and Logistic Regression Model on Test Set

Evaluation Index	Accuracy (%)	Sensitivity (%)	Specificity (%)	PPV (%)	NPV (%)	F1
RF model	88.62	88.24	92.16	55.56	97.92	0.682
BPNN model	85.37	87.10	80.00	66.67	93.10	0.727
Logistic regression model	87.90	77.78	89.62	56.00	95.90	0.652

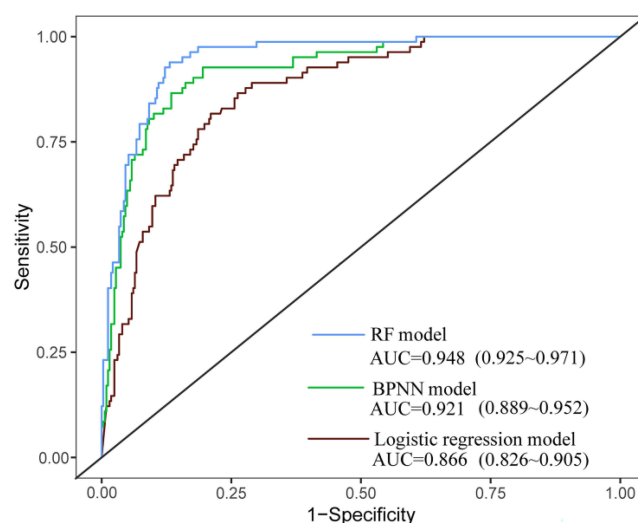**Abbreviations**: PPV, positive predictive value; NPV, negative predictive value.

Figure 2 The ROC curve of the risk assessment model predicted perimenopausal MACEs in the test set.
Abbreviations: RF, random forest; BPNN, backpropagation neural network.

decreased estrogen levels, LDL-C, Lp(a), Hcy, etc. Among them, osteoporosis or bone loss, central obesity, and Hcy were the top three important factors in the occurrence of MACE in perimenopausal women.

The decreased estrogen level, increased bone resorption, and decreased bone formation lead to vascular calcification, which leads to adverse changes such as decreased blood vessel wall elasticity and increased pulse pressure, accelerating the occurrence of cardiovascular diseases. Prior research have confirmed that osteoporosis was an independent risk factor for coronary artery disease (OR=4.42).[16] The results of this study showed that the risk of MACE in perimenopausal women with bone loss odds ratio osteoporosis was 6.15 times that of women with normal bone mass. Previous studies have shown that the prevalence of central obesity in perimenopausal women was 60.5%.[17] Additionally, a series of chain reactions caused by gestational obesity, such as gestational hypertension, insulin resistance, and abnormal lipid levels, are associated with cardiovascular adverse events. Many studies have confirmed a correlation between women who remain obese from 3 to 12 months postpartum and an increased risk of MACE.[18,19] In this study, 40.61% of perimenopausal women with central obesity were at risk for MACE. Our findings also indicate that decreased levels of Hcy, LDL-C, Lp(a), and estrogen were independent risk factors for MACE in perimenopausal women. Furthermore, Lp(a) has atherogenic properties associated with LDL-C components, and the oxidized phospholipids produced by Lp(a) binding to apolipoproteins promote inflammation and endothelial dysfunction, leading to cardiovascular disease and related complications.[20] Studies have shown that Lp(a)>30 mg/dL after menopause is an independent risk factor for MACE.[21] Zhang et al showed that a Hcy level 12% higher than the upper normal limit increased the risk of MACE by 3.4 times.[22] In addition, the level of education is influenced by various factors such as health awareness and behavior, social support, and economic status. However, individuals with higher levels of education typically exhibit better health awareness and have access to more resources, which can help reduce the risk of MACE. A nationally representative cohort study of 6198 postmenopausal women found that the highest MACE risk was observed in women with a menarche at ≥17 years and a reproductive life span ≤28 years.[23] The results of this study showed that each 1-year increase in age at menarche was associated with a 5.44-fold higher risk of MACE versus non-MACE. Similar to prior findings, our results demonstrated that for each year of increased reproductive life, the risk of MACE decreased by 3.7%.[23,24] This may reduce the MACE risk by affecting endogenous estrogen levels, improving vascular endothelial function, regulating inflammation and immune responses, and promoting metabolic health.

Diet and sleep habits are part of individual behavioral characteristics. Quesada et al found that lifestyle habits, including diet, are important risk factors for cardiovascular disease death in menopausal women.[25] Nonetheless, this study found that perimenopausal women with a preference for sweet foods had a higher risk of MACE; however, the specific intake levels require further investigation. Sleep disorders are one of the most common symptoms in

perimenopausal women, peaking in late perimenopause and continuing into late menopause. Studies have shown that 40%-60% of perimenopausal and postmenopausal women have difficulty falling asleep, and 50% of postmenopausal women with sleep disorders have a sleep time of <6h.[26] Our study found that women sleeping less than 6 hours per night had a 3.68-fold higher risk of MACE compared to those sleeping more than 6 hours. Changes such as increased cortisol levels and sympathetic nerve excitation caused by lack of sleep increase the risk of cardiovascular disease.[27] Therefore, sleep health assessment should be included as an important part of cardiovascular health management in perimenopausal women.

In this study, we constructed three prediction models using multiple machine-learning algorithms and evaluate them using sensitivity, specificity, and AUC indicators. The results showed that both the RF model and the LR model had lower F1 scores compared to the BPNN model, which might be attributed to the BPNN model's superior capability in handling complex factor relationships.[28] However, the RF model exhibited higher accuracy and sensitivity than the other two models. The AUC and accuracy metrics for all three models surpassed 0.80. Additionally, the AUC of the RF model in the training set was not significantly different from that of the BPNN model (0.948 vs 0.921), we discovered that the RF model outperformed the other models but that all showed clinical utility. Different models possess distinct advantages. The LR model excels in explaining individual factors and performs well with fewer characteristics, but it struggles to capture complex relationships. BPNN has strong nonlinear modeling and general approximation capabilities, making it suitable for complex pattern recognition tasks. However, its training process can get trapped in local minima, and it requires substantial computing resources and data. The RF model improves model accuracy and robustness by integrating multiple decision trees, effectively handling high-dimensional data and preventing overfitting with minimal data type requirements. It automatically analyzes predictor interactions and nonlinear effects, utilizing data through resampling with replacement to enhance predictive performance, resulting in more accurate, stable, and robust outcomes.

Therefore, after evaluating the performance of all constructed prediction models, the RF model was deemed optimal for forecasting MACE in perimenopausal women. Its clinical application can enhance screening rates, provide early risk alerts, and enable timely prevention, thus reducing the incidence of MACE in perimenopausal women.

However, there are some limitations. First, this study is a single-site study, the sample size may not adequately represent the broader population. Second, data on variables like age at menarche were collected via patient self-reports, potentially introducing bias. Thrid, The risk prediction model has undergone only internal validation and lacks external validation; therefore, its generalization ability remains unproven. Third, the risk prediction model has undergone only internal validation and lacks external validation, potentially limiting its practical application. In the future, we should improve and combine a variety of machine learning algorithms machine learning algorithms to develop more authoritative models, thereby improving research accuracy and generalization.

Conclusion

In summary, we constructed and validated several prediction models for MACE in perimenopausal women using machine learning algorithms, with the RF model demonstrating better performance. The top three risk factors are osteoporosis or bone loss, gestational obesity, and central obesity. Additionally, gestational obesity and central obesity are easily assessable and conducive to intervention, aiding in early screening and guiding patients in health management to reduce the risk of MACE in perimenopausal women.

Data Sharing Statement

The datasets used and/or analyzed during the current study are available from the corresponding author upon reasonable request.

Ethics Approval and Consent to Participate

This study was conducted in accordance with the Declaration of Helsinki and was approved by Binzhou Medical University Ethics Committee (approval number: 2022-375). Participation in this study requires informed consent from the subject and/or their legal guardian, and all methods are conducted in accordance with relevant guidelines and regulations.

Acknowledgments

Funding

Disclosure

The authors report no conflicts of interest in this work.

References

1. Schwarz KG, Vicencio SC, Inestrosa NC, Villaseca P, Del Rio R. Autonomic nervous system dysfunction throughout menopausal transition: a potential mechanism underpinning cardiovascular and cognitive alterations during female ageing. *J Physiol*. 2024;602(2):263–280. doi:10.1113/jp285126

2. Chinese Association of Integrated Traditional and Western Medicine obstetrics and Gynecology professional committee. Guidelines for diagnosis and treatment of climacteric syndrome with integrated Chinese and Western Medicine (2023 edition). *Chin J Pract Gynecol Obstet*. 2023;39(8):799–808. doi:10.19538/j.fk2023080109

3. Khan ZA, Janssen I, Mazzarelli JK. Serial studies in subclinical atherosclerosis during menopausal transition (from the study of women's health across the nation). *Am J Cardiol*. 2018;122(7):1161–1168. doi:10.1016/j.amjcard.2018.06.039

4. Stevenson JC. A woman's journey through the reproductive, transitional and postmenopausal periods of life: impact on cardiovascular and musculo-skeletal risk and the role of estrogen replacement. *Maturitas*. 2011;70(2):197–205. doi:10.1016/j.maturitas.2011.05.017

5. van Dijk GM, Kavousi M, Troup J, Franco OH. Health issues for menopausal women: the top 11 conditions have common solutions. *Maturitas*. 2015;80(1):24–30. doi:10.1016/j.maturitas.2014.09.013

6. Lange-Maia BS, El Khoudary SR, Crandall CJ. Pre- and early peri-menopausal physical function and risk of cardiovascular events: the study of women's health across the nation. *J Aging Health*. 2023;35(5–6):383–391. doi:10.1177/08982643221133580

7. Xi AP, Li X, Liu KX, et al. Correlation analysis of baPWV, ABI, serum Hcy, Lp(a) and cardiovascular events in perimenopausal women. *Chin Experiment Diagn*. 2022;26(6):799–803. doi:10.3969/j.issn.1007-4287.2022.06.003

8. Ahamed J, Manan Koli A, Ahmad K, Alam Jamal M, Gupta BB. CDPS-IoT: cardiovascular disease prediction system based on IoT using machine learning. *Int J Interact Multimed Artif Intell*. 2022;7(4):78–86. doi:10.9781/ijimai.2021.09.002

9. Ahamed J, Mir RN, Chishti MA. Industry 4.0 oriented predictive analytics of cardiovascular diseases using machine learning, hyperparameter tuning and ensemble techniques. *Indust Robot Int J Robotics Res Appl*. 2022;49(3):544–554. doi:10.1108/ir-10-2021-0240

10. Qian JY, Shao CY,Yu ZF. Risk factors and risk model construction of coronary atherosclerotic heart disease in perimenopausal women. *Matern Child Health Care Chin*. 2022;37(11):2100–2103. doi:10.19829/j.zgfybj.issn.1001-4411.2022.11.048

11. Xi AP, Du YH, Li X, et al. Risk factors and prediction model of atherosclerotic cardiovascular disease in perimenopausal women. *J Pract Cardio-Cerebral Pulm Vasc Dis*. 2022;30(8):49–53. doi:10.12114/j.issn.1008-5971.2022.00.202

12. Sun WJ. *Construction and Verification of Aspiration Risk Prediction Model for Inpatients with Nasal Feeding*. Chongqing Medical University; 2021.

13. Li YX, Wang GR, Yang SF, et al. Correlation between serum sex hormone levels and cardiovascular adverse events in perimenopausal patients with coronary heart disease. *Matern Child Health Care Chin*. 2021;36(12):2882–2885. doi:10.19829/j.zgfybj.issn.1001-4411.2021.12.063

14. Liu BH, Wang SY, Li L. Correlation between serum macrophage migration inhibition factor and major adverse cardiovascular events in patients with stable angina pectoris complicated with hypertension. *Chin J Evidence-Based Cardiovasc Med*. 2019;11(4):487–490. doi:10.3969/j.issn.1674-4055.2019.04.28

15. Liu RX, Fan XY. Value of platelet-lymphocyte ratio in predicting adverse cardiovascular events in patients with acute ST-segment elevation myocardial infarction after direct PCI. *Chin J Health Med*. 2020;22(5):506–508. doi:10.3969/j.issn.1674-3245.2020.05.016

16. Lee SN, Cho J-Y, Eun Y-M, Song S-W, Moon K-W. Associations between osteoporosis and coronary artery disease in postmenopausal women. *Climacteric*. 2016;19(5):458–462. doi:10.1080/13697137.2016.1200550

17. Tongdee P, Ananwattanasuk T, Benjaoran F, et al. Lipid accumulation product and index of central lipid distributions for subclinical atherosclerosis in perimenopausal/menopausal women. *J Med Assoc Thai*. 2016;99(Suppl 7):S42–8.

18. Purohit AM, Oyeka CP, Khan SS. Preventing adverse cardiovascular outcomes in pregnancy complicated by obesity. *Curr Obstet Gynecol Rep*. 2023;12(2):129–137. doi:10.1007/s13669-023-00356-9

19. Stevenson JC, Tsiligiannis S, Panay N. Cardiovascular risk in perimenopausal women. *Curr Vasc Pharmacol*. 2019;17(6):591–594. doi:10.2174/1570161116666181002145340

20. Zhao L, Sun L, Zhang Z. Lipoprotein(a) is a new prognostic factor in patients with psoriasis and coronary artery disease: a retrospective cohort study. *Lipids Health Dis*. 2023;22(1):141. doi:10.1186/s12944-023-01901-4

21. Liu S-L, Wu N-Q, Guo Y-L. Lipoprotein(a) and coronary artery disease in Chinese postmenopausal female patients: a large cross-sectional cohort study. *Postgrad Med J*. 2019;95(1128):534–540. doi:10.1136/postgradmedj-2019-136591

22. Zhang Y, Ouyang J, Zhan L, et al. Autophagy in homocysteine-induced HUVEC senescence. *Exp Ther Med*. 2023;26(1):354. doi:10.3892/etm.2023.12053

23. Chen L, Hu Z, Wang X, et al. Age at menarche and menopause, reproductive lifespan, and risk of cardiovascular events among Chinese postmenopausal women: results from a large national representative cohort study. *Front Cardiovasc Med*. 2022;9:870360. doi:10.3389/fcvm.2022.870360

24. Zhu F, Qi H, Bos M, et al. female reproductive factors and risk of new-onset heart failure: findings from UK biobank. *JACC Heart Fail*. 2023;11(9):1203–1212. doi:10.1016/j.jchf.2023.02.019

25. Quesada JA, Bertomeu-González V, Ruiz-Nodar JM, López-Pineda A, Sánchez-Ferrer F. Lifestyle and cardiovascular mortality in menopausal women: a population-based cohort study. *Rev Esp Cardiol*. 2022;75(7):576–584. doi:10.1016/j.rec.2021.10.006

26. Schaedel Z, Holloway D, Bruce D, et al. Management of sleep disorders in the menopausal transition. *Post Reprod Health*. 2021;27(4):209–214. doi:10.1177/20533691211039151

27. Khan MS, Aouad R. The effects of insomnia and sleep loss on cardiovascular disease. *Sleep Med Clin*. 2017;12(2):167–177. doi:10.1016/j.jsmc.2017.01.005

28. Jan Ben S, Dörner M, Günther MP, von Känel R, Euler S. Proof of concept: predicting distress in cancer patients using back propagation neural network (BPNN). *Heliyon*. 2023;9(8):e18328. doi:10.1016/j.heliyon.2023.e18328