Heliyon 10 (2024) e27941

Contents lists available at ScienceDirect

Heliyon

journal homepage: www.cell.com/heliyon

Research article

Risk factors and prediction models for cardiovascular complications of hypertension in older adults with machine learning: A cross-sectional study

Yixin Wu [a], Bo Xin [a], Qiuyuan Wan [a,c], Yanping Ren [b,**], Wenhui Jiang [a,*]

[a] School of Nursing, Health Science Center, Xian Jiaotong University, Xi'an, Shaanxi Province, 710061, China
[b] Department of Geriatrics, Xi'an Jiaotong University Medical College First Affiliated Hospital, Xi'an, Shaanxi Province, 710061, China
[c] Department of Obstetrics and Gynecology, Xi'an Jiaotong University Second Affiliated Hospital, Xi'an, Shaanxi Province, 710004, China

A R T I C L E I N F O

Keywords:
Hypertension
Cardiovascular complications
Risk factors
Back-propagation neural network
Random forest
Older adults

A B S T R A C T

Background: Hypertension has emerged as a chronic disease prevalent worldwide that may cause severe cardiovascular complications, particularly in older patients. However, there is a paucity of studies that use risk factors and prediction models for cardiovascular complications associated with hypertension in older adults.
Objectives: To identify the risk factors and develop prediction models for cardiovascular complications among older patients with hypertension.
Methods: A convenience sample of 476 older patients with hypertension was recruited from a university-affiliated hospital in China. Demographic data, clinical physiological indicators, regulatory emotional self-efficacy, medication adherence, and lifestyle information were collected from participants. Binary logistic regression analysis was performed to screen for preliminary risk factors associated with cardiovascular complications. Two machine learning methods, Back-Propagation neural network, and random forest were applied to develop prediction models for cardiovascular complications among the study cohort. The sensitivity, specificity, accuracy, receiver operating characteristic curve, and area under the curve (AUC) values were used to assess the performance of the prediction models.
Results: Binary logistic regression identified nine risk factors for cardiovascular complications among older patients with hypertension. The machine learning models displayed excellent performance in predicting cardiovascular complications, with the random forest model (AUC 0.954) outperforming the Back-Propagation neural network model (AUC 0.811), as confirmed by model comparison analysis. The sensitivity, specificity and accuracy of the Back-Propagation neural network model compared to the random forest model were 74.2% vs. 86.5%, 75.2% vs. 94.3%, and 74.7% vs. 90.4%, respectively.
Conclusion: The machine learning methods employed in this study demonstrated feasibility in predicting cardiovascular complications among older patients with hypertension, with the random forest model based on nine risk factors exhibiting excellent prediction performance. These models could be used to identify high-risk populations and suggest early interventions aimed at preventing cardiovascular complications in such cohorts.

* Corresponding author
** Corresponding author
E-mail addresses: ryp0071@126.com (Y. Ren), jiangwenhui@xjtu.edu.cn (W. Jiang).

https://doi.org/10.1016/j.heliyon.2024.e27941
Received 18 May 2023; Received in revised form 8 March 2024; Accepted 8 March 2024
Available online 10 March 2024

1. Introduction

Hypertension has remained one of the most severe diseases threatening human health, and it is predicted that 1.56 billion adults will be living with hypertension globally by 2025 [1]. Older adults are particularly vulnerable to hypertension, with prevalence rates of 80% in the United States and 54.6% in China [2,3]. As hypertension progresses, older patients may suffer from various complications, with the heart, brain and kidney being the main target organs. The primary complications include coronary heart disease, heart failure, aortic dissection, stroke and chronic kidney disease, etc [4,5]. In China, hypertension-related complications are responsible for around 2.1 million cardiovascular deaths and 1.2 million premature cardiovascular deaths annually, significantly affecting the quality of life of older adults and creating a major disease burden [6]. Therefore, it is essential to prevent cardiovascular complications in older patients with hypertension.

Preventing cardiovascular complications in older patients with hypertension is a global challenge facing the medical profession. The initial step toward this goal is to identify the risk factors related to cardiovascular complications. Previous studies have demonstrated that cardiovascular complications may arise from multiple factors, including demographic and disease characteristics: age [7], sex [8], age of onset [9] and blood pressure [10]. Additionally, ultrasonography findings and several laboratory test results may also be factors correlated with cardiovascular complications and include uric acid [11], pro-brain (B-type) natriuretic peptide [7], serum lipid [8], homocysteine [12], and interventricular septum thickness [13]. Furthermore, anxiety and depression [14,15], diet [16], smoking, and drinking [17,18] may also contribute to the occurrence of cardiovascular complications. Thus, cardiovascular complications are influenced by numerous risk factors, including physiological, psychological and socio-cultural aspects. However, a comprehensive evaluation of risk factors that predict the occurrence of cardiovascular complications is lacking in the literature.

The development of machine learning algorithms represents an innovative approach to evaluating the risk factors of cardiovascular complications and developing relative prediction models. Some efforts have been made previously in machine learning algorithms to assess the risk, treatment options and prognosis of hypertension, with promising results. For instance, a multilayer neural network was used to identify patients with hypertension in a highly unbalanced National Health and Nutrition Examination Surveys dataset [18], and decision trees and neural networks were used to evaluate the application prospects of different diagnoses and treatment schemes for patients with hypertension [19]. C4.5 decision tree, random forest, eXtreme Gradient Boosting (XgBoost) and Support Vector Machine (SVM) were used to predict hypertension-related complications. It was observed that the prediction performance of some novel machine learning methods for coronary heart disease was significantly better than classical logistic regression [20]. However, most existing studies have ignored sociocultural and psychological factors, measures of cardiovascular compound outcomes, and predictions for older age groups. Therefore, new methods need to be developed to identify ideal prediction models for cardiovascular complications in older patients with hypertension.

The objective of this study was to identify cardiovascular risk factors and use them in two machine learning models to predict cardiovascular complications. We aimed to identify high-risk groups of patients in advance and provide a reference for preventing the occurrence of cardiovascular complications.

1.1. Theoretical framework

The health ecological model is a theoretical model that brings a holistic ecological perspective to the field of human health. This model emphasizes the impact of environmental and individual factors on the health of individuals as multi-leveled and multi-factorial. The health ecological model consists of five levels ranging from the micro-level to the macro-level, including individual traits;

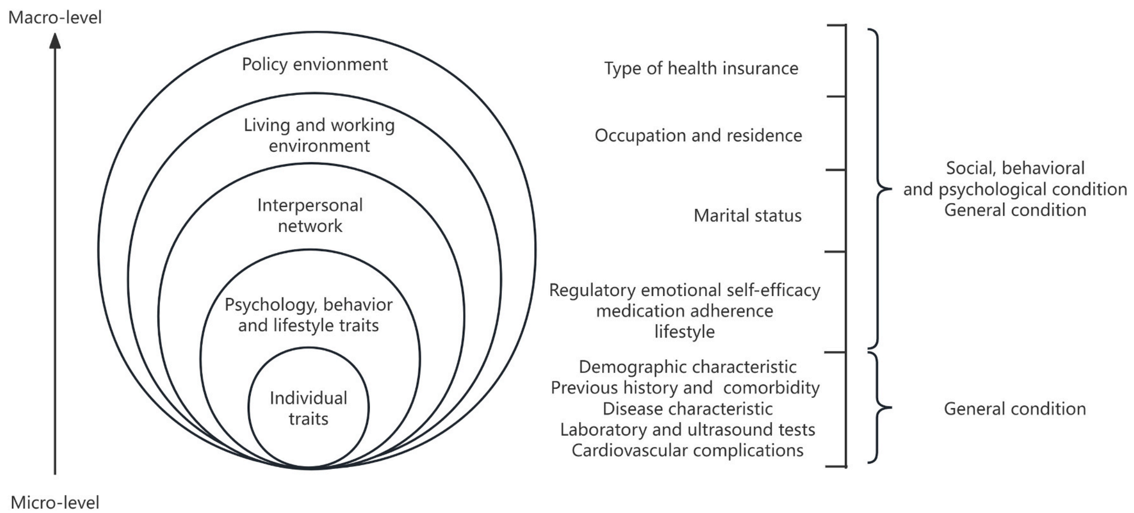

Fig. 1. Health ecological model and its five levels.

psychology, behavior, and lifestyle traits; interpersonal network; living and working environment, and policy environment (Fig. 1).

In our study, the risk factors for cardiovascular complications were collected from five different levels based on the health ecological model: 1)Individual traits corresponding to age, sex and other demographic characteristics, past medical history, cardiovascular complications, disease characteristics, laboratory and ultrasound examination, etc.; 2)Psychology, behavior, and lifestyle traits corresponding to emotional regulation self-efficacy, medication adherence and lifestyle; 3) Interpersonal network corresponding to marital status; 4) Living and working environment corresponding to residence and occupation; 5) Policy environment corresponding to the type of medical insurance.

2. Materials and methods

2.1. Study design and setting

This cross-sectional study was conducted among older patients with hypertension admitted to the First Affiliated Hospital of Xi'an Jiaotong University, a university-affiliated hospital in China, between April and August 2021.

2.2. Participants

Convenience sampling was utilized to recruit a total of 476 older patients with hypertension. The inclusion criteria for this study were: (a) patients older than 60 years or above; (b) patients diagnosed with hypertension; and (c) patients who were willing to participate in this study. Exclusion criteria included: (a) patients who were inability to communicate due to serious physical or mental illness; (b) patients who have undergone other surgical procedures during this hospitalization.

The sample size was estimated based on an incidence rate of 54.7% (5% permissible error; 95% Confidence Interval [CI]) for cardiovascular complications in older patients with hypertension that was retrieved from the Chinese National Knowledge Infrastructure(CNKI) between 2016 and 2021 through meta-analysis [21–24]. Based on the formula for calculating cross-sectional sample size, the estimated sample size was 381. Assuming a nonresponse rate of 10%, we used a final sample size of 420 for this study.

$$\mathrm{N} = \frac{Z_{1-\alpha/2}^{2} P(1-P)}{d^{2}}$$

2.3. Measurement of variables

The data in this study included medical case histories and questionnaires consisting of five parts: a general condition questionnaire, a psychological factors questionnaire, an emotion regulation questionnaire, a medication adherence questionnaire, and a lifestyle questionnaire.

2.3.1. Outcome variables

Primary outcome variable: For the purposes of this study, cardiovascular complications were defined as coronary heart disease and heart failure.

2.3.2. Risk factor variables

The general condition questionnaire included demographic characteristics, past medical history and comorbidities, disease characteristics, and the occurrence of cardiovascular complications. The demographic characteristics included sex, age, residence, occupation, health insurance, marital status, and body mass index(BMI). Past medical history and comorbidities included the course of hypertension, diabetes, and cerebrovascular disease. Cardiovascular complications included the occurrence of coronary heart disease or heart failure.

The psychological factors questionnaire included ultrasonography findings and the results of the following laboratory tests: heart rate, systolic blood pressure, diastolic blood pressure, pulse pressure, fasting blood glucose, red blood cell count, hemoglobin, serum uric acid, serum creatinine, urinary pro-brain (B-type) natriuretic peptide, creatine kinase, creatine kinase-MB, total cholesterol, triglyceride, high-density lipoprotein cholesterol, low-density lipoprotein cholesterol, fibrinogen, active partial thromboplastin time, left ventricular ejection fraction, and carotid atherosclerotic plaque.

Emotion regulation was measured using the Regulatory Emotional Self-Efficacy (RESE) scale translated into Chinese [25], a 12-item scale divided evenly into three dimensions namely, positive emotional efficacy, depression/painful emotional efficacy, and anger/anger emotional efficacy. A higher RESE score implies an elevated ability to regulate emotional self-efficacy. Cronbach's Alpha coefficient was 0.743 when the scale was applied to older patients with hypertension [26].

Medication adherence was measured by the Chinese version of the medication adherence scale (C-MMAS-8) [27,28], an 8-item scale with a total score of 8. Higher scores indicate better medication adherence. A total score of <6, 6–7, and >8 represents low adherence, medium adherence, and high adherence respectively. Cronbach's Alpha coefficient was 0.749 when the scale was applied to patients with hypertension [29].

The lifestyle questionnaire was a self-designed questionnaire to assess bad habits (smoking and drinking), dietary patterns (limitation of greasy and salty food), exercise frequency, sleep status, and defecation situation. The negative habits were "yes" or "no", the others were "never or occasionally", "generally" and "often or always".

2.4. Diagnostic criteria

Hypertension: The diagnosis of hypertension was based on three measurements of blood pressure taken on three separate days, systolic blood pressure (SBP) ≥140 mmHg and/or diastolic blood pressure (DBP) ≥90 mmHg, and/or the use of antihypertensive medications [13].

Coronary heart disease: The diagnosis of coronary heart disease was based on coronary angiography (CAG) findings showing at least one ≥50% coronary artery stenosis, or a history of myocardial infarction. Based on these clinical manifestations, signs, evaluation of risk factors, and laboratory tests, the patients were diagnosed by professional physicians [30].

Heart failure: The diagnosis of heart failure was based on various forms of breathing difficulties, lower limb edema, etc; radiography results that indicated enlargement of the heart shadow; and echocardiography findings that indicated a decrease in diastolic function or thickening of the ventricular wall. The diagnosis of heart failure was made by cardiologists.

Diabetes: The diagnosis of diabetes was based on typical diabetes symptoms such as polydipsia, polyuria, weight loss, and a fasting venous blood glucose level ≥7.0 mmol/L [12].

Cerebrovascular disease: The diagnosis of cerebrovascular disease was based on typical symptoms and signs such as dysarthria and decreased muscle strength; and computed tomography (CT) or magnetic resonance imaging (MRI) results showing the location of an infarction [31].

2.5. Data collection

Demographic characteristics, clinical indicators, and ultrasonography results were collected from the medical records of the hospital or were reported by the participants at baseline. RESE scores, medication adherence, and lifestyle were assessed on the day of admission using questionnaires. Smoking was defined as a current history of continuous smoking, or smoking cessation ≤3 months ago by the patient [32]. Alcohol consumption was defined as continuous drinking for ≥6 months and mean daily drinking of at least 50g [33]. For other lifestyles, "never or occasionally," "generally," and "often or always" represent ≤3 days per week, 4–5 days per week, and ≥6 days per week respectively.

2.6. Statistical analysis

Data analysis was performed using IBM SPSS Statistics version 22.0, and MATLAB R2021a software packages where the analysis comprised two phases: the selection of risk factors, and the development and validation of prediction models.

The risk factors for cardiovascular complications were selected by a descriptive statistical analysis performed using frequency and percentage for enumeration data and means and standard deviation (SD) for measurement data. Univariate and multivariate binary logistic regression analyses were used to screen risk factors. IBM SPSS Statistics version 22.0 was used, and a p-value less than 0.05 was considered statistically significant throughout the process. All quantitative data were converted into categorical data according to the range of medical reference values or scale classification, and the intervals for other variables were determined using standard deviations. Factors that were statistically significant in the binary logistic analysis were selected as input factors for the Back-Propagation neural network and random forest models.

The Back-Propagation neural network and random forest machine learning methods were used to develop two predictive models that were validated via MATLAB R2021a. A Back-Propagation neural network includes input, hidden, and output layers. The data were input and output by the input and output layers, respectively, and the hidden layer was responsible for the network calculation of the input data. When the output value differs significantly from the preset value, the Back-Propagation neural network turns to back propagation and updates the network weight until the error between the final output value and the preset value reaches the established requirements. The random forest method is based on a bagging algorithm that consists of three steps: the creation of a training set, the generation of a decision tree, and classification. First, random forest is a random, non-return sampling method. A decision tree is constructed from unrelated training sets and judged several times from the following nodes. Each non-leaf node is a judgment condition and each leaf node is a conclusion. Changing the value of a variable into a random number reduces the accuracy of the random forest prediction where a larger value indicates greater importance of the variable, and the importance of the variable can be sorted accordingly. Therefore, the variables can be ranked according to the calculated importance (MeanDecreaseGini) which is a more nuanced measure of importance.

2.7. Evaluation criteria

Sensitivity, specificity, and accuracy were used to validate the performance of the prediction models. Four parameters were considered: true positive (TP), true negative (TN), false positive (FP), and false negative (FN). A TP parameter denotes the number of people who have complications and where the prediction is correct. A TN signifies the number of people who are correctly predicted to be without complications. An FP is the number of people wrongly classified as having complications. An FN is the number of people misclassified as having no complications.

Sensitivity, also known as True Positive Rate (TPR) was calculated as the percentage of TP examples that were predicted correctly: $TP/(TP + FN)$;

Specificity, also known as False Positive Rate (TNR) was calculated as the percentage of TN samples that were correctly predicted: $TN/(TN + FP)$;

Table 1
Characteristics of the paticipants (N = 476).

Variables	Without cardiovascular complications (n = 208)	With Cardiovascular complications (n = 268)
Age (SD)	65.85 ± 7.68	69.91 ± 11.66
Sex (n, %)		
Male	105(50.48)	149(55.60)
Female	103(49.52)	119(44.40)
Residence (n, %)		
Urban areas	184(88.46)	203(75.75)
Rural areas	24(11.54)	65(24.25)
Occupation (n, %)		
Retirement	117(56.25)	168(62.69)
Employed	41(19.71)	16(5.97)
Farmer	18(8.65)	57(21.27)
Unemployed or Others	32(15.38)	75(10.07)
Health insurance (n, %)		
Employee basic medical insurance	158(75.96)	184(68.66)
New rural cooperative medical insurance	28(13.46)	60(22.39)
Basic medical insurance system for urban residents	15(7.21)	15(5.60)
Self-paying or other	7(3.37)	9(3.36)
Marital status (n, %)		
Married	156(75.00)	209(77.99)
Divorced	35(16.83)	38(14.18)
Widowed or unmarried	17(8.17)	21(7.83)
Family history (n, %)	71(34.13)	105(39.18)
Diabetes (n, %)	52(25.00)	108(40.30)
Cerebrovascular disease (n, %)	63(30.29)	91(33.96)
Course of hypertension (SD)	9.48 ± 8.95	16.07 ± 11.66
Body Mass Index, kg/m^2 (SD)	24.45 ± 3.06	24.63 ± 3.02
Heart Rate (SD)	78.39 ± 12.11	75.46 ± 12.88
Systolic Blood Pressure, mmHg (SD)	139.95 ± 17.68	135.22 ± 20.54
Diastolic Blood Pressure, mmHg (SD)	84.20 ± 12.55	77.92 ± 10.90
Pulse pressure, mmHg (SD)	55.76 ± 15.48	57.30 ± 17.00
Fasting blood glucose, mmol/L (SD)	6.29 ± 2.52	7.88 ± 3.65
Red blood cell, × 10^{12}/L (SD)	4.49 ± 0.49	4.30 ± 0.57
Hemoglobin, g/L (SD)	139.14 ± 16.16	133.21 ± 18.11
Serum uric acid, umol/L (SD)	332.86 ± 101.92	335.85 ± 92.03
Serum creatinine, mg/dl (SD)	62.32 ± 31.35	69.56 ± 33.92
Pro-brain (B-type) natriuretic peptide, pg/ml (SD)	186.26 ± 584.90	757.24 ± 3144.88
Creatinekinase, U/L (SD)	97.49 ± 57.64	142.62 ± 360.55
Creatine kinase-MB, U/L (SD)	13.78 ± 5.36	20.28 ± 41.51
Total cholesterol, mmol/L (SD)	4.18 ± 0.97	3.70 ± 1.03
Triglyceride, mmol/L (SD)	1.53 ± 1.16	1.35 ± 0.73
High-density lipoprotein, mmol/L (SD)	1.07 ± 0.31	1.03 ± 0.26
Low-density lipoprotein, mmol/L (SD)	2.54 ± 0.90	2.14 ± 0.88
Fibrinogen, g/L (SD)	3.30 ± 3.06	3.42 ± 3.42
Active Partial Thromboplastin time, s (SD)	28.16 ± 6.43	29.72 ± 7.96
Left ventricular ejection fraction, % (SD)	67.78 ± 6.36	64.81 ± 9.21
Left ventricular end-diastolic dimension, mm (SD)	44.92 ± 4.07	46.47 ± 5.37
Carotid atherosclerotic plaque (n, %)	90(43.27)	210(78.36)
Regulatory emotional self-efficacy(SD)	41.03 ± 5.48	37.23 ± 6.22
Medication adherence (SD)	6.08 ± 1.56	5.31 ± 1.66
Smoking (n, %)	44(21.15)	102(38.06)
Drinking (n, %)	23(11.06)	31(11.57)
Exercising (n, %)		
Never or occasionally	41(19.71)	94(35.07)
Generally	32(15.38)	38(14.18)
Often or always	135(64.90)	136(50.75)
Sufficient sleeping (n, %)		
Never or occasionally	50(24.04)	74(27.61)
Generally	39(18.75)	65(24.25)
Often or always	119(57.21)	129(48.13)
Less salty food(n, %)		
Never or occasionally	90(43.27)	92(34.33)
Generally	60(28.85)	79(29.48)
Often or always	58(27.88)	97(36.19)
Less greasy food(n, %)		
Never or occasionally	76(36.54)	62 (23.13)
Generally	55(26.44)	68 (25.37)
Often or always	77(37.02)	138 (51.50)
Defecation smoothly (n, %)		

(*continued on next page*)

Table 1 (*continued*)

Variables	Without cardiovascular complications (n = 208)	With Cardiovascular complications (n = 268)
Never or occasionally	18(8.65)	34(12.69)
Generally	15(7.21)	31(11.57)
Often or always	175(84.13)	203(75.75)

Accuracy was calculated as the percentage of correct number of samples.

TP + TN/(TP + FP + TN + FN);

We calculated the 1-specificity and sensitivity and plotted these as horizontal and vertical coordinates to obtain the receiver operating characteristic (ROC) curve. To evaluate discriminative ability, we use Swets's criteria, where the values range from 0.5 to 1.0, and 0.5–0.6 indicates bad, 0.6–0.7 indicates poor, 0.7–0.8 indicates satisfactory, 0.8–0.9 indicates good, and 0.9–1.0 indicates excellent [34].

Table 2
Univariate analysis and multivariate analysis of risk factors of elderly patients with hypertension (N = 476).

	Univariate analysis			Multivariate analysis		
	OR	95%CI	P-value	OR	95%CI	P-value
Age	1.082	1.053–1.112	< 0.001	0.832	0.510–1.358	0.462
Sex	0.814	0.566–1.171	0.267	–	–	–
Residence	2.455	1.476–4.083	0.001	2.354	0.843–6.576	0.102
Occupation						
Employed vs. Retirement	1.702	0.968–2.991	0.065	2.262	0.996–5.140	0.051
Farmer vs. Retirement	0.463	0.214–1.001	0.050	1.095	0.376–3.195	0.867
Unemployed or Others vs. Retirement	3.753	1.796–7.842	< 0.001	2.622	0.762–9.028	0.126
Types of health insurance						
NRCMI vs. EBMI	0.906	0.330–2.488	0.848	–	–	–
BMIFUR vs. EBMI	1.667	0.563–4.931	0.356	–	–	–
Self-paying or other vs. EBMI	0.778	0.230–0.634	0.686	–	–	–
Marital status						
Divorced vs. Married	1.085	0.554–2.124	0.813	–	–	–
Widowed or unmarried vs. Married	0.879	0.400–1.931	0.748	–	–	–
Family history	1.243	0.852–1.813	0.258	–	–	–
Diabetes	2.025	1.361–3.014	0.001	1.359	0.755–2.445	0.307
Cerebrovascular disease	1.183	0.802–1.746	0.397	–	–	–
course of hypertension	1.068	1.046–1.090	< 0.001	1.647	1.276–2.126	<0.001
Body Mass Index	1.020	0.961–1.083	0.517	–	–	–
Heart rate	0.982	0.967–0.996	0.012	0.469	0.207–1.064	0.070
Systolic Blood Pressure	0.987	0.978–0.997	0.009	0.600	0.341–1.053	0.075
Diastolic Blood Pressure	0.955	0.939–0.971	<0.001	0.408	0.204–0.816	0.011
Pulse pressure	1.006	0.995–1.017	0.308	–	–	–
Fasting blood glucose	1.224	1.128–1.328	<0.001	2.263	1.320–3.880	0.003
Red blood cell	0.518	0.363–0.738	<0.001	0.920	0.406–2.082	0.841
Hemoglobin	0.979	0.968–0.991	<0.001	1.645	0.573–4.719	0.355
Uric Acid	1.000	0.998–1.002	0.736	–	–	–
Serum creatinine	1.008	1.001–1.015	0.023	1.043	0.365–2.980	0.937
pro-brain (B-type) natriuretic peptide	1.001	1.000–1.001	<0.001	2.632	1.495–4.635	0.001
Creatinekinase	1.003	1.000–1.006	0.029	1.311	0.308–5.575	0.714
Creatine kinase-MB	1.023	1.003–1.045	0.026	2.649	0.836–8.388	0.098
Total cholesterol	0.627	0.520–0.757	<0.001	2.077	0.627–6.886	0.232
Triglyceride	0.810	0.665–0.986	0.036	1.453	0.802–2.634	0.218
High-density lipoprotein	0.638	0.334–1.219	0.174	–	–	–
Low-density lipoprotein	0.605	0.489–0.749	<0.001	0.982	0.455–2.118	0.963
Fibrinogen	1.011	0.955–1.072	0.699	–	–	–
Active Partial Thromboplastin time	1.031	1.004–1.058	0.024	1.785	1.024–3.112	0.041
Left ventricular ejection fraction	0.950	0.925–0.976	<0.001	1.233	0.297–5.123	0.774
Left ventricular end-diastolic dimension	1.073	1.030–1.117	0.001	1.703	0.476–6.096	0.413
Carotid atherosclerotic plaque	4.747	3.183–7.080	<0.001	3.957	2.326–6.734	<0.001
Regulatory emotional self-efficacy	0.897	0.868–0.927	<0.001	0.497	0.292–0.845	0.010
Medication adherence	0.742	0.658–0.836	< 0.001	0.504	0.351–0.723	<0.001
Smoking	2.290	1.514–3.465	< 0.001	2.013	1.145–3.538	0.015
Drinking	1.052	0.593–1.865	0.862	–	–	–
Exercising	0.674	0.545–0.835	< 0.001	1.193	0.848–1.678	0.312
Sleep	0.836	0.674–1.038	0.106	–	–	–
Salty	1.279	1.029–1.591	0.027	1.091	0.733–1.626	0.668
Greasy	1.481	1.192–1.840	< 0.001	1.249	0.805–1.936	0.321
Defecation	0.743	0.557–0.991	0.043	0.944	0.644–1.385	0.769

Abbreviation: EBMI: Employee basic medical insurance; NRCMI: New rural cooperative medical insurance; BMIFUR: Basic medical insurance for urban residents.

3. Results

3.1. Characteristics of participants

A total of 476 older patients with hypertension were finally included in the study. Their general condition, physiological indicators, psychological condition, behavior, and lifestyle are presented in Table 1. The average age of the patients was 68.12 ± 7.58 years, including 253 men (53.4%) and 223 women (46.6%). Among these 208 patients (43.7%) had essential hypertension (the group without cardiovascular complications) and 268 patients (56.3%) had cardiovascular complications (the group with cardiovascular complications). Among the complications group, 257 patients (95.90%) were diagnosed with coronary heart disease and 11 patients (4.1%) with heart failure.

3.2. Risk factors selection

All indicators were analyzed using univariate logistic regression and multifactor logistic analyses, as shown in Table 2. Binary logistic analysis extracted the following nine variables: long course of hypertension (Odds Ratio [OR]: 1.647, 95% CI: 1.276–2.126), high diastolic blood pressure (OR: 0.408, 95% CI: 0.204–0.816), high fasting blood glucose (OR: 2.263, 95% CI: 1.320–3.880), high urinary pro-brain (B-type) natriuretic peptide (OR: 2.632, 95% CI: 1.495–4.635), high activated partial thromboplastin time (OR: 1.785, 95% CI: 1.024–3.112), carotid atherosclerotic plaque (OR: 3.957, 95% CI: 2.326–6.734), high RESE score (OR: 0.497, 95% CI: 0.292–0.845), high medication adherence (OR: 0.504, 95% CI: 0.351–0.723), and smoking (OR: 2.013, 95% CI: 1.145–3.538). These risk factors were used as input variables for the machine learning models.

3.3. Evaluation and performance of prediction models

3.3.1. Back-Propagation neural network model

A total of 476 samples were randomly assigned to a training set involving 60% of all samples (286) and a test set comprising 40% of the samples (190). The input layer was nine layers, the hidden layer was ten layers and the output layer was two layers. The area under the curve (AUC) of ROC in the Back-Propagation neural network model was 0.811 (Fig. 2(a)), the predictive ability was satisfactory, the sensitivity was 74.2%, the specificity was 75.2%, and the accuracy was 74.7%.

3.3.2. Random forest model

A total of 476 samples were randomly assigned to a training set involving 60% of the samples (286) and a test set involving 40% of the samples (190) in the random forest model. The AUC value of the ROC curve in the random forest model was 0.954 (Fig. 2(b)), and this is considered excellent according to Swets's criteria. The sensitivity, specificity and accuracy were 86.5%, 94.3% and 90.4%, respectively. Table 3 summarizes the comparison of the evaluation of the two models.

One of the advantages that the random forest model offers is that it can rank the variables using their importance according to MeanDecreaseGini. The importance of each feature to the results was ranked (see Fig. 3) as follows: course of hypertension (29.40), carotid atherosclerotic plaque (24.80), medication adherence (20.26), urinary pro-brain (B-type) natriuretic peptide (19.05), fasting blood glucose (15.82), smoking (13.56), diastolic blood pressure (12.97), RESE score (12.65), and activated partial thromboplastin time (10.30).

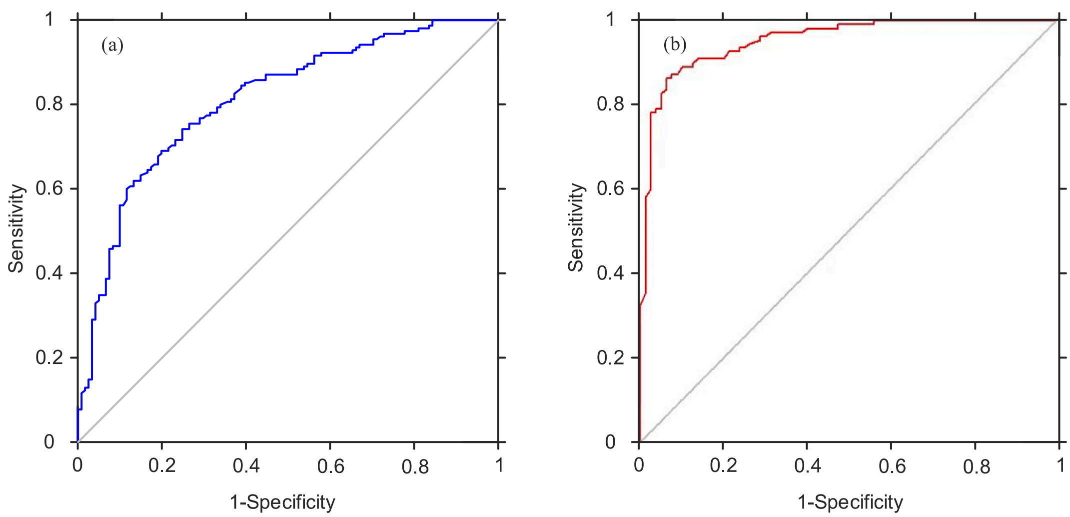

Fig. 2. Comparison of ROC curves between Back-Propagation neural network model and random forest model. (a)Back-Propagation neural network model. (b)Random forest model.

Table 3
Comparison of the evaluation of the two models (N = 476).

Types of models	AUC	Sensitivity/%	Specificity/%	Accuracy/%
Back-Propagation neural network	0.811	74.2	75.2	74.7
Random forest	0.954	86.5	94.3	90.4

4. Discussion

This study has developed and compared two novel machine learning predictive models to identify multiple risk factors for cardiovascular complications in patients with hypertension by applying the ecological model of health. The random forest model (AUC 0.954) has a better predictive performance than the Back-Propagation neural network model (AUC 0.811). It is known that the occurrence of cardiovascular complications is related to individual traits, psychological conditions, behavior and lifestyle, and the living environment. Therefore, using machine learning models to predict cardiovascular complications in older patients with hypertension is feasible, and can help to identify high-risk populations and reduce the occurrence of cardiovascular complications.

Cardiovascular complications are caused by multiple factors. It is reasonable and necessary to select risk factors based on the five levels of the ecological model of health. At the level of individual traits, increased levels of several physiological indicators in older patients with hypertension should be considered. High urinary pro-brain (B type) natriuretic peptide is a standard marker of heart failure and cardiac insufficiency (>125 pg/mL) [35,36]. High fasting blood glucose (>6.1 mmol/L) suggests that nurses should pay attention to patients with impaired fasting glucose. Carotid atherosclerotic plaque with the same mechanism as coronary heart disease should be considered for early intervention [37]. Interestingly, our current data show that low diastolic blood pressure (<90 mmHg) and high activated thromboplastin time (>31.30 s) are risk factors, and this finding differs from previous relevant studies in this field [10,38]. We speculate that this result may be due to the use of medications for blood pressure and anticoagulants. Thus, we reasoned that it was not suitable to evaluate their effects as potential risk factors. On the psychological, behavioral and lifestyle levels, patients with hypertension have a greater response to emotional pressure, negative emotions may also lead to low medication adherence and a poor prognosis for this patient cohort [39,40]. It is known that smoking and poor medication adherence are the major causes of the poor prognosis of hypertension [41,42].

In this study, the predictive accuracy of the random forest model was 90.4%, and this is better than that of the Back-Propagation neural network model. The random forest model is good at dealing with nonlinear problems and can deal with multiple variables. It can not only detect the influence of variables but also rank their importance (degree of influence). The random forest model has low sample requirements, and can still maintain a high prediction accuracy and strong anti-fitting and generalization ability for data with many missing values. By contrast, Back-Propagation neural network is a multi-layer network structure. From a mathematical perspective, a three-layer structure can approximate, in principle, any nonlinear continuous function to arbitrary accuracy, and it is therefore good at solving complex problems [43]. The Back-Propagation neural network and random forest predictive models have been constructed based on a sample dataset, and they are more suitable for high-precision prediction and classification of individual patient prognoses. Our study also demonstrated that the prediction models can show high accuracy even with a small sample size, and this suggests that it is crucial to select the appropriate characteristic variables (risk factors) for subsequent studies.

In comparison with past reports, our study focused on the prediction of coronary heart disease and heart failure as cardiovascular complications of hypertension. Du et al. [20] used random forest, K-nearest neighbor classifier, decision tree, and SVM methods to predict coronary heart disease in patients with hypertension and obtained the same risk factors as our study, namely course of hypertension and blood glucose. Their results showed that the XgBoost algorithm was the best while random forest had a high comprehensive performance. Gong et al. [44] predicted coronary heart disease in patients with hypertension based on random forest and Back-Propagation neural network models and obtained results that were similar to those of Due et al., namely the identification of the same risk factors that have been highlighted in our study: course of hypertension, activated partial thromboplastin time, blood

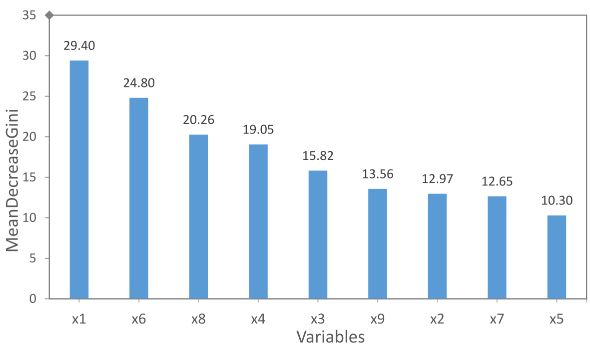

Fig. 3. Rank the importance of variables based on random forest model
Note: x1: Course; x2: Diastolic blood pressure; x3: Fasting blood glucose; x4: pro-brain (B-type) natriuretic peptide; x5: Active Partial Thromboplastin time; x6: Carotid atherosclerotic plaque; x7: Regulatory emotional self-efficacy; x8: Medication adherence; x9: Smoking.

glucose and smoking. Our research highlighted the risk factors that are important for preventing cardiovascular complications that not only reduce the disease burden on patients but also provide an important reference for predicting cardiovascular complications and identifying high-risk populations.

We acknowledge that there were several limitations to our study. The convenience sampling method may have contributed to some selection bias and limited the generalization of our results. Self-reporting measures employed in the current study may have limited the accuracy of the patient responses. It is necessary to validate these prediction models in a large sample set derived from different geographical regions of the world. In addition, future research should include geographical regions, medication levels, lifestyle habits and more risk factors during the design of larger cohort studies.

5. Conclusions

In conclusion, this study developed and validated nine risk factors and two machine learning models for predicting cardiovascular complications in older patients with hypertension. The random forest model displayed a better predictive performance with higher sensitivity and specificity compared to the Back-Propagation neural network model. The prediction models obtained in this study can be used to predict cardiovascular complications in older patients with hypertension in the future. These findings demonstrate that using machine learning methods to predict cardiovascular complications is a feasible approach that can assist health professionals in identifying high-risk groups and the early prevention of cardiovascular complications.

Ethical considerations

This study was reviewed and approved by the Ethics Board of Xi'an Jiaotong University(Approval number: NO. 2018-515) and registered in the Chinese Clinical Trial Registry (Registration number: ChiCTR1800020273). Written informed consent was obtained prior to the study, all the participants were invited to participate in this study voluntarily. The anonymity and confidentiality of the responses were protected in the study.

Funding

The study was supported by the National Social Science Fund of China (Number: 22BGL253), Key Research and Development Projects of Shaanxi Province (Number: 2021SF-190), and Fundamental Research Funds for the Central Universities (Number: SK2022157).

Data availability statement

Data will be made available on request.

CRediT authorship contribution statement

Yixin Wu: Writing – review & editing, Writing – original draft, Software, Methodology, Investigation, Data curation, Conceptualization. **Bo Xin:** Writing – original draft, Validation, Software. **Qiuyuan Wan:** Visualization, Validation. **Yanping Ren:** Resources, Conceptualization. **Wenhui Jiang:** Writing – review & editing, Project administration, Conceptualization.

Declaration of competing interest

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

Acknowledgments

The authors would like to thank the participating patients and health professionals involved in this study.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.heliyon.2024.e27941.

References

[1] H. Xia, L. Rao, H. Li, Comparison of the management and control of hypertension by public and private primary care providers in Shenzhen, China, Heliyon 7 (2) (2021) e06280, https://doi.org/10.1016/j.heliyon.2021.e06280.

[2] E.C. Lefferts, E.A. Bakker, S. Carbone, C.J. Lavie, D.C. Lee, Associations of total and aerobic steps with the prevalence and incidence of frailty in older adults with hypertension, Prog. Cardiovasc. Dis. 67 (2021) 18–25, https://doi.org/10.1016/j.pcad.2021.02.011.
[3] Q. Yu, S. Lin, J. Wu, Hypertension prevalence rates among urban and rural older adults of China, 1991-2015: a standardization and decomposition analysis, Front. Public Health 9 (2021) 713730, https://doi.org/10.3389/fpubh.2021.713730.
[4] P.J. Connelly, G. Currie, C. Delles, Sex differences in the prevalence, outcomes and management of hypertension, Curr. Hypertens. Rep. 24 (6) (2022) 185–192, https://doi.org/10.1007/s11906-022-01183-8.
[5] E.M. Ammann, E.S. O'Brien, D. Milentijevic, A.A. Kharat, D.A. Talbot, W. Canovatchel, L. Haskell, N.S. Andrawis, Characteristics, management, and blood pressure control in patients with apparent resistant hypertension in the US, Heliyon 9 (2) (2023) e13258, https://doi.org/10.1016/j.heliyon.2023.e13258.
[6] Q. Wang, L. Xu, L. Sun, J. Li, W. Qin, G. Ding, J. Zhang, J. Zhu, S. Xie, Z. Yu, C. Zhou, Rural-urban difference in blood pressure measurement frequency among elderly with hypertension: a cross-sectional study in Shandong, China, J. Health Popul. Nutr. 37 (1) (2018) 25, https://doi.org/10.1186/s41043-018-0155-z.
[7] J. Gallagher, C. Watson, S. Zhou, F. Ryan, M. Ledwidge, K. McDonald, B-type natriuretic peptide and ventricular dysfunction in the prediction of cardiovascular events and death in hypertension, Am. J. Hypertens. 31 (2) (2018) 228–234, https://doi.org/10.1093/ajh/hpx153.
[8] M.H. Wang, M. Heizhati, N.F. Li, X.G. Yao, Q. Luo, M.Y. Lin, J. Hong, Y. Ma, R. Wang, L. Sun, Y.L. Ren, N. Yue, Development and validation of a prognostic model to predict high-risk patients for coronary heart disease in snorers with uncontrolled hypertension, Front. Cardiovasc. Med. 9 (2022) 777946, https://doi.org/10.3389/fcvm.2022.777946.
[9] C. Wang, Y. Yuan, M. Zheng, A. Pan, M. Wang, M. Zhao, Y. Li, S. Yao, S. Chen, S. Wu, H. Xue, Association of age of onset of hypertension with cardiovascular diseases and mortality, J. Am. Coll. Cardiol. 75 (23) (2020) 2921–2930, https://doi.org/10.1016/j.jacc.2020.04.038.
[10] A.C. Flint, C. Conell, X. Ren, N.M. Banki, S.L. Chan, V.A. Rao, R.B. Melles, D.L. Bhatt, Effect of systolic and diastolic blood pressure on cardiovascular outcomes, N. Engl. J. Med. 381 (3) (2019) 243–251, https://doi.org/10.1056/NEJMoa1803180.
[11] G. Shen, J.Y. Huang, Y.L. Yu, L. Liu, C.L. Chen, B. Zhang, Y.Q. Huang, Y.Q. Feng, J-shaped association between serum uric acid and acute coronary syndrome in patients with essential hypertension, Postgrad. Med. 96 (1132) (2020) 73–78, https://doi.org/10.1136/postgradmedj-2019-136650.
[12] Z.Y. Zhang, X. Gu, Z. Tang, S.C. Guan, H.J. Liu, X.G. Wu, Y. Zhao, X.H. Fang, Homocysteine, hypertension, and risks of cardiovascular events and all-cause death in the Chinese elderly population: a prospective study, J. Geriatr. Cardiol. 18 (10) (2021) 796–808, https://doi.org/10.11909/j.issn.1671-5411.2021.10.005.
[13] Y. Yang, Z. Li, X. Guo, Y. Zhou, Y. Chang, H. Yang, S. Yu, N. Ouyang, S. Chen, G. Sun, Y. Hua, Y. Sun, Interventricular septum thickness for the prediction of coronary heart disease and myocardial infarction in hypertension population: a prospective study, J. Clin. Med. 11 (23) (2022), https://doi.org/10.3390/jcm11237152.
[14] S. Zhang, Y. Zhong, L. Wang, X. Yin, Y. Li, Y. Liu, Q. Dai, A. Tong, D. Li, L. Zhang, P. Li, G. Zhang, R. Huang, J. Liu, L. Zhao, J. Yu, X. Zhang, L. Yang, J. Cai, W. Zhang, Anxiety, home blood pressure monitoring, and cardiovascular events among older hypertension patients during the COVID-19 pandemic, Hypertens. Res. 45 (5) (2022) 856–865, https://doi.org/10.1038/s41440-022-00852-0.
[15] N. Graham, J. Ward, D. Mackay, J.P. Pell, J. Cavanagh, S. Padmanabhan, D.J. Smith, Impact of major depression on cardiovascular outcomes for individuals with hypertension: prospective survival analysis in UK Biobank, BMJ Open 9 (9) (2019) e024433, https://doi.org/10.1136/bmjopen-2018-024433.
[16] F.J. He, M. Tan, Y. Ma, G.A. MacGregor, Salt reduction to prevent hypertension and cardiovascular disease: JACC state-of-the-art review, J. Am. Coll. Cardiol. 75 (6) (2020) 632–647, https://doi.org/10.1016/j.jacc.2019.11.055.
[17] G. Hao, X. Wang, Z. Chen, L. Zhang, Y. Zhang, B. Wei, C. Zheng, Y. Kang, L. Jiang, Z. Zhu, J. Zhang, Z. Wang, R. Gao, Prevalence of heart failure and left ventricular dysfunction in China: the China Hypertension Survey, 2012-2015, Eur. J. Heart Fail. 21 (11) (2019) 1329–1337, https://doi.org/10.1002/ejhf.1629.
[18] F. López-Martínez, E.R. Núñez-Valdez, R.G. Crespo, V. García-Díaz, An artificial neural network approach for predicting hypertension using NHANES data, Sci. Rep. 10 (1) (2020) 10620, https://doi.org/10.1038/s41598-020-67640-z.
[19] G. Koren, G. Nordon, K. Radinsky, V. Shalev, Machine learning of big data in gaining insight into successful treatment of hypertension, Pharmacol. Res. Perspect 6 (3) (2018) e00396, https://doi.org/10.1002/prp2.396.
[20] Z. Du, Y. Yang, J. Zheng, Q. Li, D. Lin, Y. Li, J. Fan, W. Cheng, X.H. Chen, Y. Cai, Accurate prediction of coronary heart disease for patients with hypertension from electronic health records with big data and machine-learning methods: model development and performance evaluation, JMIR Med Inform 8 (7) (2020) e17257, https://doi.org/10.2196/17257.
[21] X. Liu, Y. Zhang, F. Jin, H. Liu, Q. Li, Y. Gao, R. Hou, Z. Zhang, Correlation between GPR, MHR and elderly essential hypertension with unstable angina pectoris, J. Cent. S. Univ. 46 (4) (2021) 373–378, https://doi.org/10.11817/j.issn.1672-7347.2021.200222.
[22] L. Yan, X. Ding, F. Yan, Y. Wang, Correlation between NT-proBNP level and cardiac remodeling in elderly hypertensive patients with preserved ejection fraction, J. Clin. Res. 37 (10) (2020) 1555–1558, https://doi.org/10.3969/j.issn.1671-7171.2020.10.035.
[23] X. Ma, X. Cai, W. Wang, Predictive value of serum homocysteine level and carotid artery plaque in primary hypertension patients on coronary artery disease, Medical Journal of National Defending Forces in Northwest China 37 (6) (2016) 388–391, https://doi.org/10.16021/j.cnki.1007-8622.2016.06.013.
[24] J. Zhang, The application value of early plasma N T proBN P in the prevention of acute heart failure in elderly patients with essential hypertension, International Journal of Laboratory Medicine 37 (22) (2016) 3180–3182, https://doi.org/10.3969/j.issn.1673-4130.2016.22.032.
[25] G.V. Caprara, L. Di Giunta, N. Eisenberg, M. Gerbino, C. Pastorelli, C. Tramontano, Assessing regulatory emotional self-efficacy in three countries, Psychol. Assess. 20 (3) (2008) 227–237, https://doi.org/10.1037/1040-3590.20.3.227.
[26] J. Zhou, Y.Z. Wang, X.Y. Li, M. Sun, Current situation of regulation emotional self-efficacy and its influencing factors in 310 rural elderly patients with hypertension in Changsha City, Occup. Health 36 (16) (2020) 2217–2220, https://doi.org/10.13329/j.cnki.zyyjk.2020.0597.
[27] D.E. Morisky, A. Ang, M. Krousel-Wood, H.J. Ward, Predictive validity of a medication adherence measure in an outpatient setting, J. Clin. Hypertens. 10 (5) (2008) 348–354, https://doi.org/10.1111/j.1751-7176.2008.07572.x.
[28] J. Yan, L.M. You, Q. Yang, B. Liu, S. Jin, J. Zhou, C. Lin, D.E. Morisky, Translation and validation of a Chinese version of the 8-item Morisky medication adherence scale in myocardial infarction patients, J. Eval. Clin. Pract. 20 (4) (2014) 311–317, https://doi.org/10.1111/jep.12125.
[29] T. Liu, X. Liu, S.L. Li, G.L. Liu, Impact of trans-theoretical model based home visiting on medication compliance behavior in community elderly hypertensive patients, Chin. J. Nurs. 51 (5) (2016) 629–634.
[30] X. Cheng, W. Han, Y. Liang, X. Lin, J. Luo, W. Zhong, D. Chen, Risk prediction of coronary artery stenosis in patients with coronary heart disease based on logistic regression and artificial neural network, 2022, Comput. Math. Methods Med. (2022) 3684700, https://doi.org/10.1155/2022/3684700.
[31] G. Wang, B. Fang, X. Yu, Z. Li, [Interpretation of 2018 guidelines for the early management of patients with acute ischemic stroke], Zhonghua Wei Zhong Bing Ji Jiu Yi Xue 30 (4) (2018) 289–295, https://doi.org/10.3760/cma.j.issn.2095-4352.2018.04.001.
[32] L. Steele, J. Palmer, A. Lloyd, J. Fotheringham, J. Iqbal, E.D. Grech, The impact of smoking on mortality after acute ST-segment elevation myocardial infarction treated with primary percutaneous coronary intervention: a retrospective cohort outcome study at 3 years, J. Thromb. Thrombolysis 47 (4) (2019) 520–526, https://doi.org/10.1007/s11239-019-01812-1.
[33] Y. Zhang, Y. Yu, Y. Yuan, K. Yu, H. Yang, X. Li, X. Min, C. Zhang, M. He, X. Zhang, T. Wu, Association of drinking pattern with risk of coronary heart disease incidence in the middle-aged and older Chinese men: results from the Dongfeng-Tongji cohort, PLoS One 12 (5) (2017) e0178070, https://doi.org/10.1371/journal.pone.0178070.
[34] A. Rubio-Álvarez, M. Molina-Alarcón, Á. Arias-Arias, A. Hernández-Martínez, Development and validation of a predictive model for excessive postpartum blood loss: a retrospective, cohort study, Int. J. Nurs. Stud. 79 (2018) 114–121, https://doi.org/10.1016/j.ijnurstu.2017.11.009.
[35] M. Metra, J.R. Teerlink, Heart failure, Lancet 390 (10106) (2017) 1981–1995, https://doi.org/10.1016/s0140-6736(17)31071-1.
[36] Z. Cao, Y. Jia, B. Zhu, BNP and NT-proBNP as diagnostic biomarkers for cardiac dysfunction in both clinical and forensic medicine, Int. J. Mol. Sci. 20 (8) (2019), https://doi.org/10.3390/ijms20081820.
[37] W. Chen, x. Gu, Progress in research of ultrasonic indicators for carotid atherosclerosis in diagnosis and treatment of coronary disease, Chinese Journal of Multiple Organ Diseases in the Elderly 20 (11) (2021) 862–865, https://doi.org/10.11915/j.issn.1671-5403.2021.11.181.
[38] H. Yin, X. Cheng, Y. Liang, A. Liu, H. Wang, F. Liu, L. Guo, H. Ma, Q. Geng, High perceived stress may shorten activated partial thromboplastin time and lead to worse clinical outcomes in patients with coronary heart disease, Front. Cardiovasc. Med. 8 (2021) 769857, https://doi.org/10.3389/fcvm.2021.769857.

[39] M. Burnier, E. Polychronopoulou, G. Wuerzner, Hypertension and drug adherence in the elderly, Front. Cardiovasc. Med. 7 (2020) 49, https://doi.org/10.3389/fcvm.2020.00049.

[40] A. Wiener, C.S. Rohr, N. Naor, A. Villringer, H. Okon-Singer, Emotion regulation in essential hypertension: roles of anxiety, stress, and the pulvinar, Front. Behav. Neurosci. 14 (2020) 80, https://doi.org/10.3389/fnbeh.2020.00080.

[41] R. Wilkinson, E. Garden, R.C. Nanyonga, A. Squires, F. Nakaggwa, J.I. Schwartz, D.J. Heller, Causes of medication non-adherence and the acceptability of support strategies for people with hypertension in Uganda: a qualitative study, Int. J. Nurs. Stud. 126 (2022) 104143, https://doi.org/10.1016/j.ijnurstu.2021.104143.

[42] T. Kondo, Y. Nakano, S. Adachi, T. Murohara, Effects of tobacco smoking on cardiovascular disease, Circ. J. 83 (10) (2019) 1980–1985, https://doi.org/10.1253/circj.CJ-19-0323.

[43] V. Renganathan, Overview of artificial neural network models in the biomedical domain, Bratisl. Lek. Listy 120 (7) (2019) 536–540, https://doi.org/10.4149/bll_2019_087.

[44] J. Gong, C. Du, X. Zhong, T. Xiang, H. Wang, Researches on the illness risk of essential hypertension complicated with coronary heart disease based on machine learning algorithm, Med. J. Chin. Peoples Lib. Army 45 (7) (2020) 735–741, https://doi.org/10.11855/j.issn.0577-7402.2020.07.10.