RESEARCH

Open Access

A nomogram to distinguish noncardiac chest pain based on cardiopulmonary exercise testing in cardiology clinic



Mingyu Xu^{1,2†}, Rui Li^{4†}, Bingqing Bai^{1,8†}, Yuting Liu⁵, Haofeng Zhou², Yingxue Liao², Fengyao Liu¹, Peihua Cao^{4,6,7}, Qingshan Geng^{1,3,9*} and Huan Ma^{1,10*}

Abstract

Background Psychological disorders, such as anxiety and depression, are considered to be one of the causes of noncardiac chest pain (NCCP). And these patients can be challenging to differentiate from coronary artery disease (CAD), leading to a considerable number of patients still undergoing angiography. We aim to develop a practical prediction model and nomogram using cardiopulmonary exercise testing (CPET), to help identify these patients.

Methods 1,531 eligible patients' electronic medical record data were obtained from Guangdong Provincial People's Hospital. They were randomly divided into a training dataset (N=918) and a testing dataset (N=613) at a ratio of 6:4, and 595 cases without missing data were also selected from testing dataset to form a complete dataset. The training set is used to build the model, and the testing set and the complete set are used for internal validation. Eight machine learning (ML) methods are used to build the model and the best model is finally adopted.

Results The model built by logistic regression performed the best, and among the 29 parameters, six parameters were determined to be valuable parameters for establishing the diagnostic equation and nomogram. The nomogram showed favorable calibration and discrimination with an area under the receiver operating characteristic curve (AUC) of 0.857 in the training set, 0.851 in the testing set, and 0.848 in the complete set. Meanwhile, decision curve analysis demonstrated the clinical utility of the nomogram.

Conclusions A nomogram using CPET to distinguish anxiety/depression from CAD was developed. It may optimize the disease management and improve patient prognosis.

Keywords Cardiopulmonary exercise testing, Anxiety and depression, Coronary artery disease, Diagnostic predictive models, Nomogram

[†]Mingyu Xu, Rui Li and Bingqing Bai contributed equally to this work.

*Correspondence: Qingshan Geng gengqsh@163.net Huan Ma mahuandoctor@163.com

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



© The Author(s) 2024. **Open Access** This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit http://creativecommons.org/licenses/by-nc-nd/4.0/.

Background

Chest pain is a common reason for visits to emergency departments and cardiology clinics [1, 2]. In approximately 50% of cases, patients present with noncardiac chest pain (NCCP), which occurs in the absence of an identifiable cardiac cause [3]. NCCP is often difficult to distinguish from ischemic angina [4–6]. Traditional diagnostic tools such as coronary angiography are invasive and not conducive to early screening. According to the findings derived from the Women's Ischemia Syndrome Evaluation (WISE) study, sponsored by the National Heart, Lung, and Blood Institute, over 40% of participants experienced recurring hospitalizations due to chest pain on multiple occasions. Additionally, within a follow-up period spanning 1 to 5 years, 30% of individuals underwent repeated coronary angiography, despite demonstrating "normal" coronary arteries during a prior hospitalization [7]. Excessive diagnostic testing and hospitalization resulted in not only elevated healthcare expenditures but also suboptimal treatment outcomes [8, 9]. Researchers have recognized that the symptoms in such patients may be caused by psychological factors such as anxiety and depression, possibly due to the heart-brain relationship [10-13]. Scales are frequently used to screen for anxiety and depression, but they are limited by their subjectivity and the possibility of false negatives if patients deny their psychological problems. Semi-structured interviews such as the Structured Clinical Interview for DSM disorders (SCID-I) would be accurate for diagnosis, but they are time-intensive and require involvement of a psychiatrist. Therefore, there is a pressing need in clinical practice for a more efficient, precise, and noninvasive method to distinguish between anxietydepression related NCCP and ischemic angina.

Cardiopulmonary exercise testing (CPET), a noninvasive test assessment of the functional capacity of the cardiovascular and respiratory systems during physical exertion, provides a unique insight into the independent and coupled functions of the cardiovascular, respiratory, skeletal and neurophysiologic systems. Several abnormal CPET indicators, including peak VO₂, Δ VO₂/ Δ WR and O₂ pulse, are acknowledged as indicators linked to myocardial ischemia; thus CPET is emerging as a promising instrument for the early detection and intervention of CAD [14, 15]. Compared to coronary angiography, CPET not only assesses the adequacy of myocardial perfusion but also provides a comprehensive evaluation of cardiac function, pulmonary function, and metabolic status. Furthermore, as a non-invasive procedure that does not require the use of contrast agents, CPET is particularly vital for patients with anxiety or depression who are concerned about the side effects of medical procedures. Meanwhile, previous studies have shown that patients with anxiety or depression may exhibit deviations in certain CPET parameters, such as End-tidal carbon dioxide (ETCO₂), VE/ VCO₂, CPET duration, peak respiratory exchange ratio (RER) and so on [16, 17]. However, current research has yet to explore the potential of CPET in distinguishing between anxiety-depression related NCCP and ischemic angina.

The data collected during CPET, such as heart rate, blood pressure, oxygen consumption, and CO₂ production, are extensive, and the interaction between cardiopulmonary function and psychological states is complex. Machine learning (ML) excels in handling and analyzing big data. They can identify subtle differences between NCCP and ischemic angina through CPET data and are capable of recognizing interactions among variables, which is a challenging task for traditional statistical methods. By employing ML methods, the study intends to develop and validate an optimal diagnostic model using CPET parameters from patients who underwent CPET for chest pain. This model aims to ascertain the predictive accuracy of CPET in distinguishing between anxiety/depression-related NCCP and ischemic angina, thus addressing critical gaps in current diagnostic practices.

Methods

Patients and study design

ethics committee of Guangdong Provin-The cial People's Hospital approved this retrospective study(KY2023-053-02) and waived the informed consent from the patients due to the retrospective nature of the study. All procedures carried out during the study period were performed in keeping with the Declaration of Helsinki of 1964. We conducted a comprehensive review of electronic medical records for 6,550 patients who completed CPET at Guangdong Provincial People's Hospital between January 2012 and April 2022. Patient data were extracted from the electronic medical record system, which included information on diagnoses, medical history, demographic data, and detailed reports such as results from CPET. Focusing on individuals who initially presented with complaints of chest pain, we identified 1,874 cases diagnosed with either coronary artery disease or anxiety/depression. After excluding patients with missing age, gender, and other missing values exceeding 30%, the analysis included a final analysis cohort of 1,531 patients. The baseline characteristics of the patients who were excluded due to excessive missing data can be found in Supplementary Table S2. The comparison of these characteristics with those of the patients ultimately included in the study suggests a similar distribution of data between the two groups, indicating that the reasons for data missing may be random. Then we applied a random split of 60% for the training dataset, which resulted in 918 individuals, and allocated the remaining 40% to the testing dataset, comprising 613 individuals.

The randomization process was carried out using the sample function from R base for random sampling without replacement (replace=FALSE, size=total sample size * 60%, with the random seed set to 42). The selected samples were used as the training set, and the remaining samples formed the testing set. We also selected 595 cases without missing data from the testing dataset to form a complete dataset. The case selection process is outlined in Fig. 1. It's important to note that patients who had well-known contraindications for CPET (for example, severe aortic valve stenosis, severe pulmonary hypertension, etc.) were naturally not included in the study cohort. All participants provided written informed consent before undergoing CPET. This study is reported in accordance with TRIPOD.

Cardiopulmonary exercise testing (CPET)

CPET was performed on an electronically braked bicycle ergometer (ERG 910 plus, SCHILLER, Switzerland) with breath-by-breath gas analysis using a calibrated



Fig. 1 The study flow chart. (CPET, cardiopulmonary exercise testing; LR: Logistic Regression; XGBoost: Extreme Gradient Boosting; RF: Random Forest; Bagtree: Bagged Trees; SVM: Support Vector Machine; LDA: Linear Discriminant Analysis; AUC: area under the receiver operating characteristic curve; DCA: decision curve analysis)

metabolic cart (CARDIOVIT CS-200 Office ErgoSpiro, SCHILLER, Switzerland). A 12-lead electrocardiogram and transcutaneous oxygen saturation was also continuously monitored throughout the test and blood pressure was determined manually every 2 min. The incremental exercise test consisted of 3 min of unloaded pedaling, followed by a gradual elevation in work rate of 10–25 watts per minute until symptom limitation; thereby determining the peak work rate. The incremental work rates was individually determined, adjusted based on participants' age, gender, and physical activity level, with the aim of achieving a test duration of 8–12 min.

Termination criteria for the test include not only the presence of limiting symptoms (such as angina, severe fatigue, or dyspnea) but also the following conditions: (1) achieving the target heart rate (calculated as 85% of 220 minus age); (2) achieving respiratory exchange ratio (RER) \geq 1.10; (3) reaching or exceeding a Rating of Perceived Exertion (RPE) of 17 on the Borg 6–20 scale; (4) heart rate or oxygen consumption fails to rise proportionally with the increasing workload; (5) the occurrence of either a drop in blood pressure or clear electrocardiogram abnormalities as workload progressively increases; (6) the patient requests to stop the test.

The fundamental variables chosen for analyses were based on the routine cardiopulmonary test parameters obtained during exercise, such as exercise time, workload power, rating of perceived exertion (RPE), absolute VO_2 , absolute VCO_2 , ventilation (VE), heart rate (HR), blood pressure (BP), the relationships between them and ratio of the indicators to the predicted value.

Statistical analysis

All analyses were conducted using R 4.3.1. Continuous variables are represented as mean (SD), while categorical variables are presented as numbers and relative frequencies. Between-group differences for continuous variables were assessed using the T-test, whereas categorical variables were compared using the χ^2 test. All patients were randomly divided into training set and testing set according to the proportion of 6:4. Considering the potential complex interactions between CPET variables and the relatively low computational cost of random forest, we used it to impute missing data and applied multiple imputation only on the final model for a sensitivity analysis. Analysis after multiple imputation preserves the integrity of statistical inference, making it suitable for scenarios that require further regression analysis [18]. Missing data were imputed with random forest by the missForest package separately in each of the two sets. During the imputation, we set the number of trees (ntree) to default, which is 500, to ensure the robustness. Additionally, the maximum number of iterations (maxiter) was set to 100 in our case. The training set was used for feature selection and model construction. The testing set was used to validate the models obtained from the training set.

Feature selection

Four methods, including Boruta, information gain, elastic net and genetic algorithm were used to obtain subsets of indicators for model development. We deliberately avoided using the eight machine learning methods selected in this study for feature selection, in order to prevent overfitting and data leakage, thereby minimizing bias in identifying the best-performing machine learning model. The four feature selection methods chosen were based on different principles to avoid bias from a single approach and to ensure robustness from multiple dimensions. Boruta is based on random forest algorithms that stands out for its ability to identify truly significant (genuinely relevant) features and effectively distinguish them from irrelevant ones [19, 20]. Selection based on information gain is a filter-based technique demonstrating strong performance in classification. It skillfully identifies features with decisive impact [21, 22]. Elastic net is a method that integrate the strength of Lasso and Ridge Regression. It overcomes the binary thinking of Lasso which selects only one of the correlated features, and the "all-in" approach of Ridge Regression which tends to remain all features [23]. Genetic algorithm is a wrapper-based method capable of finding the global optimal solution by avoiding local optima [24]. A total of 29 features including sex, age, BMI and 26 cardiopulmonary test parameters were included as candidates (Table 1). The common features selected by all of the four methods were ultimately utilized to build the model. The Boruta, Fselector, glmnet and caret packages in R were used for feature selection.

Model development

Model development was done by *tidymodels* package. Eight ML algorithms are employed to build models in the training set, including: Logistic Regression (LR), Extreme Gradient Boosting (XGBoost), Random Forest (RF), Bagged Trees (Bagtree), Support Vector Machine (SVM), Linear Discriminant Analysis (LDA), Decision Tree and Naive Bayes. None of the methods underwent hyperparameter tuning, except for XGBoost and RF, for which trees=1,000 was set to ensure convergence. Other parameters remained at their default settings because the default settings already yielded models with good performance in AUC on the training dataset. Supplementary Table S1 lists the functions of eight types of machine learning algorithms.

These algorithms represent a wide range of ML algorithm categories, including linear models, tree models, ensemble models, and so on. LR is suitable for binary classification like diagnostic models, and the results of

Characteristic	Total (N=1,531)	Training dataset (N=918)	Testing dataset (N=613)	P-value
Age (years)	57.5±12.5	57.3±12.8 57.7±12.0		0.543
Female Sex	440 (28.7%)	264 (28.8%)	176 (28.7%)	1
BMI	24.1 ± 3.26	24.2±3.30	24.1±3.20	0.804
Hypertension	621 (40.6%)	366 (39.9%)	9%) 255 (41.6%)	
Diabetes	284 (18.6%)	181 (19.7%) 103 (16.8%)		0.171
Diagnosis				0.555
CAD	1,206 (78.8%)	718 (78.2%)	488 (79.6%)	
Anxiety/Depression	325 (21.2%)	200 (21.8%)	125 (20.4%)	
Exercise time (minute)	447±118	445±115	450±122	
Maximum load (W)	100 ± 39.0	100 ± 38.4	101 ± 40.0	0.811
RPE	16.1±1.77	16.2±1.77	16.0 ± 1.75	0.085
RER	1.22 ± 0.143	1.22 ± 0.145	1.22±0.139	0.901
*Peak VO ₂ (mL/min)	1,290±419	$1,290 \pm 407$	$1,300 \pm 436$	0.620
*PeakVO ₂ /kg (mL/kg/min)	19.5 ± 5.53	19.5±5.37	19.6±5.75	0.588
PeakVO ₂ %pred (%)	71.0±18.8	70.8±18.6	71.5±19.3	0.460
HRrest (beat)	76.3 ± 14.0	76.4±14.3	76.1±13.5	0.652
HRpeak (beat)	134 ± 23.5	135±23.7	134±23.3	0.713
HR slope (beat/s)	0.133 ± 0.0465	0.133 ± 0.0432	0.134±0.0512	0.724
Abnormal HRrest	727 (47.5%)	426 (46.4%)	301 (49.1%)	0.325
Abnormal HR reaction	1,409 (92.0%)	855 (93.1%)	554 (90.4%)	0.063
SBPrest (mmHg)	128±19.9	128±20.1	128±19.7	0.934
DBPrest (mmHg)	77.4 ± 11.5	77.1±11.2	77.8±12.0	0.295
SBPpeak (mmHg)	177±31.7	177±31.4	177±32.2	0.996
DBPpeak (mmHg)	87.3±16.0	87.2±15.6	87.6±16.6	0.638
Abnormal DBP reaction	1,159 (75.7%)	703 (76.6%)	456 (74.4%)	0.385
Abnormal SBP reaction	957 (65.5%)	575 (62.6%)	382 (62.3%)	0.942
O ₂ pulse (mL/beat)	10.2 ± 19.4	9.69±2.81	11.0±30.4	0.304
O ₂ pulse%pred (%)	84.3±21.8	84.1±21.6	84.5±22.0	0.691
$\Delta VO_2 / \Delta WR$	9.39 ± 2.26	9.46±2.23	9.29±2.29	0.148
AT VE/VCO ₂	30.5 ± 5.18	30.3±4.99	30.8 ± 5.44	0.063
VE/VCO ₂ slope	29.2 ± 6.37	29.1±6.19	29.4±6.62	0.321
Abnormal HRR	305 (19.9%)	181 (19.7%)	124 (20.2%)	0.857
Decreased exercise tolerance	1,092 (71.3%)	667 (72.7%)	425 (69.3%)	0.176
Cardiac classification				0.973
А	719 (47.0%)	432 (47.1%)	287 (46.8%)	
В	515 (33.6%)	311 (33.9%)	204 (33.3%)	
С	262 (17.1%)	155 (16.9%)	107 (17.5%)	
D	35 (2.3%)	20 (2.2%)	15 (2.4%)	

Table 1 Characteristics of study population

Continuous variables are shown as mean±standard deviation. Discrete variables are presented as n(%)

RPE: rating of perceived exertion; RER: respiratory exchange ratio, the value of VCO₂/VO₂ at peak exercise; Peak VO₂%pred: percent predicted VO₂, calculated as actual peak VO₂/predicted VO₂ estimated by sex, age, and weight; HRrest: resting heart rate; HRpeak: peak heart rate; HR slope: the ratio between change in HR from rest to peak exercise and exercise time; Abnormal HR reaction: HRpeak-HRrest: 89; SBPrest, resting systolic blood pressure; DBPrest, resting diastolic blood pressure; DBPpeak, peak diastolic blood pressure; Abnormal DBP reaction: DBPpeak-DBPrest≤0; Abnormal SBP reaction: the SBP elevation was less than 7 times the Mets; O₂ pulse: peak VO₂ (mL/min) divided by peak HR; O₂pulse@pred: percent predicted O₂ pulse; Δ VO₂/ Δ WR: the linear relationship of oxygen consumption and work rate; AT VE/VCO₂ imute ventilation (VE)/carbon dioxide production (VCO2) in the anaerobic threshold (AT); VE/VCO₂ slope: the linear relationship of VE and VCO₂; Abnormal HRR: heart rate recovery in 1 min ≤12; Decreased exercise tolerance: peak VO₂/kg<10 or AT VO₂>14, B:16≤Peak VO₂/kg ≤20 or 11≤AT VO₂≤14, C:10≤Peak VO₂/kg<16 or 8≤AT VO₂<11, D: Peak VO₂/kg<10 or AT VO₂<8;

*Peak VO₂ was presented in both absolute (ml/min) and relative terms (mL/kg/min);

LR are easy to interpret. SVM performs well on small to medium-sized datasets, especially suitable for linearly separable data. LDA is suitable for dimensional reduction in classification by preserving key characteristics. Decision Tree is easy to interpret with a visualized tree model. Random Forest has good robustness and can effectively handle missing data. XGBoost has higher predictive power than Random Forest, with good processing speed and accuracy for large-scale and medium-high dimensional data, and can prevent overfitting. Bagtree can reduce model variance and mitigate the risk of overfitting. Naive Bayes is easy to implement, and performs well on high-dimensional data.

This study is a retrospective study, with relatively more missing data than prospective studies. The sample size included is moderate, and there is a considerable amount of potential intercorrelations between indicators. Given these, it is difficult to pre-determine which ML algorithm performs best. Therefore, we used all of them at first and then selected the most suitable one to construct the final model.

Model performance comparison

We generated Receiver Operating Characteristic (ROC) curves and calculated area under the curve (AUC) to assess each model's classification performance. Considering the possibility of overfitting, we apply the developed models to the testing set to identify the best-performing model according to AUC. We also employed other metrics to assess model performance, including sensitivity, specificity, accuracy, positive predictive value (PPV), negative predictive value (NPV), the F1 score, and the Brier score. Compared to the F1 score, AUC is insensitive to the choice of classification threshold, thus providing a more stable assessment of model performance. Unlike the Brier score, AUC does not require calibration of model output probabilities, making it more flexible for comparing models. Additionally, AUC is a concise metric for evaluating the classification ability of a model. The DeLong test is employed to compare differences in AUC between different models with the roc.test function in the pROC package.

Evaluation and interpretation of the final model

To visualize the final model, we employed the nomogram function from the *rms* package to construct a nomogram. The SHapley Additive exPlanations (SHAP) method was used to explain the final model and visualize the influence of each feature on model performance.

To exclude any bias introduced by imputation, we also calculated AUC values with the complete set, for sensitivity analysis. Calibration curves were employed to evaluate the nomogram's performance and the Hosmer-Lemeshow tests were conducted to assess the significance of its calibration. The calibration curve can assess the agreement between predicted probabilities and observed outcomes in a model. To interpret a calibration curve, one typically compares the curve against the 45-degree diagonal line (perfect calibration line), where the predicted probabilities perfectly match the actual outcomes.

We use beta regression to plot the calibration curves. It is a calibration method developed for binary classification models [25]. Calibration was performed using the cal_validate_beta function from the probably package with default parameters, and the calibration curves were plotted using the cal_plot_windowed function (window_ size=0.1, step_size=0.01). The hoslem.test function from the *ResourceSelection* package is employed to compute the Hosmer-Lemeshow Test, which uses the calibrated predicted probabilities and the actual classifications, with the remaining parameters set to their default values. By the way, the calibration curves for all eight ML models were also plotted and are presented in the supplementary materials. Then, the clinical benefit of the nomogram was evaluated by decision curve analysis (DCA).

Results

Patient characteristics

A comparison of the training and test dataset characteristics post-imputation is presented in Table 1. The training dataset and testing dataset showed no statistically significant differences in population characteristics. In the total population included for final analysis, the mean age was 57.5 years; 440(28.7%) participants were women and the mean BMI was 24.1 kg/m². Both datasets exhibit similar prevalence rates of hypertension (training set: 39.9%, testing set: 41.6%, P=0.534), diabetes (training set: 19.7%, testing set: 16.8%, P=0.171), and depression/anxiety (training set: 21.8%, testing set: 20.4%, P=0.555).

Table 2 shows the characteristics of patients with CAD and those with anxiety/depression in the general population prior to multiple imputation. The proportion of CAD patients with concomitant hypertension is 44.1%, compared to 27.4% in patients with anxiety/depression, P<0.001; whereas the co-occurrence of diabetes is 22.0% vs. 5.8%, P<0.001. Hypertension and diabetes are confounding factors affecting cardiopulmonary function; however, given that the feature selection method did not identify them as indicators, it is inclined to believe that they will not cause significant changes to the outcome.

Nomogram construction and discrimination

The features identified by the 4 feature selection methods are listed in Supplementary Table S3. Upon completion of the feature selection (see Fig. 2), seven common variables were chosen to establish the diagnostic model. Subsequently, binary classification models were constructed using eight ML methods, with the models developed on the training set and tested on the testing set. Model validation was carried out by assessing discrimination, primarily through the ROC curve and its corresponding AUC. The discrimination performance of eight models on the testing set is depicted in Fig. 3a. All the ML algorithms have performed well in terms of AUC (all >0.8 on the training set), and among them, the models constructed using XGBoost, RF, and Bagtree have achieved an AUC of 1.

Characteristic	CAD	Anxiety/	P-
	(N=1,464)	Depression (N=613)	value
Age (years)	59.8±11.2	49.3±13.4	< 0.001
Female Sex	288 (19.7%)	239 (60.5%)	< 0.001
BMI	24.4 ± 3.2	23.3 ± 3.1	< 0.001
Hypertension	658 (44.9%)	109 (27.6%)	< 0.001
Diabetes	323 (22.1%)	23 (5.8%)	< 0.001
Exercise time (minute)	445 ± 120	463±112	0.006
Maximum load (W)	98.6 ± 38.2	107±41.3	< 0.001
RPE	16.0 ± 2.0	16.4 ± 2.0	< 0.001
RER	1.2 ± 0.1	1.3 ± 0.1	< 0.001
*Peak VO ₂ (mL/min)	$1,280 \pm 417$	1,430±1,760	0.083
*PeakVO ₂ /kg (mL/kg/min)	19.1 ± 5.4	21.8 ± 6.0	< 0.001
PeakVO ₂ %pred (%)	69.8 ± 18.8	77.0 ± 18.3	< 0.001
HRrest (beat)	74.7±13.4	82.1±14.2	< 0.001
HRpeak (beat)	130 ± 22.4	148±22.6	< 0.001
HR slope (beat/s)	0.129 ± 0.04	0.147 ± 0.04	< 0.001
Abnormal HRrest	820 (56.0%)	135 (34.2%)	< 0.001
Abnormal HR reaction	89 (6.1%)	42 (10.6%)	0.003
SBPrest (mmHg)	129 ± 20.4	125 ± 17.7	< 0.001
DBPrest (mmHg)	77.2 ± 11.6	77.9±11.1	0.337
SBPpeak (mmHg)	179 ± 32.3	172 ± 29.4	< 0.001
DBPpeak (mmHg)	87.1±16.3	88.1±15.2	0.287
Abnormal DBP reaction	907 (62.0%)	250 (63.3%)	0.528
Abnormal SBP reaction	721 (49.2%)	228 (57.7%)	0.001
O ₂ pulse (mL/beat)	9.87 ± 2.8	11.4±41.9	0.51
O ₂ pulse%pred (%)	83.6 ± 22.1	86.2 ± 20.9	0.051
$\Delta VO_2 / \Delta WR$	9.32 ± 2.4	9.74 ± 1.8	< 0.001
AT VE/VCO ₂	30.8 ± 5.4	28.3 ± 3.6	< 0.001
VE/VCO ₂ slope	29.5 ± 6.7	27.5 ± 4.5	< 0.001
Abnormal HRR	214 (14.6%)	84 (21.3%)	< 0.001
Decreased exercise tolerance	861 (58.8%)	191 (48.4%)	< 0.001
Cardiac classification			< 0.001
A	554 (37.8%)	191 (48.4%)	
В	413 (28.2%)	101 (25.6%)	
С	212 (14.5%)	35 (8.9%)	
D	26 (1.8%)	0 (0%)	

Continuous variables are shown as mean \pm standard deviation. Discrete variables are presented as $n(\!\%\!)$

CAD: coronary artery disease; RPE: rating of perceived exertion; RER: respiratory exchange ratio, the value of VCO₂/VO₂ at peak exercise; Peak VO₂%pred: percent predicted VO₂, calculated as actual peak VO₂/predicted VO₂ estimated by sex, age, and weight; HRrest: resting heart rate; HRpeak: peak heart rate; HR slope: the ratio between change in HR from rest to peak exercise and exercise time; Abnormal HR reaction: HRpeak-HRrest<89; SBPrest, resting systolic blood pressure; DBPrest, resting diastolic blood pressure; DBPpeak, peak diastolic blood pressure; Abnormal DBP reaction: DBPpeak-DBPrest≤0; Abnormal SBP reaction: the SBP elevation was less than 7 times the Mets; O2 pulse: peak VO2 (mL/min) divided by peak HR; O₂pulse%pred: percent predicted O₂ pulse; ΔVO₂/ ΔWR : the linear relationship of oxygen consumption and work rate; AT VE/VCO_2: minute ventilation (VE)/carbon dioxide production (VCO2) in the anaerobic threshold (AT); VE/VCO₂ slope: the linear relationship of VE and VCO₂; Abnormal HRR: heart rate recovery in 1 min≤12; Decreased exercise tolerance: peak VO2%pred<80; Cardiac classification: according to Weber KT criterion, A:Peak VO₂/kg>20 or AT VO₂>14, B:16 \leq Peak VO₂/kg \leq 20 or 11 \leq AT VO₂ \leq 14, C:10 \leq Peak $VO_2/kg < 16 \text{ or } 8 \le AT VO_2 < 11$, D: Peak $VO_2/kg < 10 \text{ or } AT VO_2 < 8$;

*Peak VO₂ was presented in both absolute (ml/min) and relative terms (mL/kg/ min);

The ROC curves and AUC values for the models in the testing dataset are shown in Fig. 3b.

The LR model is one of the models with an AUC performance greater than or equal to 0.85, having an AUC of 0.850 (95% CI: 0.812-0.889), and another is the LDA model, with an AUC of 0.853 (95% CI: 0.814-0.891). It is worth mentioning that the performance metrics of the LR model and the LDA model are very similar, and in this study, the performance of the two models is comparable (see Supplementary Table S4). The AUCs of the other models were as follows. XGBoost: 0.799(95% CI:0.755-0.844), RF: 0.827(95% CI:0.785-0.87), Bagtree: 0.793(95% CI:0.747-0.839), SVM: 0.802(95% CI:0.752-0.852), Decision Tree: 0.821(95% CI:0.777-0.865), Naive Bayes: 0.802 (95% CI: 0.758-0.847). We chose LR to construct the final model based on its superior interpretability and high accuracy. Table 3 shows the results of the DeLong test for the AUCs of models constructed using LR and other algorithms in testing set. There is no statistically significant difference in the AUCs between LR and LDA, LR and RF, LR and Decision Tree. The DeLong test results for the AUCs of the 8 models in the training set can be found in Supplementary Table S5 and the complete discrimination metrics are listed in Supplementary Table S4. The accuracy of the LR model is 0.803, with a NPV of 0.924, a PPV of 0.511, a specificity of 0.820, a sensitivity of 0.736, and a Brier score of 0.665.

To reduce model complexity, a backward stepwise approach was applied for further feature selection in the final LR model. As shown in Fig. 4, age, female, BMI, peak VO_2/kg , resting heart rate (HRrest), VE/VCO_2 slope were significantly associated with the probability of anxiety or depression (p<0.05) in the final LR model. The contributions of each feature to the LR model were evaluated by the average SHAP values and ranked in descending order (Fig. 5). It shows that female gender is the most important predictor for distinguishing anxiety/depression from CAD, followed by age, peak VO_2 , HRrest, VE/VCO_2 slope, and BMI.

The diagnostic equation follows: is as $logit (p_{depressed anxious}) = 2.249 \cdot 0.070^* Age + 2.398^*$ Sex-0.086*BMI+0.070*Peak VO₂/kg+0.026*HRrest- $0.072*VE/VCO_2$ slope. Then, the nomogram (Fig. 6) to discriminate anxiety/depression from coronary artery disease was built based on these 6 indicators. The AUC of the model was 0.857 (95% CI: 0.826-0.888) in the training set (Fig. 7a) and 0.851 (95% CI: 0.812-0.889) in the testing set (Fig. 7b), indicating a good discrimination. In the sensitivity analysis, we selected cases without missing data from the testing dataset to form a complete dataset(N=595) and also assessed discrimination performance (Supplementary Figure S9) within this set



Fig. 2 Venn plot of feature selection



Fig. 3 The receiver operating characteristics (ROC) curves of the eight machine learning models on (a) training dataset and testing dataset (b). (LR: Logistic Regression; XGBoost: Extreme Gradient Boosting; RF: Random Forest; Bagtree: Bagged Trees; SVM: Support Vector Machine; LDA: Linear Discriminant Analysis)

(AUC=0.848, 95%CI:0.809-0.887). In the sensitivity analysis of imputing missing data using multiple imputation, the parameters and model performance of the LR model were similar to the original ones (see Supplementary Tables S6 and S7).

Calibration and clinical practicality

The overall calibration capability is good (Fig. 8) and Hosmer-Lemeshow tests showed that the P values in the testing set were 0.163. The DCA was applied to evaluate the clinical practicality of the nomogram. As shown in Fig. 9, this nomogram yields net benefit to a wide range of threshold probabilities from 10 to 100% in the testing

 Table 3
 The AUC comparison between LR and other models in testing dataset

Models	ΔAUC	95%CI	Z statistic	P value
LR~XGBoost	0.051	0.019,0.083	3.152	0.002
LR~RF	0.023	-0.001,0.047	1.914	0.056
LR ~ Bagtree	0.058	0.026,0.089	3.581	< 0.001
LR ~ SVM	0.048	0.012,0.085	2.595	0.009
LR~LDA	-0.002	-0.008,0.003	-0.843	0.399
$LR \sim Decision Tree$	0.030	-0.005,0.064	1.703	0.089
LR ~ Naive Bayes	0.048	0.023,0.073	3.748	< 0.001

LR: Logistic Regression; XGBoost: Extreme Gradient Boosting; RF: Random Forest; Bagtree: Bagged Trees; SVM: Support Vector Machine; LDA: Linear Discriminant Analysis; AUC: area under the receiver operating characteristic curve; Δ AUC: the difference of AUC between two models



Fig. 5 Feature importance of final model by SHAP values

Variables	beta	OR	95%Cl(lower)	95%Cl(upper)			P-value
(Intercept)	2.24928	9.480910	0.431782	208.178528		•	1.54e-01
Female	2.397922	11.000293	6.952023	17.405934			1.29e-24
Age	-0.070006	0.932388	0.915335	0.949760			1.06e-13
BMI	-0.085808	0.917770	0.858319	0.981339			1.20e-02
Peak VO2/kg	0.070478	1.073021	1.028117	1.119886	•		1.23e-03
HRrest	0.025535	1.025864	1.011610	1.040319			3.48e-04
VE/VCO2 slope	-0.071996	0.930535	0.889963	0.972956			1.55e-03
				1 -1	0	3 log(OR)	

Fig. 4 The forest plot of logistic regression analysis including the selected variables

set, indicating that within this range, the model is clinically practical. The nomogram works by summing the scores corresponding to all variables to obtain a total score, which is then mapped to the total score scale on the nomogram and subsequently to the probability of the expected outcome. Consider a 35-year-old woman with a BMI of 24 who presented to the clinic for chest pain. Her CPET shows a peak VO₂ of 30 mL/kg/min, a HRrest of 100 beats per minute, and a VE/VCO₂ slope of 25, resulting in a total score of 205. This indicates a 50% chance that her chest pain results from anxiety or depression, not CAD. However, if a 25-year-old with identical clinical measurements experiences chest pain, the likelihood of anxiety or depression as the cause rises to about 60%.

Discussion

To our knowledge, this study is the first to utilize a large CPET dataset to establish and validate a clinical prediction model that effectively differentiates between chest pain caused by anxiety/depression and CAD. Our study population consists of a large sample of patients presenting with chest pain. The significance of our study lies in its potential to help patients with non-cardiac chest pain avoid invasive procedures by utilizing objective, noninvasive tests for differentiation. The model consists of six variables including female gender, age, peak VO₂, HRrest, VE/VCO₂ slope, and BMI, and we have visualized the model as a nomogram. The final model selected was constructed using LR, which demonstrated relatively high accuracy. This can likely be attributed to the model's high NPV, which also explains why specificity outperforms sensitivity. While the PPV was slightly lower, a higher NPV is acceptable in this study, as the goal is to identify patients with anxiety/depression from a population of CAD patients. A higher negative predictive value ensures that CAD patients are not deprived of appropriate treatment. The nomogram includes three cardiopulmonary indicators: lower peak VO₂/kg, higher HRrest, and lower VE/VCO₂ slope, all of which associated with a higher likelihood of being diagnosed with anxiety or



Fig. 6 Nomogram for predict the probability of depression or anxiety

depression. The validation of the nomogram indicates that it possesses good discrimination and calibration capabilities.

Previous research has established that the incidence of CAD is higher in men than in women [26], while symptoms related to mental factors like chest pain are more likely to be reported in women [27]. Therefore, given these two diseases with distinctly different gender prevalence, sex has become the most important factor in distinguishing between them. As age increases, so do age-related oxidative stress, inflammation, and overall deterioration of blood vessels and myocardium. Additionally, the risk of other diseases, including diabetes, obesity, and frailty, becomes higher [28, 29]. Consequently, older patients have a significantly increased risk of developing CAD. In contrast, young patients are less likely to experience chest pain from CAD [30] and more likely to have non-cardiac chest pain [31], making age an important factor in differentiating between non-cardiac chest pain and CAD. Patients with CAD are more likely



Fig. 7 ROC curve of the predictive model constructed by LR for (a) the training dataset and (b) the testing dataset. (LR, logistic regression; AUC, area under the ROC curve)



Fig. 8 The calibration curves for the nomogram

to have cardiovascular risk factors and a higher BMI [32]. Therefore, BMI can also serve as an indicator to distinguish between anxiety/depression and CAD.

Lower peak VO₂ and a higher VE/VCO₂ slope have been proven to be predictors of poor cardiovascular outcomes [33, 34]. Our study findings suggest that patients with CAD, who have a worse prognosis compared to those with anxiety/depression (who do not have CAD),



Fig. 9 Decision curves of the nomogram

are more likely to exhibit a lower peak VO_2 [35, 36] and a higher VE/VCO₂ slope [37, 38]. In patients with CAD, the occurrence of coronary artery stenosis can lead to reduced local myocardial perfusion and an imbalance between local oxygen demand and supply, resulting in myocardial ischemia. As a result, during CPET, patients with CAD may experience an imbalance between myocardial oxygen supply and demand, as well as hemodynamic abnormalities [39]. This can manifest as an inability for VO₂ to increase proportionally with exercise power once it exceeds the myocardial ischemia threshold, resulting in a significant reduction in peak VO_2 [15]. Studies have shown that patients with anxiety or depression alone exhibit higher peak VO₂ levels compared to those with obstructive CAD [40]. Patients with symptoms but without obstructive CAD have peak VO₂ levels that are higher than those with the disease, yet lower than those of completely healthy controls [41].

An elevated VE/VCO₂ slope is often due to reduced ventilatory efficiency and is commonly seen in cardio-vascular diseases. A higher VE/VCO₂ slope value is typically associated with worsening pulmonary hemo-dynamics, increased activation of chemoreceptors and

mechanoreceptors, as well as a decline in autonomic regulation and cardiovascular function. Previous studies have confirmed that patients with coronary artery disease (CAD) have higher VE/VCO2 slopes compared to healthy individuals [42]. Our study also suggests that patients with anxiety/depression tend to have a higher HRrest. This observation is consistent with a study [43] comparing CPET results of 58 individuals with depression to 202 non-depressed individuals. The elevated HRrest may be associated with autonomic nervous system dysfunction [44–46].

Most research on CPET performance in anxiety/ depression involves patients with cardiovascular disease who also have anxiety/depression, with control groups consisting of patients without such comorbidities [40, 47]. Consequently, these studies often conclude that the coexistence of anxiety/depression is associated with indicators of a poorer prognosis. However, our study, which compares patients with CAD alone to those with anxiety/ depression, has found that anxiety and depression, when not combined with CAD, show higher peak VO₂ and lower VE/VCO₂ slope.

Previous attempts to construct diagnostic prediction models for assisting in the identification of anxiety/ depression have primarily been conducted in the field of psychiatry [48], such as in conditions like schizophrenia and Alzheimer's disease. These studies have predominantly utilized neuroimaging data from brain examinations, including head MRI and Electroencephalogram. However, patients with chest pain repeatedly visit the cardiology department, lacking a neuroimaging assessment. Therefore, finding a convenient assessment method suitable for cardiology has certain clinical value for predicting anxiety/depression or CAD. While routine psychiatric assessments are complex, and the gold standard for diagnosing CAD typically necessitates invasive procedures such as coronary angiography, our model offers an alternative. By utilizing variables from CPET, a more common cardiological examination, it provides clinicians with a more convenient, non-invasive, and reliable diagnostic tool.

Our research findings indicate that patients suffering from anxiety/depression can be identified with the help of their CPET results, in conjunction with nomogram developed in this study, as revealed in the preceding sentence. The characteristics of the nomogram, such as being intuitive, individualized, easy to use, and highly accurate, can help clinicians quickly identify patients who need to receive anxiety and depression-related treatment as soon as possible, while avoiding excessive examination burden or wasting medical resources on them. Additionally, the nomogram developed based on CPET is more objective compared to existing diagnostic assessments like the HEART score, and it significantly reduces the potential for inconsistent results due to variations in clinician experience. Regarding the use of the nomogram from this study, the threshold probability at which it is determined that imaging is not necessary can be decided by doctors based on their individual practice preferences. The decision curve indicates that if the threshold probability for the patient or doctor is greater than 10%, using the nomogram from the current study to predict anxiety/ depression provides more benefits than either imaging all patients or not imaging any patients.

An additional advantage of our study is the comparison of 8 different ML algorithms to construct models. By selecting the best-performing model, we optimized the diagnostic capability of the ML model and avoided overfitting. In the training set, XGBoost, RF, and Bagtree exhibited overfitting. XGBoost is better suited for handling nonlinear relationship, while LR and LDA are better at linear data. Better performance of LR and LDA in this study suggesting that correlation between indicators in our study is mostly linear. XGBoost is a powerful tool, but it is not as well-suited for probability calibration as linear models such as LR and LDA, which are inherently better at generating well-calibrated probability outputs. As a result, the Brier score of the XGBoost model is comparatively poorer (although all models in this study have relatively high Brier scores due to the larger proportion of negative samples). Generally speaking, Bagtree is less likely to overfitting compared to Decision Tree. However, in our study, with a moderate amount of numerical variables, the likelihood of overfitting with Decision Tree decreases. Meanwhile, Bagging is more adept at reducing the risk of overfitting in larger datasets. It is worth noting that in this study, the model with the best overall performance was the model constructed by LDA (although its Delong test with LR did not show a statistically significant difference). The good performance of LDA in this dataset may indicate that the sample size of this study is moderate and the data are mostly normally distributed. LDA can find a linear combination of features that maximizes the separation between different classes. Its objective is to maximize the between-class variance while minimizing the within-class variance. LDA assumes that the data from each class follow a Gaussian distribution with the same covariance matrix and uses Fisher's linear discriminant method to distinguish between classes. However, considering that LR does not require the normality assumption, is more widely used in binary classification and can provide an interpretable estimate of the probability of event occurrence, we choose the LR algorithm to build the final model.

In this study, logistic regression outperformed advanced models such as random forest, bagged trees, XGBoost, and decision tree, which was unexpected. Zahra Rahmatinejad [49] and colleagues also found similar results in a study of 2025 patients, where they discovered that advanced models such as Bagging and XGBoost did not outperform logistic regression in predicting in-hospital mortality rates in the emergency department. Our research, along with theirs, indicates that in practical applications, logistic regression still remains strongly competitive, especially in cases where the data volume is small, feature dimensions are low, or the relationships between variables are more linear. Although advanced modern models have significant advantages in dealing with complex and high-dimensional data, they may also lead to overfitting or be overly sensitive to data features due to their high complexity.

This study has several limitations. First, we did not include the two indicators of O₂ pulse trajectory and $\Delta VO_2/\Delta WR$ trajectory in the analysis, because they had more than 30% missing values. They are two of the three most important indicators for helping diagnosing coronary heart disease with CPET (the other one is per centpredicted peak VO₂, which has already been included in the analysis). This may result in the possibility that this nomogram is not the best one for distinguishing anxiety/depression from CPET nomograms. But the current model has shown good discrimination and calibration capabilities, and it should be quite practical in clinical settings. Second, in our sample, CAD patients were more likely to be on beta-blockers (42.4% vs. 11.9%), resulting in artificially lower HRrest. Consequently, this introduces a confounding variable that may affect the accuracy of our findings regarding the association between HRrest and anxiety/depression. Furthermore, this dataset did not collect data on the patients' lifestyle habits (such as exercise habits), therefore, the impact of this factor on the study results could not be analyzed. Third, the study used data from single center, which may not reflect the real-world variability and heterogeneity of the target population. Fourth, the study is a retrospective study, so there may be confounding factors and biases. Additionally, this study did not conduct external validation nor did it compare with existing diagnostic methods. Despite this limitation, our model shows promising performance on internal validation, suggesting its potential utility. Furthermore, we believe that our study still provides valuable insights into the predictive factors distinguishing the two groups. Future research will consider conducting prospective multicenter external validation and comparing it with existing diagnostic methods.

In practical applications, the use of the nomogram developed in this study may be limited when patients are unable to adequately cooperate during the CPET. Anxiety and depression patients may be hindered by psychological factors, resulting in peak VO₂ that are lower than the actual values. This can be assessed using the (RER) to determine whether the patient has made a sufficient

effort during the test. Additionally, commonly used medications in cardiology, such as beta-blockers, can significantly affect heart rate and potentially lead to inaccurate test results. However, in clinical practice, the standard CPET procedure typically recommends waiting 24 h after taking beta-blockers before conducting the test, thus reducing the likelihood of interference from these medications.

Conclusions

In conclusion, we have established and verified a nomogram based on clinical and CPET features to distinguish anxiety/depression from CAD. The nomogram may help these patients receive targeted treatment early and avoid unnecessary coronary angiography.

Abbreviations

NCCP	Noncardiac chest pain
CAD	Coronary artery disease
CPET	Cardiopulmonary exercise testing
ML	Machine learning
AUC	Area under the receiver operating characteristic curve
ETCO ₂	End-tidal carbon dioxide
RER	Respiratory exchange ratio
RPE	Rating of Perceived Exertion
VE	Ventilation
HR	Heart rate
BP	Blood pressure
LR	Logistic Regression
XGBoost	Extreme Gradient Boosting
RF	Random Forest
Bagtree	Bagged Trees
SVM	Support Vector Machine
LDA	Linear Discriminant Analysis
ROC	Receiver Operating Characteristic
DCA	Decision curve analysis

Supplementary Information

The online version contains supplementary material available at https://doi.or g/10.1186/s12911-024-02813-8.

Supplementary Material

Acknowledgements

We wish to thank all the study participants, research staff who participated in this research.

Author contributions

All authors contributed to the design of the study. Mingyu Xu and Yuting Liu wrote the paper and were involved in study execution. Rui Li and Peihua Cao carried out data analysis and results interpretation. Bingqing Bai, Haofeng Zhou, Yingxue Liao, Fengyao Liu participated in the process of data collection. Huan Ma and Qingshan Geng contributed to performed and supervised the research, provided fund support, analysis tools or data, and revised the paper. All authors reviewed the manuscript.

Funding

This research was supported by the grants from High-level Hospital Construction Project of Guangdong Provincial People's Hospital (DFJH201922, DFJH2020029); Guangzhou Municipal Science and Technology Program key projects(2023B03J1249); the Science and Technology Program of Guangzhou, China (Grant No. 202201011245); and China Heart House- Chinese Cardiovascular Association TCM fund(CCA-TCM-032; 202342). The funding organizations had no impact on study design, data collection, analysis, or interpretation, or decision to submit the manuscript.

Data availability

The datasets used and analyzed during the current study available from the corresponding author on reasonable request.

Declarations

Ethics approval and consent to participate

The ethics committee of Guangdong Provincial People's Hospital approved this retrospective study(KY2023-053-02) and waived the informed consent from the patients due to the retrospective nature of the study. All procedures carried out during the study period were performed in keeping with the Declaration of Helsinki of 1964.

Consent for publication

Not applicable.

Competing interests

The authors declare no competing interests.

Author details

¹Department of Cardiology, Guangdong Provincial People's Hospital (Guangdong Academy of Medical Sciences), Southern Medical University, Guangzhou, China

²Department of Internal Medicine, Guangzhou Red Cross Hospital, Jinan University, Guangzhou, China

³Department of Cardiology, Shenzhen People's Hospital, Jinan University, Shenzhen, China

⁴Department of Pulmonary and Critical Care Medicine, Zhujiang Hospital, Southern Medical University, Guangzhou, China

⁵The Second Clinical College of Jinan University, Shenzhen People's Hospital, Shenzhen, China

⁶Clinical Research Center, Zhujiang Hospital, Southern Medical University, Guangzhou, China

⁷Department of Biostatistics, School of Public Health, Southern Medical University, Guangzhou, China

⁸School of Biomedical Sciences and Engineering, South China University of Technology, Guangzhou, Guangdong, China

⁹No. 1017, Dongmen North Road, Luohu District, Shenzhen, Guangdong, China

¹⁰No. 106, Zhongshan Second Road, Yuexiu District, Guangzhou, Guangdong, China

Received: 11 March 2024 / Accepted: 9 December 2024 Published online: 24 December 2024

References

- Chen J, Oshima T, Kondo T, Tomita T, Fukui H, Shinzaki S, et al. Non-cardiac chest Pain in Japan: Prevalence, Impact, and Consultation Behavior - A Population-based study. J Neurogastroenterol Motil. 2023;29(4):446–54.
- Bahall M, Kissoon S, Islam S, Panchoo S, Bhola-Singh N, Maharaj M, et al. Patients with atypical chest Pain: Epidemiology and reported consequences. Cureus. 2024;16(1):e53076.
- Mourad G, Alwin J, Strömberg A, Jaarsma T. Societal costs of non-cardiac chest pain compared with ischemic heart disease–a longitudinal study. BMC Health Serv Res. 2013;13:403.
- Parvand M, Cai L, Ghadiri S, Humphries KH, Starovoytov A, Daniele P, et al. One-year prospective follow-up of women with INOCA and MINOCA at a Canadian women's heart centre. Can J Cardiol. 2022;38(10):1600–10.
- Reyes BJ, Hallak O, Elhabyan AK, Lucas BD, Kasem H. Angina with normal coronary arteries. JAMA. 2005;293(20):2468–9. author reply 2469.
- Watson GS. Noncardiac chest pain: a rational approach to a common complaint. JAAPA: off j Am Acad Physician Assist. 2006;19(1):20–5.
- Johnson BD, Shaw LJ, Buchthal SD, Bairey Merz CN, Kim HW, Scott KN, et al. Prognosis in women with myocardial ischemia in the absence of obstructive coronary disease: results from the national institutes of health-national heart,

lung, and blood institute-sponsored women's ischemia syndrome evaluation (WISE). Circulation. 2004;109(24):2993–9.

- Bugiardini R, Bairey Merz CN. Angina with normal coronary arteries: a changing philosophy. JAMA. 2005;293(4):477–84.
- Rutledge T, Kenkre TS, Bittner V, Krantz DS, Thompson DV, Linke SE, et al. Anxiety associations with cardiac symptoms, angiographic disease severity, and healthcare utilization: the NHLBI-sponsored women's ischemia syndrome evaluation. Int J Cardiol. 2013;168(3):2335–40.
- Gulati M, Levy PD, Mukherjee D, Amsterdam E, Bhatt DL, Birtcher KK, AHA/ ACC/ASE/CHEST et al. /SAEM/SCCT/SCMR guideline for the evaluation and diagnosis of chest pain: a report of the american college of cardiology/ american heart association joint committee on clinical practice guidelines. Circulation. 2021;144(22):e368–454.
- 11. Garroni D, Fragasso G. Heart or mind? Unexplained chest pain in patients with and without coronary disease. Heart Mind. 2018;2(1):5.
- 12. Jiang W. Neuropsychocardiology evolution and advancement of the heartmind field. Heart Mind. 2017;1(2):59.
- 13. Samuels MA. The brain-heart connection. Circulation. 2007;116(1):77-84.
- Guazzi M, Bandera F, Ozemek C, Systrom D, Arena R. Cardiopulmonary exercise testing: what is its value? J Am Coll Cardiol. 2017;70(13):1618–36.
- Guazzi M, Arena R, Halle M, Piepoli MF, Myers J, Lavie CJ. 2016 focused update: clinical recommendations for cardiopulmonary exercise testing data assessment in specific patient populations. Eur Heart J. 2018;39(14):1144–61.
- Mennitto S, Ritz T, Robillard P, France CR, Ditto B. Hyperventilation as a predictor of blood donation-related vasovagal symptoms. Psychosom Med. 2020;82(4):377–83.
- Duyan M, Vural N. Diagnostic value of end-tidal carbon dioxide in the differential diagnosis of unstable angina and non-cardiac chest pain. Am J Emerg Med. 2023;63:69–73.
- Rahmatinejad Z, Hoseini B, Reihani H, Hanna AA, Pourmand A, Tabatabaei SM, et al. Comparison of six Scoring systems for Predicting In-hospital mortality among patients with SARS-COV2 presenting to the Emergency Department. Indian J Crit Care Med. 2023;27(6):416–25.
- Kursa MB, Rudnicki WR. Feature selection with the Boruta Package. J Stat Softw. 2010;36(11):1–13.
- Speiser JL, Miller ME, Tooze J, Ip E. A comparison of Random Forest Variable selection methods for classification prediction modeling. Expert Syst Appl. 2019;134:93–101.
- 21. Lai CM, Yeh WC, Chang CY. Gene selection using information gain and improved simplified swarm optimization. Neurocomputing. 2016;218:331–8.
- 22. Shu W, Yan Z, Yu J, Qian W. Information gain-based semi-supervised feature selection for hybrid data. Appl Intell. 2023;53(6):7310–25.
- 23. Zou H, Hastie T. Regularization and variable selection via the elastic net (vol B 67, pg 301, 2005). J R Stat Soc Ser B-Stat Methodol. 2005;67:768–768.
- 24. Yang J, Honavar V. Feature Subset Selection Using a Genetic Algorithm. In: Liu H, Motoda H, editors. Feature Extraction, Construction and Selection: A Data Mining Perspective [Internet]. Boston, MA: Springer US; 1998 [cited 2024 May 13]. pp. 117–36. https://doi.org/10.1007/978-1-4615-5725-8_8
- Beyond sigmoids. How to obtain well-calibrated probabilities from binary classifiers with beta calibration [Internet]. [cited 2024 Nov 30]. https://project euclid.org/journals/electronic-journal-of-statistics/volume-11/issue-2/Beyon d-sigmoids--How-to-obtain-well-calibrated-probabilities-from/10.1214/17-EJ S1338SI.full
- 26. Khalili D, Sheikholeslami FH, Bakhtiyari M, Azizi F, Momenan AA, Hadaegh F. The incidence of coronary heart disease and the population attributable fraction of its risk factors in Tehran: a 10-year population-based cohort study. PLoS ONE. 2014;9(8):e105804.
- Carmin CN, Ownby RL, Wiegartz PS, Kondos GT. Women and non-cardiac chest pain: gender differences in symptom presentation. Arch Womens Ment Health. 2008;11(4):287–93.
- Jp MLJMMNAZ. C, G D, Cardiovascular disease in the elderly: proceedings of the European Society of Cardiology-Cardiovascular Round Table. European journal of preventive cardiology [Internet]. 2022 May 8 [cited 2024 May 28];29(10). https://pubmed.ncbi.nlm.nih.gov/35167666/
- Rodgers JL, Jones J, Bolleddu SI, Vanthenapalli S, Rodgers LE, Shah K, et al. Cardiovascular risks Associated with gender and aging. J Cardiovasc Dev Dis. 2019;6(2):19.
- Marsan RJ, Shaver KJ, Sease KL, Shofer FS, Sites FD, Hollander JE. Evaluation of a clinical decision rule for young adult patients with chest pain. Acad Emerg Med. 2005;12(1):26–31.

- Roll M, Kollind M, Theorell T. Clinical symptoms in young adults with atypical chest pain attending the emergency department. J Intern Med. 1991;230(3):271–7.
- Choi S, Kim K, Kim SM, Lee G, Jeong SM, Park SY, et al. Association of Obesity or Weight Change with Coronary Heart Disease among Young adults in South Korea. JAMA Intern Med. 2018;178(8):1060–8.
- Guazzi M, Adams V, Conraads V, Halle M, Mezzani A, Vanhees L, et al. EACPR/ AHA Scientific Statement. Clinical recommendations for cardiopulmonary exercise testing data assessment in specific patient populations. Circulation. 2012;126(18):2261–74.
- 34. Paolillo S, Agostoni P. Prognostic Role of Cardiopulmonary Exercise Testing in Clinical Practice. Ann Am Thorac Soc. 2017;14(Supplement1):S53–8.
- Belardinelli R, Lacalaprice F, Carle F, Minnucci A, Cianci G, Perna G, et al. Exercise-induced myocardial ischaemia detected by cardiopulmonary exercise testing. Eur Heart J. 2003;24(14):1304–13.
- Belardinelli R, Lacalaprice F, Tiano L, Muçai A, Perna GP. Cardiopulmonary exercise testing is more accurate than ECG-stress testing in diagnosing myocardial ischemia in subjects with chest pain. Int J Cardiol. 2014;174(2):337–42.
- Dominguez-Rodriguez A, Abreu-Gonzalez P, Gomez MA, Garcia-Baute MDC, Arroyo-Ucar E, Avanzas P, et al. Myocardial perfusion defects detected by cardiopulmonary exercise testing: role of VE/VCO2 slope in patients with chest pain suspected of coronary artery disease. Int J Cardiol. 2012;155(3):470–1.
- Mazaheri R, Shakerian F, Vasheghani-Farahani A, Halabchi F, Mirshahi M, Mansournia MA. The usefulness of cardiopulmonary exercise testing in assessment of patients with suspected coronary artery disease. Postgrad Med J. 2016;92(1088):328–32.
- Jengo JA, Oren V, Conant R, Brizendine M, Nelson T, Uszler JM, et al. Effects of maximal exercise stress on left ventricular function in patients with coronary artery disease using first pass radionuclide angiocardiography: a rapid, noninvasive technique for determining ejection fraction and segmental wall motion. Circulation. 1979;59(1):60–5.
- Zhuang H, Chen J, Xu D. Effect of anxiety and depression on cardiopulmonary exercise test in patients with coronary artery disease. Zhong Nan Da Xue Xue Bao Yi Xue Ban. 2020;45(11):1316–25.
- 41. Chaudhry S, Kumar N, Behbahani H, Bagai A, Singh BK, Menasco N, et al. Abnormal heart-rate response during cardiopulmonary exercise testing

identifies cardiac dysfunction in symptomatic patients with non-obstructive coronary artery disease. Int J Cardiol. 2017;228:114–21.

- Castello-Simões V, Minatel V, Karsten M, Simões RP, Perseguini NM, Milan JC, et al. Circulatory and ventilatory power: characterization in patients with coronary artery disease. Arq Bras Cardiol. 2015;104(6):476–85.
- Hughes JW, Casey E, Luyster F, Doe VH, Waechter D, Rosneck J, et al. Depression symptoms predict heart rate recovery after treadmill stress testing. Am Heart J. 2006;151(5):e11221–6.
- Kemp AH, Quintana DS, Gray MA, Felmingham KL, Brown K, Gatt JM. Impact of depression and antidepressant treatment on heart rate variability: a review and meta-analysis. Biol Psychiatry. 2010;67(11):1067–74.
- Chalmers JA, Quintana DS, Abbott MJA, Kemp AH. Anxiety disorders are associated with reduced heart rate variability: a meta-analysis. Front Psychiatry. 2014;5:80.
- Alvares GA, Quintana DS, Hickie IB, Guastella AJ. Autonomic nervous system dysfunction in psychiatric disorders and the impact of psychotropic medications: a systematic review and meta-analysis. J Psychiatry Neurosci: JPN. 2016;41(2):89–104.
- Milani RV, Lavie CJ, Mehra MR, Ventura HO. Impact of exercise training and depression on survival in heart failure due to coronary heart disease. Am J Cardiol. 2011;107(1):64–8.
- Shatte ABR, Hutchinson DM, Teague SJ. Machine learning in mental health: a scoping review of methods and applications. Psychol Med. 2019;49(9):1426–48.
- Rahmatinejad Z, Dehghani T, Hoseini B, Rahmatinejad F, Lotfata A, Reihani H, et al. A comparative study of explainable ensemble learning and logistic regression for predicting in-hospital mortality in the emergency department. Sci Rep. 2024;14(1):3406.

Publisher's note

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