



Developing and validating machine learning-based prediction models for frailty occurrence in those with chronic obstructive pulmonary disease

Yong Chen¹, Yonglin Yu², Dongmei Yang¹, Wenbo Zhang¹, Vasileios Kouritas³, Xiaoju Chen⁴

¹Department of Respiratory and Critical Care Medicine, The Affiliated Hospital of North Sichuan Medical College, Nanchong, China; ²Department of Stomatology, The Affiliated Hospital of North Sichuan Medical College, Nanchong, China; ³Department of Thoracic Surgery, Norfolk and Norwich University Hospital, Norwich, UK; ⁴Department of Respiratory and Critical Care Medicine, Clinical Medical College & Affiliated Hospital of Chengdu University, Chengdu, China

Contributions: (I) Conception and design: Y Chen, Y Yu; (II) Administrative support: Y Yu, W Zhang; (III) Provision of study materials or patients: Y Chen; (IV) Collection and assembly of data: Y Chen, Y Yu; (V) Data analysis and interpretation: Y Chen, Y Yu; (VI) Manuscript writing: All authors; (VII) Final approval of manuscript: All authors.

Correspondence to: Xiaoju Chen, PhD. Department of Respiratory and Critical Care Medicine, Clinical Medical College & Affiliated Hospital of Chengdu University, No. 82, North Section 2, 2nd Ring Road, North Railway Station, Hehuachi Street, Jinniu District, Chengdu 610081, China. Email: chenyongliu1991@163.com.

Background: Frailty is a medical syndrome caused by multiple factors, characterized by decreased strength, endurance, and diminished physiological function, resulting in increased susceptibility to dependence and/or death. Patients with chronic obstructive pulmonary disease (COPD) tend to be more vulnerable to frailty due to their physical and psychological burdens. Therefore, the aim of this study was to develop a reliable and accurate vulnerability risk prediction model for frailty in patients with COPD in order to improve the identification and prediction of patient frailty. The specific objectives of this study were to determine the prevalence of frailty in patients with COPD and develop a prediction model and evaluate its predictive power.

Methods: Clinical information was analyzed using data from the 2018 China Health and Retirement Longitudinal Study (CHARLS) database, and 34 indicators, including behavioral factors, health status, mental health parameters, and various sociodemographic variables, were examined in the study. The adaptive synthetic sampling technique was used for unbalanced data. Three methods, ridge regressor, extreme gradient boosting (XGBoost) classifier, and random forest (RF) regressor, were used to filter predictors. Seven machine learning (ML) techniques including logistic regression (LR), support vector machines (SVM), multilayer perceptron, light gradient-boosting machine, XGBoost, RF, and K-nearest neighbors were used to analyze and determine the optimal model. For customized risk assessment, an online predictive risk modeling website was created, along with Shapley additive explanation (SHAP) interpretations.

Results: Depression, smoking, gender, social activities, dyslipidemia, asthma, and residence type (urban *vs.* rural) were predictors for the development of frailty in patients with COPD. In the test set, the XGBoost model had an area under the curve of 0.942 (95% confidence interval: 0.925–0.959), an accuracy of 0.915, a sensitivity of 0.873, and a specificity of 0.911, indicating that it was the best model.

Conclusions: The ML predictive model developed in this study is a useful and easy-to-use instrument for assessing the vulnerability risk of patients with COPD and may aid clinical physicians in screening high-risk patients.

Keywords: Chronic obstructive pulmonary disease (COPD); frailty; prediction model; machine learning (ML); Shapley additive explanation (SHAP)

Submitted Mar 12, 2024. Accepted for publication Apr 22, 2024. Published online Apr 29, 2024.

doi: 10.21037/jtd-24-416

View this article at: <https://dx.doi.org/10.21037/jtd-24-416>

Introduction

Chronic obstructive pulmonary disease (COPD) is a common, preventable, and treatable chronic airway disease characterized by persistent airflow limitation and corresponding respiratory symptoms. It is mainly related to the abnormal inflammation of the lungs caused by harmful gases and particles (1). COPD is projected to become the third leading cause of death worldwide by 2030 (2). Additionally, it is the fourth leading cause of death in China, with its prevalence increasing every year (3). Frailty, which arises from the degeneration of several physiological systems (4), is a condition of vulnerability characterized by poor resolution of homeostasis in the body following stress and is defined as an age- and disease-related loss of health. The prevalence of frailty in older adult patients with COPD is reported to be twice as high as that in those without COPD but varies widely in the literature, ranging from 6% to 82%, which is due to the variability in assessment tools and diagnostic criteria used (5). In China, there is limited research on the prevalence of frailty in large-scale epidemiological surveys. A published study reported that the prevalence of frailty in elderly patients with COPD is approximately 10% (6). Frailty limits the ability to perform activities of daily living (ADL), reduces quality of life, and increases the risk of adverse outcomes such as falls, cognitive impairment, hospitalization, and even death. Frailty also raises the consumption of healthcare resources and places a burden on family and social care (7). A study has indicated out that COPD and frailty share common

influencing factors, such as advanced age, smoking, and socioeconomic status, as well as common pathological mechanisms, such as inflammation, an impaired immune system, and dysfunctions of the nervous system and endocrine regulation (8).

Respiratory difficulties in patients with COPD limit their physical activity and reduce the body's muscle strength and mass, which can directly or indirectly cause sarcopenia, considered to be one of the main causes of the development of frailty (9). Furthermore, frailty has been shown to be an independent risk factor for COPD disease progression (10). Early screening in vulnerable populations is important for early intervention to slow frailty onset and progression. Predictive modeling is a tool that can be used to assess the risks of frailty in patients with COPD. Few attempts have been made to develop risk prediction models to screen people at high risk of frailty, and related studies have only focused on investigating the status of frailty and the factors that influence it. Moreover, current models are solely based on healthy populations, and no machine learning (ML) prediction models have been reported concerning the development of frailty in patients with COPD. Therefore, the aim of this study was to screen the risk factors associated with the prediction of frailty in patients with COPD by using a ML approach, and to construct a risk prediction model for clinical use. We present this article in accordance with the TRIPOD reporting checklist (available at <https://jtd.amegroups.com/article/view/10.21037/jtd-24-416/rc>).

Methods

Study design and ethics

Data from this study were drawn from the 2018 China Health and Retirement Longitudinal Study (CHARLS) database (<http://charls.pku.edu.cn>). The CHARLS database provides large-sample, high-quality data and could ensure the reliability and validity of the data analysis in this study. A total of 2,745 patients were included in the analysis after participants with missing data were excluded. The study was conducted in accordance with the Declaration of Helsinki (as revised in 2013).

Data collection

Frailty

Fried *et al.*'s original definition of frailty included unintentional weight loss, weakness, self-reported

Highlight box

Key findings

- Patients with chronic obstructive pulmonary disease (COPD) tend to be more vulnerable to frailty due to their physical and psychological burden.

What is known and what is new?

- Current models are based only on healthy populations, and no machine learning predictive models for the development of frailty in patients with COPD have been reported.
- For customised risk assessment, an online predictive risk modelling website was created, along with Shapley additive explanation interpretations.

What is the implication, and what should change now?

- This computer-based approach is an effective tool for early identification and intervention of frailty in frontline clinicians and patients with COPD.

exhaustion, slow walking speed, and low physical activity (4). The diagnostic criteria for frailty from previous research were considered in this study, with fatigue, debility, inactivity, weight loss, and sluggishness being included in this definition. In our study, frailty was considered to be a binary outcome measure, with individuals categorized as either frail or not frail according to diagnostic assessments of weakness, slowness, exhaustion, low physical activity, and weight loss (4,11-14). The details of the assessments are as follows: (I) weakness was measured by asking participants if they had difficulty lifting or carrying weights over 5 kg. This self-reported item served as an indication of muscle strength and overall physical weakness. (II) For slowness or slow walking was defined as having difficulty walking 100 m or climbing several steps without stopping. This criterion for slowness was similar to what had been used in previous studies, ensuring consistency in measuring physical mobility. (III) Exhaustion or fatigue was measured using two questions from the Chinese version of the Center for Epidemiology Studies Depression Scale (CES-D). Participants were classified as having experienced fatigue if they reported feeling that all their activities in the previous week required considerable effort or if they had found it difficult to initiate activities “always or most of the time” or “occasionally or moderately”. (IV) Low physical activity was defined as failing to exercise or walk for at least 10 minutes continuously in a standard week. Although this variable differs slightly from that of Fried *et al.*'s, it was used in a similar study and proven to be effective in determining frailty. (V) Weight loss was defined as an unintentional loss of 5 or more kilograms within the previous year or having a current body mass index (BMI) of ≤ 18.5 kg/m². It has been found that weight loss serves as a better indicator of frailty compared to BMI and energy intake. By using these specific assessments, we aimed to accurately identify and categorize individuals as frail or non-frail based on five objective criteria: related to weakness, slowness, exhaustion, low physical activity, and weight loss. If three or more of these five criteria were present in an individual, this individual was considered to be frail.

Sociodemographic and behavioral factors

The sociodemographic factors analyzed in this study included age, sex, marital status, level of education, permanent residence, insurance coverage, and financial support. The categories for gender were male and female. Education level categories were upper secondary or vocational training, less than lower secondary, and tertiary.

Marital status included married, living with a spouse, single, separated, divorced, widowed, or never married. Permanent residence was classified as urban or rural were provided. Insurance and funding assistance were categorized as yes or no. Behavioral factors included social activities, smoking habits, number of cigarettes smoked daily, history of alcohol use, poor sleep quality, and nighttime sleep duration. History of alcohol use, smoking, and social interactions were classified as yes or no. Poor sleep quality was evaluated based on the response to “my sleep was restless”. Data on the total amount of sleep at night were obtained via the following question: “In the past month, what was the average amount of sleep that you had each night (in hours)?” Sleep counts were considered actual hours spent sleeping during the night.

Health status

Based on earlier research and our expertise, the following potential indicators for frailty were considered: chronic history of illnesses such as hypertension, dyslipidemia, cancer, heart disease, stroke, mental disorders, arthritis or rheumatism, liver disease, kidney disease, digestive disease, and asthma; waist circumference; grip strength; self-perceived health; ADL score; medications; vision; hearing; pain; and cognitive function. ADL was measured using The Katz Index of Independence in Activities of Daily Living (Katz ADL) (15). For measuring orientation and attention, the Telephone Interview for Cognitive Status (TICS-10) was applied (16).

Psychological factors

Mental health problems were identified as depressive symptoms and life satisfaction. To assess depression, the CES-D scale, a 10-item instrument with a total score of 30, was used. A score of 10 or higher on the CES-D indicates depression. Satisfaction with life was classified as good, fair, or poor (17).

Study procedure

Two groups were identified based on a comprehensive evaluation of the patients: COPD without frailty and COPD with frailty.

Research indicators

This study analyzed data from the 2018 CHARLS database, which included a total of 34 variables: depression, ADL score, pain, hearing, eyesight, medications, self-perceived health status, asthma, arthritis, mental disease, digestive

disease, kidney disease, stroke, heart disease, liver disease, cancer, diabetes, dyslipidemia, hypertension, sleep quality, social activities, smoking, drinking, income, insurance, residence, marital status, educational level, gender, life satisfaction, cognitive, age, smoking, and sleep time.

The measurements are presented as the median and interquartile range (IQR), and comparisons between groups were conducted with the rank sum test. Categorical variables are expressed percentages, and differences between groups were analyzed using the χ^2 test or Fisher exact test. The data set included 2,558 patients with COPD who did not have frailty and 187 people with COPD who had frailty. As the significant disparity between the two groups could have limited predictive capacity of the model, data balancing, which is widely accepted as a means to improving predictive performance, was conducted. Specifically, we used the adaptive synthetic sampling approach for data imbalance analysis (18). The balanced dataset was then randomly divided into a training set (70%) and a test set (30%). Least absolute shrinkage and selection operator (LASSO) regression analysis was used to filter 26 variables based on 10-fold cross-validation. Three methods, random forest (RF) regressor, extreme gradient boosting (XGBoost) classifier, and ridge regressor, were used to analyze variable importance ranking. The top 15 variables were selected by taking the intersection of the Wayne plots. The selected variables, including depression, smoking, gender, social activities, dyslipidemia, asthma, and residence, were used as significant predictors. Additionally, seven ML algorithms, including logistic regression (LR), support vector machines (SVM), multilayer perceptron (MLP), light gradient-boosting machine (LightGBM), XGBoost, RF, and k-nearest neighbors (KNN) were used to predict the occurrence of frailty in patients with COPD. On the training set, we used k-fold cross-validation and a resampling approach (k=10). We used the validation set to optimize the model parameters and the test set to evaluate the system performance. Three measures of model quality—discrimination, calibration and clinical utility—were used to assess the clinical value of the prediction model, and exact recall curve analysis was used to evaluate model discrimination. The degree of calibration and the difference between model predictions and actual events were measured using calibration plots. To determine clinical benefit, we used decision curve analysis (DCA), which computes the net benefit of different probability thresholds. The metrics of the confusion matrix were assessed to calculate the mean precision, accuracy, sensitivity, specificity, and F-value scores of the models.

There are limitations in the interpretation of the results of ML techniques. Lundberg's Shapley additive explanation (SHAP) method is based on game theory and is used to interpret results of any ML model (19). It is a reliable, rapid, and cost-effective approach. It is crucial to note that the SHAP method assesses the relevance of every predictor variable based on its SHAP value. A high SHAP value has a positive impact on the output of the ML model, while a low SHAP value has the opposite effect. Ultimately, a thorough analysis was completed for the assimilation of seven variables. This resulted in the creation of an online forecasting model which can be readily used by healthcare practitioners.

Statistical analysis

Each variable was considered in the comparison between the training sets and testing sets. The Mann-Whitney test was used to compare the median and IQR values of continuous variables. Chi-squared tests were used to compare the numerical and percentage values of categorical variables, which are expressed as frequencies. A two-sided P value less than 0.05 was considered to indicate a statistically significant difference. Statistical analyses were conducted using SPSS 27.0 (IBM Corp), R version 4.2.3 (The R Foundation for Statistical Computing), and Python version 3.10.1 (Python Software Foundation) software.

Results

A total of 2,745 patients with COPD were included in this study. Of these patients, 2,558 did not develop frailty while 187 patients had concomitant frailty (*Table 1*).

Analysis of factors characterizing frailty incidence in patients with COPD

In the study, the presence of frailty in patients with COPD was used as the dependent variable, and the remaining independent variables were analyzed using LASSO regression. To avoid overfitting and solve severe covariance problems, LASSO compresses variable coefficients (20) (*Figure 1A,1B*). Subsequently, 34 independent variables were reduced 26, with a minimum mean square error of 0.011 for lambda. The analysis of variable importance ranking included three methods: ridge regressor, XGBoost classifier, and RF regressor (*Figure 2A-2C*). The first 15 important variables were selected according to the

Table 1 Comprehensive comparison of baseline demographic and clinical data

Variable	All (n=2,745)	Frailty (n=2,558)	Nonfrailty (n=187)	P value
Depression, n (%)				<0.001
No	1,510 (55.009)	1,479 (57.819)	31 (16.578)	
Yes	1,235 (44.991)	1,079 (42.181)	156 (83.422)	
Pain, n (%)				<0.001
No	718 (26.157)	695 (27.170)	23 (12.299)	
Yes	2,027 (73.843)	1,863 (72.830)	164 (87.701)	
Hearing, n (%)				<0.001
Good	661 (24.080)	635 (24.824)	26 (13.904)	
Fair	1,512 (55.082)	1,413 (55.238)	99 (52.941)	
Poor	572 (20.838)	510 (19.937)	62 (33.155)	
Eyesight, n (%)				<0.001
Good	559 (20.364)	535 (20.915)	24 (12.834)	
Poor	1,363 (49.654)	1,297 (50.704)	66 (35.294)	
Good	823 (29.982)	726 (28.382)	97 (51.872)	
Medications, n (%)				<0.001
No	1,285 (46.812)	1,230 (48.084)	55 (29.412)	
Yes	1,460 (53.188)	1,328 (51.916)	132 (70.588)	
Self-perceived health status, n (%)				<0.001
Good	293 (10.674)	283 (11.063)	10 (5.348)	
Fair	1,223 (44.554)	1,190 (46.521)	33 (17.647)	
Poor	1,229 (44.772)	1,085 (42.416)	144 (77.005)	
Asthma, n (%)				<0.001
No	1,945 (70.856)	1,847 (72.205)	98 (52.406)	
Yes	800 (29.144)	711 (27.795)	89 (47.594)	
Arthritis, n (%)				<0.001
No	1,221 (44.481)	1,161 (45.387)	60 (32.086)	
Yes	1,524 (55.519)	1,397 (54.613)	127 (67.914)	
Mental disease, n (%)				<0.001
No	2,582 (94.062)	2,423 (94.722)	159 (85.027)	
Yes	163 (5.938)	135 (5.278)	28 (14.973)	
Digestive disease, n (%)				0.009
No	1,471 (53.588)	1,388 (54.261)	83 (44.385)	
Yes	1,274 (46.412)	1,170 (45.739)	104 (55.615)	
Kidney disease, n (%)				0.26
No	2,243 (81.712)	2,096 (81.939)	147 (78.610)	
Yes	502 (18.288)	462 (18.061)	40 (21.390)	

Table 1 (continued)

Table 1 (continued)

Variable	All (n=2,745)	Frailty (n=2,558)	Nonfrailty (n=187)	P value
Stroke, n (%)				<0.001
No	2,500 (91.075)	2,345 (91.673)	155 (82.888)	
Yes	245 (8.925)	213 (8.327)	32 (17.112)	
Heart disease, n (%)				<0.001
No	1,786 (65.064)	1,691 (66.106)	95 (50.802)	
Yes	959 (34.936)	867 (33.894)	92 (49.198)	
Liver disease, n (%)				0.55
No	2,360 (85.974)	2,202 (86.083)	158 (84.492)	
Yes	385 (14.026)	356 (13.917)	29 (15.508)	
Cancer, n (%)				0.009
No	2,658 (96.831)	2,483 (97.068)	175 (93.583)	
Yes	87 (3.169)	75 (2.932)	12 (6.417)	
Diabetes, n (%)				<0.001
No	2,298 (83.716)	2,161 (84.480)	137 (73.262)	
Yes	447 (16.284)	397 (15.520)	50 (26.738)	
Dyslipidemia, n (%)				<0.001
No	1,935 (70.492)	1,824 (71.306)	111 (59.358)	
Yes	810 (29.508)	734 (28.694)	76 (40.642)	
Hypertension, n (%)				<0.001
No	1,467 (53.443)	1,395 (54.535)	72 (38.503)	
Yes	1,278 (46.557)	1,163 (45.465)	115 (61.497)	
Sleep quality, n (%)				<0.001
Rarely or not at all	991 (36.102)	960 (37.529)	31 (16.578)	
Some time or a little time	469 (17.086)	445 (17.396)	24 (12.834)	
Half the time	473 (17.231)	435 (17.005)	38 (20.321)	
Most or all of the time	812 (29.581)	718 (28.069)	94 (50.267)	
Social activities, n (%)				<0.001
No	1,329 (48.415)	1,198 (46.833)	131 (70.053)	
Yes	1,416 (51.585)	1,360 (53.167)	56 (29.947)	
Smoking, n (%)				0.19
No	1,326 (48.306)	1,227 (47.967)	99 (52.941)	
Yes	1,419 (51.694)	1,331 (52.033)	88 (47.059)	
Drinking, n (%)				<0.001
No	1,873 (68.233)	1,708 (66.771)	165 (88.235)	
Yes	872 (31.767)	850 (33.229)	22 (11.765)	

Table 1 (continued)

Table 1 (continued)

Variable	All (n=2,745)	Frailty (n=2,558)	Nonfrailty (n=187)	P value
Income, n (%)				0.45
No	84 (3.060)	80 (3.127)	4 (2.139)	
Yes	2,661 (96.940)	2,478 (96.873)	183 (97.861)	
Insurance, n (%)				<0.001
No	41 (1.494)	32 (1.251)	9 (4.813)	
Yes	2,704 (98.506)	2,526 (98.749)	178 (95.187)	
Residence, n (%)				0.003
Urban	690 (25.137)	660 (25.801)	30 (16.043)	
Rural	2,055 (74.863)	1,898 (74.199)	157 (83.957)	
Marital status, n (%)				<0.001
Married	477 (17.377)	412 (16.106)	65 (34.759)	
Unmarried	2,268 (82.623)	2,146 (83.894)	122 (65.241)	
Educational level, n (%)				0.002
Below junior high school level	2,483 (90.455)	2,300 (89.914)	183 (97.861)	
High school or vocational training	235 (8.561)	232 (9.070)	3 (1.604)	
Tertiary	27 (0.984)	26 (1.016)	1 (0.535)	
Gender, n (%)				<0.001
Male	1,473 (53.661)	1,407 (55.004)	66 (35.294)	
Female	1,272 (46.339)	1,151 (44.996)	121 (64.706)	
Life satisfaction, n (%)				<0.001
Good	800 (29.144)	748 (29.242)	52 (27.807)	
Fair	1,487 (54.171)	1,415 (55.317)	72 (38.503)	
Poor	458 (16.685)	395 (15.442)	63 (33.690)	
Cognitive, median (IQR)	9.500 (5.000–13.000)	9.500 (5.500–13.000)	4.500 (1.500–9.000)	<0.001
Age, median (IQR)	65.000 (57.000–71.000)	64.000 (56.000–71.000)	70.000 (65.000–76.000)	<0.001
Smoking every day, median (IQR)	0.000 (0.000–15.000)	0.000 (0.000–18.000)	0.000 (0.000–5.000)	0.003
Sleep time, median (IQR)	6.000 (4.000–7.000)	6.000 (4.500–7.000)	5.000 (3.000–7.000)	<0.001
ADL score, median (IQR)	6.000 (5.000–6.000)	6.000 (6.000–6.000)	6.000 (5.000–6.000)	<0.001

IQR, interquartile range; ADL, activities of daily living.

intersection on Wayne plots. Seven common important variables, including depression, smoking, gender, social activities, dyslipidemia, asthma, and address, were obtained as predictors (*Figure 2D*).

Multimodel integrated analysis for classification

Seven ML methods, including LR, SVM, MLP, LightGBM,

XGBoost, RF, and KNN were trained and iterated ten times. Area under the curve (AUC) values were used for the evaluation of the models (21). LightGBM had the highest AUC value in the training set (*Figure 3A*), while XGBoost had the highest value in the validation set (*Figure 3B*). As AUC metrics focus on the predictive accuracy of the model and cannot indicate whether a model is clinically usable or determine which model may be

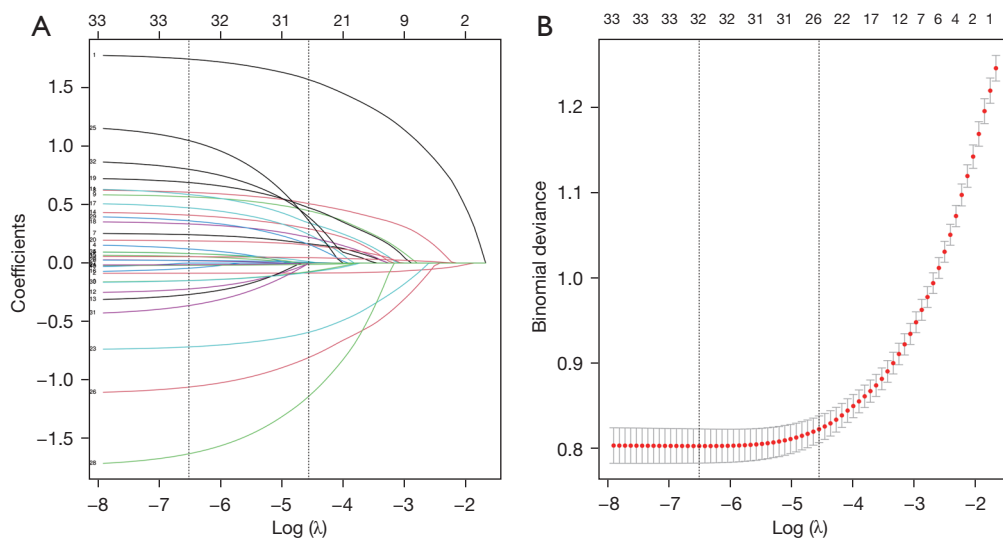


Figure 1 Presentation of the results of the LASSO regression analysis. (A) The LASSO regression model was used to select demographic and clinical characteristics. The optimum lambda resulted in coefficients other than zero. Ten-fold cross-validation using minimum criteria was used to determine the optimal parameter (lambda) in the LASSO model. (B) A virtual vertical line was drawn at the optimal value using the 1-SE criterion by plotting the partial likelihood deviation (binomial deviation) curve against log (lambda). LASSO, least absolute shrinkage and selection operator; SE, standard error.

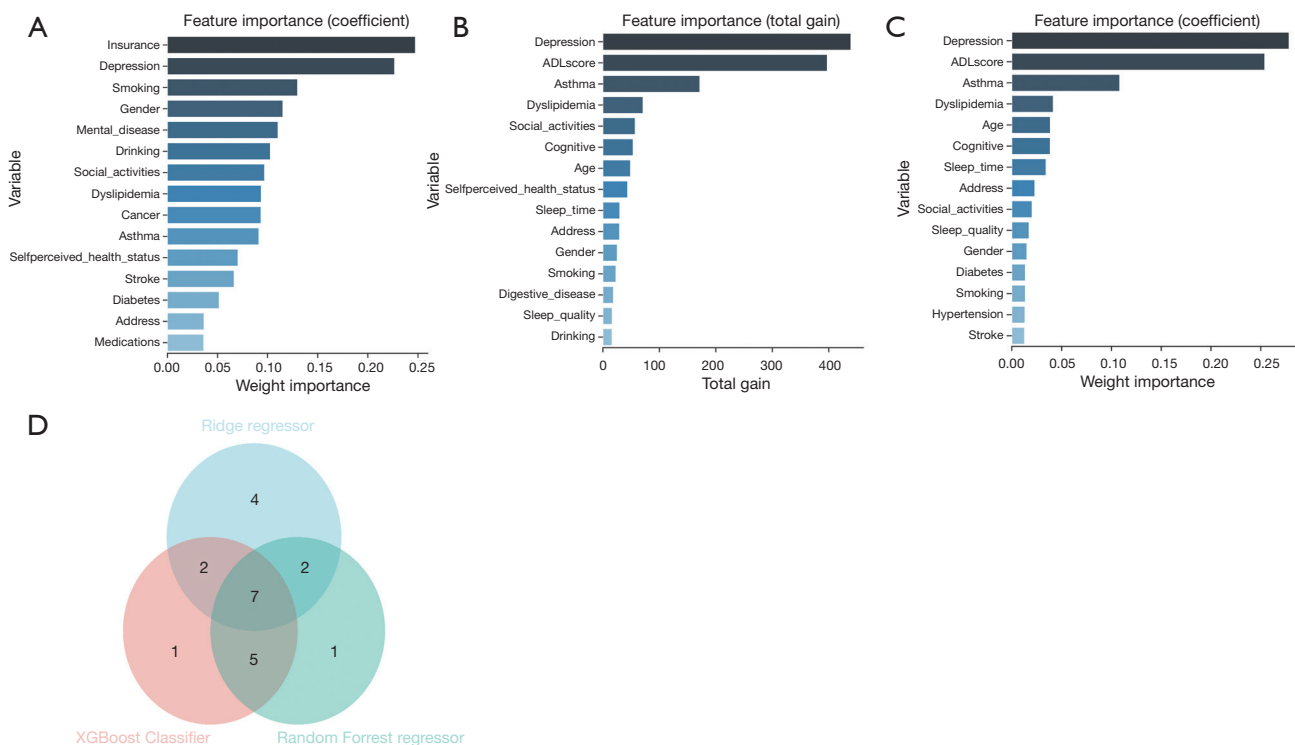


Figure 2 Presentation of the results of the screening of important predictor variables. (A-C) Ridge regressor, XGBoost classifier, and random forest regressor were used to analyze the variable importance ranking. (D) The top 15 predictive factors were selected separately, and seven commonly important predictive factors were identified via intersections in Venn's plots. XGBoost, extreme gradient boosting.

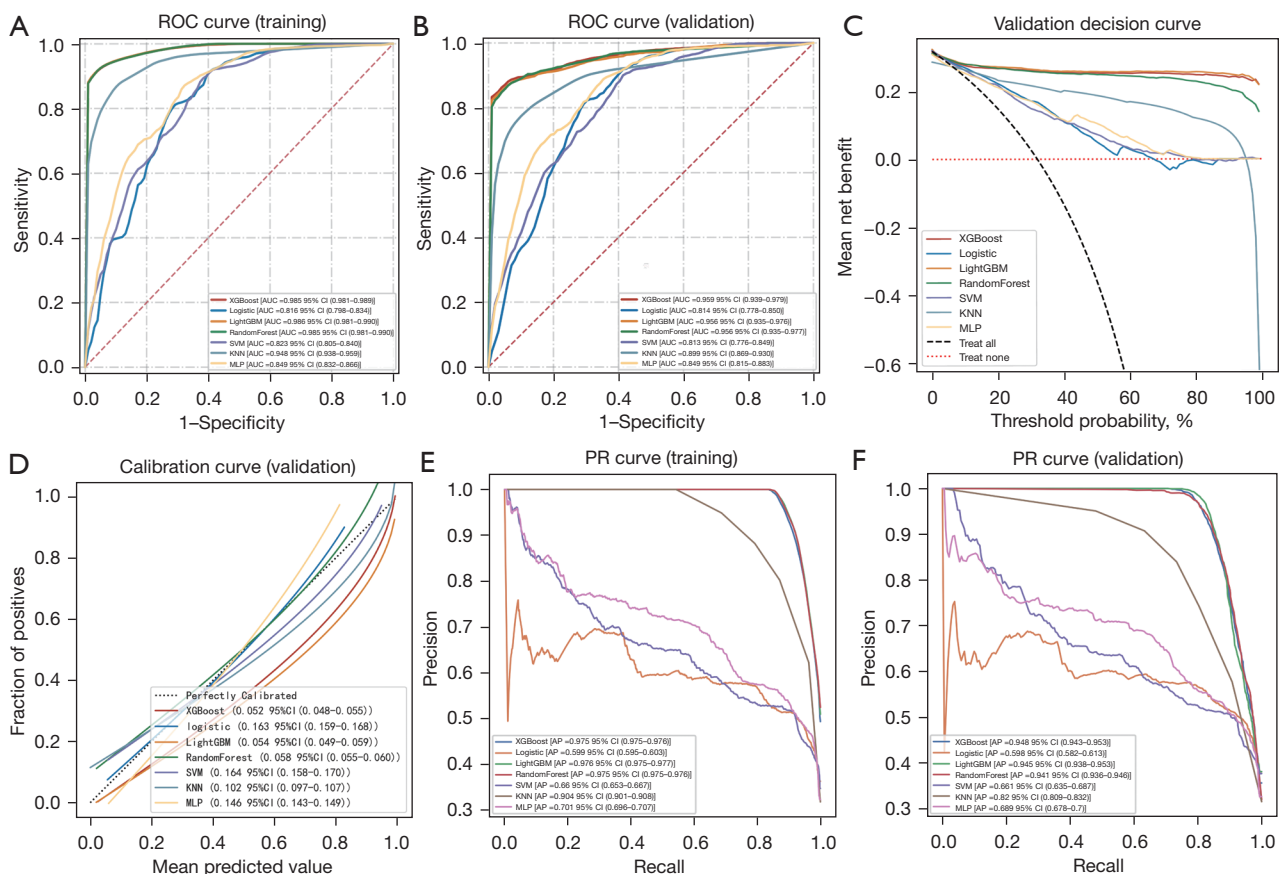


Figure 3 Analysis of the machine learning models. (A,B) ROCs and AUCs of the training and validation sets. The patients with COPD were tested 10 times at a ratio of 7:3. (C) A DCA plot of the validation set, in which all patients were hypothesized to be frail (black dashed line) or all patients were hypothesized to be nonfrail (red dashed line). Different models are represented by the remaining solid lines. (D) The dashed diagonal line represents the reference point, and the other solid lines represent the fitted models used. The smaller the number in parentheses is, the closer the fitted line to the reference line and the more accurate the predicted value of the model. A lower number in parentheses indicates that a fitted curve is more similar to a reference curve and thus more accurate. (E,F) PR curves and APs for the training set and test set. The horizontal axis indicates the recall rate, and the vertical axis indicates the precision rate. When one model's PR curve is completely covered by another's, this indicates the latter's superior performance, with higher APs indicating better model performance. The colors in the image correspond to the individual models, with their respective mean values and 95% confidence intervals displayed. ROC, receiver operating characteristic; XGBoost, extreme gradient boosting; LightGBM, light gradient-boosting machine; AP, average precision; CI, confidence interval; SVM, support vector machine; KNN, k-nearest neighbors; MLP, multilayer perceptron; AUC, area under the curve; COPD, chronic obstructive pulmonary disorder; DCA, decision curve analysis; PR, precision curve.

preferable (21,22), DCA, calibration curves, and precision-recall (PR) curves were used in this study. DCA assessed the clinical applicability of the LR and XGBoost models in improving accuracy (Figure 3C). The XGBoost and LR model predictions were more accurate according to the calibration curves (Figure 3D). In the evaluation of the clinical applicability and prediction accuracy of the XGBoost and LightGBM models, the XGBoost model showed the best performance in the training and validation

sets, with the highest average precision (AP) values in the validation set (Figure 3E,3F). A comprehensive analysis indicated that XGBoost may be relatively stable considering the high probability of overfitting in LightGBM, and thus XGBoost was selected as the optimal model (Table 2).

Building and evaluating the optimal model

The dataset reserved for training was subjected to

Table 2 Summary of the training set results for the multimodel classification

ML model	AUC (95% CI)	Cutoff (95% CI)	Accuracy (95% CI)	Sensitivity (95% CI)	Specificity (95% CI)	PPV (95% CI)	NPV (95% CI)	F1 score (95% CI)	Kappa (95% CI)
XGBoost	0.985 (0.981–0.989)	0.294 (0.274–0.314)	0.948 (0.945–0.951)	0.912 (0.905–0.919)	0.964 (0.958–0.971)	0.926 (0.912–0.940)	0.958 (0.955–0.962)	0.919 (0.915–0.922)	0.880 (0.874–0.885)
Logistic regression	0.816 (0.798–0.834)	0.296 (0.272–0.319)	0.740 (0.731–0.750)	0.824 (0.800–0.847)	0.702 (0.679–0.726)	0.563 (0.552–0.575)	0.896 (0.885–0.907)	0.668 (0.665–0.671)	0.467 (0.458–0.476)
LightGBM	0.986 (0.981–0.990)	0.328 (0.304–0.352)	0.951 (0.949–0.953)	0.908 (0.899–0.917)	0.971 (0.965–0.976)	0.938 (0.924–0.953)	0.957 (0.953–0.961)	0.923 (0.919–0.926)	0.886 (0.881–0.891)
Random forest	0.985 (0.981–0.990)	0.340 (0.324–0.356)	0.948 (0.946–0.950)	0.918 (0.912–0.924)	0.959 (0.954–0.964)	0.926 (0.914–0.938)	0.959 (0.956–0.962)	0.922 (0.918–0.926)	0.880 (0.876–0.884)
SVM	0.823 (0.805–0.840)	0.197 (0.186–0.207)	0.710 (0.702–0.719)	0.904 (0.889–0.919)	0.621 (0.603–0.638)	0.526 (0.515–0.536)	0.933 (0.924–0.941)	0.664 (0.659–0.670)	0.439 (0.428–0.450)
KNN	0.948 (0.938–0.959)	0.400 (0.400–0.400)	0.900 (0.897–0.903)	0.871 (0.863–0.880)	0.899 (0.891–0.908)	0.882 (0.870–0.894)	0.907 (0.904–0.911)	0.876 (0.871–0.881)	0.763 (0.758–0.769)
MLP	0.849 (0.832–0.866)	0.297 (0.266–0.328)	0.748 (0.737–0.758)	0.861 (0.824–0.897)	0.696 (0.664–0.728)	0.574 (0.552–0.597)	0.915 (0.901–0.929)	0.686 (0.684–0.688)	0.490 (0.480–0.500)

ML, machine learning; AUC, area under the curve; CI, confidence interval; PPV, positive predictive value; NPV, negative predictive value; XGBoost, extreme gradient boosting; LightGBM, light gradient-boosting machine; SVM, support vector machine; KNN, k-nearest neighbors; MLP, multilayer perceptron.

XGBoost regression analysis and 10-fold cross-validation. Consequently, the training set produced an average AUC (95% CI) of 0.980 (0.975–0.986), while the average AUC obtained from cross-validation of the validation set was 0.963 (0.937–0.990). In addition, the average AUC from the test set was 0.942 (0.925–0.959) (*Figure 4A–4C*). The AUC values for the training, validation, and test sets ultimately stabilized around 0.85, indicating the accurate predictive performance of the model. When the performance of the validation set in terms of AUC is lower than that of the test set or falls below 10%, it can be considered a successful model fit. The learning curve showed a strong fit between the training and validation sets, indicating overfitting and high stability (*Figure 4D*). These results indicated that for the classification modelling aspect of the dataset, the implementation of the XGBoost regression model could be used (*Table 3*).

The XGBoost model explained via the SHAP method

SHAP was used in order to demonstrate how these variables predicted frailty formation within our COPD patient model and to visualize the selected variables (23). *Figure 5A* presents the seven main characteristics of the model; each trait significance line has a variety of colored dots to indicate each patient's contribution to the outcome: red dots indicate high risk, and blue dots indicate low risk. To detect positive or negative relationships between predicted

values and target outcomes, we used SHAP values to reveal risk factors for the development of frailty in patients with COPD. The horizontal position shows the degree to which the value is high or low relative to the predicted value, and the color shows whether the variable is high (red) or low (blue) for that observation. Moreover, *Figure 5B* shows the SHAP algorithm used to obtain the significance of each predictor variable for the predicted results of the XGBoost model, and the variable significance plot lists the most significant variables in descending order. A number of features predicted whether a COPD patient was frail or not, and the SHAP values show the extent to which each feature contributed to the prediction results. Red features on the left side indicate features associated with a greater likelihood of frailty in patients with COPD, and features on the right side indicate features associated with a lower likelihood of frailty. The length of the arrows corresponds to the degree of influence on the predicted outcome: the longer the length of the arrow, the greater the impact (*Figure 5C*). A website for online predictive risk modeling was created based on the model (<https://www.xsmartanalysis.com/model/list/predict/model/html?mid=9481&symbol=3ic16994so62gh894hy8>).

Discussion

To the best of our knowledge, this is the first study to use a

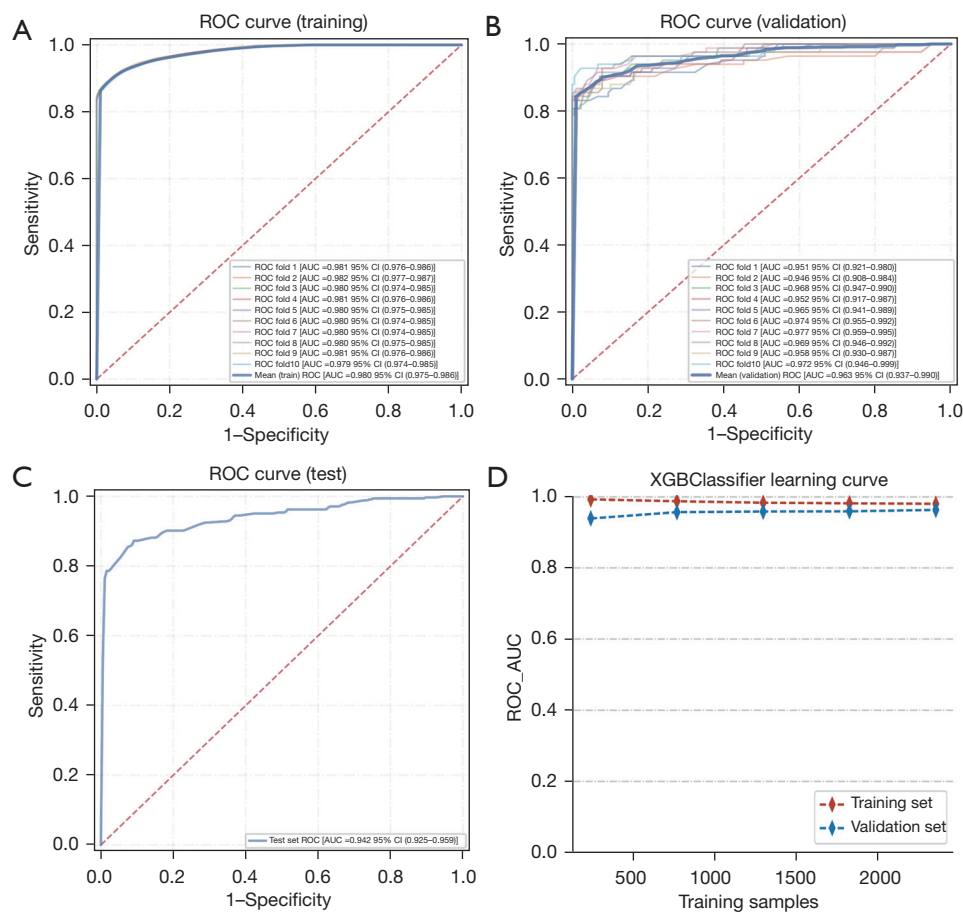


Figure 4 Training, validation, and testing of the XGBoost model. (A,B) ROC and AUC for training and validating sets. The model was trained and validated on 10% of patients in the COPD cohort. Ten different outcomes are shown as solid lines of different colors. (C) For the test set, ROC and AUC were calculated. Tests were performed on 30% of patients in the COPD cohort. (D) Learning curve. The training set is represented by the red dashed line, while the validation set is represented by the blue dashed line. Values are expressed as the mean and 95% CIs. XGBoost, extreme gradient boosting; ROC, receiver operating characteristic; AUC, area under the curve; COPD, chronic obstructive pulmonary disorder; CI, confidence interval.

ML predictive modeling approach to characterize the frailty of patients with COPD. Frailty is more common in older adults with COPD, and COPD is strongly associated with frailty. In this study, the prevalence of frailty in patients with COPD was 7%, which is similar to the prevalence of frailty reported in previous studies (24,25) but significantly lower than the study by Hirai *et al.* (26).

The development of vulnerability to COPD is complex and related to many different factors. Our study found there to be a relationship between the age of patients with COPD and their frailty. As the age of the patients increased, their level of frailty also showed an increasing trend. This suggests that with age, patients with COPD are more likely to show signs of frailty, such as physical weakness,

decreased immune function, and unstable psychological status (27), which may be related to decreased physical function, cognitive impairment, and greater dysfunction in socialization and multiple chronic diseases. In addition, our study identified an association between frailty in patients with COPD and other factors such as gender and smoking history (28), which might have interacted with the patient's age to further influence the occurrence of frailty. However, we also observed individual differences across patients of the same age. Some older patients showed better physical and cognitive functioning, whereas some younger patients showed higher levels of frailty, suggesting that age is not the only factor influencing frailty and that other factors such as genetics, lifestyle, and environment may also be critical (29).

Table 3 Summary of the validation set results for the multimodel classification

ML mode	AUC (95% CI)	Cutoff (95% CI)	Accuracy (95% CI)	Sensitivity (95% CI)	Specificity (95% CI)	PPV (95% CI)	NPV (95% CI)	F1 score (95% CI)	Kappa (95% CI)
XGBoost	0.959 (0.939–0.979)	0.294 (0.274–0.314)	0.920 (0.914–0.925)	0.871 (0.851–0.891)	0.963 (0.944–0.982)	0.875 (0.852–0.898)	0.941 (0.934–0.948)	0.872 (0.856–0.889)	0.814 (0.801–0.826)
Logistic regression	0.814 (0.778–0.850)	0.296 (0.272–0.319)	0.744 (0.731–0.757)	0.847 (0.813–0.881)	0.692 (0.654–0.730)	0.571 (0.552–0.589)	0.895 (0.880–0.910)	0.681 (0.662–0.701)	0.474 (0.454–0.494)
LightGBM	0.956 (0.935–0.976)	0.328 (0.304–0.352)	0.921 (0.911–0.931)	0.862 (0.844–0.881)	0.961 (0.944–0.978)	0.897 (0.874–0.919)	0.933 (0.927–0.938)	0.879 (0.860–0.897)	0.818 (0.796–0.840)
Random forest	0.956 (0.935–0.977)	0.340 (0.324–0.356)	0.916 (0.909–0.923)	0.869 (0.852–0.887)	0.951 (0.930–0.971)	0.876 (0.854–0.898)	0.935 (0.930–0.940)	0.872 (0.860–0.884)	0.806 (0.790–0.823)
SVM	0.813 (0.776–0.849)	0.197 (0.186–0.207)	0.693 (0.679–0.706)	0.878 (0.835–0.922)	0.634 (0.584–0.684)	0.511 (0.495–0.528)	0.925 (0.908–0.941)	0.645 (0.629–0.661)	0.410 (0.385–0.435)
KNN	0.899 (0.869–0.930)	0.400 (0.400–0.400)	0.871 (0.861–0.881)	0.773 (0.747–0.799)	0.911 (0.889–0.932)	0.840 (0.817–0.862)	0.884 (0.874–0.894)	0.804 (0.785–0.824)	0.692 (0.668–0.715)
MLP	0.849 (0.815–0.883)	0.297 (0.266–0.328)	0.737 (0.714–0.760)	0.786 (0.732–0.840)	0.782 (0.729–0.835)	0.555 (0.518–0.593)	0.921 (0.907–0.935)	0.646 (0.624–0.668)	0.471 (0.442–0.501)

ML, machine learning; AUC, area under the curve; CI, confidence interval; PPV, positive predictive value; NPV, negative predictive value; XGBoost, extreme gradient boosting; LightGBM, light gradient-boosting machine; SVM, support vector machine; KNN, K-nearest neighbors; MLP, multilayer perceptron.

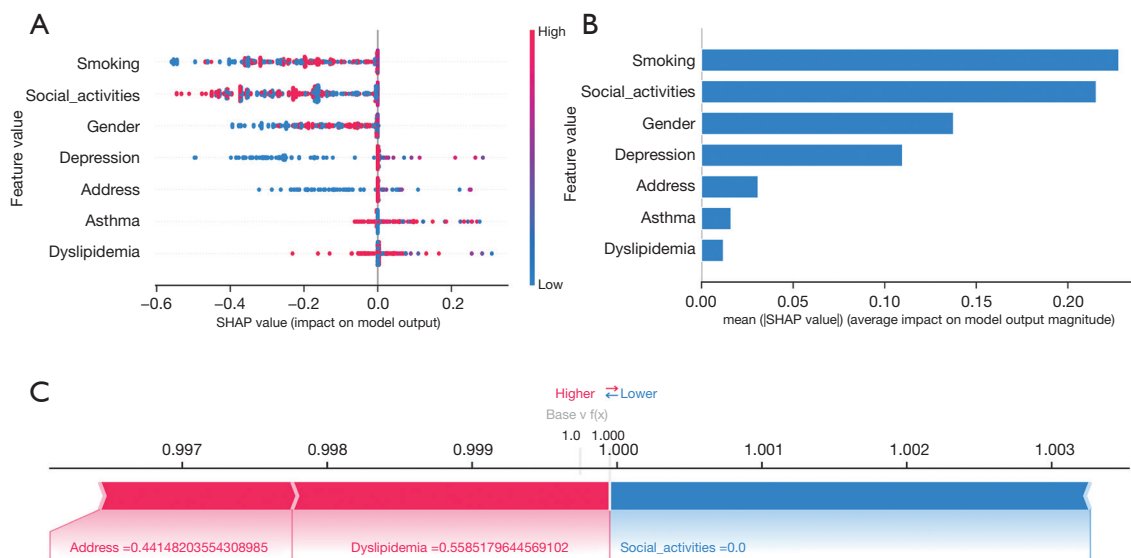


Figure 5 SHAP of the model. (A) Characteristic attributes in SHAP. The abscissa is the SHAP value, and each line denotes a feature. Higher eigenvalues are indicated by red dots, and lower eigenvalues are indicated by blue dots. (B) The SHAP-based feature importance ranking. The significance of each covariate in the creation of the final prediction model is depicted in the matrix diagram. (C) SHAP force plot. SHAP, Shapley additive explanations.

Our prediction model suggests that the relationship between cigarette smoking and frailty in patients with COPD is a complex issue that can be interpreted physiological, psychological, and social perspectives (30-33). First, the decline in lung function and physical function

due to smoking is an important factor affecting the frailty of patients with COPD. Long-term smoking leads to lung inflammation and airway narrowing, limiting patients' respiratory function and reducing their physical mobility and quality of life. Second, smoking can also have a negative

impact on patients' mental health. For one, smoking can cause dependence and withdrawal symptoms, increasing anxiety and depression; for another, long-term smoking may lead to cognitive decline. Finally, smoking also affects patients' social interactions. Due to the negative societal perception of smoking behavior, patients who smoke may face discrimination and rejection, leading to a decline in their social skills and increasing their vulnerability. In summary, there is a close relationship between smoking and the vulnerability of patients with COPD. The combination of declining lung function, reduced physical functioning, mental health problems, and limitations in social interactions due to smoking makes patients more vulnerable. Therefore, in the treatment and management of patients with COPD, attention should be paid to smoking, and active measures should be taken to help patients quit smoking in order to minimize the occurrence and further deterioration of frailty. However, further research is needed to understand and intervene in other factors affecting frailty in patients with COPD.

This study found that the occurrence of frailty in patients with COPD was related to their living conditions (34-38). First, we observed that patients with COPD in rural areas were more likely to develop frailty compared to those in urban areas. This may be partly attributed to factors such as lower economic and educational levels, as well as the insufficiency of medical resources in rural areas. Due to the lack of adequate medical facilities and healthcare professionals, patients in rural areas are often unable to receive timely and effective treatment and care, leading to disease progression and deterioration, thus increasing the risk of frailty. In addition, we found that patients with COPD in rural areas are more susceptible to environmental factors, such as poorer air quality and exposure to pollutants, which may exacerbate their conditions and increase their risk of frailty. Second, we also identified the socioeconomic factors that are associated with the occurrence of frailty among patients with COPD in urban areas. For example, factors such as low income, low educational attainment, social isolation, and psychological stress (due to factors such as the fast pace of life and intense competition) were found to be associated with a high incidence of frailty. This may be due to the higher cost of living in urban areas, which increases financial pressure and makes it difficult for patients to afford the high cost of medical care and medications. Additionally, low educational attainment and social isolation may lead to a lack of health literacy and social support for patients, further exacerbating the risk of

frailty. Additionally, we observed that some chronic disease-related factors are associated with the occurrence of frailty. For example, the severity of symptoms, the presence of comorbidities, and the long-term use of medications such as hormones in patients with COPD are associated with a high incidence of frailty. These factors may contribute to the deterioration of patients' physical condition, increasing the loss of the ability to perform activities of daily living and heightening health risks, thus rendering these patients more vulnerable to the development of frailty. In conclusion, we believe that improving frailty in patients with COPD in both rural and urban areas requires a combination of measures.

This study also revealed that frailty occurred more frequently in patients with COPD who had a variety of comorbidities. We also observed that patients with both COPD and asthma showed a higher risk of vulnerability, possibly due to the compounding effects of the two diseases (39). Vulnerability refers to an individual's susceptibility to physical, psychological, and social stress. We compared the physical functioning, psychological assessment, and social support between patients with COPD and asthma and those with uncomplicated asthma. The results indicated that patients with comorbid asthma experienced more physical limitations, higher levels of anxiety, and more severe depression according to psychological assessments, and lower social support scores. These findings suggest that patients with COPD with comorbid asthma are more likely to become frail when faced with life stressors. We further investigated factors contributing to increased frailty in patients with COPD combined with asthma. Patients with comorbid asthma experienced issues related to their treatment, including poor medication compliance, frequent acute exacerbations, and a lack of effective self-management skills. These issues may result in ineffective symptom control, increasing the patients' physical and psychological burden and vulnerability. Patients with COPD and asthma may also face problems such as social isolation and discrimination, which could lead to frailty. Frailty is linked to prognosis in those with COPD-combined asthma, and our study revealed there to be worse outcomes in patients with frailty, including increased acute exacerbations, hospitalizations, and reduced quality of life. This suggests that greater attention should be paid to frailty in the management and treatment of COPD-combined asthma. Furthermore, the prevalence of comorbid hyperlipidemia was relatively high among patients with COPD, which is consistent with previous findings

indicating its common occurrence in this population, with hyperlipidemia potentially being related to the pathogenesis of COPD (40). The literature also suggests that hyperlipidemia may contribute to increased frailty in patients with COPD through several pathways (40-42). First, hyperlipidemia can trigger a systemic inflammatory response, accelerating the progression of COPD and leading to worsening lung function. Second, it may increase the risk of lung function decline, making patients with COPD more susceptible to frailty. Furthermore, hyperlipidemia may affect cognitive function and psychological status, leading to decreased muscle function and reduced oxygenation capacity, exacerbating frailty.

Our study also identified an association of depression and socialization with the occurrence of frailty in patients with COPD, which is consistent with previous research (43). The prevalence of depression in patients with COPD is as high as 15%, which is approximately double that in those with COPD (5). A 2017 meta-analysis from several countries reported that among depressed individuals, frail individuals were 4.42 times more likely to be depressed and 4.07 times more likely to be frail (44). Moreover, depressed individuals were 4.07 times more likely to be vulnerable, the prevalence of vulnerability was 2.72 times higher in depressed patients than in nondepressed patients, and there was a 0.90 times greater probability of depression occurring in frail patients (44). These studies indicate that the risk of vulnerability significantly increases after the occurrence of depression in patients COPD although there is no such research data available in China. In line with other research (45,46), our study also found there to be a link between depression and a decline in the physical and mental health of patients, which can affect their motivation for treatment and rehabilitation, exacerbating their disease symptoms and leading to an increase in frailty. Furthermore, depression may also contribute to patients' social isolation, reducing their willingness and ability to participate in social activities, which can result in a lack of effective support and communication in their lives and increasing the likelihood of frailty. Moreover, our study identified dyspnea as the most significant symptom in patients with COPD. Specifically, as the condition of patients with COPD worsens, dyspnea due to exercise intolerance creates a vicious cycle of physical discomfort, restricted activity, and reduced social contact (47). Social activity serves as a protective factor against frailty, while comorbidities such as depression and other disorders may lead to frailty. Compared with those who rarely or never participate in social activities,

individuals who regularly participate in social activities have a reduced likelihood of becoming frail. These results are consistent with earlier research (48), indicating that socialization diminishes feelings of loneliness and social isolation and is linked with a lower frequency of frailty. In addition, social activities can help patients gain information and skills about the disease, obtain self-confidence in coping with the disease, acquire self-management skills and self-efficacy, and develop a better social support system, all of which can reduce negative feelings and distress related to the disease (49,50). This, in turn, can help patients better adapt to the disease and treatment process and reduce the incidence of frailty, there by achieving prevention of acute exacerbations and complications of COPD. For this reason, it is recommended that additional opportunities be organized for people with COPD to engage in an exchange of experiences and a variety of social activities.

A novel outcome of this study is the establishment of a predictive model for frailty in patients with COPD. As an effective and accurate assessment tool, our predictive model can help physicians identify the susceptible population at high risk of developing COPD and can provide a theoretical basis for the development of early prevention and intervention measures. This predictive model has a high degree of clinical applicability and can help to identify frail patients at high risk during the screening process.

Despite identifying promising findings in our study, we acknowledge several limitations that should be considered. Firstly, the latest version of the CHARLS database, which we used for our analysis, is lacking important predictive factors such as lung function and BMI. These missing variables could potentially affect the accuracy and comprehensiveness of our predictive model. Secondly, it is important to note that our study was conducted using data from China, and therefore, the generalizability of our findings to other regions and countries may be limited. To establish the wider relevance of our research, additional validation using external cohort data from diverse populations would be necessary. Thirdly, our study solely relied on retrospective data and did not include long-term follow-up of patients with COPD. Incorporating data from patients with extended follow-up periods would enhance the current prediction model and provide more reliable insights into the occurrence of frailty in individuals with COPD. These limitations are critically important in understanding the scope and validity of our study. To address these shortcomings, we are committed to collecting better and more comprehensive information in future research

endeavors focused on frailty in COPD patients. By doing so, we aim to improve the accuracy and applicability of our findings.

Conclusions

The study established a ML model to build a predictive model, and our results show that the XGBoost model outperformed the other techniques. In addition, developed personalized risk assessment tool for the occurrence of frailty in patients with COPD based on SHAP. This computer-based method is an effective tool for front-line clinicians and patients with COPD in the identification and intervention of frailty.

Acknowledgments

We would like to thank the investigators and patients for their support of this study and those at xsmartanalysis (<https://www.xsmartanalysis.com/model/index>) for their technical support.

Funding: This study was supported by the 2020 Municipal Applied Technology Research and Development Fund Project (No. 20YFZJ0105).

Footnote

Reporting Checklist: The authors have completed the TRIPOD reporting checklist. Available at <https://jtd.amegroups.com/article/view/10.21037/jtd-24-416/rc>

Peer Review File: Available at <https://jtd.amegroups.com/article/view/10.21037/jtd-24-416/prf>

Conflicts of Interest: All authors have completed the ICMJE uniform disclosure form (available at <https://jtd.amegroups.com/article/view/10.21037/jtd-24-416/coif>). The authors have no conflicts of interest to declare.

Ethical Statement: The authors are accountable for all aspects of the work in ensuring that questions related to the accuracy or integrity of any part of the work are appropriately investigated and resolved. The study was conducted in accordance with the Declaration of Helsinki (as revised in 2013).

Open Access Statement: This is an Open Access article distributed in accordance with the Creative Commons

Attribution-NonCommercial-NoDerivs 4.0 International License (CC BY-NC-ND 4.0), which permits the non-commercial replication and distribution of the article with the strict proviso that no changes or edits are made and the original work is properly cited (including links to both the formal publication through the relevant DOI and the license). See: <https://creativecommons.org/licenses/by-nc-nd/4.0/>.

References

1. Agustí A, Celli BR, Criner GJ, et al. Global Initiative for Chronic Obstructive Lung Disease 2023 Report: GOLD Executive Summary. *Eur Respir J* 2023;61:2300239.
2. Soler-Cataluña JJ, Izquierdo JL, Juárez Campo M, et al. Impact of COPD Exacerbations and Burden of Disease in Spain: AVOIDEX Study. *Int J Chron Obstruct Pulmon Dis* 2023;18:1103-14.
3. Yang T, Cai B, Cao B, et al. Treatment patterns in patients with stable COPD in China: analysis of a prospective, 52-week, nationwide, observational cohort study (REAL). *Ther Adv Respir Dis* 2023;17:17534666231158283.
4. Fried LP, Tangen CM, Walston J, et al. Frailty in older adults: evidence for a phenotype. *J Gerontol A Biol Sci Med Sci* 2001;56:M146-56.
5. Marengoni A, Vetrano DL, Manes-Gravina E, et al. The Relationship Between COPD and Frailty: A Systematic Review and Meta-Analysis of Observational Studies. *Chest* 2018;154:21-40.
6. He B, Ma Y, Wang C, et al. Prevalence and Risk Factors for Frailty among Community-Dwelling Older People in China: A Systematic Review and Meta-Analysis. *J Nutr Health Aging* 2019;23:442-50.
7. Dias LS, Ferreira ACG, da Silva Junior JLR, et al. Prevalence of Frailty and Evaluation of Associated Variables Among COPD Patients. *Int J Chron Obstruct Pulmon Dis* 2020;15:1349-56.
8. Walston J, Hadley EC, Ferrucci L, et al. Research agenda for frailty in older adults: toward a better understanding of physiology and etiology: summary from the American Geriatrics Society/National Institute on Aging Research Conference on Frailty in Older Adults. *J Am Geriatr Soc* 2006;54:991-1001.
9. Ershler WB, Keller ET. Age-associated increased interleukin-6 gene expression, late-life diseases, and frailty. *Annu Rev Med* 2000;51:245-70.
10. Ierodiakonou D, Kampouraki M, Poulonirakis I, et al. Determinants of frailty in primary care patients with COPD: the Greek UNLOCK study. *BMC Pulm Med*

- 2019;19:63.
11. Theou O, Cann L, Blodgett J, et al. Modifications to the frailty phenotype criteria: Systematic review of the current literature and investigation of 262 frailty phenotypes in the Survey of Health, Ageing, and Retirement in Europe. *Ageing Res Rev* 2015;21:78-94.
 12. Wu C, Smit E, Xue QL, et al. Prevalence and Correlates of Frailty Among Community-Dwelling Chinese Older Adults: The China Health and Retirement Longitudinal Study. *J Gerontol A Biol Sci Med Sci* 2017;73:102-8.
 13. Veronese N, Solmi M, Maggi S, et al. Frailty and incident depression in community-dwelling older people: results from the ELSA study. *Int J Geriatr Psychiatry* 2017;32:e141-9.
 14. Bu F, Deng XH, Zhan NN, et al. Development and validation of a risk prediction model for frailty in patients with diabetes. *BMC Geriatr* 2023;23:172.
 15. Cabañero-Martínez MJ, Cabrero-García J, Richart-Martínez M, et al. The Spanish versions of the Barthel index (BI) and the Katz index (KI) of activities of daily living (ADL): a structured review. *Arch Gerontol Geriatr* 2009;49:e77-84.
 16. Bijwaard GE, van Kippersluis H, Veenman J. Education and health: The role of cognitive ability. *J Health Econ* 2015;42:29-43.
 17. Irwin M, Artin KH, Oxman MN. Screening for depression in the older adult: criterion validity of the 10-item Center for Epidemiological Studies Depression Scale (CES-D). *Arch Intern Med* 1999;159:1701-4.
 18. Maw M, Haw SC, Ho CK. Utilizing data sampling techniques on algorithmic fairness for customer churn prediction with data imbalance problems. *F1000Res* 2021;10:988.
 19. Lundberg SM, Erion G, Chen H, et al. From Local Explanations to Global Understanding with Explainable AI for Trees. *Nat Mach Intell* 2020;2:56-67.
 20. Sauerbrei W, Royston P, Binder H. Selection of important variables and determination of functional form for continuous predictors in multivariable model building. *Stat Med* 2007;26:5512-28.
 21. Obuchowski NA, Bullen JA. Receiver operating characteristic (ROC) curves: review of methods with applications in diagnostic medicine. *Phys Med Biol* 2018;63:07TR01.
 22. Nahm FS. Receiver operating characteristic curve: overview and practical use for clinicians. *Korean J Anesthesiol* 2022;75:25-36.
 23. Xue B, Li D, Lu C, et al. Use of Machine Learning to Develop and Evaluate Models Using Preoperative and Intraoperative Data to Identify Risks of Postoperative Complications. *JAMA Netw Open* 2021;4:e212240.
 24. Limpawattana P, Putraveephong S, Inthasuwana P, et al. Frailty syndrome in ambulatory patients with COPD. *Int J Chron Obstruct Pulmon Dis* 2017;12:1193-8.
 25. Lahousse L, Maes B, Ziere G, et al. Adverse outcomes of frailty in the elderly: the Rotterdam Study. *Eur J Epidemiol* 2014;29:419-27.
 26. Hirai K, Tanaka A, Oda N, et al. Prevalence and Impact of Social Frailty in Patients with Chronic Obstructive Pulmonary Disease. *Int J Chron Obstruct Pulmon Dis* 2023;18:2117-26.
 27. Wijnant SRA, Benz E, Luik AI, et al. Frailty Transitions in Older Persons With Lung Function Impairment: A Population-Based Study. *J Gerontol A Biol Sci Med Sci* 2023;78:349-56.
 28. Gordon EH, Hubbard RE. Frailty: understanding the difference between age and ageing. *Age Ageing* 2022;51:afac185.
 29. Thillainadesan J, Scott IA, Le Couteur DG. Frailty, a multisystem ageing syndrome. *Age Ageing* 2020;49:758-63.
 30. Kojima G, Iliffe S, Walters K. Smoking as a predictor of frailty: a systematic review. *BMC Geriatr* 2015;15:131.
 31. Gonçalves RB, Coletta RD, Silvério KG, et al. Impact of smoking on inflammation: overview of molecular mechanisms. *Inflamm Res* 2011;60:409-24.
 32. Wu Z, Yue Q, Zhao Z, et al. A cross-sectional study of smoking and depression among US adults: NHANES (2005-2018). *Front Public Health* 2023;11:1081706.
 33. Martin LM, Sayette MA. A review of the effects of nicotine on social functioning. *Exp Clin Psychopharmacol* 2018;26:425-39.
 34. Fan J, Yu C, Guo Y, et al. Frailty index and all-cause and cause-specific mortality in Chinese adults: a prospective cohort study. *Lancet Public Health* 2020;5:e650-60.
 35. Ožić S, Vasiljev V, Ivković V, et al. Interventions aimed at loneliness and fall prevention reduce frailty in elderly urban population. *Medicine (Baltimore)* 2020;99:e19145.
 36. Shalini T, Chitra PS, Kumar BN, et al. Frailty and Nutritional Status among Urban Older Adults in South India. *J Aging Res* 2020;2020:8763413.
 37. Nguyen HT, Nguyen AH, Nguyen GTX. Prevalence and associated factors of frailty in patients attending rural and urban geriatric clinics. *Australas J Ageing* 2022;41:e122-30.
 38. Seo Y, Kim M, Shim H, et al. Differences in the Association of Neighborhood Environment With Physical Frailty Between Urban and Rural Older Adults: The

- Korean Frailty and Aging Cohort Study (KFACS). *J Am Med Dir Assoc* 2021;22:590-597.e1.
39. Wang X, Wen J, Gu S, et al. Frailty in asthma-COPD overlap: a cross-sectional study of association and risk factors in the NHANES database. *BMJ Open Respir Res* 2023;10:e001713.
 40. Wang Q, Wang Y, Lehto K, et al. Genetically-predicted life-long lowering of low-density lipoprotein cholesterol is associated with decreased frailty: A Mendelian randomization study in UK biobank. *EBioMedicine* 2019;45:487-94.
 41. Dao HHH, Burns MJ, Kha R, et al. The Relationship between Metabolic Syndrome and Frailty in Older People: A Systematic Review and Meta-Analysis. *Geriatrics (Basel)* 2022;7:76.
 42. Fahed G, Aoun L, Bou Zerdan M, et al. Metabolic Syndrome: Updates on Pathophysiology and Management in 2021. *Int J Mol Sci* 2022;23:786.
 43. Fearn M, Bhar S, Dunt D, et al. Befriending to Relieve Anxiety and Depression Associated with Chronic Obstructive Pulmonary Disease (COPD): A Case Report. *Clin Gerontol* 2017;40:207-12.
 44. Soysal P, Veronese N, Thompson T, et al. Relationship between depression and frailty in older adults: A systematic review and meta-analysis. *Ageing Res Rev* 2017;36:78-87.
 45. Malhi GS, Mann JJ. Depression. *Lancet* 2018;392:2299-312.
 46. Potter GG, McQuoid DR, Whitson HE, et al. Physical frailty in late-life depression is associated with deficits in speed-dependent executive functions. *Int J Geriatr Psychiatry* 2016;31:466-74.
 47. Bourbeau J. Activities of life: the COPD patient. *COPD* 2009;6:192-200.
 48. Fingerma KL, Ng YT, Huo M, et al. Functional Limitations, Social Integration, and Daily Activities in Late Life. *J Gerontol B Psychol Sci Soc Sci* 2021;76:1937-47.
 49. Thomas PA. Trajectories of social engagement and mortality in late life. *J Aging Health* 2012;24:547-68.
 50. Horgas AL, Wilms HU, Baltes MM. Daily life in very old age: everyday activities as expression of successful living. *Gerontologist* 1998;38:556-68.

Cite this article as: Chen Y, Yu Y, Yang D, Zhang W, Kouritas V, Chen X. Developing and validating machine learning-based prediction models for frailty occurrence in those with chronic obstructive pulmonary disease. *J Thorac Dis* 2024;16(4):2482-2498. doi: 10.21037/jtd-24-416