TSINGHUA SCIENCE AND TECHNOLOGY
ISSN 1007-0214 09/14 pp525–540
DOI: 10.26599/TST.2022.9010010
Volume 28, Number 3, June 2023

A Case-Finding Clinical Decision Support System to Identify Subjects with Chronic Obstructive Pulmonary Disease Based on Public Health Data

Xinshan Lin, Yi Lei, Jun Chen, Zhihui Xing, Ting Yang*, Qing Wang*, and Chen Wang*

Abstract: Chronic obstructive pulmonary disease (COPD) is a serious chronic respiratory disease. Improving the ability to identify patients with COPD in primary medical institutions is important to prevent and treat the disease. With the continuous development of medical digitization, the application of big data informatization in the medical and health fields has become possible. Recently, applying innovative technologies such as big data analysis, machine learning, and artificial intelligence-assisted decision-making in the medical field has become an interdisciplinary research hotspot. Based on the identification and diagnosis of COPD in the high-risk population, this study proposes a convenient and effective clinical decision support system to help identify patients with COPD in primary health institutions. The results of the preliminary experiments show that the proposed method is convenient and effective compared with the existing methods.

Key words: artificial intelligence; machine learning; case finding; chronic obstructive pulmonary disease (COPD); clinical decision support system (CDSS)

1 Introduction

Chronic obstructive pulmonary disease (COPD) is a serious chronic respiratory disease[1]. In relation to the high prevalence, mortality, and disease burden of COPD, low awareness and underdiagnosis in the high-risk population are disproportionate. COPD is insidious, and the rate of missed diagnosis is high[2]. Failure to timely diagnose COPD can lead to increased mortality and acute aggravation and affect the patients' health status and quality of life. Additionally, repeated hospital visits increase the medical costs and disease burden[3,4]. The early diagnosis of chronic diseases is critical at primary medical institutions. Based on global initiative for chronic obstructive lung disease (GOLD) and other large-scale epidemiological studies, in 2010, the estimated number of COPD cases was 384 million, with a global prevalence of 11.7%. Globally, approximately 3 million people die each year. As the habit of smoking increases in developing countries and populations age in high-income countries, the prevalence of COPD is expected to rise over the next 40 years, with more than 5.4 million deaths likely to occur annually by 2060 because of COPD and related diseases. Additionally, a study showed that the prevalence rate of COPD Grade 2 and above was approximately 10.1%∶11.8% in men and

• Xinshan Lin, Ting Yang, and Chen Wang are with the Department of Pulmonary and Critical Care Medicine, China-Japan Friendship Hospital, Beijing 100029, China, and also with Chinese Academy of Medical Sciences and Peking Union Medical College, Beijing 100005, China. E-mail: linxinshan@126.com; dryangting@qq.com; cyh-birm@263.net.
• Yi Lei is with the School of Software Engineering, the Faculty of Information Technology, Beijing University of Technology, Beijing 100124, China. E-mail: leiyi9345@163.com.
• Jun Chen and Zhihui Xing are with Intelligent Healthcare Unit, Baidu Inc, Beijing 100093, China. E-mail: chenjun22@baidu.com; christianahui@126.com.
• Qing Wang is with the Department of Automation, Tsinghua University, Beijing 100084, China. E-mail: qing.wang@tsinghua.edu.cn.
* To whom correspondence should be addressed.
Manuscript received: 2022-03-03; accepted: 2022-04-07

8.5% in women. According to the latest epidemiological data collected from China in 2018, the prevalence of COPD reached 8.6% in people aged older than 20 years and 13.6% in those older than 40 years[5].

Diagnostic testing for COPD in the entire population leads to overtesting and wastes medical resources, and the benefits are controversial[6]. Many studies were conducted to improve the case-finding capacity in primary healthcare institutions for COPD. Active case finding is a necessary screening method[7]. Studies showed that active case discovery was valuable for the diagnosis and intervention in populations at a high COPD risk[8] and was practiced in various countries, particularly at primary medical institutions[9, 10]. A primary care study in the United Kingdom included nonsmokers aged 40 to 79 years who were screened through a general practitioner consultation questionnaire or a mail questionnaire, and the participants reporting related respiratory symptoms were asked to undergo a spirometer test after administering a bronchodilator. The active case discovery approach resulted in a higher percentage of cases diagnosed with COPD than the opportunistic case discovery approach using conventional care (5% vs. 2%; $p < 0.0001$) and was more cost-effective (333 vs. 376 pounds per case)[11]. To effectively identify people with COPD, researchers have developed various COPD disease screening questionnaires. However, presently, investigators manually perform such case discovery work, which requires considerable manpower costs. Currently, the number of primary care physicians (PCPs) is insufficient to meet the needs of such complicated work. With the continuous development of medical digitalization, the application of big data informatization in the medical and health fields has become possible. When artificial intelligence is used to process medical data, match relevant case screening questionnaires, develop a population screening and risk assessment model, and develop a screening clinical decision support system (CDSS) for case finding, the work intensity and the cost of screening for COPD will be substantially reduced and the work efficiency will be improved[12–14].

This study aimed to develop a convenient and effective CDSS to assist in case finding and provide reliable inspection data for standardization management of COPD, improve the ability and efficiency of detecting COPD cases in primary health institutions, increase the productivity of respiratory specialists, and handle shortages of medical resources.

The main contributions of the paper are as follows.

- A CDSS was proposed to identify patients with COPD based on public health data. Since it can considerably reduce the burden on respiratory physicians (a scarce resource) and provide a large-scale COPD screening, it can be helpful and valuable in other similar medical diagnostic problems.
- A CDSS based on public health data for COPD risk assessment was constructed.
- A preliminary study was conducted to apply the proposed method in practical application scenarios. CDSS studies are valuable to improve the ability to identify patients with COPD in primary healthcare institutions. We believe that our preliminary study can serve as an important reference to develop similar health information systems to strengthen healthcare.

The study is organized as follows. Section 2 reports the related work. Section 3 introduces the framework of our system for the screening and auxiliary diagnosis of COPD. Section 4 presents the protocol and results of a preliminary study on patients with COPD, where our system was used. Section 5 discusses the results. Finally, Section 6 presents the conclusion.

2 Related Work

2.1 COPD screening

A questionnaire is the most common method for COPD screening. Scholars in various countries developed several screening questionnaires, providing a supporting tool for COPD screening and diagnosis and relevant cohort studies.

Currently, the COPD Diagnostic Questionnaire (COPD-DQ) is the most widely used worldwide. It was proposed by Price et al.[15] in 2006 and was verified for use in many populations. The diagnostic questionnaire mainly includes eight questions concerning age, body mass index (BMI), smoking status, the influence of weather on cough, daily sputum production, morning sputum production, wheezing status, and allergy status. The accumulated scores of each item provided a critical diagnostic value of COPD of 16.5 to 19.5 points. For threshold 16.5, the sensitivity was 58.7%, the specificity was 77%, and the correct classification rate was 75%. For threshold 19.5, the sensitivity was 80.4%, the specificity was 57.5%, and the correct classification rate was 61.8%.

A multicenter study[16] in Canada recruited adults with no history of asthma, COPD, or lung disease through random telephone calls and asked them if they had difficulty breathing or experienced coughing, coughing up phlegm, or wheezing in the past six months. Participants who answered yes completed COPD-DQ and COPD-related assessment tests. Patients with a COPD-DQ score of 20 or less were assessed for lung function before and after administering a diastolic agent to diagnose COPD. A total of 12 117 individuals were contacted at home and assessed for study eligibility. Of the 1260 selected patients, 910 (72%) were enrolled and underwent spirometry. Obstructive ventilation dysfunction was detected in 184 (20%) subjects, and 111 subjects were eventually diagnosed with COPD. The study confirmed undiagnosed airflow obstruction in 20% of a randomly selected group of people in Canada who reported respiratory symptoms. Questionnaires could exclude subjects at low risk but could not accurately identify subjects with undiagnosed disease.

Zhou et al.[17] from the Guangzhou Medical University developed a COPD screening questionnaire (COPD-SQ) suitable for Chinese people based on several risk factors related to epidemiological investigations in China. The questionnaire included seven items: Age, smoking index, BMI, cough, shortness of breath, family history of respiratory diseases, and biofuel exposure. Biofuel exposure is a factor that is more characteristic in China, particularly in rural areas. The questionnaire threshold was 16 points, the sensitivity was 60.6%, the specificity was 85.2%, and the correct classification rate was 82.7%.

The Chinese COPD Tiered Diagnosis and Treatment Project[18], led by the Chinese National Clinical Research Center for Respiratory Diseases and China-Japan Friendship Hospital, focuses on the promotion of tiered COPD diagnosis and treatment. With the breakthrough of strengthening the capacity building of grassroots diagnosis and treatment of COPD, it explores the mode and path of grassroots diagnosis and treatment suitable for Chinese national conditions. The screening mechanism for populations at a high COPD risk was established, and a tiered diagnosis and treatment data management system for COPD was developed. The COPD-SQ was used as a screening tool for patients with COPD, and pulmonary function examination was conducted for subjects with COPD-SQ scores of 16 points or more. Standard COPD management was conducted after diagnosis and long-term follow-up. The project is still underway. The process framework, information acquisition structure, and data management system designed by this project explored a new way to standardize the management of COPD.

2.2 Medical and health informatization

The Chinese Resident Electronic Health Record Management System is an innovative application of medical and health informatization. The system collects information from primary healthcare institutions, including basic personal information of archived residents, examination information related to cardiovascular and cerebrovascular diseases, diabetes, and other chronic diseases, and also electronic prescription information[19]. Although the information related to COPD risk factors such as smoking is collected, no specific disease record exists for patients with COPD because COPD was excluded from the national basic public health service[20].

3 Framework of the Screening and Diagnostic Support System of Populations at a High COPD Risk

3.1 Overview

This study aimed to develop a convenient and effective CDSS to assist in identifying patients with COPD in primary medical institutions. This should couple multiple types of healthcare data to provide physicians with high-quality evidence for COPD diagnosis and reduce physicians' workload. The application of this CDSS will improve the ability of early identification and diagnosis of COPD in high-risk groups in primary medical institutions and further reduce related costs. It also explores how artificial intelligence and big data technology can be applied in the medical and health fields.

Therefore, we proposed a case-finding CDSS based on public health data to identify patients with COPD (Fig. 1). Our system primarily includes screening and risk assessment systems for COPD populations.

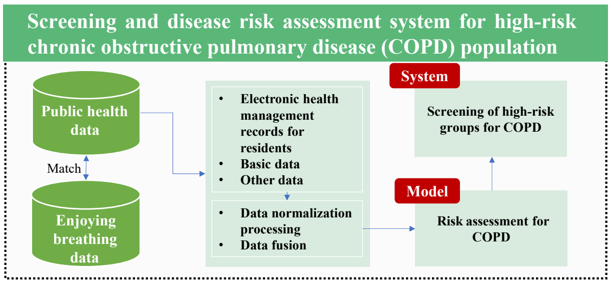

Screening and disease risk assessment system for high-risk chronic obstructive pulmonary disease (COPD) population
Public health data
Match
Enjoying breathing data
Electronic health management records for residents
Basic data
Other data
Data normalization processing
Data fusion
System
Screening of high-risk groups for COPD
Model
Risk assessment for COPD

Fig. 1 Overall framework of the screening and risk assessment system in the high-risk COPD population.

The population screening and risk assessment system processes public health data to generate risk assessments for residents with COPD. Based on the results, medical personnel may decide whether to conduct further pulmonary measurements and other tests related to diagnosing COPD.

People who are not at a high risk of COPD do not need to undergo further spirometry examination and other related tests, and this reduces the workload of many doctors. This approach allows medical personnel to identify patients with COPD more accurately, providing more time for patients who truly require their attention. In our system, every public health dataset is preprocessed, feature-extracted, and converted into a set of features, and then the results are obtained using a classification algorithm. The findings of this study can provide a reference for the respiratory department and PCPs to understand the risks and conditions of patients with COPD. The CDSS runs on cloud servers and can be integrated into the existing information systems of primary healthcare institutions. The existing information system works with several community clinics, remote rural health service stations, and other units to share medical resources through the internet. This mode of cooperation has substantially improved the quality of medical services and clinical efficiency (Figs. 1 and 2).

3.2 COPD population screening and risk assessment system

The screening and disease risk assessment system in the population at a high COPD risk proposed in this study is a low-cost, large-scale COPD risk assessment model based on public health data of Chinese residents and information from public health archives.

3.2.1 Identifying the extraction entries for public health data

We divided the COPD-SQ questionnaire entries into the following characteristics: Age, smoking, BMI, chronic cough, shortness of breath, biofuel use, and family history. Age, smoking, BMI, biofuel use, and family history were considered structured data. Chronic cough and shortness of breath were considered unstructured data.

3.2.2 Developing strategies to extract public health data

We performed a detailed analysis of the databases of the resident electronic health records and electronic medical records in the corresponding public health data analysis document and determined the corresponding entries of the COPD-SQ decomposition characteristic variables in the public health database.

The resident electronic health record database is a structured code database, while the resident outpatient medical record database is an unstructured electronic medical record database. According to the category form of the item data, we finally determined the extraction item and extraction strategy of the public health data. The automatic extraction of structured data was based on the residents' electronic health records. Natural language processing was used to extract unstructured data from the electronic medical record database.

Chronic cough: In the medical history, if the patient has chronic coughing or cough, wheezing, or

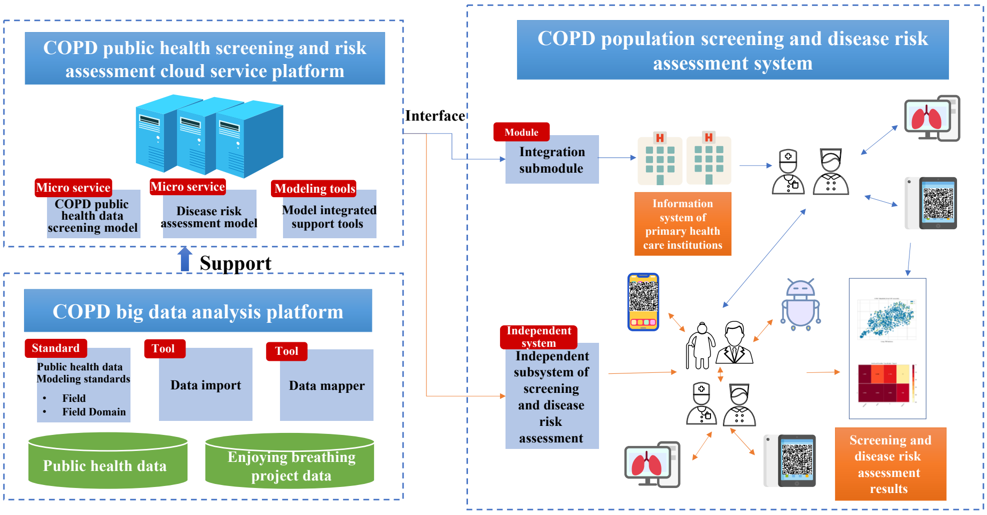

COPD public health screening and risk assessment cloud service platform
Micro service
COPD public health data screening model
Micro service
Disease risk assessment model
Modeling tools
Model integrated support tools
Interface
Support
COPD big data analysis platform
Standard
Public health data Modeling standards
Field
Field Domain
Tool
Data import
Tool
Data mapper
Public health data
Enjoying breathing project data
COPD population screening and disease risk assessment system
Module
Integration submodule
H
H
Information system of primary health care institutions
Independent system
Independent subsystem of screening and disease risk assessment
Screening and disease risk assessment results

Fig. 2 Deployment and application of the screening and risk assessment system architecture diagram in high-risk COPD population. The blue arrow indicates the integration submodule path, and the orange arrow indicates the independent subsystem path.

expectoration that is chronic or recurrent or lasts longer than 6 months and if the patient complained of chronic bronchitis, COPD, and pulmonary emphysema in the history of present illness and diagnosis, the cough score was five points. If the patient did not meet the previously mentioned criteria, the score was zero point.

Shortness of breath: In the medical history, if the patients had shortness of breath, wheezing, or other Chinese expressions for shortness of breath that is chronic or recurrent or lasted longer than six months, the shortness of breath score was three points. Those who did not meet the previously mentioned conditions received a score of zero point. Because the current electronic medical record information is insufficient to meet the more detailed classification of the severity of shortness of breath, we assigned a unified value of three points for all patients with shortness of breath and did not assign a value for the six points corresponding to severe shortness of breath in COPD-SQ. The details are shown in Table 1.

3.2.3 Proposed public health screening model

(1) Data preprocessing

Using the previously mentioned data extraction strategy, 44 input features and a label (according to the score of COPD-SQ, a score of less than 16 was assigned to the nonhigh-risk group, and the corresponding label was zero; a score of 16 or more was assigned to the high-risk group, and the corresponding label was one) were extracted from the public health data and data of Enjoying Breathing Program by unique field association matching. After analyzing all the features and deleting the fields with a missing rate greater than 90% and those highly correlated with the scoring (the scoring characteristics in Table 1), 28 input features and a label were selected, and a detailed description can be found in Table S11 in the electronic supplementary material.

Through the manual review, 1875 of 1885 patients met the inclusion criteria, and the remaining 10 patients were excluded based on the exclusion criteria. The high-risk group (238 patients) and the nonhigh-risk group (1637 patients) were screened according to COPD-SQ. To obtain the best performance model, the dataset was divided by setting the random division seed (random state) to 786, and the details of the division are shown in Table 2.

First, we divided all the data into internal and external datasets at a ratio of 9.5 : 0.5. The internal dataset was divided into training and test sets at a ratio of 7 : 3, which was mainly used for model training, model optimization, and model verification. The external dataset was mainly

Table 1 COPD-SQ scoring and data extraction strategy.

Score item	Group	Score	Data extraction method
Age (years)	40–49	0	Automatic extraction based on residents' electronic health records
	50–59	4	
	60–69	8	
	$\geqslant$ 70	11	
Smoking exposure (pack-years)	Never	0	Automatic extraction based on residents' electronic health records
	1–14	2	
	15–30	4	
	$\geqslant$ 30	5	
BMI (kg/m^2)	< 18.5	7	Automatic extraction based on residents' electronic health records
	18.5–23.9	4	
	24–27.9	1	
	$\geqslant$ 28	0	
Chronic cough	Yes	0	Natural language processing of electronic medical record information
	No	2	
Shortness of breath	None	0	Automatic extraction based on residents' electronic health records
	Shortness of breath when walking fast on flat ground or climbing a small hill	3	
	Shortness of breath when walking normally on the ground	6	
Biofuel use	Yes	1	Automatic extraction based on residents' electronic health records
	No	0	
Family history of chronic bronchitis/emphysema/COPD	Yes	3	Natural language processing of electronic medical record information
	No	0	

Table 2 Samples used in this study.

Sample group	Group	Subgroup	Number	Total
Internal dataset	Training	High-risk	150	1246
		Nonhigh-risk	1096	
	Testing	High-risk	77	535
		Nonhigh-risk	458	
External dataset	Training	High-risk	11	94
		Nonhigh-risk	83	

used for external verification of the model and did not participate in the construction of the final model. Before model construction and training, according to the data situation and preprocessing algorithm, a series of preprocessing operations were performed to obtain the best recognition performance. Missing values in categorical features were imputed with a constant "not available" value of the feature in the training dataset.

Categorical features were converted using the one-hot encoding method. Notably, in the categorical features, some features included multiple options. For example, family history characteristics included 12 options, which correspond to one type of disease history. If someone had more than one disease history at the same time, multiple choices were available, such as "2, 3", "2, 4", and "2, 4, 7". For this type of data, we do not split or reencode and consider that the combination is also a category. For continuous features, we used the Z-score and Yeo Johnson methods to transform the data. First, we normalized each feature with a mean value of 0 and a standard deviation of 1 and then mapped the data from a nondistribution to a more Gaussian-like distribution. The transformed data can balance the influence of different feature scales to the greatest extent and make the algorithm achieve the best optimization effect and convergence speed to avoid falling into the local solution to a certain extent.

Finally, these methods were used to screen the features after coding and transformation.

(a) Removal of outliers. This step was performed at the beginning of data preprocessing, using singular value decomposition to remove outliers in the training data.

(b) Ignoring the low variance. All the categorical features with insignificant variances were removed from the data.

(c) Removal of multicollinearity features. Features with intercorrelations higher than the defined threshold were removed. When two features were highly correlated with each other, the featureless correlated with the target variable was removed.

(d) Removal of completely collinear features. Perfect collinearity (features with correlation = 1) was removed from the dataset; when two features were 100% correlated, one was randomly removed from the dataset.

(2) Modeling methods and steps

The dataset in this study is structured data. For structured data, traditional machine learning models are usually used for modeling instead of deep learning models. We used Python-related libraries to select 18 machine learning models in a targeted manner and tried to identify the best performance model that meets the expectations of this study through optimization and comparison. The models mainly included support vector machine (SVM) models—linear kernel classifier; radial basis function (RBF) kernel classifier, simple tree models—decision tree (DT) classifier; random forest (RF) classifier; extreme trees (ET) classifier, regression classification models—logistic regression (LR) classifier; Ridge classifier, discriminant analysis classification models—linear discriminant analysis (LDA); quadratic discriminant analysis (QDA), gradient boosting decision tree (GBDT) models—AdaBoost classifier; gradient boosting classifier (GBC); extreme gradient boosting (XGBoost); light gradient boosting machine (LightGBM); CatBoost classifier, and other models—multi-layer perceptron (MLP) classifier; K-nearest neighbor (KNN) classifier; Gaussian process classifier (GPC); naive Bayes classifier (NB). Simultaneously, based on a single model and the top five models of comprehensive performance, four types of ensemble learning methods (bagging, boosting, blending, and stacking) were used for modeling. After the models were selected, a strictly planned process was used to train and optimize the model to obtain the best performance model. The specific process is as follows.

(a) For model preexperimentation, all 18 models were included for model training using training samples from the internal dataset. Considering that the size of the dataset was small and to avoid modeling bias caused by randomness, 10-fold cross-validation was used to evaluate the performance of each model[21]. The final performance of the model is the average of each evaluation metric of the 10-fold cross-validation. The models were trained using the default hyperparameters and finally output the results ranked according to accuracy from the highest to the lowest.

(b) To avoid wasted resources and time, the top five models evaluated in (1) for individual optimization were selected. The random grid search method was used to optimize the hyperparameters, and the optimal parameter optimization model was selected through cross-validation. The hyperparameters were tuned with the goal of accuracy, and the optimized model was tested using the test samples in the internal dataset to check the generalization ability of the model.

(c) Because of the small scale of the dataset in this study and in order to avoid data waste and improve the generalization ability of the model, the hyperparameters optimized by the top five models were used to train all the data in the internal dataset to obtain the final model. Next, the external dataset was used for the final test to obtain the test result. This result largely represented the final generalization ability of the model.

(d) Ensemble learning modeling was divided into two parts. First, the final model with the best test performance was modeled using bagging and boosting methods. The default hyperparameters, including the maximum number of features, were set to one, the maximum number of samples was set to one, and the number of base models was set to 10. In boosting, stagewise additive modeling with a multi-class exponential cost function (SAMME.R, R stands for real)[22] was used as the optimization algorithm. Unlike the SAMME algorithm, the algorithm uses weighted probability estimates to update the model and output classification probability values. In the second step, the blending and stacking integrated learning methods were used for modeling. The first two integrated learning methods used a single base model for the integrated modeling, while the blending and stacking methods used different final models in (c) as the base model for integration. The training process of the four types of ensemble learning methods was the same as that for (a) and (c). To prevent overfitting, no parameter tuning was performed. The integrated learning model was finally tested using the external datasets to compare different evaluation indicators.

Baseline model testing is a common method for new data modeling in machine learning. At the initial stage of the model, experimenting with a baseline model often reduces time by 90% and provides 90% accurate results. In our modeling process, using LR as a baseline model for rapid training and optimization helped us understand the task and the data[23].

We also explored feature selection, feature importance analysis, and model interpretability analysis. Regarding feature selection, we did not reselect the 28 features but selected the 104 features after coding using the four methods mentioned in (1) and finally retained 49 important features. The GBDT algorithm was used to analyze the feature importance and could directly provide the importance score of each attribute. We used the weight instead of gain or cover to calculate the feature importance score. In the GBDT feature importance score algorithm, the larger the improved performance of a feature to the splitting point (the closer to the root node), the larger the weight. The more the ascension trees were selected, the more important the attribute was. The performance measure could be the Gini purity of the split node selected or other metric functions. Finally, the weighted sum of the results of an attribute in all the boosting trees was averaged to obtain the importance score. Model interpretability analysis using Lundberg and Lee in 2016 displayed the SHAP (Shapley Additive exPlanations, SHAP) method[24], which can explain individual forecast methods. SHAP is based on the best Shapley value in game theory.

(3) Evaluation metrics

To select the best model for screening people at a high COPD risk, we evaluated the model using seven widely used evaluation metrics—accuracy, area under the receiver operating characteristic curve (AUC), sensitivity, precision, F1, kappa, and MCC (Matthew's correlation coefficient). The seven criteria are calculated based on the following:

$$\text{Accuracy} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{TN} + \text{FP} + \text{FN}} \tag{1}$$

$$\text{AUC} = \int_{x=0}^{1} \text{TPR}(\text{FPR}^{-1}(x))\mathrm{d}x \tag{2}$$

$$\text{Sensitivity} = \frac{\text{TP}}{\text{TP} + \text{FN}} \tag{3}$$

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}} \tag{4}$$

$$\text{F1} = 2\frac{\text{Precision} \cdot \text{Recall}}{\text{Precision} + \text{Recall}} \tag{5}$$

$$\text{Kappa} = \frac{p_0 - p_e}{1 - p_e} \tag{6}$$

$$\text{MCC} = \frac{\text{TP}\cdot\text{TN} - \text{FP}\cdot\text{FN}}{\sqrt{(\text{TP}+\text{TN})(\text{TP}+\text{FN})(\text{TN}+\text{FP})(\text{TN}+\text{FN})}} \tag{7}$$

where TP (FN) is the number of positive samples predicted to be positive (negative), and TN (FP) is the number of negative samples classified as negative (positive); the true positive rate (TPR) is equal to the previously defined recall and FPR (false-positive rate =

FP/(FP + TN)); p_0 is the sum of the number of samples from each of the correct categories divided by the total sample, representing accuracy. p_e is the sum of the product of the actual and predicted numbers for each category divided by the square of the total number of samples.

Accuracy indicates the predicted result of the percentage of total samples. AUC represents the relationship curve between TPR and FPR and the area of the curve. In medicine, AUC can comprehensively evaluate sensitivity and specificity. The sensitivity indicates the ability of the model to correctly predict the population at a high COPD risk.

(4) Results

The accuracy, AUC, sensitivity, precision, F1, kappa, and MCC of the 18 machine learning models in the training dataset are shown in Table 3. The models are ranked from high to low according to accuracy. The AUC for estimators (linear SVM and ridge) that do not support "predict proba (probability prediction)" is 0.0000. According to the results in Table 3, the overall performance of the GBDT model was better than that of the baseline model (LR) and other models, and CatBoost provided the highest accuracy of 98.96%. Although AUC is a key measure of sensitivity and specificity performance, XGBoost showed the highest AUC of 99.84%. NB showed the highest sensitivity of 95.33%.

We comprehensively considered each assessment indicator in Table 3 and screened out CatBoost, LightGBM, XGBoost, GBC, AdaBoost, and LR as candidate optimization models and the baseline model. During the optimization, the accuracy was considered the main optimization objective to maximize the screening performance of COPD-SQ, and AUC was considered the secondary optimization objective to optimize the sensitivity and specificity of the model to reduce the rate of missed diagnosis and misdiagnosis. The highest values for each indicator in the single model and ensemble model are shown in bold in Table 3. By optimization, different models showed different degrees of metrics on the test dataset. CatBoost was still the best model, with an accuracy of 99.25%. AUC had an accuracy of 99.85%, a sensitivity of 94.81%, and a precision of 1. Compared with the single and ensemble models, the comprehensive evaluation results of the stacking model were close to those of CatBoost. The external dataset test results presented in Table 4 and Fig. 3 showed that the overall generalization ability of the ensemble models was better than that of the single models.

Figure 4 shows the ranking of feature importance of a single model. After preprocessing methods such as coding and transformation, the 28 features were expanded to 104 features, and 49 features were obtained using four feature selection methods in the preprocessing. The number following the underscore in the feature name represents the category to which the category feature belongs. The features considered important by the machine learning model included cough, age, smoking status, BMI, smoking quantity, and waist-to-hip ratio

Table 3 Performance of 18 models on training samples.

Model	Accuracy	AUC	Sensitivity	Precision	F1	Kappa	MCC
CatBoost	**0.9896**	0.9980	**0.9467**	**0.9683**	**0.9560**	**0.9501**	**0.9511**
LightGBM	0.9864	0.9976	0.9200	0.9670	0.9415	0.9338	0.9351
XGBoost	0.9856	**0.9984**	0.9133	0.9661	0.9376	0.9295	0.9308
GBC	0.9855	0.9959	0.9067	**0.9719**	0.9374	0.9293	0.9304
AdaBoost	0.9839	0.9968	0.9000	0.9662	0.9307	0.9216	0.9231
DT	0.9719	0.938	0.8933	0.8841	0.8863	0.8704	0.8719
RF	0.9599	0.9846	0.7467	0.9134	0.8174	0.7952	0.8027
LR	0.9551	0.9865	0.7533	0.8631	0.7981	0.7732	0.7792
MLP	0.9542	0.9844	0.7733	0.8411	0.7998	0.7742	0.7787
Linear SVM	0.9503	0.0000	0.7533	0.8329	0.7815	0.7539	0.7608
LDA	0.9422	0.9785	0.7267	0.7865	0.7467	0.7146	0.7205
ET	0.9390	0.9688	0.6533	0.8133	0.7200	0.6864	0.6941
RBF SVM	0.9374	0.9841	0.5267	0.9294	0.6654	0.6343	0.6690
Ridge	0.9253	0.0000	0.4667	0.8723	0.5958	0.5596	0.5983
KNN	0.9189	0.9089	0.4533	0.7698	0.5579	0.5189	0.5465
GPC	0.9165	0.8978	0.3533	0.9038	0.4926	0.4586	0.5237
QDA	0.8169	0.5391	0.1733	0.2509	0.1704	0.0886	0.1022
NB	0.3283	0.6503	**0.9533**	0.1474	0.2552	0.0589	0.1526

Table 4 Performance of optimized models on test samples.

Modle type	Model	Accuracy	AUC	Sensitivity	Precision	F1	Kappa	MCC
Single model	CatBoost	**0.9925**	**0.9985**	**0.9481**	**1.0000**	**0.9733**	**0.9690**	**0.9695**
	LightGBM	0.9832	0.9963	0.8831	**1.0000**	0.9379	0.9282	0.9306
	XGBoost	0.9776	0.9929	0.8831	0.9577	0.9189	0.9059	0.9069
	GBC	0.9720	0.9961	0.8182	0.9844	0.8936	0.8776	0.8826
	AdaBoost	0.9832	0.9955	0.9091	0.9722	0.9396	0.9298	0.9305
	Baseline	0.9346	0.9661	0.6883	0.8281	0.7518	0.7145	0.7185
Ensemble model	Bagging	0.9888	0.9975	0.9221	**1.0000**	0.9595	0.9907	0.9530
	Boosting	0.9540	**0.9976**	0.9351	**1.0000**	0.9664	0.9610	0.9618
	Blending	0.9869	0.9967	0.9091	**1.0000**	0.9524	0.9448	0.9463
	Stacking	**0.9925**	0.9923	**0.9481**	**1.0000**	**0.9733**	**0.9690**	**0.9695**

(Fig. 4). Besides these features, other features accounted for a relatively small percentage of the importance rankings (less than 10%).

The model interpretability analysis is shown in Fig. 5. Figure 5a shows the influence of the characteristic value of the best single model CatBoost on prediction. The results of machine learning were consistent with those of the clinical demonstration analysis, such as in age; the higher the value, the greater the contribution to the prediction of high-risk populations. For Cough 0, the greater the value, the greater the contribution to the prediction of the nonhigh-risk population. Figure 5b shows the results of individual case analyses. We extracted 60 individual cases for analysis, and the contribution of cough and age (here, it is the converted age; the actual age is 72 years) to the predicted high-risk group was enhanced.

3.2.4 System implementation

A screening and risk assessment system for COPD was established based on the following: The system retrieves public health data from the server, conducts a risk assessment of COPD among residents, and generates a public health data COPD-SQ assessment result report. The medical staff can use the system to perform an extensive COPD screening in an efficient and convenient manner and decide whether to conduct further pulmonary function tests on residents according to the results of the COPD risk assessment given by the system.

In the pie chart of Fig. 6, red is the probability of high risk, and green is the probability of nonhigh risk. According to the threshold setting, the model determined that a certain probability value higher than 50% is a category.

4 System Validation and Results

Emeishan village, South Dulehe Town, Pinggu District, Beijing, China, was one of the first places where a public health database for residents was established, and a complete electronic health record and outpatient electronic medical record database were established. We conducted a small-scale study in this area.

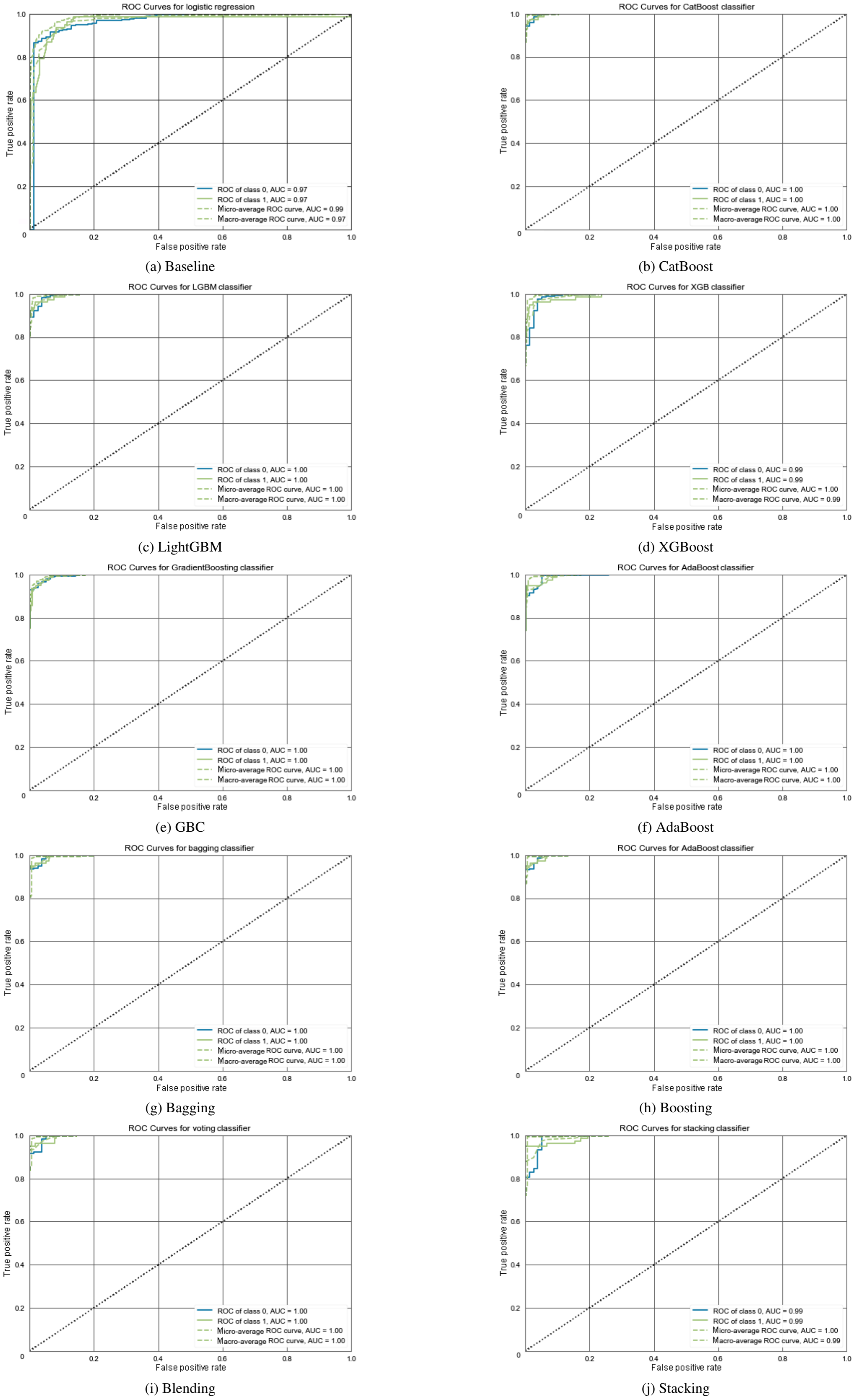
Approximately 2800 residents lived in this area, of which 1875 had electronic health records and were selected as the research objects. We used the system to screen the population at a high COPD risk in the experimental area (n = 370).

Statistical analysis. The data were expressed as means ±SD or numbers (%), as appropriate. Continuous variables were compared using a t-test or one-way analysis of variance, followed by Bonferroni's test for pairwise comparisons. Categorical data were evaluated using a chi-squared test. The accuracy of using different modalities to screen COPD could be determined using ROC curve analysis. The optimal cutoff of the selected modality was calculated using the Youden index to determine the sensitivity, specificity, positive predictive value, and negative predictive value. The characteristics of the baseline data are shown in Tables 3 and 4.

We made targeted invitations to high-risk groups. Finally, 116 people from 79 high-risk groups and 37 random groups participated in the field screening. After signing the informed consent form, the 116 visiting residents were screened using the on-site COPD-SQ questionnaire and subjected to spirometry examinations assisted by the quality control system.

Spirometry examination identified 31 patients with obstructive ventilation dysfunction (25 in the high-risk group and 6 in the random group). We summarized the relevant data of recent similar studies in Table 5.

To test the efficiency of our system, we performed an ROC curve analysis based on public health data and COPD-SQ field investigation screening results and determined whether the spirometer detected obstructive



(a) Baseline

(b) CatBoost

(c) LightGBM

(d) XGBoost

(e) GBC

(f) AdaBoost

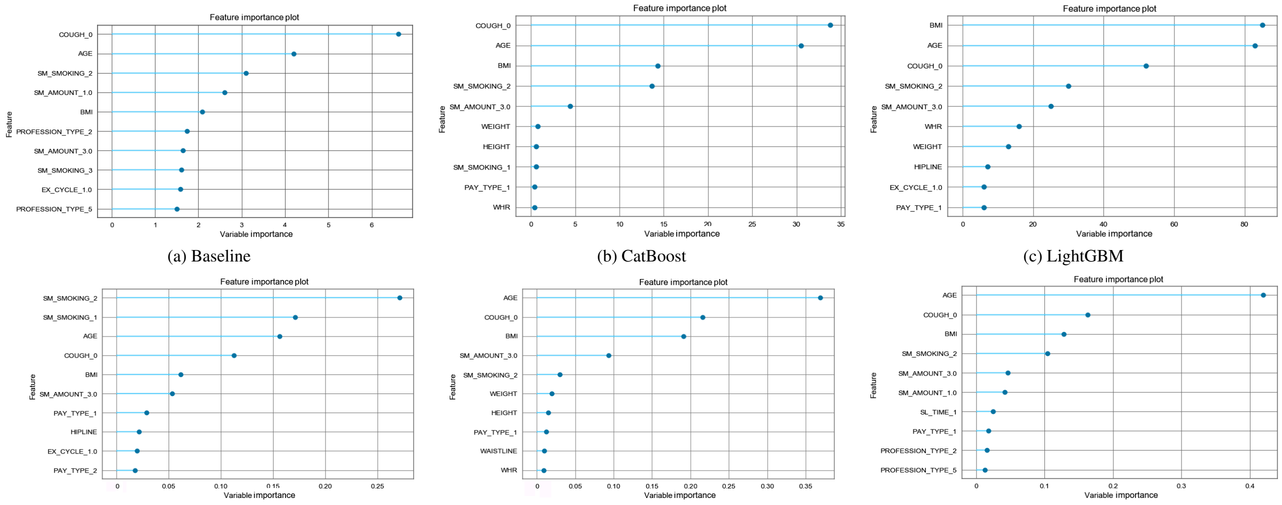
(g) Bagging

(h) Boosting

(i) Blending

(j) Stacking

Fig. 3 Comparison of model performance for detecting populations at a high COPD risk. ROC: Receiver operating characteristic curve; AUC: Area under the receiver operating characteristic curve; micro-average: Compute the metric independently for each class and then take the average (hence treating all classes equally); macro-average: Aggregate the contributions of all classes to compute the average metric.



(a) Baseline

(b) CatBoost

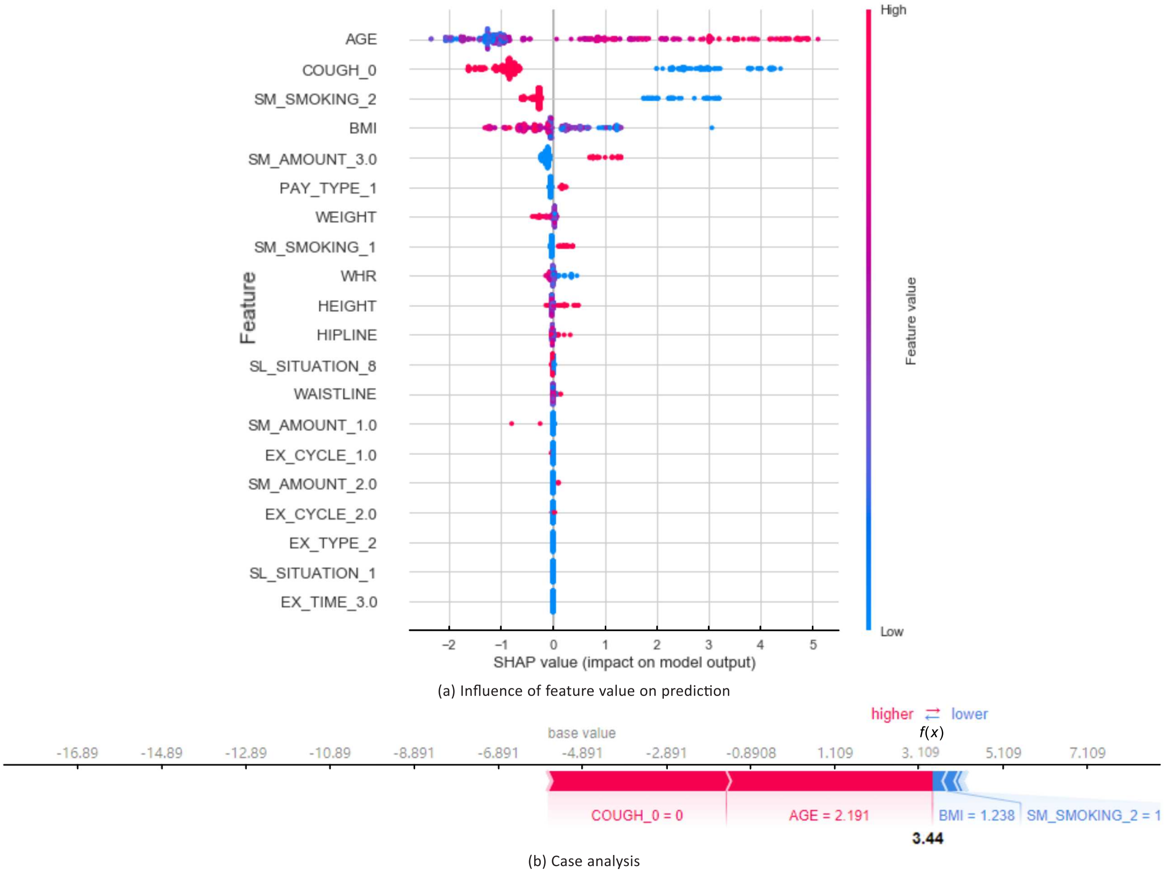
(c) LightGBM

(d) XGBoost

(e) GBC

(f) AdaBoost

Fig. 4 Ranking of feature importance of a single model. The model included 49 features; here, the top ten features were considered.



(a) Influence of feature value on prediction

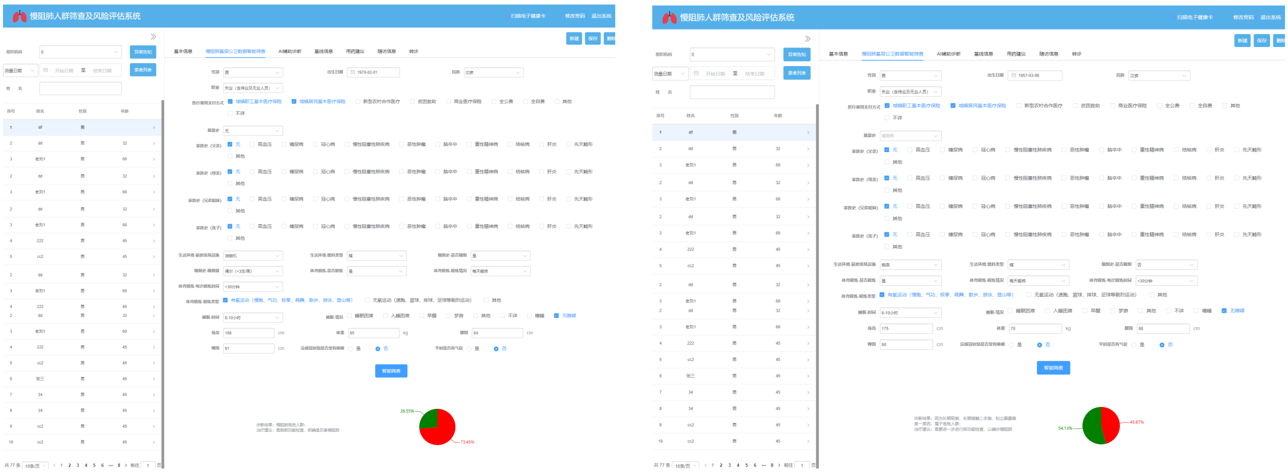
(b) Case analysis

Fig. 5 Model interpretability analysis. Each point in Fig. 5a represents an instance, and the color represents the size of the value of the feature. The horizontal axis is the value of SHAP, which considers 0 as the origin and has a positive (high-risk group of COPD) and a negative (nonhigh-risk group of COPD) impact on the judgment. Figure 5b shows the analysis of individual cases of feature interpretability, and 60 individual cases were selected for analysis.

ventilation dysfunction. The AUC of public health data screening was 0.615 (95% conference interval (CI): 0.504–0.725). The AUC of COPD-SQ field investigation screening was 0.627 (95% CI: 0.511–0.504).

5 Discussion

Presently, in all known relevant studies on the screening of COPD, the information was collected and input



(a) Influence of feature value on prediction

(b) Case analysis

Fig. 6 Interface of screening and risk assessment system for COPD patients. Figure 6a shows the first case. The patient was 50 years old and a light smoker. Diagnosis result: The patient belongs to the high-risk group of COPD. Treatment suggestion: Spirometer examination should be performed to determine whether the patient has COPD. Figure 6b shows the second condition. The patient was 63 years old and did not smoke, and the fuel type was coal. Diagnosis result: The patient belongs to the high-risk group because of long-term smoking, long-term exposure to secondhand smoke, dust exposure, and other reasons. Treatment suggestion: Spirometer examination should be performed to determine whether the patient has COPD.

Table 5 Performance of final models on the external dataset.

Modle type	Model	Accuracy	AUC	Sensitivity	Precision	F1	Kappa	MCC
Single model	CatBoost	0.9787	0.9091	0.8182	1.0000	0.9000	0.8882	0.8938
	LightGBM	0.9787	0.9091	0.8182	1.0000	0.9000	0.8882	0.8938
	XGBoost	0.9681	0.8636	0.7273	1.0000	0.8421	0.8248	0.8378
	GBC	0.9787	0.9091	0.8182	1.0000	0.9000	0.8882	0.8938
	AdaBoost	0.9787	0.9091	0.8182	1.0000	0.9000	0.8882	0.8938
	Baseline	0.9468	0.8516	0.7273	0.8000	0.7619	0.732	0.7331
Ensemble model	Bagging	0.9787	0.9091	0.8182	1.0000	0.9000	0.9787	0.8882
	boosting	0.8938	0.9091	0.8182	1.0000	0.9000	0.8882	0.8938
	Blending	0.9787	0.9091	0.8182	1.0000	0.9000	0.8882	0.8938
	Stacking	0.9787	0.9091	0.8182	1.0000	0.9000	0.8882	0.8938

manually. Because of the heavy workload and high cost, it is unsuitable for grassroots medical staff with heavy daily work to perform the daily prevention and treatment of COPD. Additionally, the questionnaire for COPD screening has a high requirement concerning the investigation skills of investigators, and short-term training cannot meet the needs of a large range of COPD screening using questionnaires. Developing an efficient and accurate screening method for the population at a high COPD riskimproves the efficiency and effect of primary medical staff in performing COPD screening[25,26].

Therefore, we developed CDSS based on COPD-SQ. All the medical data of the population were extracted from the residents' electronic health records, and the disease screening model processed the data and fed back the screening results.

In the screening model, the 10-fold cross-validation results of 18 machine learning models on the training dataset showed that the GBDT model has the best overall performance, and CatBoost had the highest accuracy (98.96%), which is close to the detection efficiency of COPD-SQ. Although the AUC is a key measure of sensitivity and specificity performance, the highest AUC of XGBoost is 0.9984, which completely avoids missed and error screenings. NB has the highest sensitivity (95.33%), butits other metrics are too low to be selected as a candidate optimization model. However, the characteristics of the NB model revealed that some metrics in the data substantially contributed to screening high-risk groups. We comprehensively considered each evaluation metric and screened CatBoost, LightGBM, XGBoost, GBC, AdaBoost, and LR as candidate optimization models and baseline models.

We further tested the previously mentioned models on the test and external datasets. According to the comprehensive performance ranking of various evaluation metrics in Table 3, we selected five models (CatBoost, LightGBM, XGBoost, GBC, and AdaBoost) as the final optimization objects. During the optimization, the accuracy was considered the main optimization objective to maximize the screening performance of COPD-SQ, and AUC was considered the secondary optimization objective to optimize the sensitivity and specificity of the model to reduce the rate of missed diagnosis and misdiagnosis. Through the optimization, various metrics of different models on the test dataset have been improved. CatBoost remains the best model for comprehensive evaluations, with an accuracy of 99.25%, nearly 6% higher than the baseline model after optimization, an AUC of 99.85%, and a sensitivity of 94.81%. The precision of CatBoost at this time is one, indicating that the nonhigh-risk population is not diagnosed as a high-risk population, and the misdiagnosis rate is zero. Compared with the single models and ensemble learning models, the comprehensive evaluation result of the stacking model is close to CatBoost. The external dataset test results show that the overall generalization ability of the ensemble learning model is better than that of the single model.

The test results of several models were the same because of the small number of samples in the external dataset (94 patients) and unbalanced sampling of positive and negative samples. However, the verification results can reflect the generalization ability of the final model to a certain extent.

Simultaneously, we conducted a study on the importance of model features and found that the important features included cough, age, smoking status, BMI, smoking quantity, and the waist-to-hip ratio. Besides these features, the other features accounted for a relatively small percentage of the importance ranking (less than 10%). These results also indicate that the high-risk factors for COPD considered by the algorithm are consistent with the important indicators considered by experienced clinicians. Additionally, some features after the ranking of important features can be used as additional observation indicators for clinicians to study and verify the clinical correlation.

We used the final model as the kernel and developed a COPD screening CDSS based on public health data. In the external validation of the system, we analyzed the electronic health records of 1875 patients, of which 370 (19.73%) required further spirometry examination according to model screening. Of 370 patients, 79 (21.35%) underwent spirometry. Of 79 patients, 25 (31.64%) were found to have obstructive ventilation dysfunction. We compared the results with previous studies on COPD case findings. Compared with COPD-SQ, the efficiency of our systematic screening when detecting obstructive ventilation dysfunction was approximately at the same level, as shown in Fig. 7. In Table 6, the pooled results of several similar studies are shown.

The China Pulmonary Health Study adopted multistage stratified cluster random sampling to conduct field investigations on Chinese residents and collect information related to respiratory diseases. Obstructive ventilatory dysfunction was observed in 13.5% of Chinese individuals aged older than 20 years[2]. Large-scale epidemiological investigations only apply to the cross-sectional investigation of the disease. This method is unsuitable for daily COPD screening work. Therefore, scholars worldwide have conducted a series of case-finding studies using questionnaire screening. The undiagnosed COPD and asthma population study in Canada[16] identified a COPD risk population using a random telephone survey. A total of 12 117 individuals were surveyed by telephone. A total of 1260 (10.40%) were selected for spirometry examination, and 910 (72%) were registered and underwent spirometry. Obstructive ventilation dysfunction was detected in 184 subjects (20% of the study participants), and 111 patients were eventually diagnosed with COPD. The Chinese

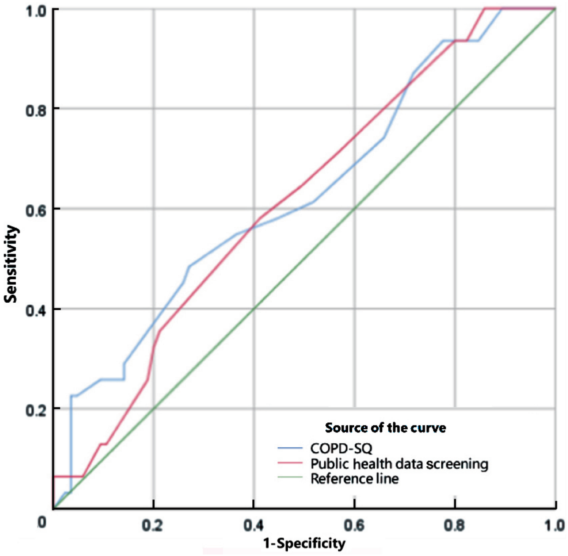


Fig. 7 Diagnostic accuracy according to ROC curve analysis.

Table 6 Summary of our research and related studies.

Study	Investigation method	Sample size	Number of patients requiring spirometry examination (rate)	Actual number of spirometry examinations performed (rate)	Number of patients with obstructive ventilation dysfunction (rate)
The China Pulmonary Health study[2]	Field investigation	50 991	50 991 (100%)	50 991 (100%)	6866 (13.47%)
Undiagnosed COPD and asthma population study[16]	Telephone interview	12 117	1260 (10.40%)	910 (72.22%)	184 (20.22%)
Chinese COPD Tiered Diagnosis and Treatment Project[18]	Field investigation	1 008 518	191 498 (18.99%)	63 523 (33.17%)	20 700 (31.59%)
Our study	Big data screening	1875	370 (19.73%)	79 (21.35%)	25 (31.64%)

COPD Tiered Diagnosis and Treatment Project[1, 27] is based on community units. PCPs from primary medical institutions performed an on-site COPD-SQ questionnaire for residents aged older than 40 years in the communities. The project is still in progress. By November 2019, 1 008 518 people were screened using questionnaires, of whom 191 498 (18.99%) had COPD-SQ scores of 16 and 63 523 (33.17%) underwent spirometry examinations. Of 63 523 people, 20 700 (31.59%) were diagnosed with obstructive ventilation dysfunction.

Comparing the results of the previously mentioned studies, we found that our study is more efficient. The screening of high-risk groups accurately identified the population that required further examination and avoided the waste of medical resources. Because of the high screening accuracy, the new model of big data screening significantly reduces the cost of screening. Additionally, our screening model significantly reduces unnecessary crowding because the initial screening effort is performed by CDSS analysis of available public health data. This finding is valuable in the context of the current COVID-19 epidemic.

Our study has limitations. We did not perform bronchodilation tests or other tests to confirm COPD because of the limitations of the conditions. Although using prebronchodilator values has been shown to overestimate the prevalence of COPD[27, 28], no difference was found in the diagnostic accuracy for COPD between them. Hoesein showed that bronchodilation tests have little value in diagnosing COPD in older symptomatic populations[29]. Comparing the detection rate of obstructive ventilation dysfunction in spirometry examinations in similar studies indicates the effectiveness of this protocol in identifying patients with COPD to a certain extent. Another limitation is the small sample size, which may limit the generalization of the proposed method. Therefore, whether our findings apply to other populations remains to be determined. Our scheme has completed the preliminary model verification and system experiment. The next step concerns the pilot area data of the Chinese COPD Tiered Diagnosis and Treatment Project to verify the model and effectiveness of the system in a larger scope.

6 Conclusion

Strengthening the capacity to identify patients with COPD at primary medical institutions is critical to reducing the prevalence and burden of the disease. To improve the performance of PCPs, the screening, diagnosis, and standardized management of COPD must be covered. How to effectively identify populations at a high COPD risk and standardize spirometry examinations remain to be further solved. The application of innovative technologies such as big data analysis, machine learning, and artificial intelligence-assisted decision-making can effectively improve the early identification and diagnosis of populations at a high COPD risk at primary medical institutions, providing new solutions for standardized management of COPD and reducing the impact of COPD on residents' health.

Acknowledgment

This work was supported by the Major Research Program of the National Natural Science Foundation of China (No. 91843302).

Electronic Supplementary Material

Supplementary materials including

- the hyperparameters of the optimized classification models, and
- the characteristics of the study subjects are available in the online version of this article at https://doi. org/10.26599/TST.2022.9010010

References

[1] W. Li, T. Yang, and C. Wang, Current status and progress of prevention and treatment of chronic obstructive pulmonary disease in China, (in Chinese), *J. Chin. Res. Hosp.*, vol. 7, no. 5, pp. 78–84, 2020.

[2] C. Wang, J. Y. Xu, L. Yang, Y. J. Xu, X. Y. Zhang, C. X. Bai, J. Kang, P. X. Ran, H. H. Shen, F. Q. Wen, et al., Prevalence and risk factors of chronic obstructive pulmonary disease in China (the China Pulmonary Health [CPH] study): A national cross-sectional study, *Lancet*, vol. 391, no. 10131, pp. 1706–1717, 2018.

[3] S. Haroon, R. E. Jordan, D. A. Fitzmaurice, and P. Adab, Case finding for COPD in primary care: A qualitative study of the views of health professionals, *Int. J. Chron. Obstruct. Pulmon. Dis.*, vol. 10, pp. 1711–1718, 2015.

[4] National Health Commission of the People's Republic of China, Healthy China initiative (2019–2030), (in Chinese), http://www.nhc.gov.cn/guihuaxxs/s3585u/201907/e9275fb95d5b4295be8308415d4cd1b2.shtml, 2019.

[5] World Health Organization, Projections of mortality and causes of death, 2016 and 2060, http://www.who.int/healthinfo/global_burden_disease/projections/en/, 2020.

[6] A. Kaplan and M. Thomas, Screening for COPD: The gap between logic and evidence, *Eur. Respir. Rev.*, vol. 26, no. 143, p. 160113, 2017.

[7] National Health Commission of the People's Republic of China, Work plan for monitoring chronic diseases and nutrition in the Chinese population (trial), http://www.chinanutri.cn/tzgg_6537/tzgg_102/201412/t2014123_108847.html, 2014.

[8] D. Y. Zhang, Y. Gao, W. H. Jian, M. Yao, J. P. Zheng, and N. S. Zhong, Feasibility and suggestions on the promotion of pulmonary function test in primary health care institutions, (in Chinese), *Chin. Gen. Prac.*, vol. 23, no. 29, pp. 3638–3643, 2020.

[9] K. L. Hon, E. Leung, J. L. Tang, C. M. Chow, T. F. Leung, K. L. Cheung, and P. C. Ng, Premorbid factors and outcome associated with respiratory virus infections in a pediatric intensive care unit, *Pediatr. Pulmonol.*, vol. 43, no. 3, pp. 275–280, 2008.

[10] Y. Gao and J. P. Zheng, Develop standardized lung function training to help prevent and control chronic respiratory diseases, (in Chinese), *Chin. J. Pract. Intern. Med.*, vol. 39, no. 5, pp. 481–484, 2019.

[11] R. E. Jordan, P. Adab, A. Sitch, A. Enocson, D. Blissett, S. Jowett, J. Marsh, R. D. Riley, M. R. Miller, B. G. Cooper, et al., Targeted case finding for chronic obstructive pulmonary disease versus routine practice in primary care (TargetCOPD): A cluster-randomised controlled trial, *Lancet Respir. Med.*, vol. 4, no. 9, pp. 720–730, 2016.

[12] P. J. P. Poels, T. R. J. Schermer, R. P. Akkermans, A. Jacobs, M. Van Den Bogart-Jansen, B. J. A. M. Bottema, and C. Van Weel, General practitioners' needs for ongoing support for the interpretation of spirometry tests, *Eur. J. Gen. Pract.*, vol. 13, no. 1, pp. 16–19, 2007.

[13] A. A. Montgomery, T. Fahey, T. J. Peters, C. Macintosh, and D. J. Sharp, Evaluation of computer based clinical decision support system and risk chart for management of hypertension in primary care: Randomised controlled trial, *BMJ*, vol. 320, no. 7236, pp. 686–690, 2000.

[14] R. T. Sutton, D. Pincock, D. C. Baumgart, D. C. Sadowski, R. N. Fedorak, and K. I. Kroeker, An overview of clinical decision support systems: Benefits, risks, and strategies for success, *NPJ Digit. Med.*, vol. 3, no. 1, p. 17, 2020.

[15] D. B. Proce, D. G. Tinkelman, R. J. Nordyke, S. Isonaka, and R. J. Halbert, Scoring system and clinical application of COPD diagnostic questionnaires, *Chest*, vol. 129, no. 6, pp. 1531–1539, 2006.

[16] M. Preteroti, G. A. Whitmore, K. L. Vandemheen, J. M. Fitzgerald, C. Lemière, L. P. Boulet, E. Penz, S. K. Field, S. Gupta, R. A. Mcivor, et al., Population-based case-finding to identify subjects with undiagnosed asthma or COPD, *Eur. Respir. J.*, vol. 55, no. 6, p. 2000024, 2020.

[17] Y. M. Zhou, S. Y. Chen, J. Tian, J. Y. Cui, X. C. Li, W. Hong, Z. X. Zhao, G. P. Hu, F. He, R. Qiu, et al., Development and validation of a chronic obstructive pulmonary disease screening questionnaire in China, *Int. J. Tuberc. Lung Dis.*, vol. 17, no. 12, pp. 1645–1651, 2013.

[18] C. B. Jia, K. Huang, C. Y. Zhang, F. Fang, F. Dong, X. Y. Gu, H. T. Niu, S. W. QuMu, X. X. Ren, W. Li, et al., Status survey of the diagnosis and management ability of chronic obstructive pulmonary disease in eighty cities in China, (in Chinese), *Chin. J. Clin.*, vol. 49, no. 6, pp. 669–671, 2021.

[19] National Health Commission of the People's Republic of China, Work plan for basic framework and data standards for health records (trial), (in Chinese), http://www.nhc.gov.cn/wjw/gfxwj/201304/a281f9a650d644afa5f76e114074a91c.shtml, 2009.

[20] National Health Commission of the People's Republic of China, Guidelines on the Construction of Regional Health Information Platform Based on Health Records, http://www.nhc.gov.cn/guihuaxxs/s10741/200906/d15e616ef81d4815babc7c9fd4636a09.shtml, 2009.

[21] R. Kohavi, A study of cross-validation and bootstrap for accuracy estimation and model selection, in *Proc. 14th Int. Joint Conf. Artificial Intelligence*, Montreal, Canada, 1995, pp. 1137–1143.

[22] T. Hastie, S. Rosset, J. Zhu, and H. Zou, Multi-class AdaBoost, *Stat. Interface*, vol. 2, no. 3, pp. 349–360, 2009.

[23] A. K. Sahoo, C. Pradhan, and H. Das, Performance evaluation of different machine learning methods and deep-learning based convolutional neural network for health decision making, in *Nature Inspired Computing for Data Science*, M. Rout, J. Rout, and H. Das, eds. Cham, Switzerland: Springer, 2020, pp. 201–212.

[24] S. Lundberg and S. I. Lee, A unified approach to interpreting model predictions, arXiv preprint arXiv: 1705.07874, 2017.

[25] F. J. Martinez, A. E. Raczek, F. D. Seifer, C. S. Conoscenti, T. G. Curtice, T. D'Eletto, C. Cote, C. Hawkins, and A. L. Phillips, Development and initial validation of a self-scored COPD population screener questionnaire (COPD-PS), *COPD*, vol. 5, no. 2, pp. 85–95, 2008.

[26] D. Kotz, P. Nelemans, C. P. van Schayck, and G. J. Wesseling, External validation of a COPD diagnostic questionnaire, *Eur. Respir. J.*, vol. 31, no. 2, pp. 298–303, 2008.

[27] P. S. Bakke, E. Rönmark, T. Eagan, F. Pistelli, I. Annesi-Maesano, M. Maly, M. Meren, P. Vermeire, J. Vestbo, G. Viegi, et al., Recommendations for epidemiological studies

on COPD, *Eur. Respir. J.*, vol. 38, no. 6, pp. 1261–1277, 2011.

[28] J. G. Hansen, L. Pedersen, K. Overvad, Ø. Omland, H. K. Jensen, and H. T. Sørensen, The prevalence of chronic obstructive pulmonary disease among Danes aged 45–84 years: Population-based study, *COPD*, vol. 5, no. 6, pp. 347–352, 2008.

[29] F. A. A. Mohamed Hoesein, P. Zanen, A. P. E. Sachs, T. J. M. Verheij, J. W. J. Lammers, and B. D. L. Broekhuizen, Spirometric thresholds for diagnosing COPD: 0.70 or LLN, pre- or post-dilator values?, *COPD*, vol. 9, no. 4, pp. 338–343, 2012.

Zhihui Xing received the PhD degree in control science and engineering from Beihang University, Beijing, China, in 2016. She is currently a senior software engineer at the Intelligent Healthcare Unit, Baidu Inc, Beijing, China. Her research interests include data processing and deep learning algorithms and applications, and her developed techniques have been applied in web searching and CDSS areas.

Yi Lei received the MS degree from North University of China, Taiyuan, China, in 2018. He is currently pursuing the PhD degree at the School of Software Engineering, the Faculty of Information Technology, Beijing University of Technology, Beijing, China. His research interests include data mining, machine learning, and big data.

Jun Chen received the BS degree and the PhD degree in software engineering from Tsinghua University, Beijing, China, in 2012 and 2017, respectively. He is currently a technical lead at the Intelligent Healthcare Unit, Baidu Inc, Beijing, China. He has published and presented research papers in major journals and conferences, including the ACL, AAAI, IJCAI, ACM Multimedia, IEEE TIP, IEEE TKDE, and IEEE ICDE. He has served on program committees for international conferences and journals, including NAACL-HLT 2021, ACL 2020, AMIA 2019, and IET Electronics Letters. His research interests include machine learning for health care, natural language processing (NLP) applications, and personalization and recommendation systems.

Xinshan Lin received the BM degree in clinical medicine from Shandong University, Jinan, China, in 2013, and the MM degree in clinical medicine from Shandong Academy of Medical Sciences, Jinan, China, in 2016. He is currently pursuing the MD degree in respiratory and critical care medicine at Chinese Academy of Medical Sciences and Peking Union Medical College, Beijing, China. He participated in the Chinese COPD Tiered Diagnosis and Treatment Project and was responsible for data management and operation, quality control, and related CDSS system research and development. His main research directions include application and exploration of artificial intelligence and big data technology in the management of chronic respiratory diseases.

Qing Wang received the PhD degree in modern control engineering from Tsinghua University, Beijing, China, in 2006. He is a research fellow at the Department of Automation, Tsinghua University, Beijing, China. His research interests include web services technology, data mining, and machine learning, particularly in health care. Recently, his research has focused on the application of big data technology in medical services.

Ting Yang received the BM degree in clinical medicine and BM and MD degrees in respiratory and critical care medicine from the Capital Medical University, Beijing, China, in 1992, 2001, and 2005, respectively. She is currently the chief physician of the Department of Pulmonary and Critical Care Medicine, China-Japan Friendship Hospital, Beijing, China, and a professor at the Chinese Academy of Medical Sciences and Peking Union Medical College, Beijing, China. She is a member of the COPD Professional Group of the Chinese Respiratory Medicine Association, vice chairman of the COPD Specialty Committee of Respiratory Physicians Branch of the Chinese Medical Doctor Association, vice president of the Chinese COPD Association, member and secretary-general of the Respiratory Disease Prevention and Professional Committee of the Chinese Preventive Medical Association, and vice chairman and secretary-general of the China Grassroots Respiratory Disease Prevention and Control Alliance.

Chen Wang received the BM degree in clinical medicine and the MD degree in respiratory and critical care medicine from Capital Medical University, Beijing, China, in 1985 and 1991, respectively. He is an academic and vice president of the Chinese Academy of Engineering, president of Chinese Academy of Medical Sciences and Peking Union Medical College, Beijing, China, a foreign academic member of the American National Academy of Medical Sciences, the director of the National Respiratory Clinical Research Center of China, and a member of the National Health Science Expert Bank of China. He is a specialist in respiratory and critical care medicine.