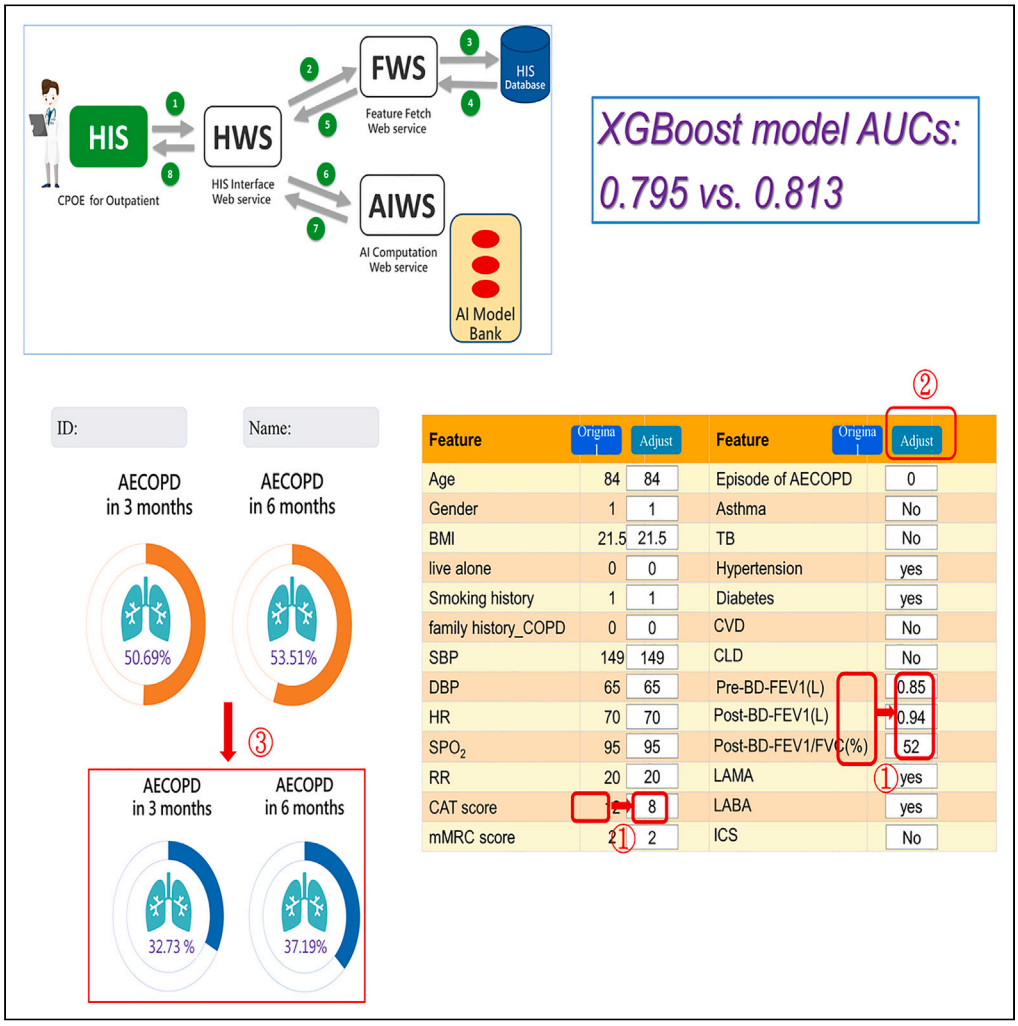
iScience

CellPress
OPEN ACCESS

Article

Machine learning approaches for practical predicting outpatient near-future AECOPD based on nationwide electronic medical records



Kuang-Ming Liao, Kuo-Chen Cheng, Mei-I Sung, Yu-Ting Shen, Chong-Chi Chiu, Chung-Feng Liu, Shian-Chin Ko

chungfengliu@gmail.com (C.-F.L.)
737005@mail.chimei.org.tw (S.-C.K.)

Highlights

XGBoost excels in predicting outpatient AECOPD risk (AUCs: 0.795/0.813 for 3/6 months)

Explainability analysis identifies key predictors of near future outpatient AECOPD

Outperforms traditional methods for near future outpatient AECOPD risk prediction

Interactive tool for physicians effectively predicts near future outpatient AECOPD risk

Liao et al., iScience 27, 109542
April 19, 2024
https://doi.org/10.1016/j.isci.2024.109542

iScience

CellPress
OPEN ACCESS

Article

Machine learning approaches for practical predicting outpatient near-future AECOPD based on nationwide electronic medical records

Kuang-Ming Liao,[1,2,8] Kuo-Chen Cheng,[7,8] Mei-I Sung,[3] Yu-Ting Shen,[3] Chong-Chi Chiu,[4,5,6] Chung-Feng Liu,[3,9,*] and Shian-Chin Ko[7,*]

SUMMARY

In this research, we aimed to harness machine learning to predict the imminent risk of acute exacerbation in chronic obstructive pulmonary disease (AECOPD) patients. Utilizing retrospective data from electronic medical records of two Taiwanese hospitals, we identified 26 critical features. To predict 3- and 6-month AECOPD occurrences, we deployed five distinct machine learning algorithms alongside ensemble learning. The 3-month risk prediction was best realized by the XGBoost model, achieving an AUC of 0.795, whereas the XGBoost was superior for the 6-month prediction with an AUC of 0.813. We conducted an explainability analysis and found that the episode of AECOPD, mMRC score, CAT score, respiratory rate, and the use of inhaled corticosteroids were the most impactful features. Notably, our approach surpassed predictions that relied solely on CAT or mMRC scores. Accordingly, we designed an interactive prediction system that provides physicians with a practical tool to predict near-term AECOPD risk in outpatients.

INTRODUCTION

Chronic obstructive pulmonary disease (COPD) is a prevalent and debilitating respiratory disease that affects millions of individuals worldwide. Early detection and management of acute exacerbations of COPD (AECOPD) is crucial for improving patient outcomes and reducing healthcare costs. COPD is a heterogeneous lung disease characterized by chronic airway symptoms, included cough with or without sputum, dyspnea and exacerbations due to abnormalities of the airways (bronchitis, bronchiolitis) and/or alveoli (emphysema) that cause airway and systemic inflammatory disease and lead to persistent, not fully reversible airflow obstruction.[1] Patient with COPD may experience acute exacerbation (AE) which characterized by worsening respiratory symptoms that need additional therapy. COPD, ranking as the third leading cause of death globally, accounted for 3.3 million deaths in 2019. Between 2009 and 2019, the worldwide death toll from COPD increased by 14.1%, attributed to factors such as urbanization, air pollution, and tobacco use. This health burden is disparately distributed among nations, with 90% of COPD-related deaths concentrated in low-income and middle-income countries.[2] AECOPD associated with disease progression, increased comorbidities and mortality and it also represent the largest component of the medical expenditure and socioeconomic burden of COPD.[3] A population-based cohort study in patients with COPD and first-ever exacerbations requiring hospitalizations in Taiwan found that 4% COPD patients died during the hospitalization and 22% of hospital survivors were dead at one year after discharge.[4]

In order to improve the clinical outcome and decrease AECOPD in Taiwan, the Bureau of National Health Insurance (NHI) and Joint Commission of Taiwan (JCT) conducted a Pay for Performance (P4P) program for COPD in April 2017.[5] Our previous study showed that patients enrolled in the national pay-for-performance program for COPD can improve in pulmonary function and symptom score across all patients with COPD. The decreased number of exacerbations was observed in COPD Groups C and D.[6] In another work, we used patient characteristics, laboratory data, comorbidities and adopted machine learning (ML) methods to early predict acute respiratory failure, ventilator dependence, and mortality in inpatients with COPD in Taiwan. The ML models had good predictive performance and offering physicians an applicable decision-making tool.[7]

[1]Department of Internal Medicine, Chi Mei Medical Center, Chiali, Tainan 722013, Taiwan
[2]Department of Nursing, Min-Hwei Junior College of Health Care Management, Tainan 73658, Taiwan
[3]Department of Medical Research, Chi Mei Medical Center, Tainan 710402, Taiwan
[4]Department of General Surgery, E-Da Cancer Hospital, I-Shou University, Kaohsiung 82445, Taiwan
[5]School of Medicine, College of Medicine, I-Shou University, Kaohsiung 82445, Taiwan
[6]Department of Medical Education and Research, E-Da Cancer Hospital, I-Shou University, Kaohsiung 82445, Taiwan
[7]Department of Pulmonary Medicine, Chi Mei Medical Center, Tainan 710402, Taiwan
[8]These authors contributed equally
[9]Lead contact
*Correspondence: chungfengliu@gmail.com (C.-F.L.), 737005@mail.chimei.org.tw (S.-C.K.)
https://doi.org/10.1016/j.isci.2024.109542

In this study, our aim was to evaluate the risk of AE of chronic obstructive pulmonary disease (AECOPD) in the near future, specifically at 3 and 6 months, utilizing advanced artificial intelligence (AI) and machine learning (ML) techniques. We collected electronic medical records from outpatients with COPD who participated in a government pay-for-performance program at two hospitals in Taiwan. We used five ML algorithms to build predictive models for 3-month and 6-month AECOPD, and interpreted the feature importance using SHapley Additive exPlanations (SHAP) analysis.[8] We also developed an interactive prediction system based on the best models, which can be used as an applicable decision-making tool for physicians.

Predicting AEs of chronic obstructive pulmonary disease (COPD) is of significant importance both from a public health perspective and in terms of patient management. The ability to anticipate and manage COPD exacerbations can lead to improved outcomes, reduced healthcare costs, and enhanced overall well-being for affected individuals.

From a public health perspective, COPD exacerbations contribute substantially to hospital admissions, emergency department visits, and healthcare resource utilization. Predictive models that identify individuals at higher risk of exacerbations allow for targeted interventions and preventive measures, potentially reducing the overall burden on healthcare systems.[9] Anticipating COPD exacerbations enables better resource allocation by healthcare providers and public health agencies. This, in turn, can aid in planning and implementing strategies to address the anticipated increase in healthcare demand during periods of heightened exacerbation risk.[10]

Regarding patient management, early identification of COPD exacerbations allows for prompt intervention and management, leading to improved symptom control and a better quality of life for COPD patients.[11] Predictive models can assist healthcare providers in developing personalized care plans for COPD patients based on their individual risk profiles. This may involve targeted medication adjustments, closer monitoring, or specific interventions to prevent exacerbations.[12]

The significance of this study lies in its potential to improve the diagnosis, treatment, and prevention of AECOPD in outpatients with COPD. With the help of our developed AI prediction system, physicians can identify high-risk patients who may develop AECOPD in the near future, so that they have more time to plan targeted interventions to prevent exacerbation of the disease. This study fills a gap in the current research on the use of ML and AI in COPD diagnosis and treatment, and provides a valuable tool for clinicians to improve patient outcomes.

RESULTS

From September 1, 2018, to September 30, 2022, a total of 11,555 samples were initially collected for this study. However, after performing data cleaning and applying the exclusion criteria, electronic medical records (EMRs) of 11,046 samples were used for model building and analysis. Among these samples, there were 561 AECOPD cases within three months and 825 AECOPD cases within six months. The low missing data rate in our study can be attributed to Taiwan's commitment to COPD treatment, especially in Chi Mei hospitals under the Pay-for-Performance Program, which ensures thorough record-keeping.

Demographics

The characteristics of the patients and the significance of the two durations (3 months vs. 6 months) are presented in Tables 1 and 2. A total of 11,046 cases were included in the analysis, with an average age of 71.7 years. The majority of the patients were male (85.7%). Among the cases, 561 patients experienced AECOPD within 3 months, and 825 patients experienced AECOPD within 6 months.

Machine learning modeling results

The model performance for each predicting outcome is summarized in Table 3. In the model for predicting AECOPD within 3 months, the Voting model achieved the highest AUC value of 0.799, followed by the XGBoost, Logistic Regression, Random Forest, and SVM with AUC values of 0.795, 0.791, 0.789, and 0.785, respectively. In the model for predicting AECOPD within 6 months, the XGBoost model achieved the highest AUC value of 0.813, followed by the Random Forest, Logistic Regression, SVM, and MLP Classifier models with AUC values of 0.812, 0.792, 0.783, and 0.783 respectively. Additionally, employing ensemble machine learning strategies, particularly Stacking and Voting techniques, did not significantly enhance the model's predictive quality.

In real-world patient-level scenarios, it becomes crucial to ensure that our predictive models are well-calibrated. Inaccuracies in individual predicted probabilities could potentially lead to incorrect decisions made by healthcare professionals. To assess the calibration of our models, we generated calibration plots that visualize the alignment between observed and predicted case states within distinct absolute probability subgroups or bins. A calibration curve that closely follows the diagonal line signifies a higher degree of calibration for the corresponding model. Our assessment, as depicted in Figure 1, indicates that both of the predictive models meet the calibration criterion well within acceptable limits. As a result, these models can be considered suitable for implementation in a prediction system.

Interpreting the feature importance to the models

In order to better interpret the "Black-box" ML models, we conducted SHAP analysis for each best AI model (excluding the ensemble models) to understand the contribution of each feature to the associated outcome. SHAP analysis provides insights into how individual features affect the prediction. Figures 2 and 3 illustrate the SHAP plots for the best models, providing both global and importance explanations for the 3-month and 6-month models, respectively.

Figure 2A represents the SHAP values of each feature used by the model to predict 3-month AECOPD. The x axis corresponds to the SHAP values, indicating the impact of each feature on the prediction. In the global view, it is observed that a lower value for "Episode of AECOPD"

Table 1. Baselines of data and significance (3-month AECOPD)

Feature	Overall N = 11046	AECOPD within 3 months: No N = 10485	AECOPD within 3 months: Yes N = 561	p-Value
Gender				
Female, n (%)	1583 (14.3)	1484 (14.2)	99 (17.6)	0.025
Male, n (%)	9463 (85.7)	9001 (85.8)	462 (82.4)	
Age, mean (SD)	71.7 (9.4)	71.7 (9.4)	72.9 (9.5)	0.003
BMI, mean (SD)	24.6 (5.6)	24.7 (5.7)	23.3 (4.4)	<0.001
Live alone, n (%)	1059 (9.6)	1013 (9.7)	46 (8.2)	0.284
Smoking history, n (%)	8844 (80.1)	8391 (80.0)	453 (80.7)	0.718
SBP, mean (SD)	131.0 (18.0)	131.1 (17.9)	130.8 (20.0)	0.804
DBP, mean (SD)	75.6 (12.1)	75.7 (11.9)	75.1 (13.9)	0.362
HR, mean (SD)	82.6 (21.3)	82.4 (21.5)	87.7 (15.8)	<0.001
SpO2, mean (SD)	96.2 (2.1)	96.3 (2.0)	95.0 (3.1)	<0.001
RR, mean (SD)	18.8 (2.4)	18.7 (2.4)	19.9 (2.7)	<0.001
Family history_COPD, n (%)	1591 (14.4)	1507 (14.4)	84 (15.0)	0.739
Medical history				
Asthma, n (%)	845 (7.6)	800 (7.6)	45 (8.0)	0.796
TB, n (%)	1115 (10.1)	1048 (10.0)	67 (11.9)	0.156
Hypertension, n (%)	5997 (54.3)	5666 (54.0)	331 (59.0)	0.024
Diabetes, n (%)	2261 (20.5)	2105 (20.1)	156 (27.8)	<0.001
CVD, n (%)	4756 (43.1)	4484 (42.8)	272 (48.5)	0.009
CLD, n (%)	1356 (12.3)	1298 (12.4)	58 (10.3)	0.171
Pre-BD-FEV1(L), mean (SD)	1.4 (0.6)	1.4 (0.6)	1.1 (0.5)	<0.001
Post-BD-FEV1(L), mean (SD)	1.5 (0.6)	1.5 (0.6)	1.2 (0.7)	<0.001
Post-BD-FEV1/FVC(%), mean (SD)	52.9 (13.2)	53.1 (13.1)	48.4 (14.1)	<0.001
CAT score, mean (SD)	6.7 (5.1)	6.5 (4.9)	11.5 (6.7)	<0.001
mMRC score, mean (SD)	1.2 (0.9)	1.2 (0.9)	2.0 (1.2)	<0.001
Episode of AECOPD, mean (SD)	1.2 (2.2)	1.1 (2.0)	3.5 (3.8)	<0.001
Medication (Inhaled bronchodilator)				
Non, n (%)	180 (1.6)	170 (1.6)	10 (1.8)	<0.001
LAMA, n (%)	1447 (13.1)	1411 (13.5)	36 (6.4)	
LAMA+LABA, n (%)	4778 (43.3)	4585 (43.7)	193 (34.4)	
LABA+ICS, n (%)	1426 (12.9)	1371 (13.1)	55 (9.8)	
LAMA+LABA+ICS, n (%)	3215 (29.1)	2948 (28.1)	267 (47.6)	

The significance of categorical variables was evaluated using the Chi-square test method, while that of continuous variables was evaluated using Student's *t* test method. Non: No regular or routine medication of inhaled bronchodilator prescribed, except for pro re nata (PRN) medicine such as Short-Acting Beta2-Agonist (SABA) to be used as needed.

(indicating no previous AECOPD) is associated with a decrease in the predicted outcome (not obtaining AECOPD within 3 months). On the other hand, smaller mMRC scores (blue dots) also imply a decrease in the predicted outcome, although there are some red dots to the left of the 0-point (SHAP value 0), suggesting that slightly higher mMRC scores can also lead to AECOPD within 3 months. Additionally, SpO2 (visible in blue) is associated with a decrease in the predicted outcome. Similarly, Figure 3A presents the explanation for predicting 6-month AECOPD using SHAP plots.

Moreover, the absolute SHAP value of each feature can directly represent the importance of each feature on the associated outcome. As shown in Figures 2B and 3B, the top five features with the greatest impact on both 3- and 6-month AECOPD are episode of AECOPD, mMRC score, CAT score, RR and ICS.

Table 2. Baselines of data and significance (6-month AECOPD)

Feature	Overall N = 11046	AECOPD within 6 months		p-Value
		No N = 10221	Yes N = 825	
Gender				
Female, n (%)	1583 (14.3)	1446 (14.1)	137 (16.6)	0.059
Male, n (%)	9463 (85.7)	8775 (85.9)	688 (83.4)	
Age, mean (SD)	71.7 (9.4)	71.7 (9.4)	72.9 (9.2)	<0.001
Height, mean (SD)	162.0 (8.0)	162.1 (8.0)	161.0 (7.9)	<0.001
Weight, mean (SD)	64.5 (13.0)	64.8 (13.0)	60.4 (12.7)	<0.001
BMI, mean (SD)	24.6 (5.6)	24.7 (5.7)	23.3 (4.4)	<0.001
Live alone, n (%)	1059 (9.6)	988 (9.7)	71 (8.6)	0.351
Smoking history, n (%)	8844 (80.1)	8160 (79.8)	684 (82.9)	0.038
SBP, mean (SD)	131.0 (18.0)	131.1 (17.9)	130.3 (19.6)	0.276
DBP, mean (SD)	75.6 (12.1)	75.7 (11.9)	75.1 (13.8)	0.248
HR, mean (SD)	82.6 (21.3)	82.3 (21.6)	87.0 (15.6)	<0.001
SpO2, mean (SD)	96.2 (2.1)	96.3 (2.0)	95.1 (3.1)	<0.001
RR, mean (SD)	18.8 (2.4)	18.7 (2.4)	19.8 (2.6)	<0.001
Family history_COPD, n (%)	1591 (14.4)	1470 (14.4)	121 (14.7)	0.863
Asthma, n (%)	845 (7.6)	778 (7.6)	67 (8.1)	0.644
TB, n (%)	1115 (10.1)	1016 (9.9)	99 (12.0)	0.067
Hypertension, n (%)	5997 (54.3)	5496 (53.8)	501 (60.7)	<0.001
Diabetes, n (%)	2261 (20.5)	2019 (19.8)	242 (29.3)	<0.001
CVD, n (%)	4756 (43.1)	4361 (42.7)	395 (47.9)	0.004
CLD, n (%)	1356 (12.3)	1266 (12.4)	90 (10.9)	0.235
Pre-BD-FEV1(L), mean (SD)	1.4 (0.6)	1.4 (0.6)	1.1 (0.5)	<0.001
Post-BD-FEV1(L), mean (SD)	1.5 (0.6)	1.5 (0.6)	1.2 (0.6)	<0.001
Post-BD-FEV1/FVC(%), mean (SD)	52.9 (13.2)	53.2 (13.1)	48.3 (14.0)	<0.001
CAT score, mean (SD)	6.7 (5.1)	6.4 (4.8)	11.0 (6.5)	<0.001
mMRC score, mean (SD)	1.2 (0.9)	1.2 (0.9)	2.0 (1.1)	<0.001
Episode of AECOPD, mean (SD)	1.2 (2.2)	1.1 (2.0)	3.2 (3.6)	<0.001
Medication (Inhaled bronchodilator)				
Non, n (%)	180 (1.6)	166 (1.6)	14 (1.7)	<0.001
LAMA, n (%)	1447 (13.1)	1394 (13.6)	53 (6.4)	
LAMA+LABA, n (%)	4778 (43.3)	4498 (44.0)	280 (33.9)	
LABA+ICS, n (%)	1426 (12.9)	1347 (13.2)	79 (9.6)	
LAMA+LABA+ICS, n (%)	3215 (29.1)	2816 (27.6)	399 (48.4)	

The significance of categorical variables was evaluated using the Chi-square test method, while that of continuous variables was evaluated using Student's *t* test method. Non: No regular or routine medication of inhaled bronchodilator prescribed, except for pro re nata (PRN) medicine such as Short-Acting Beta2-Agonist (SABA) to be used as needed.

Feature reduction experiment

Based on the SHAP plots of the XGBoost model, we conducted a feature reduction experiment to prioritize the removal of less important features based on their feature importance ranking.

For the 3-month AECOPD outcome, we found that removing SBP, DBP, HR, Asthma, and CLD did not significantly differ from the full-feature model in terms of the DeLong test result. However, further removal of one feature, age, led to a significant decline in model quality (p value 0.002 in the DeLong test). Hence, in future applications, especially when medical resources are limited or data collection is challenging, the 21-feature model may be considered.

Table 3. Model performance

Model	Algorithm	Accuracy	Sensitivity	Specificity	PPV	NPV	AUC	p-Value
3-month AECOPD	XGBoost	0.730	0.732	0.73	0.127	0.981	0.795	
	Logistic Regression	0.728	0.726	0.728	0.125	0.98	0.791	0.741
	Random Forest	0.712	0.714	0.712	0.117	0.979	0.789	0.444
	SVM	0.706	0.720	0.705	0.115	0.979	0.785	0.345
	MLP Classifier	0.716	0.714	0.716	0.118	0.979	0.761	0.033
	Voting[a]	0.709	0.732	0.708	0.118	0.980	0.799	0.607
	Stacking[a]	0.714	0.702	0.715	0.116	0.978	0.774	0.002
6-month AECOPD	XGBoost	0.732	0.73	0.732	0.180	0.971	0.813	
	Random Forest	0.723	0.722	0.723	0.174	0.97	0.812	0.945
	Logistic Regression	0.705	0.722	0.704	0.165	0.969	0.792	0.007
	SVM	0.71	0.702	0.711	0.164	0.967	0.783	<0.001
	MLP Classifier	0.712	0.706	0.712	0.165	0.968	0.783	<0.001
	Voting[a]	0.731	0.738	0.73	0.181	0.972	0.812	0.896
	Stacking[a]	0.729	0.73	0.729	0.179	0.971	0.802	0.071

AUC: area under the receiver operating characteristic curve; PPV: Positive Predictive Value; NPV: Negative Predictive Value; AECOPD: acute exacerbation of chronic obstructive pulmonary disease.
p-value was conducted by DeLong test.
[a]Ensemble machine learning methods.

Similarly, for the 6-month AECOPD outcome, removing SBP and asthma did not significantly differ from the original full-feature model. However, further removal of one feature, DBP, resulted in a significant decline in model quality (p value <0.001 in the DeLong test). Therefore, in future applications, especially when medical resources are limited or data collection is challenging, the 24-feature model may be preferred.

Performance comparison with mMRC score and CAT score

The mMRC (Modified Medical Research Council) Dyspnea Scale[13] serves as a tool for evaluating the level of breathlessness or dyspnea in individuals, particularly those with respiratory conditions like chronic obstructive pulmonary disease (COPD). Widely employed in both clinical research and practice, the mMRC Dyspnea Scale quantifies the impact of breathlessness on a patient's daily life and functional capacity. The scale is graded from 0 to 4, with higher scores indicating more severe breathlessness.

The CAT (COPD Assessment Test)[14] is a questionnaire designed to assess the impact of chronic obstructive pulmonary disease (COPD) on a patient's health-related quality of life. It is a patient-reported instrument that measures the symptoms and their impact on daily life. The CAT consists of eight items that cover various aspects of COPD, including cough, sputum, chest tightness, breathlessness, sleep, energy, activities, and confidence in leaving home. Patients rate each item on a scale from 0 to 5, with a higher score indicating a greater impact of COPD on their life. The total CAT score is the sum of the individual item scores, ranging from 0 to 40. The higher the CAT score, the more severe the impact of COPD on the patient's health-related quality of life.

As the mMRC score and CAT score are well-known tools used for assessing COPD outcomes, we conducted a comparison of our models with mMRC scores and CAT scores alone, respectively. The results, presented in Table 4, indicate that our ML models performed significantly better than either the mMRC score or the CAT score alone for predicting AECOPD.

Interactive prediction system development and user pilot evaluation

To assess the feasibility and acceptability of the built ML models, we proceeded to develop an AI risk prediction system utilizing the two best models. This system was deployed in outpatient clinics with the aim of assisting physicians in their decision-making process.

Based on the AI software framework proposed by Liu et al.,[15] we improved three web service programs, named as HIS interface web service (HWS), Feature fetch web service (FWS), and AI computation web service (AIWS), to implement this prediction function, as shown in Figure 4. Eight sending/receiving messages are executed while a prediction is triggered by physicians through the existing hospital information systems. The ML models were developed using the Python programming language, while the web service programs was developed using MS Visual Studio.

- HWS

The HWS receives prediction requests from the existing HIS (outpatient CPOE system for physicians), forwards them to FWS to fetch the necessary feature values for the AECOPD predictive model, then forwards them to AIWS to obtain the prediction result. Finally, the HWS sends the prediction result (e.g., 50.69% of 3-month AECOPD risk) back to the HIS for display to physicians.

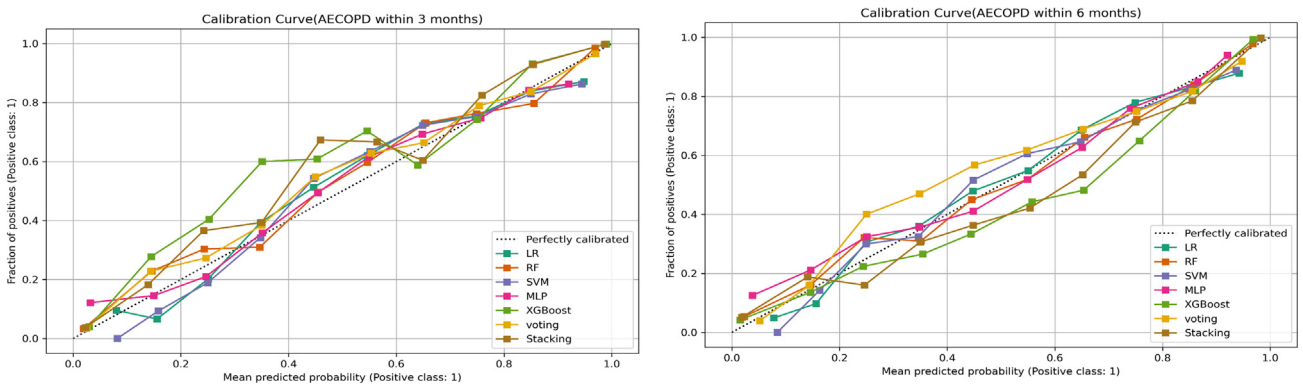


Figure 1. Calibration curves

- FWS

The FWS receives requests from HWS, retrieves the feature values of the COPD patient (such as gender, age, BMI, SBP, SpO2, family history of COPD, etc.) for the predictive model, and returns them to HWS.

- AIWS

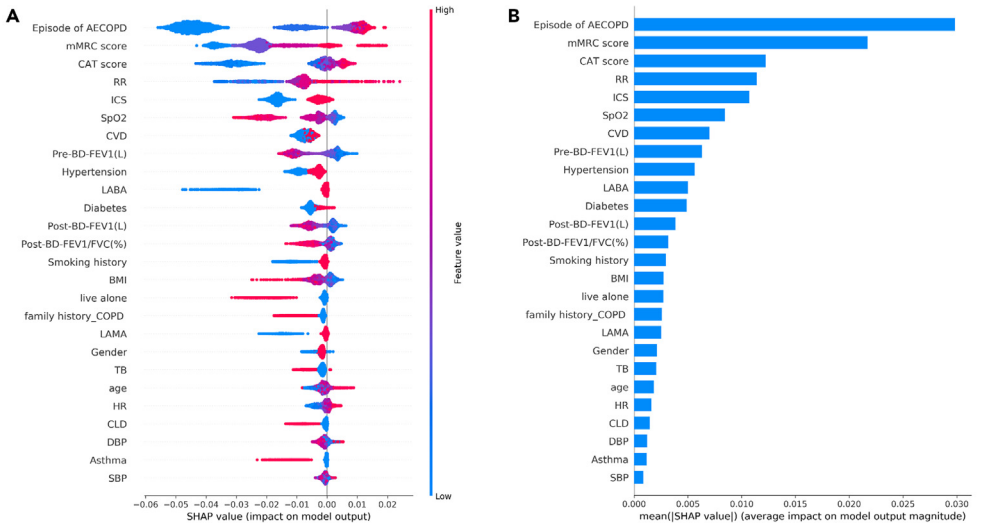
The AIWS receives feature values from HWS and forwards them to the corresponding AECOPD AI model from Chi Mei's existing model bank, which stores various ML models for clinical use. Subsequently, the AIWS conducts the AI prediction and sends back the result (risk probability) to HWS for display to physicians.

The AI risk prediction system can smoothly integrate with the existing subsystems of the Hospital Information System (HIS), including the lab system, physician order system, nursing system, and others. This integration allows for the automatic retrieval of real-time values for each feature (as denoted by "FWS" in Figure 4) without the need for manual input when calculating the predictive probability from the associated model in the model bank (as denoted by "AIWS" in Figure 4). We have successfully implemented a severe adverse outcome predictive system for hospitalized COPD patients using an earlier version of the web service framework.[16] Figure 5 provides a snapshot of the AI risk prediction system (i.e., HWS in Figure 4), where a probability of $\geq$50% indicates a high probability of AECOPD occurrence. Notably, we have designed an interactive adjustment function ("Adjust" button in Figure 5), allowing physicians to input new feature values and obtain updated predicted results. This feature proves helpful in simulating potential changes in a patient's condition, enabling physicians to determine specific treatment strategies to improve the patient's condition proactively. Figure 5 demonstrates the steps (1, 2, 3) taken to adjust the features and the resulting change in AECOPD risk. Based on the case in Figure 5A, the AECOPD risks decreased from 50.69% to 53.51% (more than 50%, tending to occur the outcome) to 32.73% and 37.19% (less than 50%, tending not to occur the outcome) after adjusting feature values with higher feature importance. It implies that implementing suitable interventions to modify these important features may significantly help prevent AECOPD occurrence in patients. Interestingly, in the case of Figure 5B, the use of LAMA alone without LABA did not appear to significantly affect the risk of AECOPD. However, follow-up rigorous studies are needed to confirm and explore this finding.

After the system went live, we demonstrated the AI system to three physicians for pilot use, receiving extremely positive feedback. Their experiences and suggestions were analyzed to evaluate user acceptance of the system. We proposed four structured questions (using a five-point scale where one point indicated strongly disagree and five points indicated strongly agree): (1) Is the operation simple? (2) Does it have clinical utility? (3) Is the provided information accurate and worth referencing? (4) Would you be willing to use it? Additionally, we encouraged them to provide other comments. The survey results showed that they had a positive attitude toward the predictive system (with mean scores of 4.2, 4.0, 4.3, and 4.1 for usability, practicality, accuracy and intention to use, respectively). Physicians recognized the system as a valuable intelligent tool for identifying high-risk patients on an individual level. By considering the risk probabilities provided by the system, physicians can proactively plan appropriate treatment options, thus optimizing medical resource utilization. Anticipated implementation of this system is expected to significantly enhance the quality and efficiency of COPD care, promoting an increased willingness for subsequent follow-up visits among COPD patients.

DISCUSSION

To the best of our knowledge, our study represents one of the pioneering applications of AI and a national pay-for-performance program for COPD in practical predictions of near-future occurrences of AECOPD. By utilizing readily available daily clinical data and incorporating important features such as comorbidities, ICS use, and AECOPD history, we developed predictive models using five different ML algorithms. Our best models were validated to outperform the well-known mMRC and CAT score alone, leading to their selection for further development as a predictive system for clinical use. This finding highlights the potential of these models to serve as promising tools in enhancing COPD management and providing more accurate predictions of AECOPD in a clinical setting. Furthermore, the SHAP technique provided insights into the interpretability of the AI models, highlighting the significance of episode of AECOPD, mMRC score, CAT score, respiratory rate (RR), and



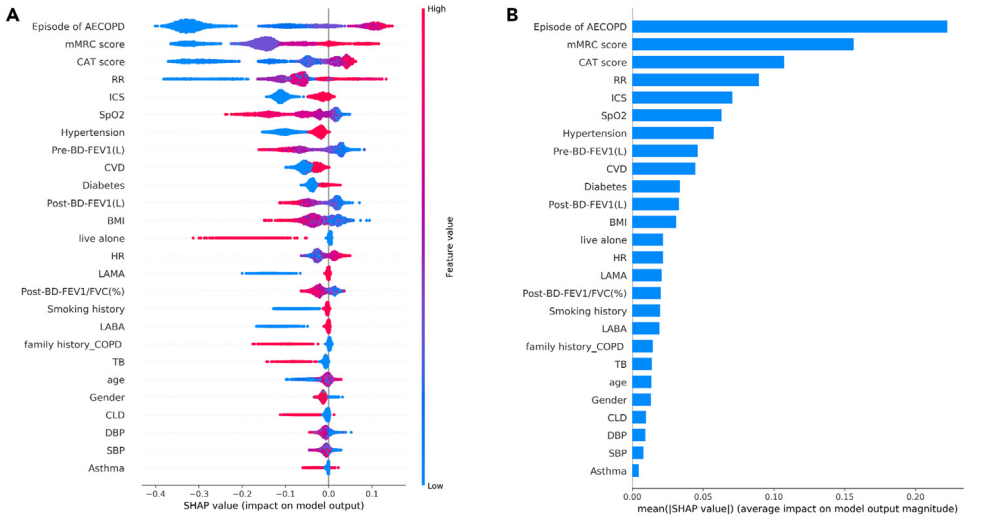
SHAP value of each feature for global explanation

Absolute SHAP value of each feature for importance explanation

Figure 2. SHAP plot for best model of 3-month AECOPD (XGBoost)
Note. SHAP: SHapley Additive exPlanations.

inhaled corticosteroid (ICS) use in predicting AECOPD for both the 3-month and 6-month time frames. This study demonstrates a comprehensive and robust application of AI and ML techniques in predicting AECOPD using national pay-for-performance data. For COPD patients at a high risk of exacerbation, inhaled corticosteroids may be added, especially for those with a history of frequent exacerbations, particularly if there is an eosinophil count exceeding 300 cells/μL. Additionally, all COPD patients require pulmonary rehabilitation, smoking cessation, vaccinations, management of comorbidities, and regular follow-up.

A previous retrospective observational cohort study conducted in England and Wales included 58,589 COPD patients aged 40 years and above, utilizing electronic medical records. This study found an independent risk of AECOPD during the follow-up period, with a higher risk observed in patients with frequent episodes or a history of prior AEs and increasing dyspnea scores.[17] Consistent with these findings, our study also identified a history of exacerbations as the most strongly associated characteristic with AECOPD. We observed a clear relationship between exacerbations in the prior 12 months and the occurrence of prospective exacerbations at 3 to 6 months. Patients who did not experience exacerbations in the prior 12 months had the lowest risk of future exacerbations, while those with one prior exacerbation had a higher



SHAP value of each feature for global explanation

Absolute SHAP value of each feature for importance explanation

Figure 3. SHAP plot for best model of 6-month AECOPD (XGBoost)
Note. SHAP: SHapley Additive exPlanations.

Table 4. Performance comparison of the proposed models with CAT score and mMRC score

Method	Accuracy	Sensitivity	Specificity	PPV	NPV	AUC	Delong test
3-month AECOPD							
ML model(XGBoost)	0.730	0.732	0.730	0.127	0.981	0.795	
mMRC score	0.739	0.616	0.712	0.110	0.971	0.709	<0.001
CAT score	0.704	0.625	0.708	0.103	0.972	0.735	<0.001
6-month AECOPD							
ML model(XGBoost)	0.732	0.730	0.732	0.180	0.971	0.813	
mMRC score	0.744	0.597	0.756	0.165	0.959	0.714	<0.001
CAT score	0.700	0.601	0.708	0.143	0.956	0.727	<0.001

ML: machine learning; AUC: area under the receiver operating characteristic curve; AECOPD: acute exacerbation of chronic obstructive pulmonary disease; mMRC: modified Medical Research Council; CAT: COPD Assessment Test.

risk compared to those with none. Our results align with previous studies, further supporting the association between prior exacerbations and the increased risk of future exacerbations.[18]

In our study, we focused specifically on patients with severe exacerbations of COPD necessitating emergency room visits and hospitalization. We found that the most powerful predictor of future exacerbations in COPD patients is patients' history of previous exacerbations.[19] That is, patients who have experienced frequent exacerbations in the past are more likely to experience exacerbation again in the future. While this method is commonly used in clinical settings, it has some subjectivity and may lead to missed diagnoses due to potential underreporting or variations in reporting.[20] Studies have confirmed a link between poor quality of life in COPD patients and an increased risk of future exacerbations. Patients with a lower quality of life are more vulnerable to exacerbations, and assessing their quality of life can be an essential tool in predicting exacerbation risk.[14]

We found that the mMRC score was an independent risk factor for future severe exacerbations. This finding is consistent with another study that enrolled sixty COPD patients and reported significantly higher mMRC score in the frequent exacerbator group.[21] Similarly, Cote et al.[22] reported higher mMRC scale in COPD patients with repeated AEs compared to those with a single exacerbation over a 24-month period. Our study provides further evidence supporting the predictive value of mMRC score for future severe exacerbations.

Previous research has demonstrated that mMRC scores and the number of exacerbations experienced in the previous year can predict hospitalization in the intensive care unit.[23] Another study conducted in Korea, which included patients from 37 tertiary referral hospitals, reported higher mMRC scores in the AECOPD group compared to those without AECOPD during the follow-up period.[24] Our findings align with these studies, highlighting the significant association between AECOPD and impaired health status. The identified risk factors, including the episode of AECOPD and mMRC scores, are often considered unmodifiable.

The CAT score is a recommended tool by the Global Initiative for Chronic Obstructive Lung Disease (GOLD) for evaluating and quantifying the impact of COPD symptoms on a patient's health. It assesses the severity of COPD and evaluates symptoms such as cough, phlegm, and dyspnea. A higher CAT score is associated with an increased risk of future exacerbations.[25,26] The CAT score shows significant correlations

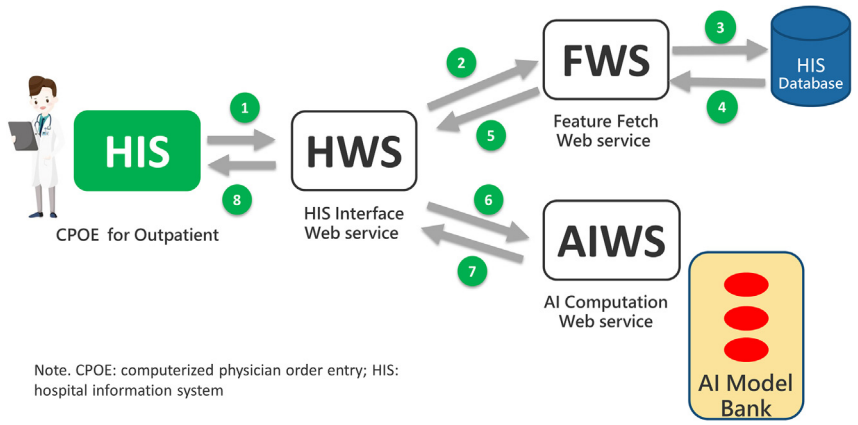
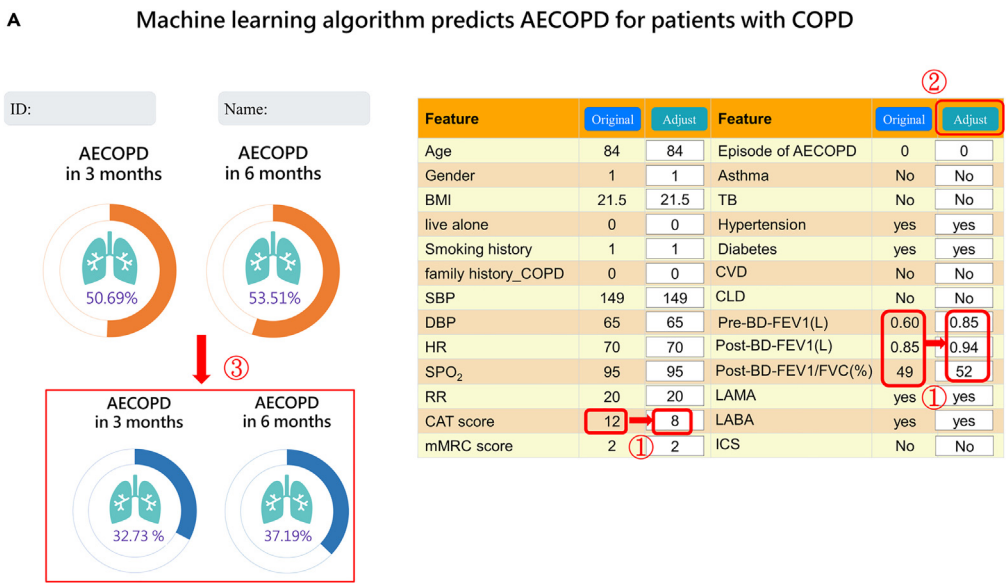


Figure 4. AI web service framework and interactive procedure

Note. CPOE: computerized physician order entry; HIS: hospital information system.



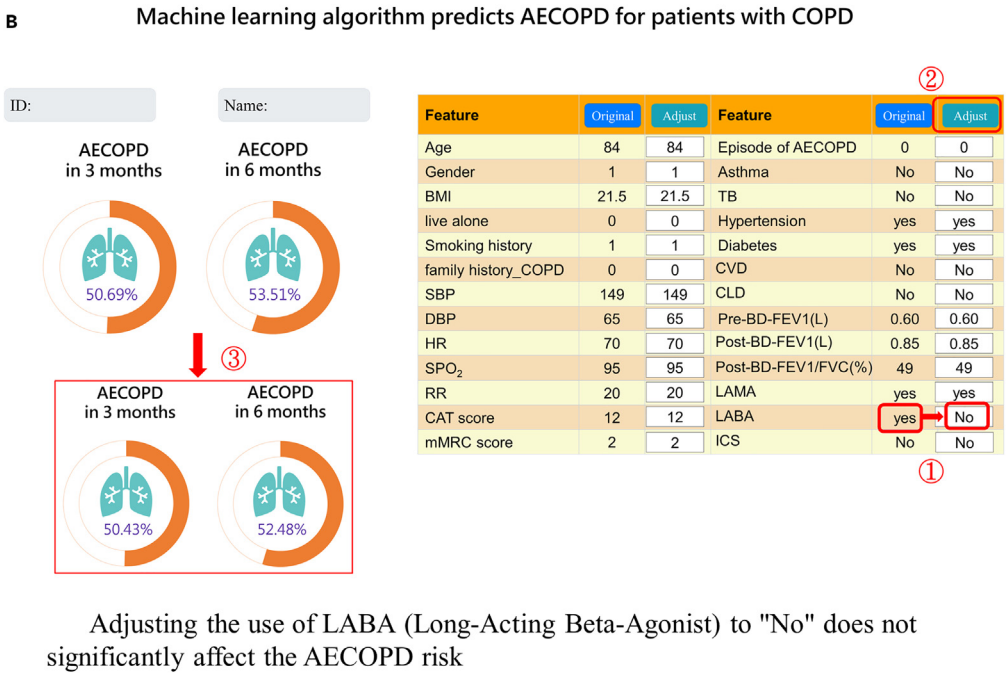


Figure 5. Snapshot of the AI risk prediction system with feature adjustment examples

with respiratory rate, pulse rate, forced vital capacity (FVC), and forced expiratory volume in the first second (FEV1).[27] These correlations suggest that the CAT score can reflect physiological changes and lung function in COPD patients. The CAT score remains relevant during AECOPD and correlates well with pulmonary function test parameters even during these exacerbation periods.[28] Overall, the CAT score appears to be a valuable tool in the management of COPD patients. It aids in assessing disease severity, evaluating symptoms, and identifying the risk of exacerbations.[29]

By integrating the mMRC score, CAT score, patient history of exacerbations, and quality of life into our AI models and system, healthcare professionals can accurately identify individuals at a higher risk of experiencing future exacerbations in COPD. This comprehensive approach enables a more holistic understanding of each patient's condition and risk factors, leading to better-informed treatment decisions.

In a study conducted by Kor et al.,[30] they also utilized a COPD Pay-for-Performance Program database at Changhua Christian Hospital from 2017 to 2019 and enrolled 606 patients with COPD. They developed four ML models, including support vector machine (SVM), random forest (RF), gradient boosting machine (GBM), and extreme gradient boosting (XGB), to predict first-time AECOPD, and the best-performing model was selected. They further employed an explainable approach based on ML and SHAP analysis to assess the risk factors of exacerbation and generate individual explanations for the model's decisions. They identified patients' CAT scores and symptoms of wheezing as the two most important features for predicting first-time AECOPD, based on the highest SHAP values.

Our research builds upon Kor et al.[31]'s study, resulting in even more valuable findings. Firstly, we enrolled a larger sample size of 11,046 COPD cases compared to their 606 cases. Secondly, we employed five ML algorithms, namely logistic regression, random forest, SVM, XGBoost, and MLP, expanding the range of models used. In contrast, they utilized four ML models. Thirdly, our study not only predicted AECOPD but also considered the history of AECOPD, which was not analyzed in their study. Fourthly, we selected a reduced number of features (26 in total) to enhance clinical applicability, whereas they utilized 38 features. Fifthly, our predictions of risk within 30 or 60 days may be more practical and useful for physicians, allowing them enough time to plan careful interventions to prevent AECOPD in their patients. Lastly, we implemented an interactive prediction system integrated into the existing HIS to aid physicians in decision-making, demonstrating the feasibility of our ML models.

Additionally, in Kor et al.[31]'s study, a patient's history of AECOPD was not suitable for predicting the risk of first-time AECOPD, as they suspected that some patient records may lack information regarding prior AECOPD episodes. In contrast, we ensured that each patient had accurate information regarding their prior AECOPD episodes by including one year of electronic medical records (EMR) data before the observation period. Our study integrated ML-based modeling with vital signs, laboratory data, comorbidities, and medication information. The results showed that XGBoost achieved the highest prediction accuracy, with an AUC of 0.813.

Our study proves that integrating AI/ML into clinical practice enhances the prediction of near-future AECOPD, leading to fewer emergency department visits and hospitalizations. The early identification of COPD exacerbations facilitates swift intervention and management, resulting in enhanced symptom control and an improved quality of life for individuals with COPD. Predictive models play a crucial role in aiding healthcare providers to formulate personalized care plans tailored to the individual risk profiles of COPD patients. This may encompass targeted adjustments to medication, closer monitoring, or specific interventions designed to prevent exacerbations. Our predictive system confirms the real-world clinical utility of AI/ML, paving the way for transformative COPD management and improved patient outcomes. We identified key risk factors for AECOPD within 3- and 6-month periods, offering crucial insights for physicians to implement targeted interventions. Integrating this intelligent technology into hospital information systems advances COPD management, marking a significant shift toward data-driven, patient-centric care. The adoption of AI/ML heralds a future of optimized and personalized healthcare.

Limitations of the study

While our study provides valuable insights, it is important to acknowledge several limitations. First, the follow-up period of 6 months was relatively short, preventing a thorough evaluation of long-term exacerbation episodes in AECOPD patients. However, our findings highlight the occurrence of AECOPD in a substantial number of diagnosed COPD patients with varying degrees of airflow limitation. Second, although we utilized AI and ML techniques to investigate the relationship between comorbidities and AECOPD, the assessment of exacerbation rates specifically related to comorbidities was not fully explored. Future studies should aim to comprehensively assess the impact of comorbidities on AECOPD occurrences.

Third, unobservable factors such as patients' medication adherence may have influenced the change in outcome variables and could potentially confound the accuracy of our results. Fourthly, our study primarily focused on the short-term effects of the COPD Pay-for-Performance program in southern Taiwan. It is essential to conduct further observations to assess the long-term effects of the program and its applicability to a broader COPD population. Lastly, the generalizability of our study findings may be limited to other country populations due to the unique characteristics of Taiwan's national healthcare system and the COPD pay-for-performance model.

STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- KEY RESOURCES TABLE
- RESOURCE AVAILABILITY
  - Lead contact
  - Materials availability
  - Data and code availability
- EXPERIMENTAL MODEL AND STUDY PARTICIPANT DETAILS
  - Feature and outcome variables
  - Model building and evaluation

- METHOD DETAILS
- QUANTIFICATION AND STATISTICAL ANALYSIS

SUPPLEMENTAL INFORMATION

Supplemental information can be found online at https://doi.org/10.1016/j.isci.2024.109542.

ACKNOWLEDGMENTS

This research was funded by Chi Mei Medical Center, grant number CMFHR11139.

AUTHOR CONTRIBUTIONS

K.-M.L. and K.-C.C. were responsible for the initial conceptualization of the study. The methodology was primarily designed by C.-F.L., while the formal analysis was executed by both Y.-T.S. and M.-I.S. Essential resources for the research were provided by S.-C.K. and Y.-T.S. The original draft of the manuscript was prepared by K.-M.L. and C.-F.L. Further reviews and edits to enhance the writing were made by C.-C.C. and S.-C.K. All authors have meticulously read and unanimously agreed to the published version of the manuscript.

DECLARATION OF INTERESTS

The authors declare no conflict of interest.

Received: September 14, 2023
Revised: February 7, 2024
Accepted: March 18, 2024
Published: March 20, 2024

REFERENCES

1. Global Initiative for Chronic Obstructive Lung Disease (GOLD). Global Strategy for Prevention, Diagnosis and Management of COPD: (2023). Report. https://goldcopd.org/2023-gold-report-2.
2. Chen, S., Kuhn, M., Prettner, K., Yu, F., Yang, T., Bärnighausen, T., Bloom, D.E., and Wang, C. (2023). The global economic burden of chronic obstructive pulmonary disease for 204 countries and territories in 2020-50: a health-augmented macroeconomic modelling study. Lancet. Glob. Health *11*, e1183–e1193.
3. MacLeod, M., Papi, A., Contoli, M., Beghé, B., Celli, B.R., Wedzicha, J.A., and Fabbri, L.M. (2021). Chronic obstructive pulmonary disease exacerbation fundamentals: Diagnosis, treatment, prevention and disease impact. Respirology *26*, 532–551. https://doi.org/10.1111/resp.14041.
4. Ho, T.W., Tsai, Y.J., Ruan, S.Y., Huang, C.T., Lai, F., and Yu, C.J.; HINT Study Group (2014). In-hospital and one-year mortality and their predictors in patients hospitalized for first-ever chronic obstructive pulmonary disease exacerbations: a nationwide population-based study. PLoS One *9*, e114866.
5. Lin, C.H., Yu, C.J., Wang, H.C., Lin, M.C., and Cheng, S.L. (2020). The beneficial effect of COPD pay-for-performance program in Taiwan. In International Perspectives on Pulmonary and Critical Care Medicine (American Thoracic Society International Conference), p. A6566.
6. Cheng, K.C., Lai, C.C., Wang, C.Y., Wang, C.M., Ho, C.H., Sung, M.I., Hsing, S.C., Liao, K.M., and Ko, S.C. (2022). The Impact of the Pay-for-Performance Program on the Outcome of COPD Patients in Taiwan after One Year. Int. J. Chron. Obstruct. Pulmon. Dis. *17*, 883–891.
7. Chen, S., Shi, Y., Hu, B., and Huang, J. (2023). A Prediction Model for In-Hospital Mortality of Acute Exacerbations of Chronic Obstructive Pulmonary Disease Patients Based on Red Cell Distribution Width-to-Platelet Ratio. Int. J. Chron. Obstruct. Pulmon. Dis. *18*, 2079–2091.
8. Lundberg, S.M., and Lee, S.I. (2017). A Unified Approach to Interpreting Model Predictions. In Proceedings of the 31st International Conference on Neural Information Processing Systems (NIPS 2017), pp. 4768–4777.
9. Suissa, S., Dell'Aniello, S., and Ernst, P. (2012). Long-term natural history of chronic obstructive pulmonary disease: severe exacerbations and mortality. Thorax *67*, 957–963.
10. Müllerova, H., Maselli, D.J., Locantore, N., Vestbo, J., Hurst, J.R., Wedzicha, J.A., Bakke, P., Agusti, A., and Anzueto, A. (2015). Hospitalized exacerbations of COPD: risk factors and outcomes in the ECLIPSE cohort. Chest *147*, 999–1007.
11. Soler-Cataluña, J.J., Martínez-García, M.A., Román Sánchez, P., Salcedo, E., Navarro, M., and Ochando, R. (2005). Severe acute exacerbations and mortality in patients with chronic obstructive pulmonary disease. Thorax *60*, 925–931.
12. Wedzicha, J.A., Calverley, P.M.A., Seemungal, T.A., Hagan, G., Ansari, Z., and Stockley, R.A.; INSPIRE Investigators (2008). The prevention of chronic obstructive pulmonary disease exacerbations by salmeterol/fluticasone propionate or tiotropium bromide. Am. J. Respir. Crit. Care Med. *177*, 19–26.
13. Bestall, J.C., Paul, E.A., Garrod, R., Garnham, R., Jones, P.W., and Wedzicha, J.A. (1999). Usefulness of the medical research council (MRC) dyspnoea scale as a measure of disability in patients with chronic obstructive pulmonary disease. Thorax *54*, 581–586.
14. Jones, P.W., Harding, G., Berry, P., Wiklund, I., Chen, W.H., and Kline Leidy, N. (2009). Development and first validation of the COPD Assessment Test. Eur. Respir. J. *34*, 648–654.
15. Liu, C.F., Huang, C.C., Wang, J.J., Kuo, K.M., and Chen, C.J. (2021). The Critical Factors Affecting the Deployment and Scaling of Healthcare AI: Viewpoint from an Experienced Medical Center. Healthcare (Basel) *9*, 685.
16. Liao, K.M., Liu, C.F., Chen, C.J., and Shen, Y.T. (2021). Machine Learning Approaches for Predicting Acute Respiratory Failure, Ventilator Dependence, and Mortality in Chronic Obstructive Pulmonary Disease. Diagnostics *11*, 2396.
17. Müllerová, H., Shukla, A., Hawkins, A., and Quint, J. (2014). Risk factors for acute exacerbations of COPD in a primary care population: a retrospective observational cohort studyBMJ. Open *4*, e006171.
18. Hurst, J.R., Vestbo, J., Anzueto, A., Locantore, N., Müllerova, H., Tal-Singer, R., Miller, B., Lomas, D.A., Agusti, A., Macnee, W., et al. (2010). Susceptibility to exacerbation in chronic obstructive pulmonary disease. N. Engl. J. Med. *363*, 1128–1138.
19. Jenkins, C.R., Celli, B., Anderson, J.A., Ferguson, G.T., Jones, P.W., Vestbo, J., Yates, J.C., and Calverley, P.M.A. (2012). Seasonality and determinants of moderate and severe COPD exacerbations in the TORCH study. Eur. Respir. J. *39*, 38–45.
20. Byng, D., Lutter, J.I., Wacker, M.E., Jörres, R.A., Liu, X., Karrasch, S., Schulz, H., Vogelmeier, C., and Holle, R. (2019). Determinants of healthcare utilization and costs in COPD patients: first longitudinal results from the german

COPD cohort COSYCONET. COPD *14*, 1423–1439.

21. Lee, S.J., Lee, S.H., Kim, Y.E., Cho, Y.J., Jeong, Y.Y., Kim, H.C., Kim, J.H., You, J.J., Yoon, C.H., Lee, J.D., and Hwang, Y.S. (2012). Clinical Features according to the Frequency of Acute Exacerbation in COPD. Tuberc. Respir. Dis. *72*, 367–373.
22. Cote, C.G., Dordelly, L.J., and Celli, B.R. (2007). Impact of COPD exacerbations on patient-centered outcomes. Chest *131*, 696–704.
23. Koç, Ç., and Şahin, F. (2022). What Are the Most Effective Factors in Determining Future Exacerbations, Morbidity Weight, and Mortality in Patients with COPD Attack? Medicina (Kaunas) *58*, 163.
24. Kim, J.K., Lee, S.H., Lee, B.H., Lee, C.Y., Kim, D.J., Min, K.H., Kim, S.K., Yoo, K.H., Jung, K.S., and Hwang, Y.I. (2016). Factors associated with exacerbation in mild- to-moderate COPD patients. Int. J. Chron. Obstruct. Pulmon. Dis. *11*, 1327–1333.
25. Lee, S.-D., Huang, M.-S., Kang, J., Lin, C.H., Park, M.J., Oh, Y.M., Kwon, N., Jones, P.W., and Sajkov, D.; Investigators of the Predictive Ability of CAT in Acute Exacerbations of COPD PACE Study (2014). The COPD assessment test (CAT) assists prediction of COPD exacerbations in high-risk patients. Respir. Med. *108*, 600–608.
26. Karloh, M., Fleig Mayer, A., Maurici, R., Pizzichini, M.M.M., Jones, P.W., and Pizzichini, E. (2016). The COPD Assessment Test: What Do We Know So Far?: A Systematic Review and Meta-Analysis About Clinical Outcomes Prediction and Classification of Patients Into GOLD Stages. Chest *149*, 413–425.
27. Bansode, P.M., Kumar, C., and M, S. (2022). Usefulness of Cat Score in Patients with Stable Copd and Acute Exacerbation of Copd and it's Co-Relation with PFT. J. Assoc. Physicians India *70*, 11–12.
28. Rovina, N., Symiakakis, M., Tsioka, A., Travlos, A., Vlachos, K., Koutsoukou, A., and Koulouris, N. (2013). COPD assessment test (CAT) in acute exacerbation of COPD and in the long term follow up of COPD patients. Eur. Respir. J. *42*, P780.
29. Lee, S.D., Huang, M.S., Kang, J., Lin, C.H., Park, M.J., Oh, Y.M., Kwon, N., Jones, P.W., and Sajkov, D.; Investigators of the Predictive Ability of CAT in Acute Exacerbations of COPD PACE Study (2014). The COPD assessment test (CAT) assists prediction of COPD exacerbations in high-risk patients. Respir. Med. *108*, 600–608.
30. Kor, C.T., Li, Y.R., Lin, P.R., Lin, S.H., Wang, B.Y., and Lin, C.H. (2022). Explainable Machine Learning Model for Predicting First-Time Acute Exacerbation in Patients with Chronic Obstructive Pulmonary Disease. J. Pers. Med. *12*, 228.
31. R Core Team (2021). R: A Language and Environment for Statistical Computing (R Foundation of Statistical Computing).
32. Buuren, S.v., and Groothuis-Oudshoorn, K. (2011). Mice: multivariate imputation by chained equations in R. J. Stat. Soft. *45*, 1–67. https://doi.org/10.18637/jss.v045.i03.
33. Chawla, N.V., Bowyer, K.W., Hall, L.O., and Kegelmeyer, W.P. (2002). SMOTE: Synthetic minority over-sampling technique. J. Artif. Int. Res. *16*, 321–357.
34. Zhang, L., Wang, Z., Zhou, Z., Li, S., Huang, T., Yin, H., and Lyu, J. (2022). Developing an ensemble machine learning model for early prediction of sepsis-associated acute kidney injury. iScience *25*, 104932.

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Deposited data		
Anonymized patient data	Chi Mei Medical Center & Chi Mei regional hospital	N/A
Software and algorithms		
Python codes for modeling	Intelligent Healthcare Center, Chi Mei Medical Center	https://github.com/Yuting0305/IDHChiMeiMedicalCenter
Python software version 3.10.9	Python Software Foundation	https://www.python.org/
Python software package: scikit-learn 1.2.1	https://scikit-learn.org/stable/	https://scikit-learn.org/stable/
Python software package: shap 0.41.0	Scott Lundberg. Revision 1cf6838f.	https://shap.readthedocs.io/en/latest/
Python software package: NumPy 1.23.5	https://numpy.org/	https://numpy.org/
Python software package: XGBoost 1.7.5	Xgboost developers. Revision 096047c5	https://xgboost.readthedocs.io/en/stable/
Python software package: Matplotlib 3.7.1	TheMatplotlib development team	https://matplotlib.org/
Python software package:Pandas 1.5.3	pandas via NumFOCUS, Inc. Hosted by OVHcloud.	https://pandas.pydata.org/
Python software package : imbalanced-learn 0.10.1	The imbalanced-learn developers. Created using Sphinx 6.0.0.	https://imbalanced-learn.org/stable/
R: A language and environment for statistical computing	R Core Team[31]	https://www.r-project.org/
mice (R package)	Van Buuren and Groothuis-Oudshoorn[32]	https://cran.r-project.org/web/packages/mice/index.html

RESOURCE AVAILABILITY

Lead contact

Further information and requests for data access should be directed to and will be fulfilled by the lead contact, Chung-Feng Liu (chungfengliu@gmail.com).

Materials availability

This paper did not generate new unique materials.

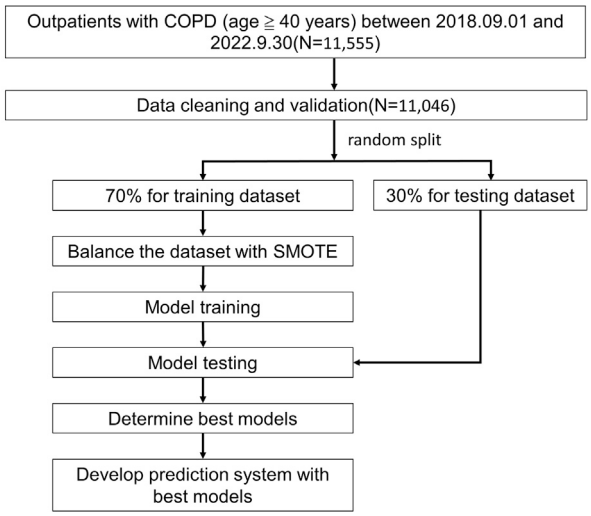
Data and code availability

- The anonymized data utilized in this study are not eligible for deposition in a public repository. However, the dataset can be made available upon reasonable request to the lead contact, adhering to the terms of the IRB and licensing agreements.
- All original code used for modeling in this study has been deposited on GitHub and is publicly available as of the date of publication. (GitHub: https://github.com/Yuting0305/IDHChiMeiMedicalCenter).
- Any additional information required to reanalyze the data reported in this paper is available from the lead contact upon request.

EXPERIMENTAL MODEL AND STUDY PARTICIPANT DETAILS

This retrospective study collected electronic medical records (EMR) from outpatients with COPD who participated in a government pay-for-performance (P4P) program at two hospitals of Chi Mei Medical Group in Taiwan (1 medical center and 1 regional hospital). The study period spanned from September 1, 2018, to September 30, 2022. The inclusion criteria for patients were those who were initially diagnosed within the range of J41-J44 (ICD-10-CM) and were aged 40 years or older at the time of diagnosis, enrolled in the P4P program, and had completed a minimum of three follow-up visits. Patients with incomplete or missing records were excluded.

The study adhered to the guidelines of the Declaration of Helsinki and received approval from the Institutional Review Board of Chi Mei Medical Center (IRB No. 11203-020; April 10, 2023). Informed consent from patients was waived due to the retrospective nature of the study. A total of 11,555 patients' data were included for analysis. Below figure illustrates the research flow.



Note. COPD: chronic obstructive pulmonary disease; SMOTE: synthetic minority over-sampling technique; HIS: hospital information system

Research flow

Feature and outcome variables

The predicted outcome of interest in this study is occurrence of AECOPD. Specifically, we aimed to predict the occurrence of AECOPD in COPD patients with ICD10 diagnoses J41-J44 or J12-J18, that is, those patients who experienced AECOPD from inpatient or emergency department within three and six months after outpatient visit.

Furthermore, based on literature evidence, [e.g., 6,7,16] clinical experience and usually available in clinical, we selected 26 features that affect the two outcome variables for ML modeling, including demographics of gender, age, body mass index (BMI), smoking history, live alone, family history of COPD; vital signs of systolic blood pressure(SBP), diastolic blood pressure(DBP), pulse, respiratory rate(RR), SpO2,Heart Rate(HR), pre-bronchodilator forced expiratory volume in one second(Pre-BD-FEV1(L)), post-bronchodilator forced expiratory volume in one second (post-BD-FEV1(L)), and Post-BD-FEV1 to forced vital Capacity ratio(FVC)(%); evaluation scores of modified Medical Research Council (mMRC) and COPD Assessment Test (CAT); medication of LAMA(Long-acting muscarinic antagonists), LABA(Long-acting beta2-agonist) and ICS(Inhaled corticosteroid); and disease history of Asthma, pulmonary tuberculosis(TB), hypertension, diabetes mellitus(DM), cardiovascular disease(CVD), chronic liver disease(CLD), and the history of AECOPD episodes. The history of AECOPD episodes was defined as the total count of a patient's emergency room visits or hospitalizations due to diagnosed COPD within one year prior to the clinic visit.

Model building and evaluation

In the process of constructing our predictive models, we followed a systematic approach, starting with data collection based on the specified criteria. Data cleaning and preprocessing were integral components of this endeavor. Data cleaning encompassed the identification and handling of missing values, erroneous entries (including those related to lung function, vital signs, medical history, CAT score, and mMRC score), and the removal of extreme outliers. Extreme outliers were defined as systolic blood pressure (SBP) values greater than or equal to 300 or less than 1, diastolic blood pressure (DBP) values greater than or equal to 200 or less than 1, and oxygen saturation (SPO2) values greater than 100.

To maximize model performance, we utilized all the selected features to build prediction models (full-feature model) without conducting a specific feature selection process beforehand. The dataset underwent minority-outcome stratification and was then randomly divided into a training dataset (70%) for model training, and a testing dataset (30%) for model testing. Due to the data imbalance with fewer positive classes (outcomes to be predicted) in the clinic, we employed the SMOTE method (synthetic minority over-sampling technique)[33] to improve data balance in the training dataset. The test dataset was not subjected to SMOTE preprocessing.

We employed five ML algorithms to construct the predictive models: (1) Logistic regression (LR), (2) Random forest (RF), (3) Support Vector Machine (SVM), (4) XGBoost, and (5) Multilayer perceptron (MLP). The implementation was carried out using the Python programming language and the Scikit-learn ML toolkits. Statistical analysis involved t-tests for numerical variables and Chi-square tests for categorical variables were used to R programming language.

For hyper-parameter (see Table S1) tuning and model construction, we utilized grid search with 5-fold cross-validation on the training dataset to determine the best models. After constructing the model, we assessed its performance on the test dataset using metrics including accuracy, sensitivity, specificity, and AUC (area under the receiver operating characteristic curve). Additionally, to optimize our predictive

model's efficacy, we employed ensemble machine learning strategies, notably stacking[34] and voting techniques. The model with the highest AUC value was considered the best model and selected for developing a practical prediction system.

METHOD DETAILS

This section elaborates on the methodology used in our study, covering the experimental setup, participant details, the variables involved, and the approach to building and evaluating the model.

QUANTIFICATION AND STATISTICAL ANALYSIS

Model building was conducted using the Python programming language on the scikit-learn platform. The graphic abstract and all figures were exported using Microsoft PowerPoint. Additionally, SHAP (SHapley Additive exPlanations) and Calibration plots were generated with specific Python packages. All statistical analyses were executed within Python programs.