Original Research

A Web-Based Risk Calculator for Psychological Distress in Stable Chronic Obstructive Pulmonary Disease: Development and Validation Study

Hui Chen, BSC[1*] Fengli Zuo, MSN[2*] Jie Zhang, BSC[1*] Feili Li, MSN[3] Fangrong Liu, BSC[1] Xu Tian, PhD[4] Rongying Tang, MSN[5]

[1]Outpatient Department, Chongqing University Cancer Hospital, Chongqing, China

[2]Department of Pulmonary Diseases, Chongqing Traditional Chinese Medicine Hospital, Chongqing, China

[3]Department of Nursing, Chongqing Traditional Chinese Medicine Hospital, Chongqing, China

[4]Division of Sci-Tech and International Affairs, Chongqing Traditional Chinese Medicine Hospital, Chongqing, China

[5]Department of Breast Cancer Center, Chongqing University Cancer Hospital, Chongqinq, China

**Contributed to the work equally*

Address correspondence to:
Xu Tian, PhD
Division of Sci-Tech and Foreign Affairs
Chongqing Traditional Chinese Medicine Hospital
No. 6, the 7th Branch of Panxi Road, Liangjiang New District
Chongqing 400021, China
Email: xu.tian@alumni.urv.cat; yxtx880919@hotmail.com

Rongying Tang, MSN
Department of Breast Cancer Center
Chongqing University Cancer Hospital
No. 181, Hanyu Road, Shapingba District
Chongqing 400030, China
Email: CQCH_TangRY@163.com

Running Head: **Machine Learning Distress Risk Prediction in COPD**

Copyright *Chronic Obstructive Pulmonary Diseases: Journal of the COPD Foundation* ©2026
Published online November 12, 2026 https://doi.org/10.15326/jcopdf.2026.0808

Keywords: chronic obstructive pulmonary disease; psychological distress; trajectory analysis; machine learning; prediction model

Abbreviations: **COPD**=chronic obstructive pulmonary disease; **GOLD**=Global Initiative for Chronic Obstructive Lung Disease; **FEV1**=forced expiratory volume in 1 second; **FVC**=forced vital capacity; **GMM**=growth mixture modeling; **DT**=Distress Thermometer; **CAT**=COPD Assessment Test; **mMRC**=modified Medical Research Council dyspnea scale; **COPD-Q**=Chronic Obstructive Pulmonary Disease Knowledge Questionnaire; **SCSQ**=Simplified Coping Style Questionnaire; **PSSS**=Perceived Social Support Scale; **PSS-10**=10-item Perceived Stress Scale; **SSCI-8**=8-item Stigma Scale for Chronic Illness; **STROBE**=Strengthening the Reporting of Observational Studies in Epidemiology; **LASSO**=least absolute shrinkage and selection operator; **ML**=machine learning; **SVM**=support vector machine; **KNN**=k-nearest neighbors; **RF**=random forest; **LightGBM**=light gradient boosting machine; **XGBoost**=extreme gradient boosting; **MLP**=multilayer perceptron; **AUC**=area under the receiver operating characteristic curve; **AUROC**=area under the receiver operating characteristic curve; **PPV**=positive predictive value; **NPV**=negative predictive value; **ROC**=receiver operating characteristic; **DCA**=decision curve analysis; **SHAP**=SHapley Additive exPlanations; **AIC**=Akaike Information Criterion; **BIC**=Bayesian Information Criterion; **aBIC**=adjusted Bayesian Information Criterion; **LMR**=Lo-Mendell-Rubin; **BLRT**=bootstrap likelihood ratio test; **AvePP**=average posterior probabilities; **GEE**=generalized estimating equations; **HADS** – Hospital Anxiety and Depression Scale

***Funding Support*:** This study was supported by the Chongqing Medical Scientific Research Project (Joint project of Chongqing Health Commission and Science and Technology Bureau) (Grant 2026QNXM010), the Natural Science Foundation of Chongqing (Grant CSTB2025NSCQ-GPX1096), and the 2025 Annual Technological Innovation Project of Shapingba District, Chongqing Municipality (Grant 2025141)

Date of Acceptance*:** October 16, 2026 | ***Published Online Date: November 12, 2026

Citation: Chen H, Zuo F, Zhang J, et al. A web-based risk calculator for psychological distress in stable chronic obstructive pulmonary disease: development and validation study. *Chronic Obstr Pulm Dis.* 2026; Published online November 12, 2026. https://doi.org/10.15326/jcopdf.2026.0808

Copyright *Chronic Obstructive Pulmonary Diseases: Journal of the COPD Foundation* ©2026

Abstract

Background: Patients with stable chronic obstructive pulmonary disease (COPD) experience psychological distress that evolves in dynamic and heterogeneous patterns. Conventional cross-sectional screening and baseline-only prediction models fail to capture longitudinal heterogeneity, and few tools integrate trajectory analysis with machine learning (ML) for individualized risk stratification. This study aimed to identify longitudinal trajectories of psychological distress and develop, validate, and deploy a web-based ML risk calculator for clinical use.

Methods: We enrolled 323 stable COPD outpatients from three tertiary hospitals in Chongqing, China, between March and December 2024. Baseline data included sociodemographic, clinical, and key psychosocial variables. Psychological distress was reassessed at 1 week, 1 month, 3 months, and 6 months. Growth mixture modeling (GMM) identified distinct longitudinal trajectory classes. The trajectory outcome was dichotomized for prediction: patients in the persistent low-distress group were defined as low-risk, and all others as high-risk. Candidate predictors were screened using univariate analysis and least absolute shrinkage and selection operator (LASSO) regression. Eight ML algorithms were trained, internally validated, and compared, and the best-performing model was selected for final development.

Results: GMM revealed four trajectories: persistent high distress (8.0%), transient threshold-exceeding distress with rapid decline (15.2%), persistent near-threshold distress with gradual decline (65.0%), and persistent low distress (11.8%). Using the persistent low-distress group as reference, nine predictors were retained: symptom burden, number of children, negative coping, payment method, perceived stress, acute exacerbation frequency, dyspnea severity, educational level, and gender. The XGBoost model achieved the best-balanced performance, with AUC values of 0.859 (training), 0.686 (validation), and 0.739 (test), indicating moderate discriminative ability. The model showed high positive predictive value but limited negative predictive value, constraining its utility for ruling out distress. A web-based calculator was

Copyright *Chronic Obstructive Pulmonary Diseases: Journal of the COPD Foundation* ©2026
Published online November 12, 2026 https://doi.org/10.15326/jcopdf.2026.0808

subsequently developed for clinical implementation.

Conclusion: Stable COPD patients follow four distinct longitudinal trajectories of psychological distress. The XGBoost model provides moderate predictive performance as an exploratory risk-stratification tool. The web-based calculator may support clinicians in early identification of individuals at elevated risk of persistent psychological distress, but further validation in independent, diverse cohorts is required before widespread clinical adoption.

Pre-Proof

Copyright *Chronic Obstructive Pulmonary Diseases: Journal of the COPD Foundation* ©2026
Published online November 12, 2026 https://doi.org/10.15326/jcopdf.2026.0808

Introduction

Chronic obstructive pulmonary disease (COPD) remains a major global health burden, characterized by persistent respiratory symptoms and progressive airflow limitation.[1] Worldwide, the prevalence of COPD is projected to exceed 600 million by 2050,[2] with annual deaths anticipated to surpass 5.4 million by 2060.[3] In China, more than 100 million people live with COPD,[4] posing substantial challenges to public health and socioeconomic systems.

Beyond physical impairment, the chronic and progressive nature of COPD exerts considerable psychological strain. Patients with COPD face a significantly higher risk of psychological distress - a broad, transdiagnostic state of overall emotional suffering - than the general population.[5] Cross-sectional studies report clinically significant distress in 42.9% of community-dwelling COPD patients and in 64% of those experiencing acute exacerbations,[6] indicating that psychological distress is substantial in the stable state and further elevated during acute exacerbations. According to Bonanno's framework,[7] exposure to chronic illness elicits heterogeneous long-term adaptation patterns rather than uniform responses. Distinct longitudinal distress trajectories have been identified in cancer, cardiovascular diseases, and other chronic conditions,[8-10] yet longitudinal evidence specifically mapping dynamic trajectories of psychological distress in stable COPD remains limited.[11-13] Most existing studies rely on cross-sectional assessments or simple mean-change models that overlook heterogeneity,[11,12] and few characterize how distress evolves during outpatient self-management after clinical stabilization.[13]

We therefore hypothesize that psychological distress in stable COPD follows heterogeneous longitudinal trajectories across 6 months of follow-up. Growth mixture modeling (GMM) is uniquely suited to identify unobserved latent subgroups with distinct temporal patterns.[14] Unlike traditional repeated-measures methods that assume homogeneous change, GMM captures heterogeneity in level, slope, and curvature,[15] making it ideal for delineating diverse adaptation pathways to chronic illness.

Copyright Chronic Obstructive Pulmonary Diseases: Journal of the COPD Foundation ©2026
Published online November 12, 2026 https://doi.org/10.15326/jcopdf.2026.0808

However, GMM results are not directly translatable to bedside risk assessment: they describe group-level patterns rather than generating individual-level predictions, and their clinical application requires advanced statistical expertise.[16] Machine learning (ML) offers a pragmatic solution by converting trajectory-defined risk into a clinically usable prediction tool.[17] Compared with conventional regression, ML better models non-linear relationships and high-dimensional predictor interactions, which are common in psychosocial and clinical data. Moreover, explainable ML approaches such as SHapley Additive Explanation (SHAP) improve transparency and clinical interpretability.[18] Hence, in this study, we employed ML techniques to predict the risk of psychological distress utilizing the classification results of GMM along with covariates. This dual approach of GMM and ML techniques not only provides a refined understanding of psychological distress but also enables the prediction of individual risk levels.[16]

The present study addresses the currently available gaps by: (1) characterizing the longitudinal evolution of psychological distress in stable COPD over 6 months; (2) developing a ML-based risk prediction model for psychological distress; and (3) building a user-friendly web-based calculator for clinical use.

Methods

Study design

We conducted a longitudinal observational study in outpatient clinics of three tertiary hospitals in Chongqing, China. The study was reported in accordance with the STROBE guideline for cohort studies.

Participants and setting

Assessments were scheduled at five time points: baseline (index clinic visit), 1 week, 1 month, 3 months, and 6 months. The one-week time point was included to capture distress during the early transition from clinic-supervised management to home-based self-care, when emotional strain was anticipated to be greatest. Patients were eligible if

Copyright Chronic Obstructive Pulmonary Diseases: Journal of the COPD Foundation ©2026
Published online November 12, 2026 https://doi.org/10.15326/jcopdf.2026.0808

they: (1) had a confirmed diagnosis of COPD per Global Initiative for Chronic Obstructive Lung Disease (GOLD) criteria (post-bronchodilator FEV1/FVC <0.70); (2) were in clinically stable condition (no acute exacerbation or hospitalization in the prior 4 weeks); (3) were aged ≥40 years; (4) had sufficient cognitive and communication ability to complete questionnaires; and (5) provided electronic informed consent. Exclusion criteria: (1) acute exacerbation or hospitalization within 4 weeks before enrollment; (2) severe comorbidities that could interfere with study outcomes, such as severe cardiovascular disease; (3) diagnosed severe mental illness or documented cognitive impairment; (4) development of an acute exacerbation of COPD during the 6-month follow-up period; and (5) participation in concurrent structured pulmonary rehabilitation or psychological intervention trials.

Sample size

Sample size was estimated using G*Power software.[19] With $\alpha = 0.05$, $\beta = 0.10$ and an assumed effect size $f = 0.10$, a minimum of 154 participants was required. Accounting for an expected 30% attrition rate in longitudinal follow-up, we aimed to recruit at least 200 patients at baseline.

Data collection

A total of 398 clinically stable outpatients with COPD were recruited between March and December 2024. At baseline (T1), participants completed a questionnaire battery covering sociodemographic characteristics, psychological distress, COPD knowledge, symptom burden, dyspnea severity, coping strategies, social support, perceived stress, and stigma. After data cleaning, 386 participants provided valid baseline data (12 were excluded for invalid or incomplete baseline questionnaires). Of these, 63 participants were lost to follow-up at the first follow-up assessment, with no further attrition thereafter. Consequently, 323 participants completed all five assessments and constituted the final analytic sample. Follow-up surveys were administered via the SoJump platform.

Instruments

Copyright Chronic Obstructive Pulmonary Diseases: Journal of the COPD Foundation ©2026
Published online November 12, 2026 https://doi.org/10.15326/jcopdf.2026.0808

Socio-demographic and clinical characteristics

A purpose-designed form was used to capture participants' sociodemographic and clinical data, including age, gender, marital status, number of children, place of residence, educational level, payment method, smoking and alcohol consumption history, monthly household income, body mass index (BMI), frequency of exercise, disease duration, Global Initiative for Chronic Obstructive Lung Disease (GOLD) stage, acute exacerbation frequency, and presence of comorbidities (recorded as any versus none).

Psychological distress

Psychological distress was measured using the Distress Thermometer (DT), a single-item visual analog scale ranging from 0 (no distress) to 10 (extreme distress).[20] Originally developed for psycho-oncology and endorsed as a routine screening tool, the DT has been validated in Chinese populations.[21] Because the optimal cut-off value is population-specific, we applied the ≥5 threshold that we previously derived and validated specifically in patients with stable COPD,[22] rather than carrying the oncology cut-off over unchanged. A DT score of ≥5 was therefore treated as a positive screen requiring further assessment, not as a psychiatric diagnosis.

Symptom burden

Symptom burden was evaluated using the COPD Assessment Test (CAT),[23] a GOLD-recommended instrument for symptom assessment. The CAT comprises eight items, each scored from 0 to 5, yielding a total score range of 0–40. Higher total scores reflect a greater symptom burden.

Dyspnea severity

The modified Medical Research Council (mMRC) dyspnea scale was employed to grade breathlessness, with scores ranging from 0 (only during strenuous exercise) to 4 (breathlessness at rest).[24] Higher scores indicate more severe dyspnea.

COPD knowledge

The COPD Knowledge Questionnaire (COPD-Q) consists of 13 items covering clinical

Copyright Chronic Obstructive Pulmonary Diseases: Journal of the COPD Foundation ©2026
Published online November 12, 2026 https://doi.org/10.15326/jcopdf.2026.0808

manifestations, risk factors, medication use, oxygen therapy, prevention, and disease outcomes.[25] Each correct answer receives a score of 1, while incorrect or "do not know" responses receive 0. Total scores thus range from 0 to 13, with higher scores denoting better disease-related knowledge.

Coping style

Coping style was assessed using the Simplified Coping Style Questionnaire (SCSQ), a 20-item scale comprising two subscales: positive coping (12 items) and negative coping (8 items).[26] Each item is rated on a 0–3 scale, producing subscale scores of 0–36 for positive coping and 0–24 for negative coping. Higher scores on each subscale indicate greater use of that particular coping style.

Social support

The Perceived Social Support Scale (PSSS),[27] adapted for Chinese populations by Jiang,[28] contains 12 items measuring three domains: family, friends, and significant others. Items are answered on a 7-point Likert scale, with total scores ranging from 12 to 84. Higher scores reflect stronger perceived social support.

Perceived stress

The 10-item Perceived Stress Scale (PSS-10) was used to assess stress experienced over the preceding month.[29] Each item is scored on a 5-point scale, giving a total score between 0 and 40, with higher scores indicating greater perceived stress. The Chinese version of the PSS-10 has demonstrated sound psychometric properties.[30]

Perceived stigma

Perceived stigma was measured using the 8-item Stigma Scale for Chronic Illness (SSCI-8). Items are rated from 1 (never) to 5 (always), yielding total scores between 8 and 40.[31] Higher scores denote a higher level of perceived stigma. The Chinese version has established reliability and validity in chronic illness populations.[32]

Data Analysis

Descriptive and univariate analysis

Continuous variables are reported as mean ± standard deviation, and categorical

Copyright Chronic Obstructive Pulmonary Diseases: Journal of the COPD Foundation ©2026
Published online November 12, 2026 https://doi.org/10.15326/jcopdf.2026.0808

variables as frequencies (percentages). A complete-case analysis was used in this study: participants who did not complete all five assessments were excluded from the analyses, and no data imputation was performed. Univariate analyses, employing the Mann-Whitney U test, chi-square test, or Fisher-Freeman-Halton test as appropriate, were performed to identify potential factors associated with psychological distress. All analyses were conducted using SPSS version 26.0 (IBM Corp., Armonk, NY, USA), and a two-tailed p-value <0.05 was considered statistically significant.

Longitudinal trajectory pattern analysis

GMM was performed in Mplus 8.8 to identify distress trajectories over time. Model selection followed established recommendations[15,33] based on: (1) fit indices (Akaike Information Criteria [AIC], Bayes Information Criteria [BIC], and sample-size adjusted BIC [aBIC], with lower values indicating better fit; (2) likelihood ratio tests (Lo-Mendell-Rubin [LMR] test and adjusted LMR likelihood-ratio test [LRT]), with significant p-values favoring the *K*-class model over *K*-1; (3) classification accuracy (entropy near 1.0, average posterior probabilities [AvePP] >0.70, and class proportions >5%). The final model was chosen based on statistical fit, parsimony, and interpretability.

Feature selection

Consistent with prior recommendations,[34] participants following a "persistent low distress" trajectory were classified as non-cases, while all others were classified as cases. Predictor selection was performed using least absolute shrinkage and selection operator (LASSO) regression, which shrinks coefficients of irrelevant predictors to zero.[35] The penalty parameter (lambda) was optimized via 10-fold cross-validation (LassoCV), and predictors with non-zero coefficients were retained for further analysis.

Machine learning model development

Eight machine learning algorithms were trained and compared: logistic regression, support vector machine (SVM), k-nearest neighbors (KNN), random forest (RF), light gradient boosting machine (LightGBM), extreme gradient boosting (XGBoost),

Copyright Chronic Obstructive Pulmonary Diseases: Journal of the COPD Foundation ©2026
Published online November 12, 2026 https://doi.org/10.15326/jcopdf.2026.0808

decision tree, and multilayer perceptron (MLP). Model implementation was carried out in Python 3.7 using the xgboost 1.2.1, lightgbm 3.2.1, and scikit-learn 0.22.1 libraries. The dataset was randomly partitioned into training (70%) and testing (30%) sets, with bootstrap resampling applied during the training phase. Model performance on the validation set was evaluated using the area under the receiver operating characteristic curve (AUROC), accuracy, sensitivity, specificity, positive predictive value (PPV), negative predictive value (NPV), and F1 scores. Clinically relevant thresholds were defined as: AUROC ≥0.70,[36] sensitivity and specificity ≥0.70, PPV and NPV ≥0.70, and F1 scores ≥0.75. Calibration was assessed by plotting predicted distress incidence against observed outcomes.

Model optimization and evaluation

Five-fold cross-validation was employed to ensure generalizability. Discriminative ability was evaluated via ROC analysis and the resulting AUC. Calibration plots were used to compare predicted probabilities with observed outcomes. Decision curve analysis (DCA) was performed to estimate the clinical net benefit.[37] Feature importance was interpreted using SHAP, where larger absolute SHAP values indicate a greater contribution to the prediction.[38]

Ethical considerations

The study protocol was approved by the Institutional Review Board of Chongqing Traditional Chinese Medicine Hospital (Approval No. 2024-KY-HY-35). The study was conducted in accordance with the Declaration of Helsinki. Written informed consent was obtained from all participants prior to enrolment. Participation was voluntary, and all data were kept confidential.

Results

Participant characteristics

Of the 398 patients who completed the baseline survey, 12 were excluded due to invalid or incomplete responses. The remaining 386 individuals were invited to participate in

Copyright Chronic Obstructive Pulmonary Diseases: Journal of the COPD Foundation ©2026
Published online November 12, 2026 https://doi.org/10.15326/jcopdf.2026.0808

the follow-up assessments, of whom 63 declined after the first wave. No patient was lost from the subsequent follow-up points, a total of 323 patients therefore completed all study measures and were included in the final analytical sample. Table 1 summarizes the baseline demographic and clinical characteristics. The sample was predominantly female (79.3%), with a mean age of 63.7 years (SD = 9.0) and an average disease duration of 13.3 years (SD = 4.1). The majority lived in urban areas (81.1%) and had completed only primary education or less (68.1%). Monthly household income was below ¥3,000 for 39.3% of participants, and 72.8% relied on out-of-pocket payment for healthcare. Clinically, 75.6% of patients were classified as GOLD stages III or IV, and the mean annual frequency of acute exacerbations was 2.51 (SD = 1.08). Comorbidities were present in 52.3% of the cohort. A history of smoking was reported by 51.4%, alcohol use by 47.1%, and only 19.8% engaged in regular physical exercise. Body mass index (BMI) distribution was balanced across underweight, normal, and overweight categories.

Baseline psychosocial variables

At baseline, the mean Distress Thermometer (DT) score was 3.86 (SD = 1.50). The mean COPD Assessment Test (CAT) score was 20.07 (SD = 7.49), indicating a substantial symptom burden. The mean modified Medical Research Council (mMRC) dyspnea score was 1.96 (SD = 0.77). Additional baseline measures included the COPD-Q (mean = 6.38, SD = 4.01), coping style (mean = 30.15, SD = 13.06), perceived social support (mean = 14.17, SD = 1.86), perceived stress (mean = 19.34, SD = 7.94), and perceived stigma (mean = 23.60, SD = 6.59).

Longitudinal trend of psychological distress

The mean DT scores across the five time points were as follows: 3.86 (SD = 1.50) at T1, 4.72 (SD = 1.17) at T2, 4.34 (SD = 1.72) at T3, 4.06 (SD = 1.28) at T4, and 3.76 (SD = 1.20) at T5. A generalized estimating equation (GEE) model with distress as the dependent variable and time as the independent factor revealed a significant main effect of time (Wald $\chi^2 = 1143.046$, $p < 0.001$), indicating that distress levels varied

Copyright Chronic Obstructive Pulmonary Diseases: Journal of the COPD Foundation ©2026
Published online November 12, 2026 https://doi.org/10.15326/jcopdf.2026.0808

significantly over the study period.

Identification of four distinct distress trajectories

GMM was applied to test one- to five-class solutions (Table 2). All models yielded entropy values above 0.9. The BLRT was significant for all class solutions, whereas the LMR test was significant for one- to four-class models. The AIC, BIC, and aBIC decreased progressively from one to four classes but increased from four to five classes. Based on statistical fit indices, interpretability, and parsimony, the four-class model was selected (Figure 1). The average latent class probabilities for most likely class membership ranged from 0.996 to 1.000 (Table 3), indicating excellent classification accuracy. With the DT cut-off of ≥5 defining clinically significant distress, the four trajectories were labeled as follows: (1) persistent high distress (scores ≥5 throughout follow-up), (2) transient threshold-exceeding distress with rapid decline (scores ≥5 early in follow-up that then fell below 5), (3) persistent near-threshold distress with gradual decline (scores of approximately 3.5-4.5, just below the clinical threshold), and (4) persistent low distress (scores below 3.5). Among these, the persistent low distress trajectory was the only one that achieved sustained recovery to a low-distress state, and it therefor served as the reference ("non-case") group; the other three trajectories, which did not achieve this recovery, were pooled as "cases." The persistent near-threshold trajectory was classified as a case on the basis of its time course rather than its absolute level, since distress that persisted just below the clinical threshold throughout follow-up signified incomplete adaptation rather than recovery, and therefore a state warranting monitoring.

Feature selection for trajectory classification

Univariate analyses identified 15 variables that differed significantly across the four trajectory classes: gender, age, residence, education, payment method, number of children, disease duration, GOLD stage, acute exacerbation frequency, CAT score, mMRC dyspnea score, SCSQ-Positive score, SCSQ-Negative score, PSS-10 score, and SSCI-8 score (Table 4). These variables were entered into a least absolute shrinkage

Copyright Chronic Obstructive Pulmonary Diseases: Journal of the COPD Foundation ©2026
Published online November 12, 2026 https://doi.org/10.15326/jcopdf.2026.0808

and selection operator (LASSO) regression with 10-fold cross-validation. Nine predictors retained non-zero coefficients: gender, number of children, education, payment method, acute exacerbation frequency, CAT score, mMRC score, SCSQ-Negative score, and PSS-10 score (Figure 2).

Comparative performance of classification models

Table 5 and Figure 3 summarize the performance of eight ML models on the training and validation sets. The XGBoost model achieved the best predictive performance overall. Receiver operating characteristic (ROC) curves (Figures 3A, 3B) suggested that KNN and RF were more prone to overfitting, whereas XGBoost exhibited greater stability. Calibration plots (Figure 3C) demonstrated good agreement between predicted probabilities and observed outcomes. A forest plot (Figure 3D) provides a summary of ROC results across all models.

Optimization of the XGBoost model

Performance metrics for the final XGBoost model are presented in Table 6. The AUC values were 0.859 for the training set, 0.686 for the validation set, and 0.739 for the test set, indicating moderate discriminative ability (Figures 4A–4C). The learning curve (Figure 4D) showed convergence of training and validation errors, suggesting no underfitting or overfitting. The calibration curve (Figure 4E) indicated close alignment between predicted and observed probabilities. DCA (Figure 4F) revealed that the XGBoost model provided greater net benefit than either the “treat-all” or “treat-none” strategies across a range of threshold probabilities.

Model interpretation

SHAP summary plots (Figure 5A) illustrate the contribution of each feature to the model’s predictions. Bar graphs (Figure 5B) rank feature importance by mean absolute SHAP values. Individual patient-level explanations are provided for two example cases (Figures 5C, 5D), detailing how specific feature values shift the predicted risk of psychological distress upward or downward.

Model presentation

Copyright Chronic Obstructive Pulmonary Diseases: Journal of the COPD Foundation ©2026
Published online November 12, 2026 https://doi.org/10.15326/jcopdf.2026.0808

The final XGBoost model was implemented as an interactive web application using the R Shiny package. This tool allows clinicians to input patient data and obtain an estimated risk score for psychological distress, thereby supporting clinical decision-making. A preview of the user interface is shown in Figure 6.

Discussion

This study delineated four distinct longitudinal trajectories of psychological distress among patients with stable COPD and further developed a ML-based predictive model to identify individuals at elevated risk. Among eight ML algorithms evaluated, XGBoost demonstrated superior predictive performance. To facilitate bedside or outpatient application, we also constructed a web-based risk calculator.

Psychological distress in COPD is well recognized to correlate with diminished health-related quality of life, suboptimal self-management, and heightened risks of acute exacerbations, hospital admissions, and mortality.[39] Although earlier cross-sectional studies have documented the prevalence and associated factors of distress,[11-13] its evolution over time has remained poorly characterized. In our cohort, distress levels rose shortly after the clinic visit (within one week) and then gradually subsided over a six-month period—a pattern analogous to that reported in patients with oesophageal cancer undergoing radiotherapy[40] and in ovarian cancer patients.[41] From the perspective of stress-adaptation theory,[42] the initial increase likely reflects an acute stress response triggered by the transition from direct clinical supervision to home-based self-care, a shift accompanied by reduced environmental security, heightened illness awareness following the comprehensive baseline assessment (including spirometry and symptom questionnaires), and the emergence of new self-management demands. The subsequent decline may represent successful adaptation, mediated by cognitive reappraisal of the illness and a growing sense of self-efficacy.[43] Given that the one-week time point emerged as an inflection point for distress, we propose a dual-phase supportive strategy: structured symptom monitoring

Copyright Chronic Obstructive Pulmonary Diseases: Journal of the COPD Foundation ©2026
Published online November 12, 2026 https://doi.org/10.15326/jcopdf.2026.0808

coupled with dyspnea management training at the time of discharge, followed by proactive tele-assessment one week after the clinic visit.

Using GMM, we identified four heterogeneous adaptation pathways. A small subgroup (8.0%) exhibited persistently high distress and may benefit from targeted psychosocial interventions, such as mindfulness-based programmes[44] or cognitive behavior therapy.[45] Another subgroup (15.2%) showed transient distress that exceeded the clinical threshold only during the care transition; for these patients, timely interventions—for instance, mindfulness-based breathing exercises initiated at discharge[46] could enhance emotional regulation. The largest subgroup maintained subthreshold distress that gradually declined yet remained near the clinical cutoff for three months, highlighting the need for sustained health education and psychological support to avert progression. The predominance of this near-threshold class (65.0%) suggests that, for most patients with stable COPD, distress persists as a chronic, low-grade emotional strain that fluctuates just below the clinical threshold rather than remitting fully, reflecting the persistent yet partially manageable burden of living with a progressive chronic illness. Only a minority (11.8%) sustained consistently low distress levels; for this group, continued emotional and informational reinforcement is warranted to preserve psychological well-being.

Early recognition of high-risk patients enables targeted interventions aimed at improving both psychological and clinical outcomes. Although tools such as the DT and the Hospital Anxiety and Depression Scale (HADS) are valuable for cross-sectional screening, they lack the capacity to forecast future distress. Conventional modelling approaches often struggle with high-dimensional data and non-linear relationships.[47] ML algorithms, by contrast, can accommodate a broader array of predictors and model complex interactions more effectively,[48] d which explains their growing adoption in medical prognosis. In the present study, the XGBoost-based model achieved area under the curve (AUC) values of 0.859, 0.686, and 0.739 in the training, validation, and test sets, respectively, indicating moderate discriminative performance rather than high

Copyright Chronic Obstructive Pulmonary Diseases: Journal of the COPD Foundation ©2026
Published online November 12, 2026 https://doi.org/10.15326/jcopdf.2026.0808

accuracy. Calibration and DCA provided preliminary support for reliability, yet the marked decline from training to validation AUC suggests overfitting, consistent with observations in other modest-sample ML prediction studies. Furthermore, the model yielded high PPV but very low NPV, meaning it can effectively identify high-risk individuals but cannot reliably exclude low risk, which constrains its clinical utility as a rule-out tool. To address the "black box" concern commonly raised with ML, we applied SHapley Additive exPlanations (SHAP) to elucidate predictor contributions at the individual level,[49] thereby enabling clinicians to tailor interventions based on modifiable factors highlighted by the model. Notably, several core predictors (CAT, mMRC, and PSS-10) overlapped with variables defining GMM trajectory classes, which introduces circularity that may partially inflate predictive performance; this analytical constraint should be considered when interpreting model validity.

A key strength of this study is the novel integration of GMM and ML to capture dynamic distress trajectories while providing individualized risk prediction—an advance over conventional statistical approaches. Nevertheless, several limitations should be acknowledged. First, the convenience sample was recruited from a single region (three hospitals in Chongqing) and was predominantly female (79.3%), which limits the generalizability of findings to broader, more sex-balanced COPD populations; 63 participants (16.3%) were also lost to follow-up, and although the retained sample supported the planned analyses, differential attrition cannot be fully excluded. Second, the six-month follow-up is relatively brief; longer observation would be needed to fully delineate long-term trajectories. Third, the sample size (n = 323) is below the n >500 heuristic recommended for latent class models and is modest for robust ML model development and validation, which may contribute to overfitting and performance instability. Fourth, distress was measured with a screening instrument rather than a gold-standard diagnostic interview, and comorbidity was recorded as a binary (any versus none) variable, so the severity of distress and the number of comorbidities could not be determined. Fifth, all data were self-reported, introducing possible recall and

Copyright Chronic Obstructive Pulmonary Diseases: Journal of the COPD Foundation ©2026
Published online November 12, 2026 https://doi.org/10.15326/jcopdf.2026.0808

social desirability biases. Sixth, LASSO feature selection was performed on the full dataset prior to data splitting, which raises risks of data leakage without nested cross-validation. Finally, the conversion of multi-class trajectories into binary outcome simplifies complex longitudinal patterns and may reduce conceptual validity, and unmeasured confounders, such as medication adherence, family dynamics, and genetic factors, may influence distress trajectories.

Conclusions

This study identified four distinct longitudinal trajectories of psychological distress in patients with stable COPD. We further developed an ML-based predictive model with moderate, exploratory predictive accuracy for identifying distress risk. The model has been translated into an online risk calculator to assist healthcare professionals in early recognition of high-risk patients and delivery of targeted psychological support. Widespread clinical application requires further validation in larger, independent, and demographically diverse cohorts before definite clinical utility can be established.

Implications for clinical practice

The web-based risk calculator offers a practical tool for early stratification of stable COPD patients at elevated risk of psychological distress. By integrating routinely available predictors—including symptom burden, dyspnea severity, negative coping style, and perceived stress—the tool categorizes patients into distinct risk tiers. In clinical use, a clinician enters the patient's nine predictor values into the interface, and the calculator returns an estimated probability of belonging to an elevated-distress trajectory to guide triage. Several retained predictors (female sex, lower educational level, and a larger number of children) are non-modifiable and instead identify subgroups that warrant closer monitoring, whereas the modifiable predictors (symptom burden, dyspnea severity, negative coping style, and perceived stress) are actionable targets for intervention. A high predicted risk should prompt a formal psychological

Copyright Chronic Obstructive Pulmonary Diseases: Journal of the COPD Foundation ©2026
Published online November 12, 2026 https://doi.org/10.15326/jcopdf.2026.0808

assessment (for example, the PHQ-9 for depression or the GAD-7 for anxiety) or referral to psychological support, whereas scores in the border zone around the decision threshold should trigger brief follow-up monitoring rather than immediate referral, because the model's limited negative predictive value means it cannot reliably rule out distress. This stratification enables more efficient allocation of limited psychosocial support resources, ensuring that interventions are directed toward those most likely to benefit. The online interface supports point-of-care use during routine follow-up visits, facilitating timely clinical decisions and fostering a more personalized, proactive approach to managing psychological comorbidity in COPD.

Implications for future research

Further external validation in larger, multicenter, sex-balanced cohorts is required to confirm the generalizability of the prediction model across diverse clinical settings and demographic groups. Extending follow-up beyond six months would clarify long-term distress trajectories and the durability of risk predictions. Research should also explore the mechanisms linking predictors such as negative coping and perceived stress to persistent distress, potentially uncovering more precise intervention targets. Methodological improvements including nested cross-validation, stricter anti-leakage protocols, and reduction of predictor-outcome circularity are needed to enhance model robustness. Finally, implementation studies examining the integration of the risk calculator into clinical workflows and its impact on patient-centered outcomes (e.g., quality of life, exacerbation rates, and healthcare utilization) are essential to translate this tool into tangible clinical benefit.

Copyright *Chronic Obstructive Pulmonary Diseases: Journal of the COPD Foundation* ©2026
Published online November 12, 2026 https://doi.org/10.15326/jcopdf.2026.0808

Acknowledgements

The authors would like to thank the participants for contributing to the study. This work is supported by the Extreme Smart Analysis platform (http://xsmartanalysis.cn/). We would also like to thank Deepseek V3.2 for its valuable assistance in refining our academic writing.

Contributions

H.C., R.Y.T., J.Z., F.R.L. and X.T.: Conceptualization; H.C., R.Y.T. and J.Z.: Writing - original draft, Data Collection, Formal analysis, Methodology; H.C., R.Y.T., J.Z., and F.L.Z.: Data Collection, Data curation; F.R.L. and X.T.: Validation; Visualization; F.L.L.: Supervision and Methodology; F.R.L. and X.T.: Writing - review & editing; All authors reviewed the manuscript.

Consent to participate

Informed consent was obtained from all individual participants included in the study.

Consent for publication

Not applicable.

Data availability statement

The data that support the findings of this study are available from the corresponding author upon reasonable request.

Conflict of interest

The authors declare no competing interests.

Pre-Proof

Copyright *Chronic Obstructive Pulmonary Diseases: Journal of the COPD Foundation* ©2026
Published online November 12, 2026 https://doi.org/10.15326/jcopdf.2026.0808

References

1. Celli B, Fabbri L, Criner G, et al. Definition and Nomenclature of Chronic Obstructive Pulmonary Disease: Time for Its Revision. *Am J Respir Crit Care Med.* 2022;206(11):1317-1325.
2. Boers E, Barrett M, Su JG, et al. Global Burden of Chronic Obstructive Pulmonary Disease Through 2050. *JAMA Network Open.* 2023;6(12):e2346598-e2346598.
3. Agustí A, Celli BR, Criner GJ, et al. Global Initiative for Chronic Obstructive Lung Disease 2023 Report: GOLD Executive Summary. *Eur Respir J.* 2023;61(4).
4. Wang C, Xu J, Yang L, 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.* 2018;391(10131):1706-1717.
5. Wagena EJ, Arrindell WA, Wouters EF, van Schayck CP. Are patients with COPD psychologically distressed? *Eur Respir J.* 2005;26(2):242-248.
6. Andenaes R, Kalfoss MH. Psychological distress in hospitalized patients with chronic obstructive pulmonary disease. *Eur J Epidemiol.* 2004;19(9):851-859.
7. Bonanno GA, Westphal M, Mancini AD. Resilience to loss and potential trauma. *Annu Rev Clin Psychol.* 2011;7:511-535.
8. Wang X, Zou W, Du Q, et al. Longitudinal Trajectories of Psychological Distress and Its Influence Factors in Young and Middle-Aged Patients With Colorectal Cancer. *J Adv Nurs.* 2025;81(7):3929-3942.
9. Li J, Yin C, Ye M, Kang L, Zhu J, Huang L. Psychological Distress in Surgical Lung Cancer Patients: Trajectories, Influencing Factors and Impacts on Quality of Life. *Journal of clinical nursing.* 2025.
10. Zhu YQ, Sun X, Wen J. Longitudinal study on the trajectory of psychological pain in pulmonary tuberculosis patients. *Chinese Evidence-based Nursing.* 2025;11(10):1992-1997.
11. Wang C, Yan J, Ma C. Psychological distress and its associated factors among patients with chronic obstructive pulmonary disease in Hunan, China: a cross-sectional study. *Sci Rep.* 2023;13(1):5199.
12. Lan M, Yang L, Zhang H, Su A, Yin Q, Li J. A Structural Equation Model of the Relationship Between Symptom Burden, Psychological Resilience, Coping Styles, Social Support, and Psychological Distress in Elderly Patients With Acute Exacerbation Chronic Obstructive Pulmonary Disease in China. *Asian Nurs Res (Korean Soc Nurs Sci).* 2024;18(3):231-237.
13. Tian X, Yi L, Liu X, et al. Relationship Between Social Support, Symptom Burden, Dyspnea, Perceived Stress, Perceived Stigma, Coping Styles, and Psychological Distress in Patients with Stable COPD: A Structural Equation Model. *Int J Chron Obstruct Pulmon Dis.* 2025;20:2183-2198.
14. Miettunen J, Nordström T, Kaakinen M, Ahmed AO. Latent variable mixture modeling in psychiatric research--a review and application. *Psychol Med.*

Copyright Chronic Obstructive Pulmonary Diseases: Journal of the COPD Foundation ©2026
Published online November 12, 2026 https://doi.org/10.15326/jcopdf.2026.0808

2016;46(3):457-467.
15. Ram N, Grimm KJ. Growth Mixture Modeling: A Method for Identifying Differences in Longitudinal Change Among Unobserved Groups. *Int J Behav Dev.* 2009;33(6):565-576.
16. Tian X, Li H, Li F, Jiménez-Herrera MF, Ren Y, Shang H. Development and validation of a web-based calculator for determining the risk of psychological distress based on machine learning algorithms: A cross-sectional study of 342 lung cancer patients. *Supportive care in cancer : official journal of the Multinational Association of Supportive Care in Cancer.* 2024;33(1):63.
17. Gao L, Cao Y, Cao X, et al. Machine learning-based algorithms to predict severe psychological distress among cancer patients with spinal metastatic disease. *Spine J.* 2023;23(9):1255-1269.
18. Ning Y, Ong MEH, Chakraborty B, et al. Shapley variable importance cloud for interpretable machine learning. *Patterns (N Y).* 2022;3(4):100452.
19. Faul F, Erdfelder E, Lang AG, Buchner A. G*Power 3: a flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behav Res Methods.* 2007;39(2):175-191.
20. Roth AJ, Kornblith AB, Batel-Copel L, Peabody E, Scher HI, Holland JC. Rapid screening for psychologic distress in men with prostate carcinoma. *Cancer.* 1998;82(10):1904-1908.
21. Tang LL, Zhang YN, Pang Y, Zhang HW, Song LL. Validation and reliability of distress thermometer in chinese cancer patients. *Chin J Cancer Res.* 2011;23(1):54-58.
22. Tian X, Liu X, Gan X, Jimenez-Herrera MF, Shang H, Ren Y. Validation of the Distress Thermometer for detecting psychological distress in patients with stable chronic obstructive pulmonary disease: optimal cut-off score and influencing factors. *BMC Pulm Med.* 2025;25(1):377.
23. Finch S, Laska IF, Abo-Leyah H, Fardon TC, Chalmers JD. Validation of the COPD Assessment Test (CAT) as an Outcome Measure in Bronchiectasis. *Chest.* 2020;157(4):815-823.
24. Bestall JC, Paul EA, Garrod R, Garnham R, Jones PW, Wedzicha JA. Usefulness of the Medical Research Council (MRC) dyspnoea scale as a measure of disability in patients with chronic obstructive pulmonary disease. *Thorax.* 1999;54(7):581-586.
25. Maples P, Franks A, Ray S, Stevens AB, Wallace LS. Development and validation of a low-literacy Chronic Obstructive Pulmonary Disease knowledge Questionnaire (COPD-Q). *Patient Educ Couns.* 2010;81(1):19-22.
26. Folkman S, Lazarus RS. Coping as a mediator of emotion. *Journal of Personality and Social Psychology.* 1988;54(3):466-475.
27. Blumenthal JA, Burg MM, Barefoot J, Williams RB, Haney T, Zimet G. Social support, type A behavior, and coronary artery disease. *Psychosom Med.* 1987;49(4):331-340.

Pre-Proof

Copyright *Chronic Obstructive Pulmonary Diseases: Journal of the COPD Foundation* ©2026
Published online November 12, 2026 https://doi.org/10.15326/jcopdf.2026.0808

28. Zhang Y, Jia Y, MuLaTiHaJi M, et al. A cross-sectional mental-health survey of Chinese postgraduate students majoring in stomatology post COVID-19 restrictions. *Front Public Health.* 2024;12:1376540.
29. Cohen S, Kamarck T, Mermelstein R. Perceived stress scale. Measuring stress: a guide for health and social scientists. *J Health Soc Behav.* 1994;10(2):1-2.
30. Liu X, Zhao Y, Li J, Dai J, Wang X, Wang S. Factor Structure of the 10-Item Perceived Stress Scale and Measurement Invariance Across Genders Among Chinese Adolescents. *Front Psychol.* 2020;11:537.
31. Molina Y, Choi SW, Cella D, Rao D. The stigma scale for chronic illnesses 8-item version (SSCI-8): development, validation and use across neurological conditions. *Int J Behav Med.* 2013;20(3):450-460.
32. Jiang N, Zhang Y-X, Zhao J, et al. The mediator role of stigma in the association of mindfulness and social engagement among breast cancer survivors in China. *Supportive Care in Cancer.* 2022;30(6):5007-5015.
33. Kim SY. Determining the Number of Latent Classes in Single- and Multi-Phase Growth Mixture Models. *Struct Equ Modeling.* 2014;21(2):263-279.
34. Fu H, Si L, Guo R. What Is the Optimal Cut-Off Point of the 10-Item Center for Epidemiologic Studies Depression Scale for Screening Depression Among Chinese Individuals Aged 45 and Over? An Exploration Using Latent Profile Analysis. *Front Psychiatry.* 2022;13:820777.
35. Vasquez MM, Hu C, Roe DJ, Chen Z, Halonen M, Guerra S. Least absolute shrinkage and selection operator type methods for the identification of serum biomarkers of overweight and obesity: simulation and application. *BMC Med Res Methodol.* 2016;16(1):154.
36. Mandrekar JN. Receiver operating characteristic curve in diagnostic test assessment. *Journal of thoracic oncology : official publication of the International Association for the Study of Lung Cancer.* 2010;5(9):1315-1316.
37. Vickers AJ, Elkin EB. Decision curve analysis: a novel method for evaluating prediction models. *Med Decis Making.* 2006;26(6):565-574.
38. Lundberg SM, Erion G, Chen H, et al. From local explanations to global understanding with explainable AI for trees. *Nature machine intelligence.* 2020;2(1):56-67.
39. Yu T, Frei A, Ter Riet G, Puhan MA. Impact of Stressful Life Events on Patients with Chronic Obstructive Pulmonary Disease. *Respiration.* 2018;95(2):73-79.
40. Li J, Yang HM, Li XF, Zhang BY, Yao Y. Trajectories of psychological distress and influencing factors in patients undergoing radiotherapy for oesophageal cancer. *Chinese Journal of Nursing Education.* 2024;21(10):1255-1263.
41. Yu Q, Weng ZZ, HUang F, Li Y, Chen XQ. Longitudinal study on the level of psychological distress and its influencing factors in ovarian cancer patients. *Chinese Evidence-based Nursing.* 2023;9(7):1290-1293.
42. Slade L, Pulinilkunnil T. The MiTF/TFE Family of Transcription Factors:

Copyright *Chronic Obstructive Pulmonary Diseases: Journal of the COPD Foundation* ©2026
Published online November 12, 2026 https://doi.org/10.15326/jcopdf.2026.0808

Master Regulators of Organelle Signaling, Metabolism, and Stress Adaptation. *Mol Cancer Res.* 2017;15(12):1637-1643.

43. Lim KE, Kim SR, Kim HY, Kim SR, Lee YC. Self-management model based on information-motivation-behavioral skills model in patients with chronic obstructive pulmonary disease. *J Adv Nurs.* 2022;78(12):4092-4103.
44. Tian X, Liao Z, Yi L, Tang L, Chen G, Jiménez Herrera MF. Efficacy and mechanisms of 4-week MBSR on psychological distress in lung cancer patients: A single-center, single-blind, longitudinal, randomized controlled trial. *Asia Pac J Oncol Nurs.* 2023;10(1):100151.
45. Arefian M, Asgari-Mobarake K. Mindfulness-integrated Cognitive Behavioral Therapy reduces pain and psychological distress, and improves equanimity, hope and post-traumatic growth during breast cancer treatment: A pilot randomized controlled trial. *Eur J Oncol Nurs.* 2025;76:102881.
46. Fan J, Chang Y, Cheng S, Liang B, Qu D. Effect of breathing exercises on patients with interstitial lung disease: A systematic review and meta-analysis. *Qual Life Res.* 2024;33(9):2335-2347.
47. Wu X, Zhai F, Chang A, Wei J, Guo Y, Zhang J. Application of machine learning algorithms to predict osteoporosis in postmenopausal women with type 2 diabetes mellitus. *J Endocrinol Invest.* 2023;46(12):2535-2546.
48. Shaabanpour Aghamaleki F, Mollashahi B, Nosrati M, Moradi A, Sheikhpour M, Movafagh A. Application of an Artificial Neural Network in the Diagnosis of Chronic Lymphocytic Leukemia. *Cureus.* 2019;11(2):e4004.
49. Lundberg SM, Nair B, Vavilala MS, et al. Explainable machine-learning predictions for the prevention of hypoxaemia during surgery. *Nature biomedical engineering.* 2018;2(10):749-760.

Pre-Proof

Copyright *Chronic Obstructive Pulmonary Diseases: Journal of the COPD Foundation* ©2026
Published online November 12, 2026 https://doi.org/10.15326/jcopdf.2026.0808

Table 1. Socio-demographic and clinical characteristics of study participants (n = 323).

Variable	Mean	SD
Age, years	63.71	8.96
Disease duration, years	13.28	4.09
Number of acute exacerbations	2.51	1.08
Variable	**Number of patients**	**Proportion, %**
Gender		
Male	67	20.7
Female	256	79.3
Residence		
Rural	61	18.9
Urban	262	81.1
Educational level		
Primary school or below	220	68.1
Junior high	39	12.1
Senior high	31	9.6
University or above	33	10.2
Marital status		
Single	83	25.7
Married	158	48.9
Divorced/Widowed	82	25.4
Payment method		
Private payment	235	72.8
Partial medical insurance	44	13.6
Complete medical insurance	44	13.6
Quantity of children		
0	46	14.2
1	31	9.6

Pre-Proof

Copyright *Chronic Obstructive Pulmonary Diseases: Journal of the COPD Foundation* ©2026
Published online November 12, 2026 https://doi.org/10.15326/jcopdf.2026.0808

≥2	246	76.2
Household monthly income, Chinese yuan		
<3,000	127	39.3
3,000-8,000	95	29.4
>8,000	101	31.3
Smoking history		
No	157	48.6
Yes	166	51.4
Drinking history		
No	171	52.9
Yes	152	47.1
Co-morbidity		
No	154	47.7
Yes	169	52.3
Frequency of exercise		
Never	62	19.2
Occasionally	197	61.0
Regularly	64	19.8
BMI, kg/m^2		
<18.4	107	33.2
18.4-23.8	108	33.4
>23.8	108	33.4
GOLD degree		
I	38	11.8
II	41	12.7
III	216	66.9
IV	28	8.7

COPD, chronic obstructive pulmonary disease; GOLD, global initiative for chronic

Pre-Proof

Copyright *Chronic Obstructive Pulmonary Diseases: Journal of the COPD Foundation* ©2026
Published online November 12, 2026 https://doi.org/10.15326/jcopdf.2026.0808

obstructive lung disease; BMI, body mass index.

Pre-Proof

Copyright *Chronic Obstructive Pulmonary Diseases: Journal of the COPD Foundation* ©2026
Published online November 12, 2026 https://doi.org/10.15326/jcopdf.2026.0808

Table 2. Model fit indices for trajectory patterns with different classes.

Class	AIC	BIC	aBIC	LMR(p)	BLRT(p)	Entropy	Sample proportion (%) per class
1	4125.777	4163.548	4131.829	n.a.	n.a.	n.a.	n.a.
2	3940.250	3989.359	3948.125	<0.001	<0.001	0.996	62.209/37.791
3	3809.960	3870.403	3819.652	0.006	<0.001	0.992	20.124/62.229/17.647
4	3121.098	3192.873	3132.608	0.012	<0.001	0.997	8.050/15.170/65.015/11.765
5	3127.098	3210.206	3140.425	0.879	<0.001	0.997	8.050/15.170/65.015/0.000

AIC, Akaike Information Criterion; BIC, Bayesian Information Criterion; aBIC, adjusted Bayesian Information Criterion; LMR, Lo–Mendell–Rubin; BLRT, bootstrap likelihood ratio test; n.a., not applicable.

Pre-Proof

Copyright *Chronic Obstructive Pulmonary Diseases: Journal of the COPD Foundation* ©2026
Published online November 12, 2026 https://doi.org/10.15326/jcopdf.2026.0808

Table 3. Classification probabilities for the most likely trajectory class membership by latent class.

Latent class	Latent class membership			
	C1 (26)	C2 (49)	C3 (210)	C4 (38)
C1	0.998	0.000	0.000	0.002
C2	0.000	1.000	0.000	0.000
C3	0.000	0.000	1.000	0.000
C4	0.004	0.000	0.000	0.996

Pre-Proof

Copyright *Chronic Obstructive Pulmonary Diseases: Journal of the COPD Foundation* ©2026
Published online November 12, 2026 https://doi.org/10.15326/jcopdf.2026.0808

Table 4. Comparison of different characteristics of 323 participants by different trajectory classes.

Variable	C1 (N=26)	C2 (N=49)	C3 (N=210)	C4 (N=38)	*Statistics*	*P*-value
Age, yearss	64.50±6.65	64.31±7.28	62.29±9.74	59.61±10.26	9.646[a]	**<0.001**
Disease duration, years	12.81±2.55	13.47±2.41	12.97±4.26	11.63±5.80	7.298[a]	**0.008**
Number of acute exacerbations	2.77±0.51	2.82±0.44	2.46±1.13	1.84±1.39	11.210[a]	**<0.001**
CAT	25.65±5.64	24.57±3.29	19.20±7.87	15.21±5.06	20.009[a]	**<0.001**
mMRC	2.35±0.63	2.39±0.57	1.89±0.78	1.53±0.60	13.337[a]	**<0.001**
COPD-Q	6.42±4.12	5.71±4.00	6.37±3.96	7.29±4.09	1.105[a]	0.347
SCSQ-Positive	14.27±5.50	14.82±4.15	19.70±7.04	22.16±5.55	15.514[a]	**<0.001**
SCSQ-Negative	16.27±4.26	14.92±3.14	11.52±5.31	8.95±3.46	18.881[a]	**<0.001**
SCSQ-Total	30.58±15.73	28.67±12.69	30.55±13.03	29.58±12.04	0.307[a]	0.820
PSSS-Family	4.62±1.10	4.67±1.13	4.74±1.04	4.53±1.01	0.499[a]	0.683
PSSS-Friends	4.65±1.09	4.63±1.04	4.80±1.12	5.05±1.23	1.157[a]	0.326
PSSS-Others	4.42±0.90	4.78±1.07	4.70±1.08	4.66±1.10	0.668[a]	0.572
PSSS-Total	13.69±1.69	14.08±1.72	14.24±1.96	14.24±1.55	0.717[a]	0.542
PSS-10	23.88±6.97	24.35±4.13	18.49±8.25	14.50±5.75	17.090[a]	**<0.001**
SSCI-8	27.58±4.46	26.69±4.70	22.97±6.91	20.39±5.41	11.421[a]	**<0.001**

Copyright *Chronic Obstructive Pulmonary Diseases: Journal of the COPD Foundation* ©2026
Published online November 12, 2026 https://doi.org/10.15326/jcopdf.2026.0808

Gender					31.682[c]	**<0.001**
Male	26 (100.00%)	49 (100.00%)	159 (75.71%)	22 (57.90%)		
Female	0 (0.00%)	0 (0.00%)	51 (24.29%)	16 (42.11%)		
Residence					27.051[c]	**<0.001**
Rural	0 (0.00%)	0 (0.00%)	47 (22.38%)	14 (36.84%)		
Urban	26 (100.00%)	49 (100.00%)	163 (77.62%)	24 (63.16%)		
Educational level					50.518[c]	**<0.001**
Primary school or below	26 (100.00%)	49 (100.00%)	130 (61.91%)	15 (39.47%)		
Junior high	0 (0.00%)	0 (0.00%)	30 (14.29%)	9 (23.68%)		
Senior high	0 (0.00%)	0 (0.00%)	24 (11.43%)	7 (18.42%)		
University or above	0 (0.00%)	0 (0.00%)	26 (12.38%)	7 (18.42%)		
Marital status					5.183[b]	0.159
Single	10 (38.46%)	9 (18.37%)	56 (26.67%)	8 (21.05%)		
Married	10 (38.46%)	24 (48.98%)	107 (50.95%)	17 (44.74%)		
Divorced/Widowed	6 (23.08%)	16 (32.65%)	47 (22.38%)	13 (34.21%)		
Payment method					44.375[c]	**<0.001**
Private payment	26 (100.00%)	49 (100.00%)	142 (67.62%)	18 (47.37%)		

Pre-Proof

Copyright *Chronic Obstructive Pulmonary Diseases: Journal of the COPD Foundation* ©2026
Published online November 12, 2026 https://doi.org/10.15326/jcopdf.2026.0808

Partial medical insurance	0 (0.00%)	0 (0.00%)	38 (18.10%)	6 (15.79%)		
Complete medical insurance	0 (0.00%)	0 (0.00%)	30(14.29%)	14 (36.84%)		
Quantity of children					39.015[c]	**<0.001**
0	0 (0.00%)	0 (0.00%)	34 (16.19%)	12 (31.58%)		
1	0 (0.00%)	0 (0.00%)	24 (11.43%)	7 (18.42%)		
≥2	26 (100.00%)	49 (100.00%)	152 (72.38%)	19 (50.00%)		
Household monthly income, Chinese yuan					1.837[b]	0.607
<3,000	11 (42.31%)	22 (44.90%)	80 (38.10%)	14 (36.84%)		
3,000-8,000	7 (26.92%)	16 (32.65%)	61 (29.05%)	11 (28.95%)		
>8,000	8 (30.77%)	11 (22.45%)	69 (32.86%)	13 (34.21%)		
Smoking history					2.886[b]	0.410
No	9 (34.62%)	27 (55.10%)	102 (48.57%)	19 (50.00%)		
Yes	17 (65.38%)	22 (44.90%)	108 (51.43%)	19 (50.00%)		
Drinking history					1.330[b]	0.722
No	11 (42.31%)	26 (53.06%)	114 (54.29%)	20 (52.63%)		

Pre-Proof

Copyright *Chronic Obstructive Pulmonary Diseases: Journal of the COPD Foundation* ©2026
Published online November 12, 2026 https://doi.org/10.15326/jcopdf.2026.0808

Yes	15 (57.69%)	23 (46.94%)	96 (45.71%)	18 (47.37%)		
Co-morbidity					2.750[b]	0.432
No	11 (42.308%)	23 (46.939%)	106 (50.476%)	14 (36.842%)		
Yes	15 (57.692%)	26 (53.061%)	104 (49.524%)	24 (63.158%)		
Frequency of exercise					0.191[b]	0.979
Never	9 (34.615%)	17 (34.694%)	33 (15.714%)	3 (7.895%)		
Occasionally	8 (30.769%)	13 (26.531%)	144 (68.571%)	32 (84.211%)		
Regularly	9 (34.615%)	19 (38.776%)	33 (15.714%)	3 (7.895%)		
BMI, kg/m^2					1.774[b]	0.621
<18.4	6 (23.077%)	14 (28.571%)	72 (34.286%)	15 (39.474%)		
18.4-23.8	9 (34.615%)	19 (38.776%)	71 (33.810%)	9 (23.684%)		
>23.8	11 (42.308%)	16 (32.653%)	67 (31.905%)	14 (36.842%)		
GOLD degree					13.488[c]	**0.004**
I	0 (0.000%)	0 (0.000%)	30 (14.286%)	8 (21.053%)		
II	0 (0.000%)	0 (0.000%)	29 (13.810%)	12 (31.579%)		
III	26 (100.000%)	49 (100.000%)	130 (61.905%)	11 (28.947%)		

Pre-Proof

Copyright *Chronic Obstructive Pulmonary Diseases: Journal of the COPD Foundation* ©2026
Published online November 12, 2026 https://doi.org/10.15326/jcopdf.2026.0808

IV	0 (0.000%)	0 (0.000%)	21 (10.000%)	7 (18.421%)

COPD, chronic obstructive pulmonary disease; GOLD, global initiative for chronic obstructive lung disease; BMI, body mass index. In this table, superscript a indicates F-value, b indicates χ^2 value, and c indicates results from the Fisher-Freeman-Halton test.

Pre-Proof

Copyright *Chronic Obstructive Pulmonary Diseases: Journal of the COPD Foundation* ©2026
Published online November 12, 2026 https://doi.org/10.15326/jcopdf.2026.0808

Table 5. Predictive performance of the eight machine learning algorithms in the training and validation sets for psychological distress among patients with stable CODP.

Models	AUC	Accuracy	Sensitivity	Specificity	PPV	NPV	F1 score
Training set							
Logistic	0.780	0.611	0.574	0.895	0.977	0.222	0.720
SVM	0.876	0.929	0.951	0.764	0.968	0.684	0.959
KNN	1.000	1.000	1.000	1.000	1.000	1.000	1.000
RF	1.000	1.000	1.000	1.000	1.000	1.000	1.000
LightGBM	0.697	0.732	0.743	0.652	0.9430	0.271	0.827
XGBoost	0.828	0.737	0.727	0.809	0.969	0.297	0.826
Decision Tree	0.705	0.724	0.731	0.679	0.946	0.257	0.822
MLP	0.631	0.527	0.484	0.848	0.969	0.196	0.597
Validation set							
Logistic	0.753	0.589	0.554	0.85	0.963	0.203	0.703
SVM	0.651	0.759	0.804	0.425	0.912	0.229	0.853
KNN	0.582	0.579	0.579	0.575	0.910	0.158	0.706
RF	0.695	0.746	0.804	0.318	0.898	0.182	0.847
LightGBM	0.608	0.706	0.737	0.479	0.918	0.165	0.813
XGBoost	0.704	0.672	0.684	0.579	0.925	0.210	0.777
Decision Tree	0.561	0.678	0.712	0.411	0.906	0.141	0.794
MLP	0.689	0.536	0.495	0.839	0.969	0.193	0.615

COPD, chronic obstructive pulmonary disease; AUC, area under the curve; PPV, positive predictive value; NPV, negative predictive value; SVM, support vector machine; KNN, K-Nearest Neighbor; RF, random forest; GBM, gradient boosting machine; XGBoost, extreme gradient boosting; MLP, multilayer perceptron.

Pre-Proof

Copyright *Chronic Obstructive Pulmonary Diseases: Journal of the COPD Foundation* ©2026
Published online November 12, 2026 https://doi.org/10.15326/jcopdf.2026.0808

Table 6. Performance of the XGBoost classification model for predicting the risk of psychological distress among patients with stable COPD.

Models	AUC	Accuracy	Sensitivity	Specificity	PPV	NPV	F1 score
Training set	0.859	0.765	0.757	0.823	0.970	0.326	0.848
Validation set	0.686	0.703	0.728	0.517	0.924	0.182	0.808
Test set	0.739	0.753	0.802	0.364	0.908	0.19	0.852

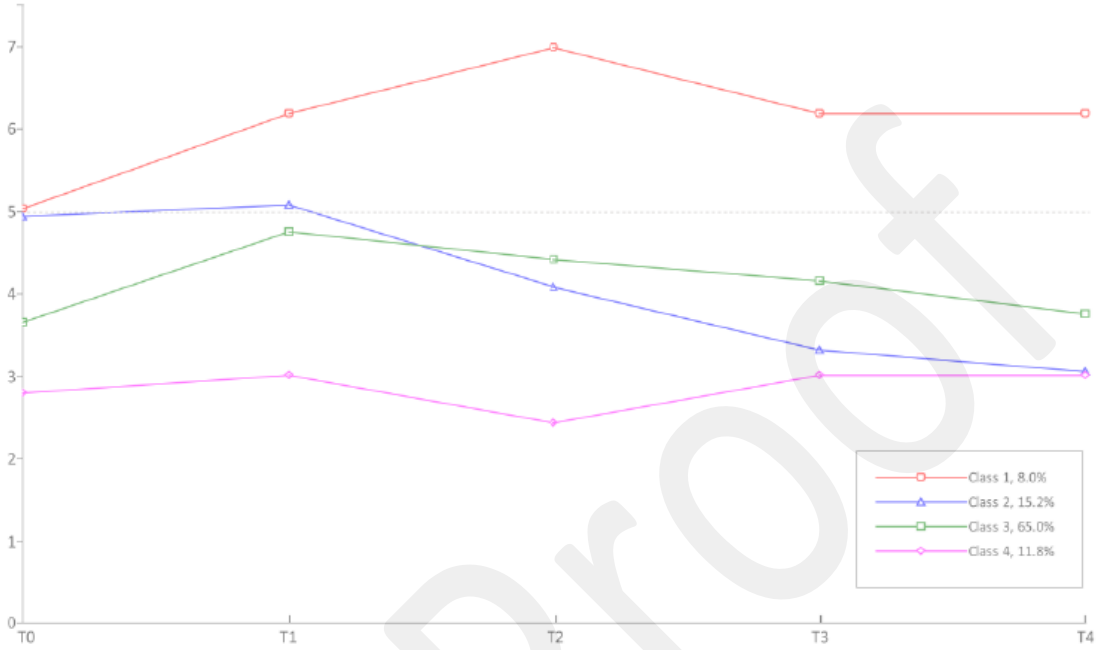
COPD, chronic obstructive pulmonary disease; AUC, area under the curve; PPV, positive predictive value; NPV, negative predictive value. XGBoost, extreme gradient boosting.

Pre-Proof

Copyright *Chronic Obstructive Pulmonary Diseases: Journal of the COPD Foundation* ©2026
Published online November 12, 2026 https://doi.org/10.15326/jcopdf.2026.0808

Figure legends

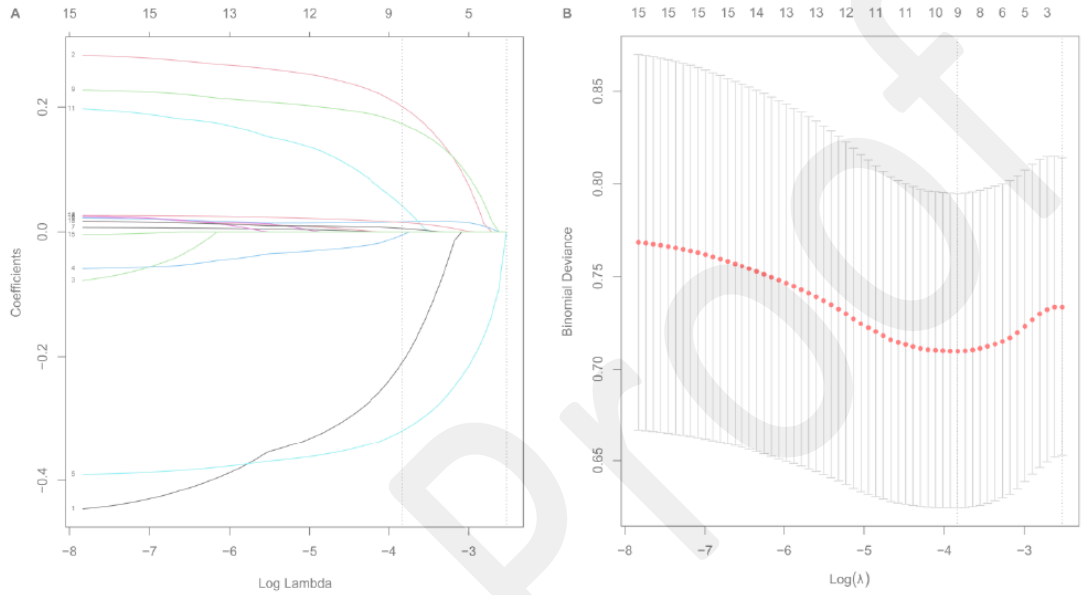
Figure 1. Four longitudinal trajectory patterns of psychological distress in patients with stable COPD. COPD, chronic obstructive pulmonary disease.



Pre-Proof

Copyright *Chronic Obstructive Pulmonary Diseases: Journal of the COPD Foundation* ©2026
Published online November 12, 2026 https://doi.org/10.15326/jcopdf.2026.0808

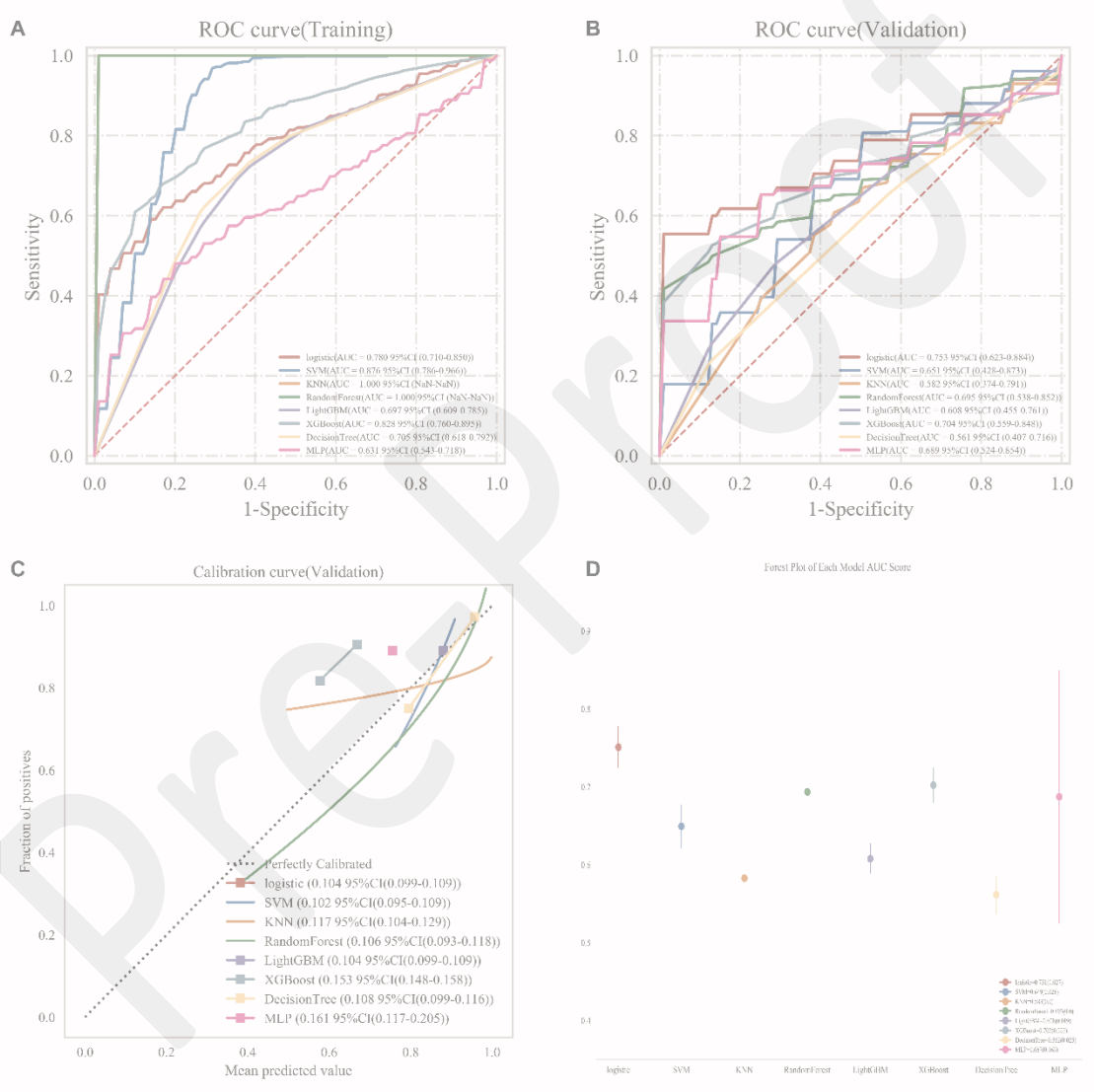
Figure 2. Feature selection based on the LASSO regression. Coefficient profile (A) and penalty plot (B). The minimum criteria and the one standard error of the minimum criteria were chosen as the optimal values for the drawn dotted vertical lines.



Pre-Proof

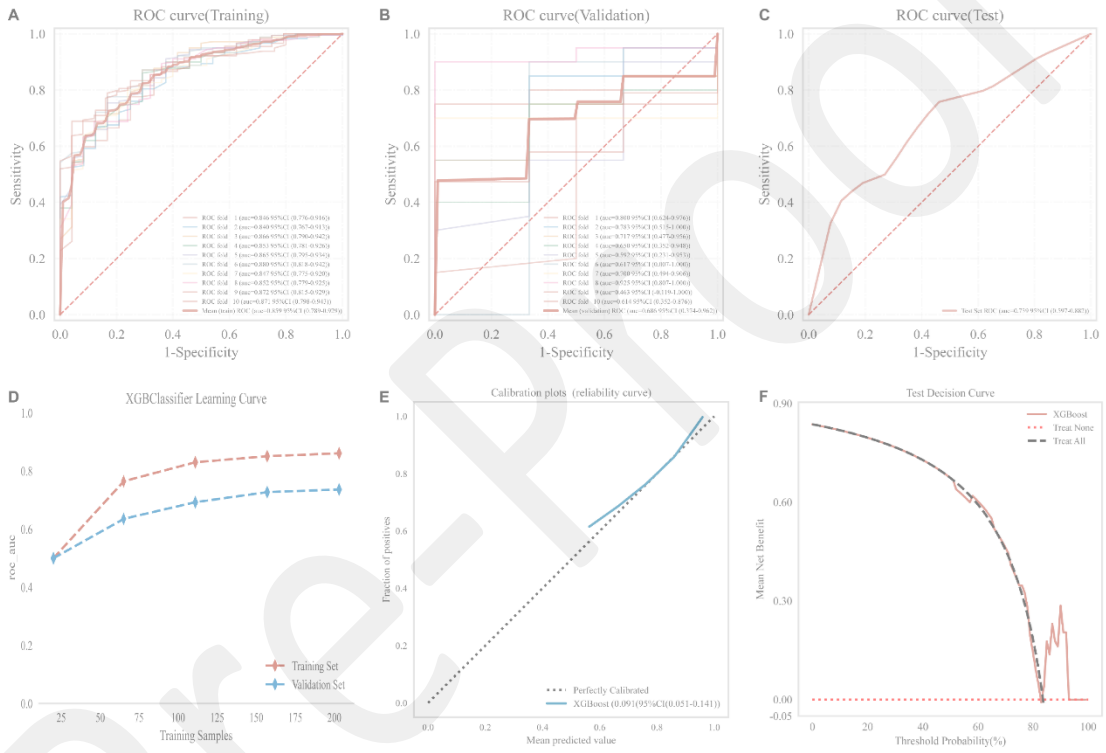
Copyright *Chronic Obstructive Pulmonary Diseases: Journal of the COPD Foundation* ©2026
Published online November 12, 2026 https://doi.org/10.15326/jcopdf.2026.0808

Figure 3. Construction and comparison of eight ML models. ROC curves of eight ML algorithms for predicting psychological distress in the train (A) and validation (B) sets as well as calibration curves (C) and forest plot of various models (D). ROC, receiver operating characteristic; ML, machine learning.



Copyright *Chronic Obstructive Pulmonary Diseases: Journal of the COPD Foundation* ©2026
Published online November 12, 2026 https://doi.org/10.15326/jcopdf.2026.0808

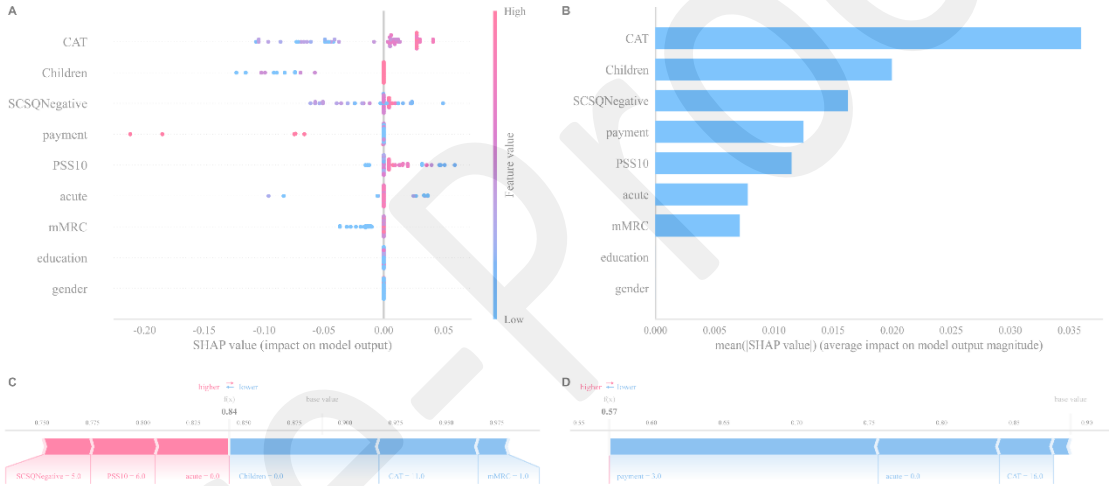
Figure 4. Development and evaluation of XGBoost classification model. ROC curves of XGBoost using the 5-fold cross-validation in the training (A), validation (B) and test (C) sets, respectively. D-F represent machine learning curve, calibration curve for XGBoost, and decision curve for XGBoost classification model, respectively. ROC, receiver operating characteristic.



Pre-Proof

Copyright *Chronic Obstructive Pulmonary Diseases: Journal of the COPD Foundation* ©2026
Published online November 12, 2026 https://doi.org/10.15326/jcopdf.2026.0808

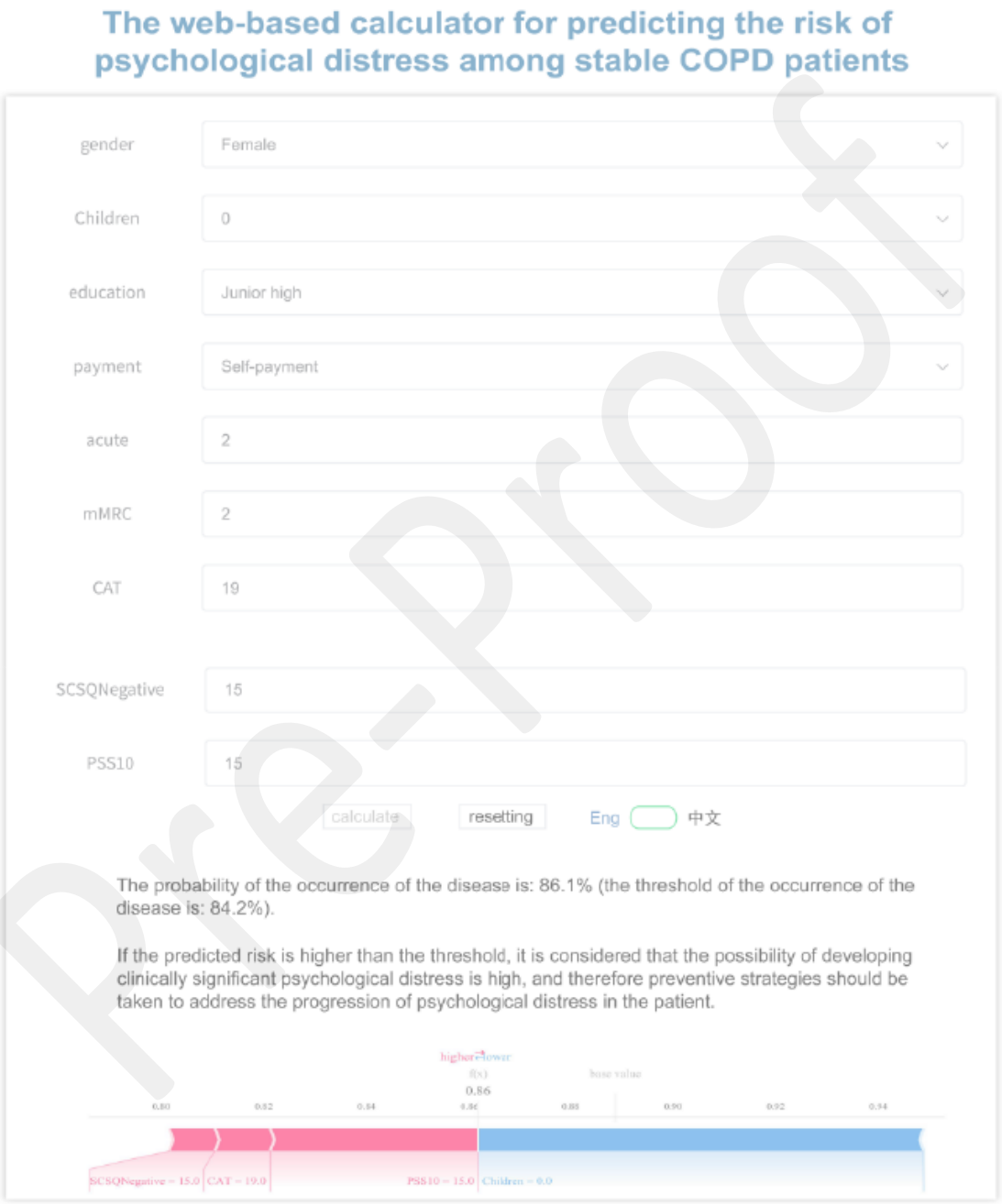
Figure 5. Interpretation of the results of XGBoost classification model based on the SHAP method. SHAP summary chart (A) and bar chart (B). The left dot plot represents the direction of contribution of each value of each variable, with red indicating larger values and blue indicating lower values of each variable. The right bar plot represents the importance of the variables and their overall contribution to the model predictions. C and D indicate that the SHAP scores explain the predicted risk of psychological distress in two subjects.



Pre-Proof

Copyright *Chronic Obstructive Pulmonary Diseases: Journal of the COPD Foundation* ©2026
Published online November 12, 2026 https://doi.org/10.15326/jcopdf.2026.0808

Figure 6. The web-based calculator for predicting the risk of psychological distress among patients with stable COPD. COPD, chronic obstructive pulmonary disease.



Pre-Proof

Copyright *Chronic Obstructive Pulmonary Diseases: Journal of the COPD Foundation* ©2026
Published online November 12, 2026 https://doi.org/10.15326/jcopdf.2026.0808