

<https://doi.org/10.1038/s41540-025-00489-y>

Deep learning for detecting and early predicting chronic obstructive pulmonary disease from spirogram time series



Shuhao Mei^{1,2,10}, Xin Li^{3,10}, Yuxi Zhou^{1,4}✉, Jiahao Xu¹, Yong Zhang⁴✉, Yuxuan Wan¹, Shan Cao⁵, Qinghao Zhao⁶, Shijia Geng⁷, Junqing Xie⁸, Shengyong Chen¹✉ & Shenda Hong^{2,9}✉

Chronic Obstructive Pulmonary Disease (COPD) is a chronic lung condition characterized by airflow obstruction. Current diagnostic methods primarily rely on identifying prominent features in spirometry (Volume-Flow time series) to detect COPD, but they are not adept at predicting future COPD risk based on subtle data patterns. In this study, we introduce a novel deep learning-based approach, DeepSpiro, aimed at the early prediction of future COPD risk. DeepSpiro consists of four key components: SpiroSmoother for stabilizing the Volume-Flow curve, SpiroEncoder for capturing volume variability-pattern through key patches of varying lengths, SpiroExplainer for integrating heterogeneous data and explaining predictions through volume attention, and SpiroPredictor for predicting the disease risk of undiagnosed high-risk patients based on key patch concavity, with prediction horizons of 1–5 years, or even longer. Evaluated on the UK Biobank dataset, DeepSpiro achieved an AUC of 0.8328 for COPD detection and demonstrated strong predictive performance for future COPD risk (p -value < 0.001). In summary, DeepSpiro can effectively predict the long-term progression of COPD disease.

Chronic Obstructive Pulmonary Disease (COPD) is a progressively worsening lung disease that leads to difficulty breathing, limited activity, and a decline in quality of life^{1,2}. As the disease progresses, COPD may also increase the risk of cardiovascular diseases³ and even lead to premature death. Therefore, timely and accurate COPD detection is crucial to reduce patient health risks⁴. Previous studies have shown a strong correlation between detecting the disease at an early stage and the success of its treatment. Failing to identify the disease during this crucial period will worsen its severity^{5,6}.

Clinical diagnosis often involves identifying COPD patients by determining if the FEV1/FVC ratio is below 70%^{7–9}. Researchers have discovered that the FEV1/FVC ratio method is not always accurate when used with people of different ages^{10,11}. This means that some patients might not get the right diagnosis at the right time, missing out on early and personalized treatment options.

In recent years, researchers have used deep learning technologies¹² to identify COPD characteristics by analyzing the Volume-Flow curve. This

method helps to overcome the limitations of traditional approaches but cannot effectively predict an individual's latent risk of developing COPD. Moreover, introducing deep learning technology has resulted in models lacking transparency, making it difficult to gain the trust of medical professionals and patients¹³. Therefore, developing an artificial intelligence interpretable algorithm that can accurately detect patients with COPD and early predict an individual's latent risk of COPD is crucial for slowing disease progression and preventing patient mortality.

However, most existing methods for COPD risk analysis mainly face the following four challenges:

- In order to determine the degree of airflow obstruction, it is necessary to generate Volume-Flow curves based on the original Time-Volume curves. Nevertheless, current techniques for generating Volume-Flow curves can yield unstable curves, potentially resulting in incorrect predictions by the model.
- The length of the Volume-Flow curve varies among individuals due to differences in exhalation durations. Nevertheless, existing techniques

¹Department of Computer Science, Tianjin University of Technology, Tianjin, China. ²National Institute of Health Data Science, Peking University, Beijing, China.

³Department of Rehabilitation Medicine, Beijing Tsinghua Changgung Hospital, School of Clinical Medicine, Tsinghua University, Beijing, China. ⁴DCST, BNRist, RIIT, Institute of Internet Industry, Tsinghua University, Beijing, China. ⁵Department of Biotherapy, The Second Hospital of Tianjin Medical University, Tianjin, China. ⁶Department of Cardiology, Peking University People's Hospital, Beijing, China. ⁷HeartVoice Medical Technology, Hefei, China. ⁸Centre for Statistics in Medicine and NIHR Biomedical Research Centre Oxford, University of Oxford, Oxford, UK. ⁹Institute for Artificial Intelligence, Peking University, Beijing, China. ¹⁰These authors contributed equally: Shuhao Mei, Xin Li. ✉e-mail: joy_yuxi@pku.edu.cn; zhangyong05@tsinghua.edu.cn; csy@tjtu.edu.cn; hongshenda@pku.edu.cn

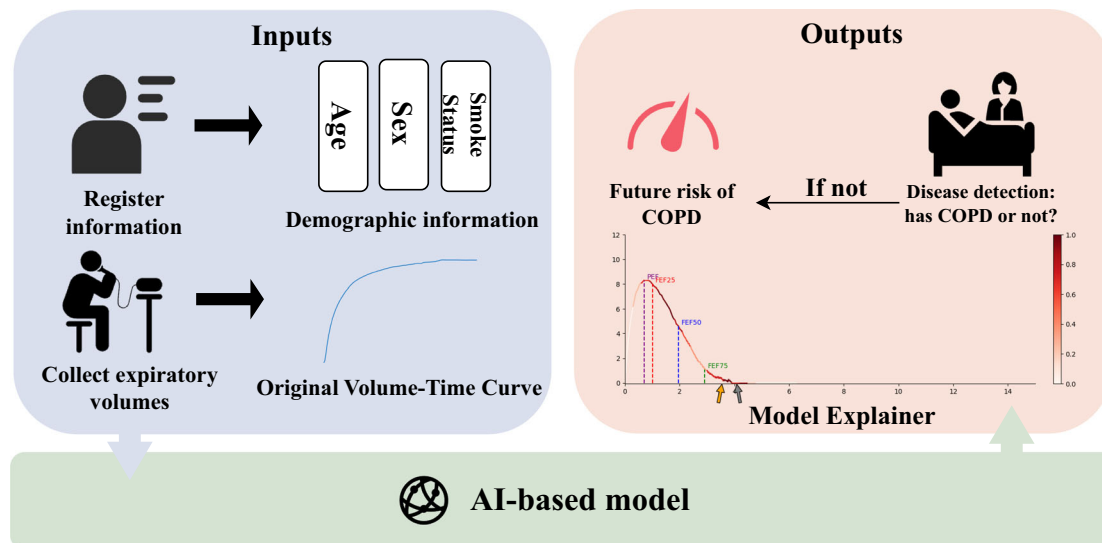


Fig. 1 | Our input module uses the raw Time–Volume curve time series collected from hospitals and patient demographic data, which are then passed into our AI-based model module. The AI-based model module is divided into four tasks (see the section “Overview” for details). After processing through the AI-based model module, the output data is handled by the output module. If the AI-based model

diagnoses the individual as a COPD patient, we will output their diagnosis results and the interpretability figure of the model. If the AI-based model diagnoses the individual as a non-COPD patient, we will output their risk of developing COPD over the next 1–5 years.

commonly handle Volume–Flow curves with varying lengths by either filling in missing values or truncating through downsampling. The former can readily add extraneous noise, whereas the latter could sacrifice important data dependencies.

- Existing deep learning models are still black-box models that can only produce detection outcomes without offering an explainer for those results. Models lacking transparency may find it challenging to gain the trust of medical professionals and patients.
- Currently, existing methods can only detect patients who have already had COPD based on obvious characteristics displayed on the spirogram (In this article, the spirogram specifically involves measuring Volume–Flow curve time series). However, these methods fail to early predict an individual’s probability of COPD in the future based on changes in the spirogram.

To address the aforementioned challenges, we propose DeepSpiro, a method based on deep learning for early prediction of future COPD risk (Fig. 1). Specifically, this paper makes four major contributions:

- We apply a method for constructing Volume–Flow curves guided by Time–Volume instability smoothing (SpiroSmoother), which uses a curve smoothing algorithm to precisely enhance the stability of the Volume–Flow curve while retaining the essential physiological information from the original Volume–Flow data.
- We develop a COPD identification method based on learning from varied-length key patch variability-pattern (SpiroEncoder). This algorithm can dynamically calculate the “key patch” number that is best suited for each time series data patch, thereby unifying the time series representation and extracting key physiological information from the original high-dimensional dynamic sequence into a unified low-dimensional time series representation.
- We propose a method for explaining the model based on volume attention and heterogeneous feature fusion (SpiroExplainer). This method combines the probability of having COPD with demographic data like age, sex, and probability. This information is then used as input for the model, which outputs a COPD risk assessment and offers an explainer for the model’s decisions regarding the individual.
- We developed a method for predicting the risk of COPD based on the variability pattern of key patch concavity (SpiroPredictor) for the first time. Our method can precisely forecast the probability of disease onset

in undiagnosed high-risk patients over the next 1–5 years and beyond. Additionally, it can accurately categorize these patients, thereby addressing the current deficiency in early predicting future COPD risk.

Our work has made COPD detection and early risk prediction more accurate, ultimately contributing to improved clinical decision-making and patient prognosis. By providing interpretable results and predicting future risks, DeepSpiro has the potential to become a valuable early screening tool. This could help delay disease progression and may potentially reduce patient mortality.

Results

Model detection and prediction performance

In this study, we aim to detect COPD and validate DeepSpiro superiority by comparing it to a ResNet18 baseline model. Previously, Justin Cosentino et al. proposed a baseline model in a paper published in Nature Genetics¹². It was trained and evaluated on the UK Biobank dataset, with evaluation metrics including AUROC, AUPRC, and F1-score. To ensure a fair comparison, we used the same dataset and evaluation criteria. In the COPD detection task, DeepSpiro achieved an AUROC of 83.28%, an AUPRC of 35.70%, and an F1-score of 39.50% as shown in Fig. 2. DeepSpiro outperformed the baseline model in these metrics by 1.16%, 1.24%, and 1.36%, respectively.

Table 1 presents a comparison of different metrics prior to and following signal enhancement. Based on the experimental findings, our signal enhancement technique effectively improves the stability of the Volume–Flow curve without compromising the essential physiological information of the original Volume–Flow data. This enhancement greatly aids in evaluating patients with COPD.

DeepSpiro emphasizes data integrity and the preservation of key information, in contrast to the Volume–Flow ResNet18 method (Baseline model). Based on the experimental findings, DeepSpiro betters the Volume–Flow ResNet18 method in all categories, as evidenced by higher values of key metrics such as AUROC, AUPRC, and F1-score.

In terms of parameters, DeepSpiro also performs exceptionally well. In order to demonstrate the specific advantages of our proposed DeepSpiro with respect to certain parameters, a comparative study is carried out between DeepSpiro and the Volume–Flow ResNet18 model. Specifically, Pytorch is used to build both models and utilize the top library to calculate both models’ parameter count and floating-point

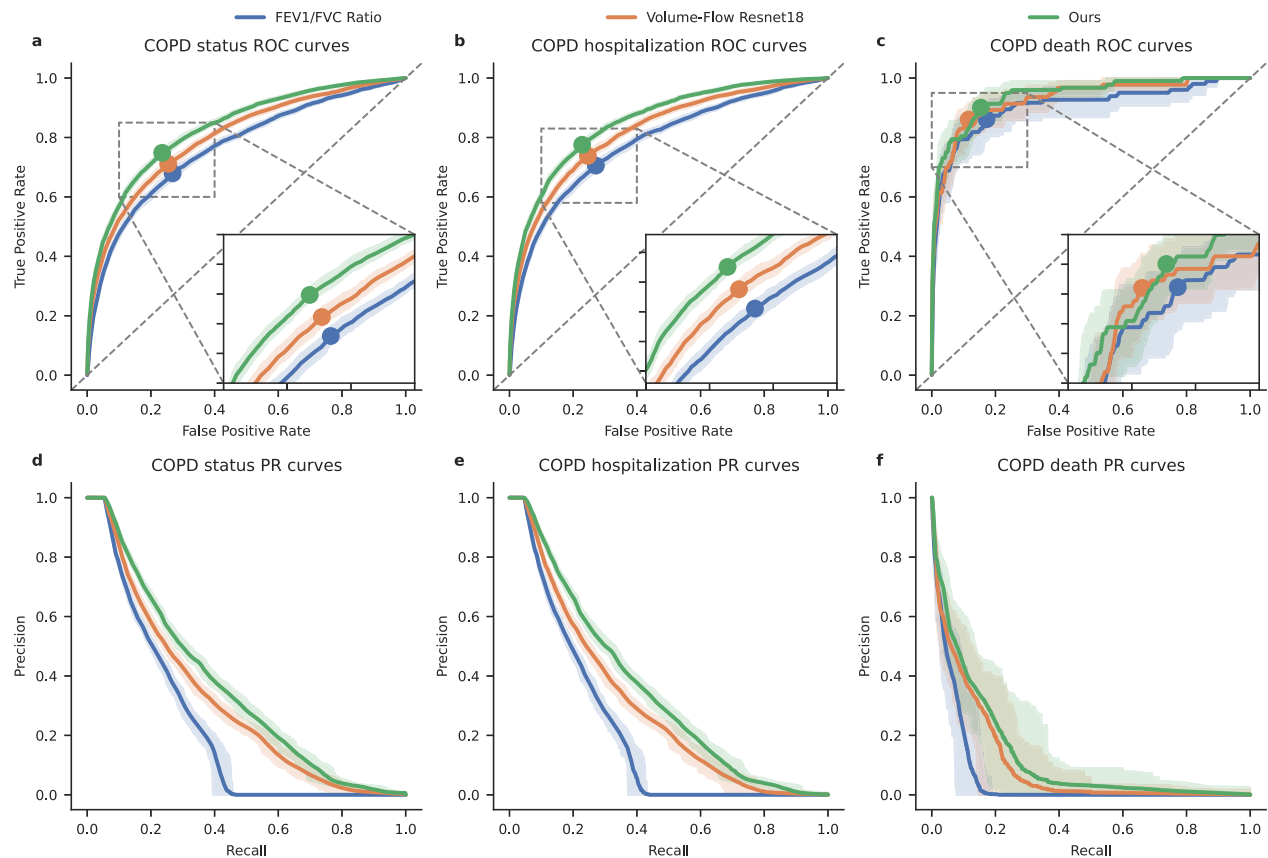


Fig. 2 | Evaluation comparison. **a–c** Comparison of receiver operating characteristic (ROC) curves for three prediction methods—DeepSpiro COPD risk predictions, Volume-Flow ResNet18 COPD risk predictions, and FEV1/FVC ratio-based risk—across three evaluation tasks: All COPD patients (left), future COPD-related hospitalization (center), and COPD-related death (right). **d–f** Comparison of precision-recall (PR) curves across three evaluation tasks—All COPD patients (left), future COPD-related hospitalization (center), and COPD-related death (right)—for three prediction methods: DeepSpiro COPD risk predictions, Volume-Flow ResNet18 COPD risk predictions, and FEV1/FVC ratio-based risk. The error bars represent bootstrapped 95% confidence intervals (n = 100 bootstrapping samples).

Table 1 | Evaluation comparison table

Category	Method	AUROC	AUPRC	F1-Score
All	FEV1/FVC ⁴⁰	0.7771	0.2266	0.3143
	Volume-Flow ResNet18 ¹²	0.8212	0.3446	0.3814
	DeepSpiro (non-smooth)	0.8315	0.3537	0.3913
	DeepSpiro	0.8328	0.3570	0.3950
Hospitalization	FEV1/FVC ⁴⁰	0.7925	0.2079	0.3024
	Volume-Flow ResNet18 ¹²	0.8382	0.3247	0.3703
	DeepSpiro (non-smooth)	0.8527	0.3453	0.3871
	DeepSpiro	0.8538	0.3490	0.3900
Death	FEV1/FVC ⁴⁰	0.9242	0.0393	0.1206
	Volume-Flow ResNet18 ¹²	0.9374	0.0887	0.2026
	DeepSpiro (non-smooth)	0.9419	0.1129	0.2211
	DeepSpiro	0.9419	0.1165	0.2304

This table presents the AUROC, AUPRC, and F1 scores for detecting COPD using four different methods: the FEV1/FVC method, the ResNet18 model, DeepSpiro trained on unsmoothed curves, and DeepSpiro trained on smoothed curves.

operations. The calculation results are shown in Supplementary Table 4. The results indicate that DeepSpiro’s parameter count and floating-point operations are significantly lower than those of the Volume-Flow ResNet18 model.

By leveraging demographic information, the model can gain a more comprehensive understanding of the patient’s background¹⁴ and on this basis, improve its prediction accuracy. Additionally, incorporating FEV1/FVC, the gold standard for COPD diagnosis, into the model can enhance the model’s capability in diagnosing COPD. To more comprehensively reflect the improvement in the model’s diagnostic capability by demographic information and the FEV1/FVC diagnostic gold standard^{15–17}, we concatenate the probabilities outputted by module II (Fig. 8II), demographic information, and the FEV1/FVC diagnostic gold standard into a new vector and input it into a logistic regression model. Figure 3 is a nomogram¹⁸, from which we can intuitively see that the added demographic information and the FEV1/FVC diagnostic gold standard enhance the model’s diagnostic accuracy. Therefore, the model explainer method based on volume attention and heterogeneous feature fusion has a promotive effect on the diagnosis of COPD.

In summary, traditional pulmonary function assessment indicators cannot fully reflect the disease’s complexity and patients’ overall health status. As shown in Table 1, DeepSpiro is more advanced than traditional pulmonary function assessment indicators and the Volume-Flow ResNet18 model.

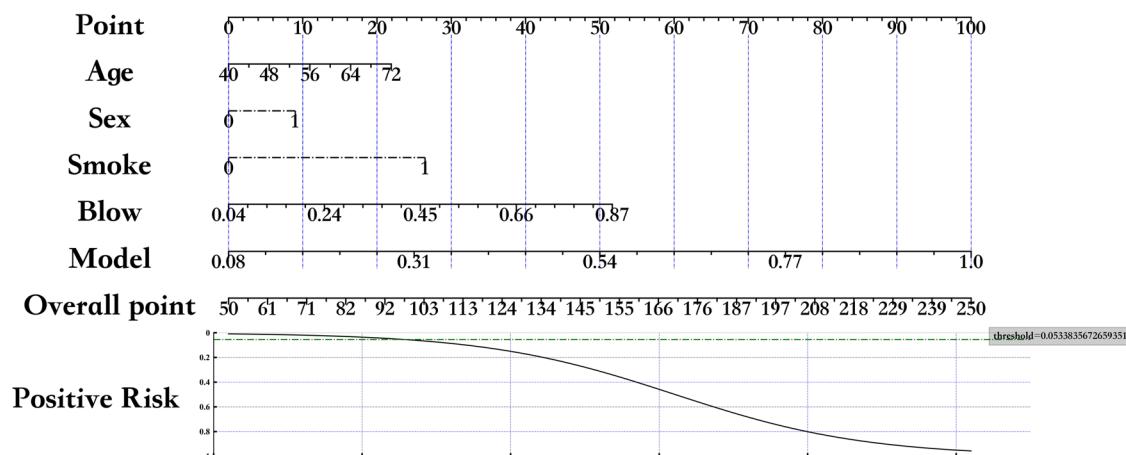


Fig. 3 | The nomogram of COPD detection. The nomogram illustrates the contribution of demographic information and the FEV1/FVC diagnostic gold standard to the model's diagnostic accuracy for COPD. The nomogram allows for visual estimation of the probability of COPD diagnosis by assigning weighted scores to

each variable. The “threshold” in the figure represents the cutoff value at which the predicted probability indicates a positive diagnosis for COPD. For instance, a threshold of 0.5 means that a predicted probability greater than 0.5 would be considered indicative of COPD.

Future risk prediction analysis

To better explain the results, we categorize the curves into two groups: early-phase curves (PEF~FEF25 and FEF25~FEF50) and late-phase curves (FEF50~FEF75 and FEF75+). As shown in Fig. 4a, by observing the trend of changes in the two types of curves, we can see a significant difference in the degree of concavity. The concavity of the early-phase curves decreases as disease risk decreases, while the concavity of the late-phase curves increases as disease risk decreases. By observing Fig. 4a, we find that individuals with higher disease risk (e.g., 1 year) are more likely to experience curve collapse in the early phases, whereas individuals with lower risk (e.g., Non-COPD) tend to exhibit curve collapse in the late phases. As the risk of disease increases, the phase in which the curve collapse occurs tends to shift earlier.

To more intuitively demonstrate the relationship between disease risk and the timing of curve collapse, we designed a concavity trend measurement model. This model uses the formula: $\mathcal{C}(PEF \sim FEF25) + \mathcal{C}(FEF25 \sim FEF50) - \mathcal{C}(FEF50 \sim FEF75) - \mathcal{C}(FEF75+)$ to measure the concavity trend of the curve. From the formula, it can be seen that a larger concavity trend value indicates that the curve collapse (the phase where expiratory flow rate drops most significantly) occurs in the earlier phases, whereas a smaller concavity trend value indicates that the curve collapse occurs in the later phases.

Figure 4b includes the future disease risk of high-risk and low-risk populations. The model predicts individuals as a high-risk group for those identified as at risk of disease and as a low-risk group for those predicted not to have the disease. We plotted the changes in disease risk for both groups as curves, showing the proportion diagnosed with COPD over time. The X-axis represents the time elapsed since the pulmonary function test, and the Y-axis represents the proportion of patients diagnosed with COPD at a specific timestamp within the population, i.e., the risk of disease. From the figure, it can be seen that from the time of the pulmonary function test, the disease risk for the model-predicted high-risk group shows an exponential increase over time, while the disease risk for the low-risk group remains near zero, essentially unchanged with a slight and weak increase. High-risk and low-risk groups also exhibit significant differences, with a p -value of <0.001 .

In order to obtain the expected Volume-Flow curve for different periods, we used the k -median algorithm to extract features from the model for samples corresponding to different risk periods (1 year, 2 years, 3 years, 4 years, 5+ years, Non-COPD). We then determined the median sample for each risk period and used it as the expected Volume-Flow curve for that stage. Similarly, to obtain the expected Volume-Flow curve for different subgroups, we subdivided the groups according to their classifications and used k -median analysis to find the median sample for each subgroup, using it to represent the expected Volume-Flow curve for that subgroup.

As shown in Fig. 4c, as the onset time progresses, the concavity of the patient's Volume-Flow curve decreases year by year. This trend also indicates that the impact of the disease on lung function is more significant in the later stages, and the closer it is to the onset time, the more apparent the disease becomes.

From the predictive distribution probabilities over future times and the future disease risk scenarios for high-risk and low-risk populations, it can be seen that DeepSpiro can effectively predict the future development trends of the disease, thereby providing better treatment and management plans for patients.

Result of subgroup analysis

To further explore the diagnostic accuracy of the model across different subsets of the population, subgroup analyses for various age groups, sex, and smoking statuses are conducted.

The subgroup analysis results for smokers illustrate the model's application scenarios. Among patients with COPD, the prevalence rate in smokers is higher than in non-smokers¹⁹. For our model, as shown in Fig. 5a, the prediction probability for smokers is higher than for non-smokers, and the prediction range for smokers is also wider than that for the non-smoking population. This is consistent with clinical significance. Regarding the p -value, the model demonstrates a strong discriminative ability between the two groups.

The subgroup analysis results for sex showcase the model's application scenarios. The 2018 China Adult Lung Health Study, which surveyed 50,991 individuals across 10 provinces and cities, has shown that the incidence rate among men is higher than that among women, with men at 11.9% and women at 5.4%. For our model, as shown in Fig. 5b, the prediction probability for men is higher than for women, and the prediction range for men is also greater than that for women. This aligns with clinical significance. Regarding the p -value, the model demonstrates a strong discriminative ability between the two groups.

Age is a significant factor influencing the development of COPD. Typically, the high-incidence age range for COPD is between 45 and 80 years. Within this age range, patients' lung functions tend to decline gradually. Particularly after age 45, the trend of declining lung function becomes more apparent. After reaching 80 years, lung function further declines, and patients' immune systems weaken, making them relatively more susceptible to diseases. Therefore, prevention and early identification of COPD are especially important for this age group. We divide patients into Youth (18–44), Middle (45–55), and Elderly (55 and above) for subgroup analysis, as shown in Fig. 5c. The results indicate that DeepSpiro's prediction probability for smokers is higher than for non-smokers, and the prediction

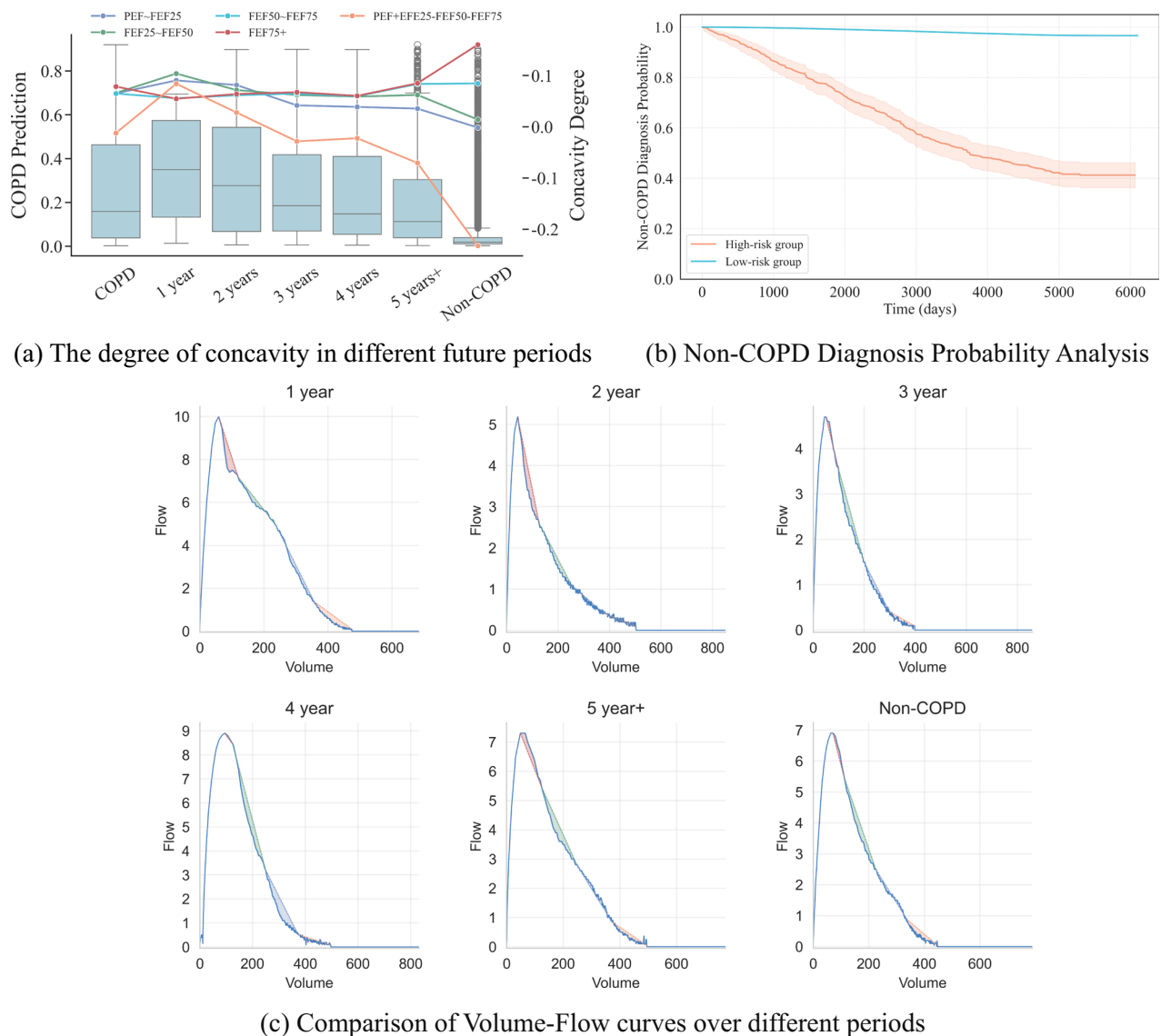


Fig. 4 | Future COPD prediction overview. **a** The left vertical axis represents the predicted probability of COPD, while the right vertical axis indicates the concavity degree based on the directed area metric for each phase. Each plot displays the mean concavity degree for the respective phase, illustrating how the concavity changes over time in relation to COPD risk. **b** This figure illustrates the probabilities of not being diagnosed with COPD over time for the high-risk and low-risk groups, as

predicted by the model. The X-axis represents the time since the pulmonary function test, while the Y-axis shows the probability of not being diagnosed with COPD at each time point. Due to right censoring, not all high-risk patients are diagnosed within the observation period, resulting in probabilities that remain above zero. **c** As the onset time progresses, the concavity of the patient's Volume-Flow curve decreases year by year.

range for smokers is also larger than for the non-smoking population, consistent with clinical significance. Regarding the *p*-value, the model shows strong discriminative ability between the two groups.

Additionally, we compare the AUROC of DeepSpiro, the Volume-Flow ResNet18 model, and the FEV1/FVC metric across different subgroups, as shown in Table 2. Across all subgroups, DeepSpiro's AUROC is the highest, demonstrating DeepSpiro's applicability in different subgroups.

We also conducted a subgroup analysis for individuals across different future periods. As shown in Fig. 5d, as the onset time progresses, an individual's concavity measure gradually decreases. Compared to non-smokers, smokers show significantly higher lung function concavity measures, further confirming the impact of smoking on COPD. This trend is consistent with clinical observations, highlighting the association between long-term smoking and lung function decline.

As shown in Fig. 5e, as the onset time progresses, an individual's concavity measure gradually decreases. Compared to females, males show significantly higher lung function concavity measures, which are typically related

to higher smoking rates, different lifestyle habits, and physiological differences in men. This aligns with the expected results from clinical research.

As shown in Fig. 5f, as the onset time progresses, an individual's concavity measure gradually decreases. Compared to younger patients, older patients show significantly higher lung function concavity measures, which may be related to natural lung function decline with age, weakened immunity, and other factors. This is consistent with clinical observations and expected results. (Note: Due to the small sample size in the Youth (18–44) group, it was not possible to subdivide and compare across different stages, so only the Middle (45–55) and Elderly (55+) groups were compared.)

Result of SHAP analysis

To further analyze whether the model's predictions hold clinical significance, we conduct a SHAP analysis²⁰.

As shown in the beeswarm plot in Fig. 6a, the relative importance of features is revealed. The figure shows that the higher the COPD risk

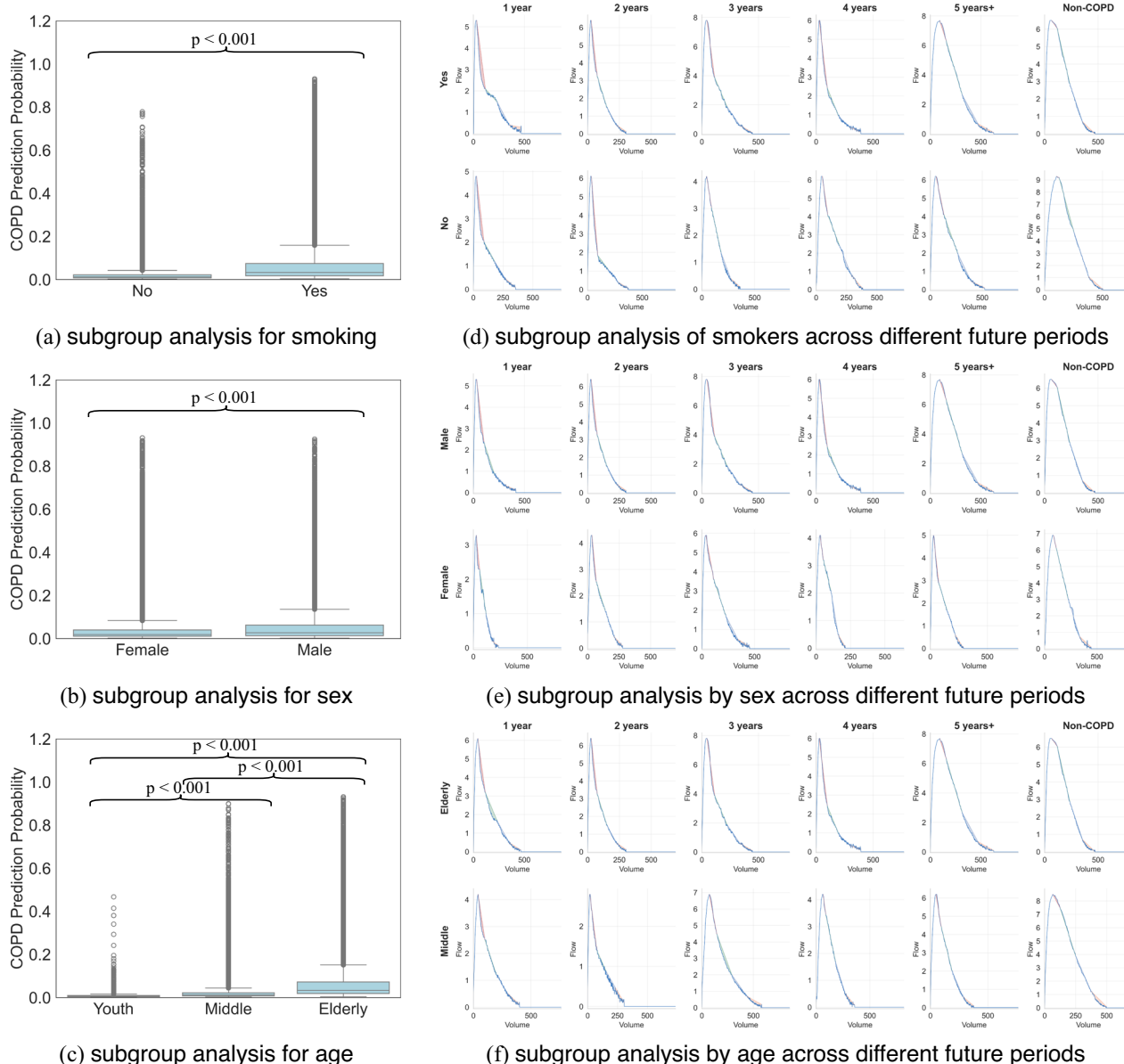


Fig. 5 | Subgroup analysis overview. **a** The subgroup analysis for smoking. **b** The subgroup analysis by sex. **c** The subgroup analysis by age. **d** As the onset time progresses, an individual's concavity measure gradually decreases. Compared to non-smokers, smokers show significantly higher lung function concavity measures. **e** As the onset time progresses, an individual's concavity measure gradually

decreases. Compared to females, males show significantly higher lung function concavity measures. **f** As the onset time progresses, an individual's concavity measure gradually decreases. Compared to younger patients, older patients show significantly higher lung function concavity measures.

value, the greater its impact on the model. Smoking, being male, and older age all influence the model's judgment. DeepSpiro's findings are consistent with related research conclusions, indicating that smokers, males, and older patients have a higher risk of being predicted as having COPD.

We examine the relationship between smoking and risk values as well as age and risk values and present our results with dependency graphs. Figure 6b shows the relationship between smoking and the risk of COPD. From the figure, it is evident that the predicted risk value of DeepSpiro is closely related to smoking status. The higher the risk value, the larger the proportion of smokers. This aligns with clinical significance. For COPD patients, smokers are more susceptible to the disease.

Figure 6c illustrates the relationship between age and the risk of COPD. From the figure, it can be seen that the predicted risk value of DeepSpiro is closely related to age. The higher the risk value, the older the age. This aligns

with clinical significance. For COPD patients, older patients are more susceptible to the disease.

Model explainer analysis

Healthcare professionals often struggle to accurately identify and locate disease information directly from long sequence data in COPD classification tasks. To address this issue, we utilize SpiroExplainer, automatically focusing on the diseased regions of the Volume-Flow data.

The proposed method can automatically capture anomalies from long sequence Volume-Flow data, providing specific patches with abnormalities. Figure 7 shows that the model can automatically focus on key patches that distinguish COPD patients. Brighter colors indicate higher attention in the corresponding areas, while darker colors indicate lower attention.

An interesting phenomenon is that for non-COPD patients, as shown in Fig. 7a, the Volume-Flow curve in the FEF₂₅-FEF₅₀ area appears "full,"

Table 2 | Comparison of subgroup evaluations

Group	Class	Model	AUROC
Sex	Male	FEV1/FVC ⁴⁰	0.7719
		ResNet18 ¹²	0.8166
		DeepSpiro	0.8335
	Female	FEV1/FVC ⁴⁰	0.7629
		ResNet18 ¹²	0.8196
		DeepSpiro	0.8329
Smoke	Yes	FEV1/FVC ⁴⁰	0.7774
		ResNet18 ¹²	0.8223
		DeepSpiro	0.8266
	No	FEV1/FVC ⁴⁰	0.6857
		ResNet18 ¹²	0.7428
		DeepSpiro	0.7620
Age	Youth	FEV1/FVC ⁴⁰	0.6960
		ResNet18 ¹²	0.7344
		DeepSpiro	0.7634
	Middle	FEV1/FVC ⁴⁰	0.7325
		ResNet18 ¹²	0.7779
		DeepSpiro	0.8022
	Elderly	FEV1/FVC ⁴⁰	0.7622
		ResNet18 ¹²	0.8024
		DeepSpiro	0.8210

ResNet18 in the table refers to Volume-Flow ResNet18. This table shows the performance of different methods in detecting COPD across various subgroup analyses. These methods include the FEV1/FVC method, the ResNet18 model, and DeepSpiro trained on smoothed Volume-Flow curves.

without the small airway collapse phenomenon near FEF₇₅ at the tail of the curve. DeepSpiro primarily focuses on these two areas near FEF₂₅–FEF₅₀ and FEF₇₅, which match the characteristics of the non-COPD patient group. On the other hand, for COPD patients, as shown in Fig. 7d, the curve's tail near the FEF₇₅ area collapses due to a small airway collapse resulting from COPD. DeepSpiro focuses on the area near FEF₇₅, aligning with the characteristics of the COPD patient group. Thus, DeepSpiro can effectively differentiate and assess the features of different populations.

In analyzing cases where the model made incorrect predictions, we found that for some individuals who might have asthma, as shown in Fig. 7b, due to insufficient inhalation, the curve after FEF₇₅ shows a weak exhalation characteristic similar to the small airway collapse seen in COPD patients. Since the model mainly focuses on the area near FEF₇₅, it leads to lower prediction accuracy for these asthma patients. For some individuals who might have COPD, as shown in Fig. 7c, due to uneven exhalation effort or mid-breath inhalation by COPD patients, the Volume-Flow curve might present double or multiple peaks, appearing “full” in the FEF₂₅–FEF₅₀ range similar to non-COPD patients. DeepSpiro primarily focuses on areas near FEF₂₅–FEF₅₀ and FEF₇₅, which also leads to lower prediction accuracy for these COPD patients. We believe the model's predictions are medically interpretable, and due to the complexity and overlap of disease characteristics, it is challenging to avoid prediction errors completely.

Discussion

Recent research on COPD has primarily focused on three areas: detection, prediction, and genetic studies. The application of machine learning and deep learning has achieved significant progress in disease detection. Researchers have used advanced deep learning models, such as convolutional neural networks (CNNs), to analyze various types of data like e-nose signals²¹, spirometry data¹², lung sounds²², and CT images²³, playing a vital role in COPD detection and classification. To improve diagnostic accuracy,

efforts have also been made to integrate multi-source data, such as combining multi-omics data like proteomics and transcriptomics with disease-specific protein/gene interaction information²⁴ or integrating imaging data with questionnaire data²⁵, providing a more comprehensive diagnostic foundation. In data processing, specialized preprocessing workflows have been developed to eliminate noise and enhance dataset quality, such as denoising and feature extraction for lung sounds²⁶. In the field of prediction, machine learning models have been employed to forecast COPD disease progression and survival rates^{27,28}, including the use of deep neural networks to predict COPD stages²⁹, and applying anomaly detection methods to identify COPD manifestations in chest CT scans³⁰, which have shown remarkable effectiveness in detecting lung function impairment and disease severity. In genetic research, bioinformatics approaches have been used to identify key genes and genetic markers associated with COPD^{31–34}, deepening our understanding of COPD pathogenesis and providing valuable insights for early detection and potential therapeutic target development.

In this paper, we designed a deep-learning method for detecting and early-predicting COPD from the Volume-Flow curve time series. Specifically, we use SpiroSmoother, thereby enhancing the stability of the Volume-Flow curve. We propose SpiroEncoder, which unifies the temporal representation of “key patches” and achieves the conversion of key physiological information from high to low dimensions. Utilizing SpiroExplainer, we fuse diagnostic probabilities with demographic information to improve the accuracy of COPD risk assessment and the model's explainer. The results and explainer of the model can provide patients with timely COPD risk reports. Moreover, we propose SpiroPredictor, by incorporating key information such as the degree of concavity, the model can accurately predict the future disease probability of high-risk patients who have not yet been diagnosed, thereby enabling early intervention and treatment to slow the progression of the disease. Our proposed DeepSpiro model has two primary functions. First, it can detect COPD with high performance, achieving AUC and other performance metrics that surpass those of current SOTA models. Second, our model can predict the risk of developing COPD over the next 1–5 years using key patch concavity information, an innovative method we have proposed for the first time. Experimental results show that our novel method has high accuracy in predicting future COPD risk.

Our work does have some limitations. First, although DeepSpiro has shown good performance in detecting and predicting COPD using spirogram time series in the UK Biobank, its generalizability to other population groups remains uncertain. The characteristics of the spirogram may vary among different populations, which could potentially affect the model's performance.

Second, while DeepSpiro demonstrates high accuracy in predicting COPD using spirogram time series in a research setting, further validation in real clinical environments is crucial. The transition from research to practical application may present unforeseen challenges, such as variations in data quality, increased noise, and differences in clinical workflows. These factors could impact the model's performance, necessitating additional adjustments and optimizations in future work.

Third, although our model demonstrated high AUROC values across most prediction intervals, its performance in terms of AUPRC was less satisfactory, particularly for predicting COPD onset within 1–2 years. We attribute this to the inherent class imbalance in the dataset (see Supplementary Table 1), where the proportion of new COPD cases is extremely low. Since AUPRC is highly sensitive to class imbalance, it may underestimate the true performance of the model in such cases. To investigate whether this issue stemmed from our methodology or the inherent challenges of the task, we compared our model with the approach proposed by Cosentino et al.¹². The comparison results (see Supplementary Table 2 and Supplementary Fig. 1) revealed that under similar experimental conditions, their model also exhibited low AUPRC values, highlighting the general difficulty of identifying rare COPD onset events in highly imbalanced datasets. Nevertheless, we observed that for certain prediction windows, such as those exceeding 5 years, our method achieved comparable or

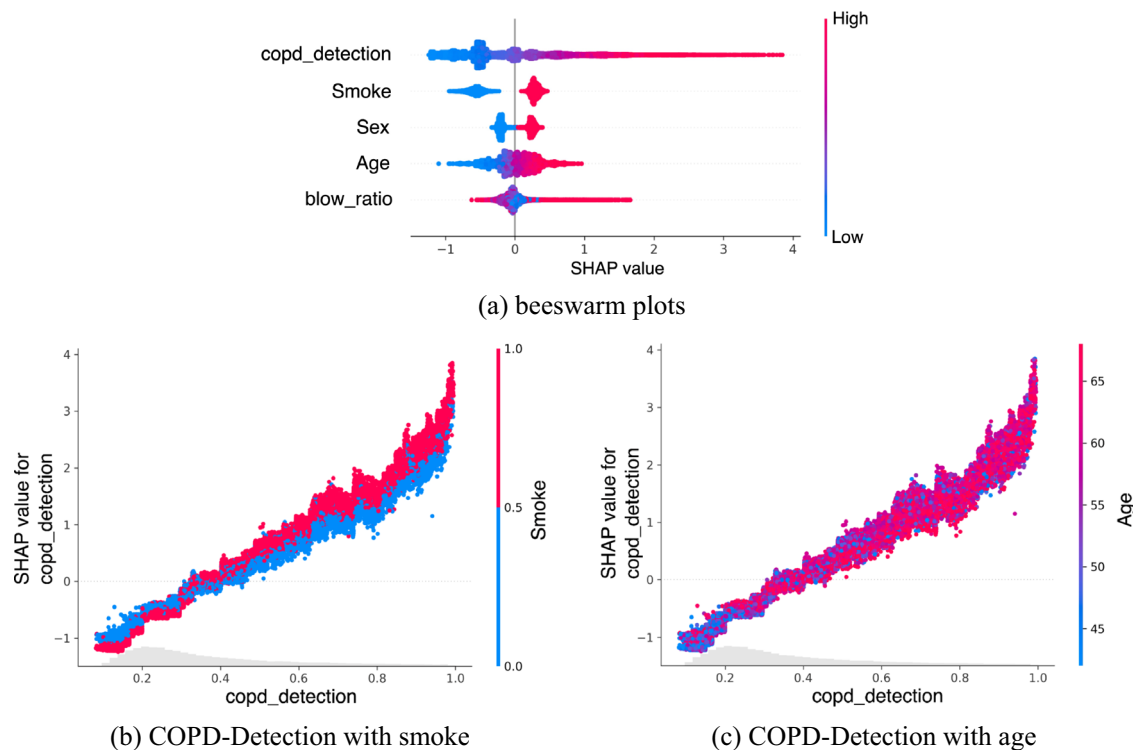


Fig. 6 | SHAP analysis overview. **a** Brighter colors mean that the feature has a more positive impact on the model's predictions. The blow ratio represents the FEV1/FVC value, an important indicator of lung function. **b** Relationship between smoking status and predicted risk of COPD. The figure illustrates how the risk value predicted by DeepSpiro correlates with smoking status, indicating a higher risk value is associated with a larger proportion of smokers. This result is consistent with clinical

findings, as smokers are more likely to develop COPD. **c** Relationship between age and predicted risk of COPD. The figure demonstrates that the predicted risk value of DeepSpiro increases with age, indicating that older individuals have a higher risk of developing COPD. This observation is consistent with clinical findings, which show that older patients are more susceptible to COPD.

superior AUPRC values. This indicates that incorporating concavity features from the volume-flow curve alongside demographic information helps capture critical features over longer time horizons. From a clinical perspective, the importance of early COPD detection often outweighs the risk of false positives. Individuals flagged as high-risk can undergo further confirmatory assessments, such as chest imaging. Additionally, the robust AUROC performance of our model across various prediction windows demonstrates its strong discriminative capability, making DeepSpiro a valuable early screening tool. This, in turn, has the potential to improve patient care quality and optimize resource allocation in real-world medical practice.

In the future, we will focus on training and validating the model using datasets from diverse geographic locations and population backgrounds. This will help assess the impact of demographic and environmental factors on model performance and improve its generalizability. Additionally, we plan to develop user-friendly software tools that allow doctors to conveniently apply the model in clinical settings and obtain reliable and understandable results to assist in diagnosing and treating patients. Finally, we aim to improve the model's generalizability by incorporating more high-quality datasets, enabling it to provide effective predictions in a wider range of scenarios.

Methods

Problem definition

For a COPD dataset $D = \{X, Y\}$ composed of N instances, where $X = \{x_1, x_2, \dots, x_p, \dots, x_N\}$ represents the monitoring data for N instances, and $Y = \{y_1, y_2, \dots, y_p, \dots, y_N\}$ represents the COPD disease labels for the N instances. The disease label for the j th instance, $y_j \in \{0, 1\}$, with 0 indicating no disease and 1 indicating the presence of disease. The monitoring data $x_j = \{s_j, d_j, a_j\}$ includes the spirogram data s_j , demographic information $d_j = \{dg_j, da_j, ds_j,$

$\dots\}$, and key concavity information $a_j = \{a_{j,pef-fef25}, a_{j,fe25-fe50}, a_{j,fe50-fe75}, a_{j,fe75+}\}$. Here, the spirogram data is a varied-length sequence $s_j = \{s_{j,1}, s_{j,2}, \dots, s_{j,p}, \dots\}$ representing airflow variation over time during exhalation. In the demographic information, dg_j denotes sex, da_j represents age, and ds_j indicates smoking status. The key patch concavity information includes $a_{j,pef-fe25}$ for the concavity information between PEF and FEF25, $a_{j,fe25-fe50}$ for the concavity information between FEF25 and FEF50, $a_{j,fe50-fe75}$ for the concavity information between FEF50 and FEF75, and $a_{j,fe75+}$ for the concavity information beyond FEF75.

DeepSpiro takes the monitoring data X of N instances as input and produces predicted disease labels $\hat{Y} = \{\hat{y}_1, \hat{y}_2, \dots, \hat{y}_j, \dots, \hat{y}_N\}$, explainer $\hat{E} = \{\hat{e}_1, \hat{e}_2, \dots, \hat{e}_j, \dots, \hat{e}_N\}$, COPD risk values R and the risk coefficients contributed by each monitoring input dimension $R_s, R_{dg}, R_{da}, R_{ds}$, etc., as well as the probabilities of future disease occurrence for high-risk undiagnosed COPD patients $F_{copd}, F_{1year}, F_{2year}, F_{3year}, F_{4year}, F_{5year+}, F_{non-copd}$. Here, $y_j \in \{0, 1\}$, where 0 indicates no disease, and 1 indicates the presence of disease.

Overview

Figure 8 illustrates the architecture of DeepSpiro, where the blocks represent different components of the framework. DeepSpiro is primarily divided into two tasks: COPD detection and early COPD prediction. The inputs to DeepSpiro always include spirometry data and demographic information (such as age, gender, smoking status, etc.). For the COPD detection task, the model outputs interpretability data and a COPD risk score. For the early prediction task, it outputs the probability of the individual developing COPD within the next 1–5 years. It is important to note that DeepSpiro only proceeds with the early prediction task if the detection task determines that the individual does not currently have COPD. Therefore, if the detection results indicate that the patient has COPD, the model's final output is

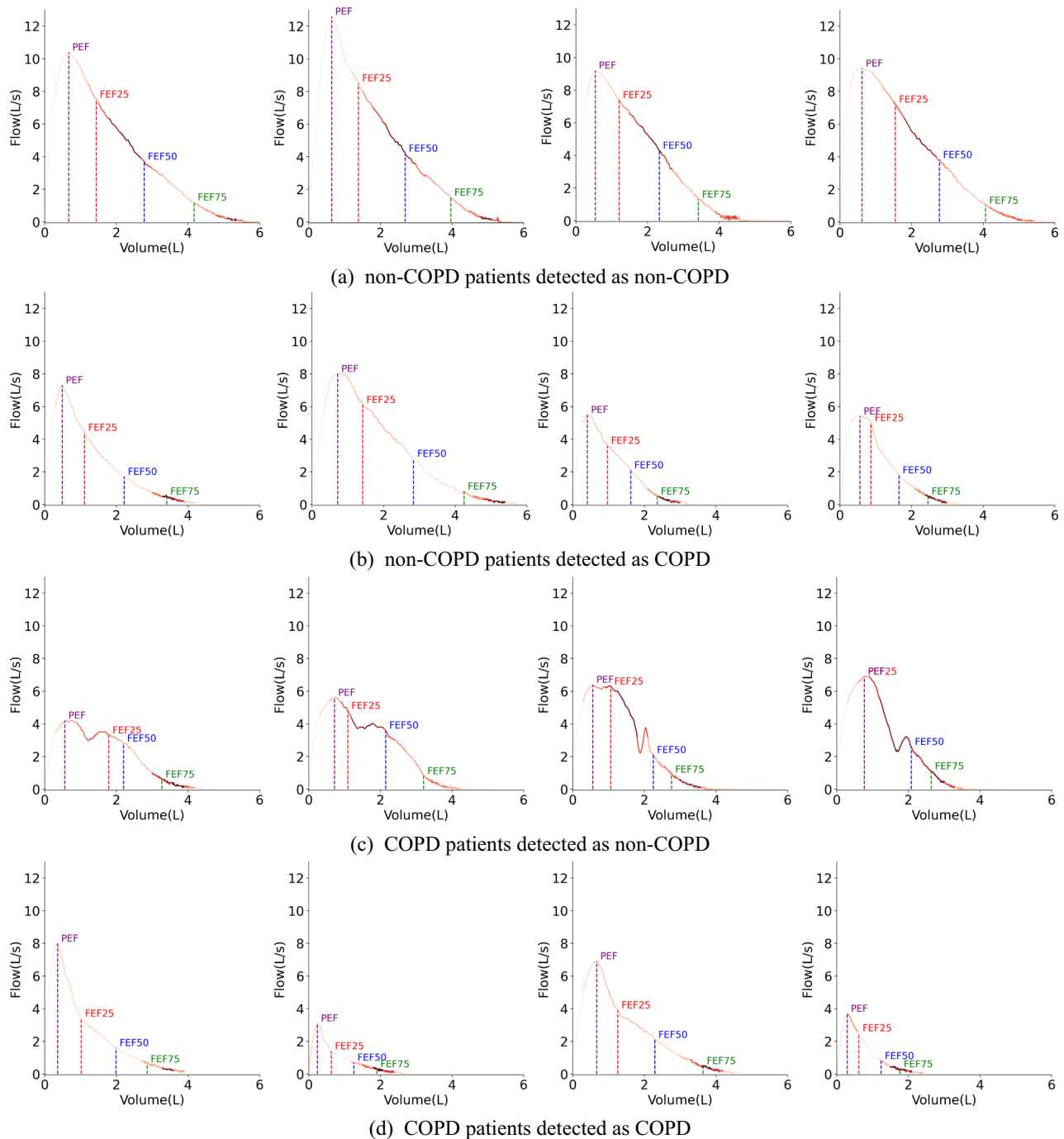


Fig. 7 | The explainer of the model. The brighter the color, the more attention the model pays. **a** A model interpretability figure for cases where non-COPD patients detected as non-COPD by the model. **b** A model interpretability figure for cases where non-COPD patients detected as COPD by the model. **c** A model

interpretability figure for cases where COPD patients detected as non-COPD by the model. **d** A model interpretability figure for cases where COPD patients detected as COPD by the model.

whether the patient has COPD. If not, the output includes both the current COPD status and the probability of developing COPD in the next 1–5 years.

DeepSpiro can be divided into four main modules: (1) SpiroSmoother (see Fig. 8I): constructing Volume-Flow curves guided by Time-Volume instability smoothing: This enhances the stability of key physiological information in the original Volume-Flow data. (2) SpiroEncoder (see Fig. 8II): extracting critical features from the variability-pattern of varied-length key patches: This dynamically identifies “key patches” and unifies the time-series representation, achieving the conversion of key physiological information from high-dimensional to low-dimensional. (3) SpiroExplainer (see

Fig. 8III): explaining model based on volume attention and heterogeneous feature fusion: This combines diagnostic probabilities with demographic information to assess COPD risk and provides an explainer of the model’s decisions. (4) SpiroPredictor (see Fig. 8IV): predicting the risk of COPD based on the variability pattern of key patch concavity: This integrates key concavity information to assess COPD risk for various future periods.

Specifically, the original spirogram data is processed with signal enhancement technology to obtain smoothed Volume-Flow data. The smoothed Volume-Flow data is used as input in DeepSpiro for varied-length time series data in order to extract essential physiological information.

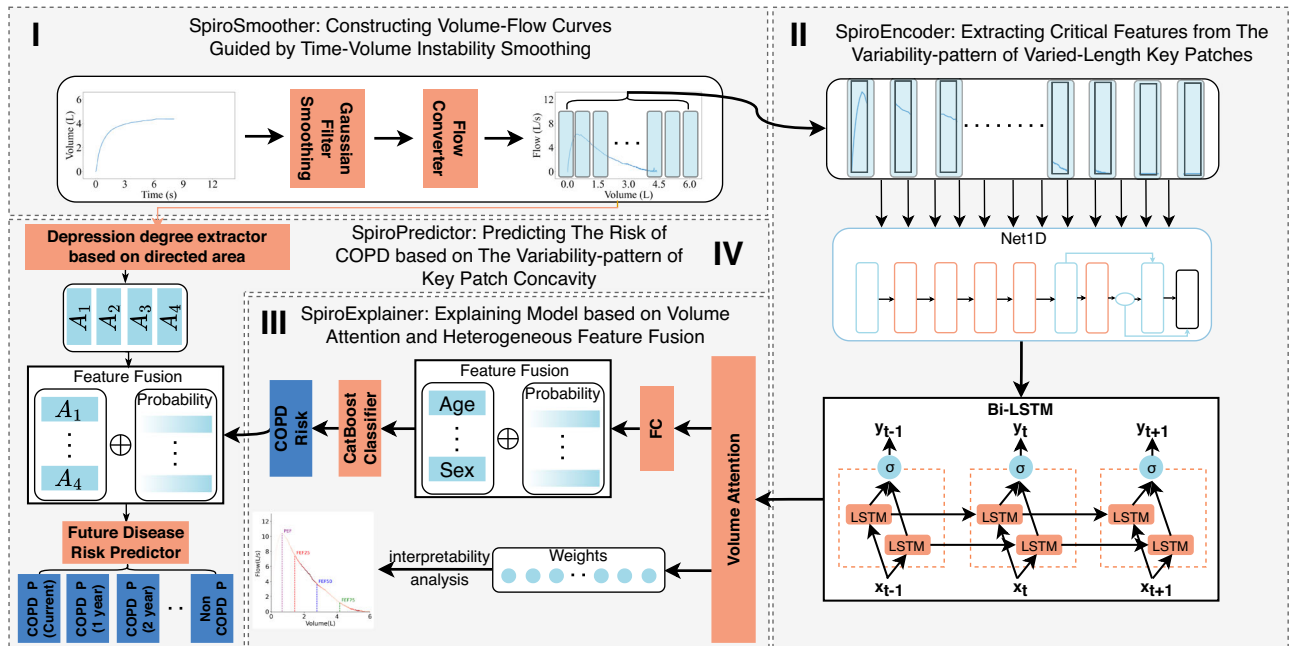


Fig. 8 | Framework overview. It is divided into four modules. In the first module, we process the varied-length original volume curve and convert it into a smoothed Volume-Flow curve. In the second module, we extract features from the varied-length Volume-Flow curve. In the third module, combined with demographic

information, we output the COPD detection results and model explainer. In the fourth module, based on concavity information, we output the risk values for COPD at different future periods.

Subsequently, the model combines the important physiological details and relevant demographic information using volume attention and heterogeneous feature fusion techniques. This process enables the model to output a COPD risk assessment and provide an explainer for its decisions. By using the COPD risk assessment values and concavity information obtained from the Volume-Flow signals in the feature fusion model, we can make precise predictions about the probability of future disease in high-risk patients who have not yet received a diagnosis.

SpiroSmoother: constructing Volume-Flow curves guided by Time-Volume instability smoothing

The current state-of-the-art (SOTA) method for detecting COPD from the Volume-Flow curve using deep learning has been proposed by J. Cosentino et al.¹². This method involves converting original Time-Volume curves into Volume-Flow curves, which can result in instability in the curves (as shown in Fig. 9C). Extensive experimental validation has been demonstrated that these instabilities can impact the final risk assessment for COPD patients. To address this, we employ SpiroSmoother. This approach precisely enhances the stability of the original Volume-Flow curves while preserving key physiological signals within the original Volume-Flow data.

As shown in Fig. 9C, we observe instability when converting the Time-Volume curve into a Time-Flow curve. Therefore, we believe that smoothing the original Time-Volume curve could help enhance the stability of both the Time-Flow curve and the Volume-Flow curve.

Gaussian filtering is employed to implement instability smoothing guidance for the original Time-Volume curves. This method achieves smoothing by taking a weighted average of the areas around the data points, where the weight of each data point is determined by a Gaussian function. For a given data point x_i its smoothed value y_i can be defined as

$$y_i = \frac{\sum_{j=-k}^k x_{i+j} \cdot g(j)}{\sum_{j=-k}^k g(j)} \quad (1)$$

In this formula, $g(j)$ represents the Gaussian function, k is the filter's window size, and x_{i+j} represents the original data point and its adjacent points. The

formula calculates the weighted average of data points around x_i , using weights derived from the Gaussian function $g(j)$. This function is centered at x_i , and its standard deviation determines both the range and the degree of smoothing. Through this method, Gaussian filtering effectively smooths the data while preserving important feature information.

To convert the smoothed Time-Volume curve into a Volume-Flow curve, we utilize the finite difference method to approximate the first derivative of the volume data with respect to time. Simply put, it involves calculating the volume change rate at consecutive time points to obtain corresponding flow data. The flow $Q(t)$ can be obtained by calculating the derivative of volume $V(t)$ with respect to time t , where flow $Q(t)$ is defined as

$$Q(t) \approx \frac{V(t + \Delta t) - V(t)}{\Delta t} \quad (2)$$

where $V(t)$ represents a set of Time-Volume curve data, Δt is the time interval between two adjacent time points, and t denotes the time points.

We linearly interpolate the calculated flow data to ensure that the Time-Flow curve is consistent with the original Time-Volume curve in terms of time data. In this way, we can construct the Volume-Flow curves $\mathcal{F}(v)$ using the Time-Flow curves $Q(t)$ and the Time-Volume curves $V(t)$. The Volume-Flow curves $\mathcal{F}(v)$ is defined as

$$\mathcal{F}(v) = Q(V^{-1}(v)) \quad (3)$$

As shown in Fig. 9c, the Time-Flow curve and the Volume-Flow curve have been constructed using the Time-Volume instability smoothing guidance technique and have effectively eliminated the original fluctuations, significantly enhancing the data's stability. In subsequent experiments, we have used these stable Volume-Flow curves to replace the original data for model training. Experimental results have confirmed that the curves processed through instability smoothing guidance demonstrate superior performance in risk assessment.

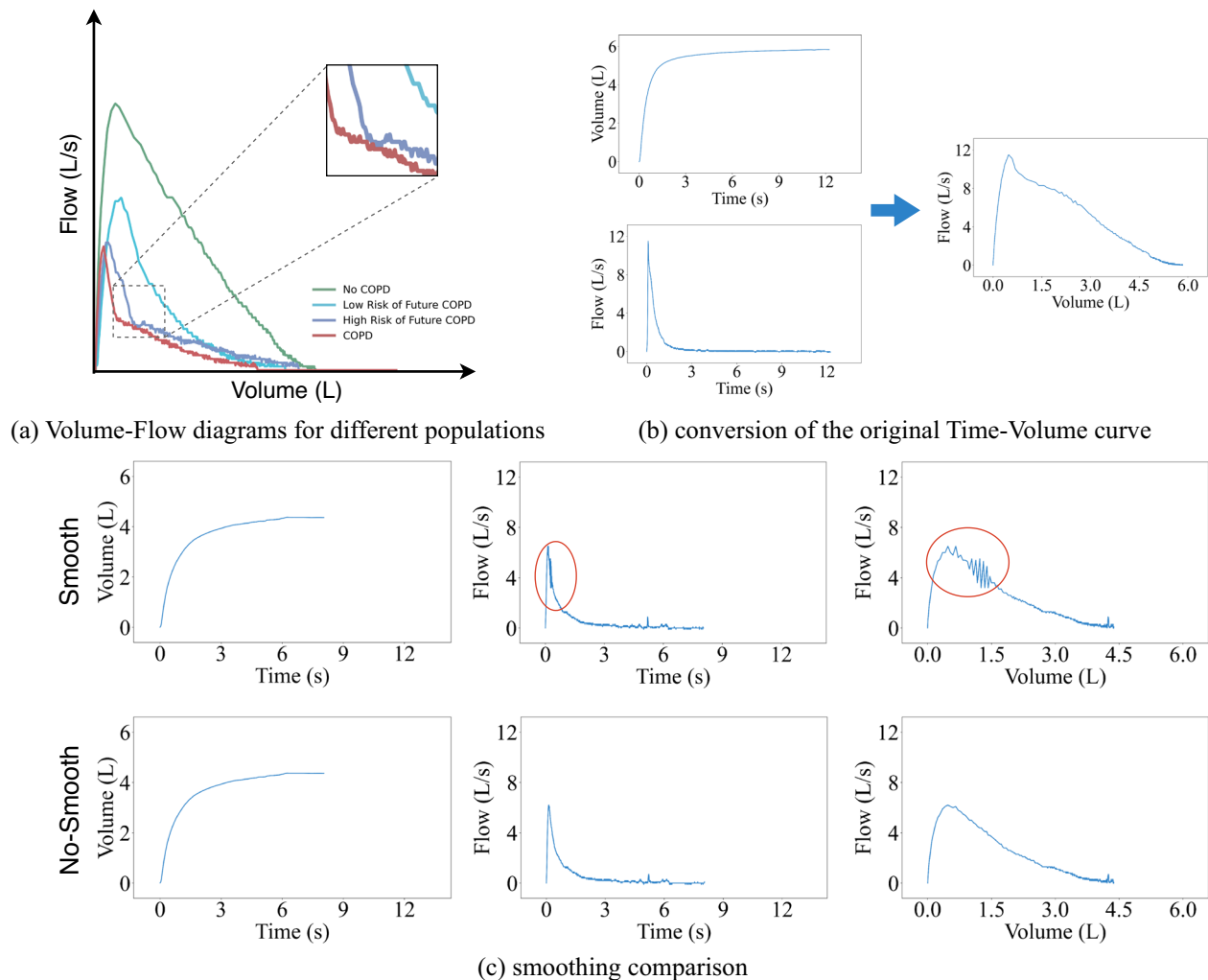


Fig. 9 | Spirometry curve and smoothing effects overview. **a** Examples of Volume-Flow diagrams for different populations. **b** The conversion of the original Time-Volume curve. To obtain the degree of airflow limitation, we need to use finite difference methods on the original Time-Volume curve to calculate the

corresponding Volume-Flow curve. **c** Smoothing comparison, the original curve shows fluctuations in some areas when converting to Time-Flow and Volume-Flow curves. After smoothing, the curves in the same positions become more stable.

SpiroEncoder: extracting critical features from the variability-pattern of varied-length key patches

Existing methods for modeling varied-length time series data generally use missing value imputation or downsampling truncation. The former introduces external noise, while the latter will likely lose meaningful data dependencies. We aim to preserve the key signals of the original data and adapt them to model input. To this end, we propose SpiroEncoder.

First, we need to reconstruct the original data. We have adopted an adaptive temporal decomposition method, which dynamically calculates the most suitable “key patches” count for each time series patch, accurately capturing the key Volume-Flow information within each time patch. By introducing the hyperparameter k , we can calculate the number of key patches based on the ratio of the sequence length to k , ensuring that each patch contains enough information for subsequent analysis. For a given length of time series L , the number of key patches S can be defined as

$$S = \left\lceil \frac{L}{k} \right\rceil \quad (4)$$

where $\lceil \cdot \rceil$ represents the ceiling function, ensuring that meaningful data dependencies are not lost even when L is not divisible by k .

We can obtain the number of key patches for each sequence using the above-mentioned method. Thus, we can divide each sequence into several

key patches. We feed all the key patches into the Net1D network³⁵ to extract the spatial features between the patches and the temporal features within each patch. The Net1D network extracts spatial features for us. We must compress the high-dimensional spirogram space into the same low-dimensional key patch to adapt these features for our subsequent modules.

First, we determine the length of the most extended sequence in the dataset and use this length along with the preset hyperparameter k . Calculate the number of key patches all sequences should have, thereby normalizing the temporal representation of the data. For the longest sequence length L_{\max} in the dataset, the maximum number of patches is defined as

$$N_{\max} = \left\lceil \frac{L_{\max}}{k} \right\rceil \quad (5)$$

We perform encoding for the key features of each sequence. This encoding marks the key patches and records them as a mask. For each sequence, we create a mask vector. Parts of the vector where key patches exist are masked as 1, and parts extending from the end of the existing key patches to the maximum number of patches are masked as 0.

We construct a zero tensor $O \in \mathbb{R}^{N \times S \times C}$, where N represents the number of samples, S represents the number of key patches, and C represents the number of output channels. Subsequently, we selectively apply these encodings to the zero tensor O using the mask vector. For each sample,

the mask vector indicates where to insert the features of key patches. Specifically, if the mask vector is one at a certain position, then we insert the encoding at the corresponding position in the zero tensor O ; if the mask vector is 0 at a position, the zero tensor remains unprocessed at that position. (This part also constitutes the padding of 0s for the sequence.) Thus, the zero tensor O transforms into a fixed-length sequence containing encoding information, and we refer to the transformed zero tensor O as the feature tensor of key patches O . For each sample i and its key patches j , we use the mask vector $M_{i,j}$ to decide whether to apply encoding. $O_{i,j,:}$ can be defined as

$$O_{i,j,:} = M_{i,j} \cdot E_{i,j,:} \quad (6)$$

where the features extracted by Net1D for each key patch are $E_{i,j,:} \in \mathbb{R}^C$.

Because the process of constructing the key patch feature tensor O involves padding with zeros, this padded information may introduce unnecessary noise when captured by the Bidirectional Long Short-Term Memory Network (Bi-LSTM). To address this, we aim to effectively ignore this portion of padded information before entering the Long Short-Term Memory Network.

First, we calculate the effective length Length_i for each sequence, which is the number of valid key patches in the sequence i . The effective length Length_i can be defined as

$$\text{Length}_i = \sum_{j=1}^S M_{i,j} \quad (7)$$

Then, we create a new tensor $P \in \mathbb{R}^{T \times C}$, where $T = \sum_{i=1}^N \text{Length}_i$, that is, the sum of the effective lengths of all sequences, and C is the number of output channels. Following the sequence order, we copy the features of valid key patches from the key patch feature tensor O into the tensor P . After the above steps, the tensor P contains only valid key patch features.

After the tensor P enters the Bidirectional Long Short-Term Memory network (Bi-LSTM)³⁶, the variability-pattern relationships among key patches are obtained.

SpiroExplainer: explaining model based on volume attention and heterogeneous feature fusion

Existing deep learning models still function as black boxes, capable only of producing diagnostic outcomes without offering an explainer for those results. In order to establish credibility with medical professionals and patients, the model must offer a decision-making process that is more open and easily understood by both parties. Therefore, through the dynamic volume attention integration method, we accurately highlight the time patches crucial for model predictions using a method for explaining models based on volume attention and heterogeneous feature fusion.

Data enters the first linear transformation layer as it passes through the volume attention layer. This layer provides an initial feature representation for each time step, which is specifically manifested as

$$X' = W_1 X + b_1 \quad (8)$$

Here, W_1 and b_1 are the weight and bias of the linear layer, respectively.

To perform a nonlinear transformation, We follow up by applying the Swish activation function to the linearly transformed data X' . The Swish activation function helps to improve the model's ability to capture complex features. The transformed data after applying Swish activation function is given by

$$X'' = X' \cdot \sigma(X') \quad (9)$$

where $\sigma(X') = \frac{1}{1+e^{-X'}}$ is the sigmoid function applied element-wise to X' . Next, we apply a bilinear transformation to the activated data X'' by using a

bilinear weight matrix. This is specifically represented as

$$X''' = X'' W_{\text{bil}} \quad (10)$$

where W_{bil} is the bilinear weight matrix.

After passing through the bilinear transformation, the data X''' goes through another linear layer to calculate the attention scores S for each time step. Finally, the S scores are normalized using the *softmax* function to obtain the volume attention weights.

We combine the volume attention information with the smoothed Time-Volume curve data. The decision-making process within the deep neural network model can be transformed into visualized graphs. The graphs provide a clear representation of the specific data regions that the model prioritizes when making decisions, thereby improving the transparency of the model and bolstering the credibility of its decision-making process.

Research indicates that there is a significant correlation between smoking, age, and other demographic information with COPD. Therefore, we aim to incorporate some demographic information to enhance the effectiveness of our model. For this purpose, after passing through the volume attention layer, we first conduct an initial COPD assessment through a fully connected layer, resulting in $\hat{P} = \{\hat{p}^1, \hat{p}^2, \dots, \hat{p}^j, \dots, \hat{p}^N\}$, where \hat{P} represents a series of COPD assessment values for the samples, with $\hat{p}^j \in [0, 1]$ denoting the assessment value for the j th sample. In the subsequent steps, we use these evaluation values to generate the feature $F_{\text{probability}}$, representing the probability of disease. Specifically, $F_{\text{probability}}$ is obtained by applying the *Softmax* function to \hat{P} , as follows:

$$F_{\text{probability}} = \text{Softmax}(\hat{P}) \quad (11)$$

Additionally, we extract features F_{struct} from demographic and other structured data. Then, we combine these features with $F_{\text{probability}}$ using the following fusion operation:

$$F_{\text{fusion}} = F_{\text{probability}} \oplus F_{\text{struct}} \quad (12)$$

where \oplus denotes the concatenation of features.

By utilizing the Gradient Boosting framework (Catboost) to process the fused features F_{fusion} , we predict the risk of COPD. The COPD risk assessment obtained in this manner outperforms the assessment results derived solely from unstructured varied-length Volume-Flow curve data.

SpiroPredictor: predicting the risk of COPD based on the variability-pattern of key patch concavity

In future risk prediction tasks, we used the degree of concavity in individual Volume-Flow graphs as one of the features for prediction. As shown in Fig. 9a, high-risk individuals and COPD patients exhibit more significant concavity, while low-risk individuals exhibit fewer concavity than high-risk individuals and COPD patients. The following provides a detailed explanation of how this feature of concavity is extracted.

Since the spirogram itself is not strictly monotonic, each phase consists of a set of concave and convex curve segments (such as the FEF50–FEF75 phase in Fig. 10a), making it difficult to simply judge the degree of collapse in the curve. Therefore, we have defined a concavity-based directed area metric to approximate the degree of collapse in such non-monotonic curves, in order to characterize the overall concave-convex nature exhibited by the curve.

First, given the Volume-Flow curve $\mathcal{F}(v)$ for a particular phase s , we define its baseline $BL(s, v)$ as the straight line connecting the starting and ending points of the curve in that phase, where the starting volume of the phase curve is $B(s)$ and the ending volume is $G(s)$. The slope of the baseline

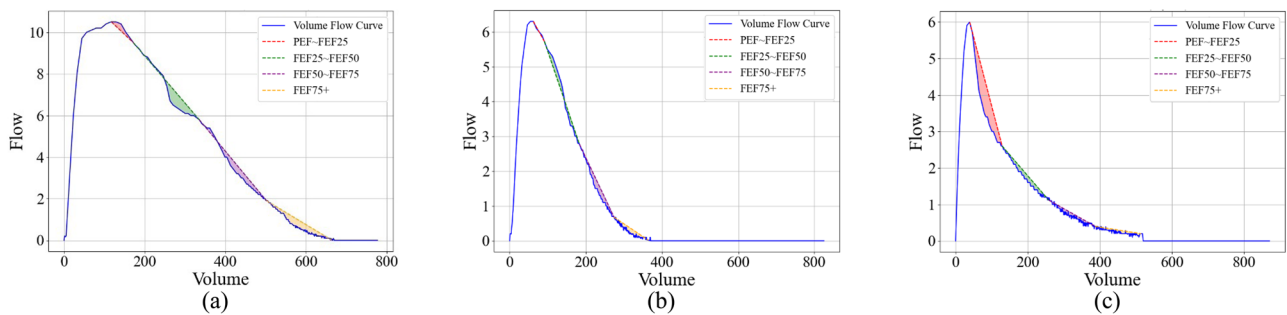


Fig. 10 | Volume-Flow curves illustrating phase-specific concavity and airway collapse. **a** A representative Volume-Flow curve divided into four phases: early (PEF-FEF25), mid-early (FEF25-FEF50), mid-late (FEF50-FEF75), and late (FEF75+), highlighting concave and convex segments relative to the baseline.

b Volume-Flow curve of a healthy individual, showing airway collapse occurring in the late phase (FEF75+). **c** Volume-Flow curve of a COPD patient, showing early-stage airway collapse in the early phase (PEF-FEF25).

$BL(s, v)$, denoted as $m(s)$, can be defined as:

$$m(s) = \frac{\mathcal{F}((G(s)) - \mathcal{F}((B(s)))}{G(s) - B(s)} \quad (13)$$

The intercept $b(s)$ of the baseline $BL(s)$ can be defined as

$$b(s) = \mathcal{F}(B(s)) - m(s) \times B(s) \quad (14)$$

Thus, the baseline BL for a given phase s can be expressed as: where v is a specific volume value on the Volume-Flow curve. Next, we define a concavity measure $\mathcal{C}(s)$ based on directed area, which is the directed area of phase s :

$$\mathcal{C}(s) = \sum_{v=B(s)}^{G(s)} (BL(s, v) - \mathcal{F}(v)) \quad (15)$$

Therefore, if the curve of phase s is convex, meaning that the Volume-Flow curve of this phase lies above its baseline (as in the PEF-FEF25 phase in Fig. 10a), the directed area is negative, i.e., the concavity measure is negative. Conversely, if the curve of phase s is concave, meaning that the Volume-Flow curve of this phase lies below its baseline (as in the FEF25-FEF50 phase in Fig. 10a), the directed area is positive, i.e., the concavity measure is positive. According to the above definition, the more concave a curve is, the larger the concavity measure, and conversely, the more convex a curve is, the smaller the concavity measure. The concavity measure of a given phase represents the concavity information for that phase.

Furthermore, for Volume-Flow curves in some phases that are neither strictly monotonic increasing nor monotonic decreasing, the total concavity measure of phase s depends on the sum of the concave area (positive value) and the convex area (negative value). If the concave area in phase s is larger, the concavity measure is greater (positive value), whereas if the convex area is larger, the concavity measure is smaller (negative value). For example, in the FEF50-FEF75 phase in Fig. 10a, the concave area is much larger than the convex area. Thus, by calculation, the concavity measures $\mathcal{C}(\text{FEF50} - \text{FEF75})$ based on the directed area are positive, indicating that this phase exhibits a significant concave trend rather than a convex one.

If we divide the spirogram into four phases: early (PEF-FEF25), mid-early (FEF25-FEF50), mid-late (FEF50-FEF75), and late (FEF75+), we find that in healthy individuals, the spirogram typically remains full during the early phases, and collapse does not occur until the late phase (such as in the FEF75+ phase, as shown in Fig. 10b). In contrast, COPD patients often experience early-stage collapse (such as in the PEF-FEF25 phase, as shown in Fig. 10c). Therefore, we aim to use the proposed concavity measure based on directed area to study the relationship between the phase in which collapse occurs and future disease risk.

We fused the COPD Risk output from the SpiroExplainer module with the key concavity information of the Volume-Flow curve and input this combined data into a future disease risk predictor for assessing future risk. Our future disease risk predictor can be classifiers like Catboost³⁷, Xgboost³⁸, Random Forest³⁹, etc. The model, having been obtained in this way, accurately predicts the probability of illness in the next 1–5 years, and beyond for high-risk patients who have not yet been diagnosed.

Dataset and preprocessing

The data used in our study was obtained from the UK Biobank. We specifically focused on information from 453,558 patients who underwent their initial pulmonary function test (The test outputs the Time-Volume time-series curve). It is worth noting that due to factors such as genetics, environment, or lifestyle, the spirogram characteristics may vary among different population groups. Therefore, in our data processing, we chose to conduct experiments only on the largest population group in the UK Biobank, which is of European descent. The DeepSpiro model relies on detailed and precise patterns in spirogram time series to detect and predict COPD. Therefore, incomplete or distorted data may lead to errors in the model's predictions. To mitigate the issue of inaccurate predictions caused by poor data quality, we have implemented several data preprocessing measures. These measures ensure that only high-quality and reliable data is used for model training and validation, thereby enhancing the robustness and accuracy of the DeepSpiro model.

The original Time-Volume (expiratory volumes over a period of time) measurements are extracted from UK Biobank field 3066, which contains expiratory volumes recorded every 10 ms in ml. The Time-Volume measurements are used from each participant's first visit. To ensure the Time-Volume data are valid, we consult UK Biobank field 3061; if the value in field 3061 is either 0 or 32, the expiration is considered valid. If multiple valid expirations are available, we choose the first one in sequence. To control the quality of the expirations, we review the UK Biobank fields 3062 (FVC), 3063 (FEV1), and 3064 (PEF); if any of these are in the top or bottom 0.5% of the observed values, any expiratory data will discard from that patient. The original expiratory volume measurements are converted from milliliters to liters, and the corresponding flow curves are calculated by approximating the first derivative with respect to time using finite differences. As shown in Fig. 9b, the Time-Flow and Time-Volume curves combine to generate a one-dimensional Volume-Flow curve time series.

We refer to the labeling methods in Justin Cosentino et al.'s research¹². Specifically, the labels are generated by combining information from various sources in the UK Biobank. We use a binary COPD label that is determined through self-reporting, hospital admissions, and primary care reports to train DeepSpiro. Self-reported COPD status is derived from codes 1112, 1113, and 1472 in the UK Biobank field 20002. Hospital-reported COPD status comes from codes like J430 and others in the UK Biobank field 41270, and codes like 4920 and others in field 41271. The COPD status from

Table 3 | The following table lists the required fields and corresponding codes during label extraction

Field Id	Code	Type
20002	1112, 1113, 1472	Self-report
41270	J430, J431, J432, J438	Hospitalization
	439J, J440, J441, J448	
	J449	
41271	4920, 4928, 4929, 496X	Hospitalization
42040	J430, J431, J432, J438	Primary care
	439J, J440, J441, J448	
	J449	

A code suffixed with 'X' stands for any code starting with the figures preceding the X.

Table 4 | The number of COPD cases under different categories

Category	Nums	COPD case
All	348,039	19,308
Hospitalization	348,039	16,254
Death	24,105	4243

primary care reports is required using TRUD to map the UK Biobank's gp-clinical to read codes from UK Biobank field 42040, with mapped fields including codes like J430, among others. For specific extraction of field codes, please refer to Table 3. In addition, death information is used in subsequent analyses. Death information is recorded in UK Biobank field 40000. If any value is present in this field, we consider the patient to have died.

After completing the necessary data processing, the data were categorized into three groups: All (the complete dataset), Hospitalization (patients reporting hospitalization), and Death (patients who passed away). The number of COPD cases in each category is summarized in the Table 4.

Currently, the dataset includes a total of 348,039 participants, divided into training and testing sets at an 8:2 ratio. Specifically, the training set contains 278,431 participants, while the testing set comprises 69,608 participants.

Evaluations

To evaluate the performance of this method, we used a validation set to calculate the model's area under the receiver operating characteristic curve (AUROC), area under the precision-recall curve (AUPRC), and F1 scores.

AUROC: The area under the receiver operating characteristic curve (AUROC) is a widely used metric for evaluating the performance of classification models, especially in binary classification problems. It is calculated by varying the classification threshold to determine the true positive rate ($TPR = \frac{TP}{TP+FN}$) and the false positive rate ($FPR = \frac{FP}{FP+TN}$). The ROC curve shows the relationship between the true positive rate and the false positive rate. The value of AUROC is the area under the ROC curve, which can be approximated as $AUROC = \int_0^1 R(F) dF$, where F represents the false positive rate and $R(F)$ is the corresponding true positive rate. Numerical methods generally estimate the FPR values f_1, f_2, \dots, f_n and the corresponding TPR values t_1, t_2, \dots, t_n . The trapezoidal area approximation of the AUROC is: $AUROC \approx \sum_{i=1}^{n-1} \frac{(f_{i+1}-f_i) \times (t_i+t_{i+1})}{2}$.

AUPRC: The area under the precision-recall curve (AUPRC) is a commonly used metric to evaluate models for classification problems. The PR curve is derived by varying the classification threshold to calculate precision ($\text{precision} = \frac{TP}{TP+FP}$) and recall ($\text{recall} = TPR = \frac{TP}{TP+FN}$). It can be approximated as $\int P(r) dR(r)$, where $P(r)$ represents precision, $R(r)$ denotes recall, and r represents the decision threshold.

F1 score: The F1 score is a crucial metric for evaluating the accuracy of a model. It is the harmonic mean of precision and recall ($F1 = 2 \times \frac{\text{precision} \times \text{recall}}{\text{precision} + \text{recall}}$), providing a single metric that considers both the accuracy and completeness of the model.

These metrics collectively help us comprehensively assess the model's performance across different aspects. In this experiment, the AUROC, AUPRC, and F1 scores are calculated using functions from the scikit-learn library and code from Justin Cosentino et al.¹².

Compared methods

To assess the effectiveness of DeepSpiro, we compared it against several baseline methods, including:

- **FEV1/FVC ratio:** The FEV1/FVC ratio is the GOLD standard in clinical practice for determining whether an individual has COPD. If an individual's ratio is less than 70%, they are considered to have COPD. Therefore, this method also be used as our baseline.
- **ResNet18 (Nature Genetics):** The ResNet18 convolutional neural network (CNN) is capable of efficiently learning features from large-scale datasets. In a paper published in Nature Genetics, Justin Cosentino et al. utilized this model for the task of COPD detection¹². Therefore, we also adopt this model as the primary baseline.

Data availability

Data from the UK Biobank, which is available after the approval of an application at <https://www.ukbiobank.ac.uk>. UK Biobank received ethical approval from the National Information Governance Board for Health and Social Care and the National Health Service North West Centre for Research Ethics Committee (Ref: 21/NW/0157).

Code availability

Our code is publicly available at <https://github.com/yudaleng/COPD-Early-Prediction>.

Received: 11 June 2024; Accepted: 30 December 2024;

Published online: 15 February 2025

References

1. Agustí, A. et al. Global initiative for chronic obstructive lung disease 2023 report: gold executive summary. *Am. J. Respir. Crit. Care Med.* **207**, 819–837 (2023).
2. Ferrera, M. C., Labaki, W. W. & Han, M. K. Advances in chronic obstructive pulmonary disease. *Annu. Rev. Med.* **72**, 119–134 (2021).
3. Maeda, T. & Dransfield, M. T. Chronic obstructive pulmonary disease and cardiovascular disease: mechanistic links and implications for practice. *Curr. Opin. Pulm. Med.* **30**, 141–149 (2024).
4. Stolz, D. et al. Towards the elimination of chronic obstructive pulmonary disease: a lancet commission. *Lancet* **400**, 921–972 (2022).
5. Aaron, S. D. et al. Early diagnosis and treatment of COPD: the costs and benefits of case-finding. *Am. J. Respir. Crit. Care Med.* **209**, 928–937 (2024).
6. MacLeod, M. et al. Chronic obstructive pulmonary disease exacerbation fundamentals: Diagnosis, treatment, prevention and disease impact. *Respirology* **26**, 532–551 (2021).
7. Torén, K. et al. The ratio FEV1/FVC and its association to respiratory symptoms—a Swedish general population study. *Clin. Physiol. Funct. imaging* **41**, 181–191 (2021).
8. Bhatt, S. P. et al. Discriminative accuracy of FEV1: FVC thresholds for COPD-related hospitalization and mortality. *JAMA* **321**, 2438–2447 (2019).
9. Hoesein, F. A. M., Zanen, P. & Lammers, J.-W. J. Lower limit of normal or FEV1/FVC < 0.70 in diagnosing copd: an evidence-based review. *Respir. Med.* **105**, 907–915 (2011).
10. Bhatt, S. P. et al. Fev1/fvc severity stages for chronic obstructive pulmonary disease. *Am. J. Respir. Crit. Care Med.* **208**, 676–684 (2023).

11. Medbø, A. & Melbye, H. Lung function testing in the elderly—can we still use FEV1/FVC < 70% as a criterion of COPD? *Respir. Med.* **101**, 1097–1105 (2007).
12. Cosentino, J. et al. Inference of chronic obstructive pulmonary disease with deep learning on raw spirograms identifies new genetic loci and improves risk models. *Nat. Genet.* **55**, 787–795 (2023).
13. Fernandez-Quilez, A. Deep learning in radiology: ethics of data and on the value of algorithm transparency, interpretability and explainability. *AI Ethics* **3**, 257–265 (2023).
14. Sun, Y. et al. Covid-19 diagnosis based on swin transformer model with demographic information fusion and enhanced multi-head attention mechanism. *Expert Syst. Appl.* **243**, 122805 (2024).
15. Mirza, S., Clay, R. D., Koslow, M. A. & Scanlon, P. D. COPD guidelines: a review of the 2018 gold report. *Mayo Clin Proc.* **93**, 1488–1502 (2018).
16. Han, M. K. et al. From gold 0 to pre-COPD. *Am. J. Respir. Crit. Care Med.* **203**, 414–423 (2021).
17. Venkatesan, P. Gold copd report: 2024 update. *Lancet Respir. Med.* **12**, 15–16 (2024).
18. Hong, H. & Hong, S. simpleNomo: a Python package of making nomograms for visualizable calculation of logistic regression models. *Health Data Sci.* **3**, 0023 (2023).
19. Yang, I. A., Jenkins, C. R. & Salvi, S. S. Chronic obstructive pulmonary disease in never-smokers: risk factors, pathogenesis, and implications for prevention and treatment. *Lancet Respir. Med.* **10**, 497–511 (2022).
20. Alabdullah, A. A. et al. Prediction of rapid chloride penetration resistance of metakaolin based high strength concrete using light GBM and XGBoost models by incorporating SHAP analysis. *Constr. Build. Mater.* **345**, 128296 (2022).
21. Avian, C., Mahali, M. I., Putro, N. A. S., Prakosa, S. W. & Leu, J.-S. Fx-net and purenet: convolutional neural network architecture for discrimination of chronic obstructive pulmonary disease from smokers and healthy subjects through electronic nose signals. *Comput. Biol. Med.* **148**, 105913 (2022).
22. Dianat, B. et al. Classification of pulmonary sounds through deep learning for the diagnosis of interstitial lung diseases secondary to connective tissue diseases. *Comput. Biol. Med.* **160**, 106928 (2023).
23. Almeida, S. D. et al. Prediction of disease severity in COPD: a deep learning approach for anomaly-based quantitative assessment of chest ct. *Eur. Radiol.* **34**, 4379–4392 (2024).
24. Zhuang, Y. et al. Deep learning on graphs for multi-omics classification of COPD. *PLoS ONE* **18**, e0284563 (2023).
25. Zhu, Z. et al. Development and application of a deep learning-based comprehensive early diagnostic model for chronic obstructive pulmonary disease. *Respir. Res.* **25**, 167 (2024).
26. Wu, Y. et al. Deep CNN for COPD identification by multi-view snapshot integration of 3d airway tree and lung field. *Biomed. Signal Process. Control* **79**, 104162 (2023).
27. Yin, H. et al. A machine learning model for predicting acute exacerbation of in-home chronic obstructive pulmonary disease patients. *Comput. Methods Prog. Biomed.* **246**, 108005 (2024).
28. Willer, K. et al. X-ray dark-field chest imaging for detection and quantification of emphysema in patients with chronic obstructive pulmonary disease: a diagnostic accuracy study. *Lancet Digit. Health* **3**, e733–e744 (2021).
29. Yin, C. et al. Fractional dynamics foster deep learning of COPD stage prediction. *Adv. Sci.* **10**, 2203485 (2023).
30. Kumar, S. et al. A novel multimodal framework for early diagnosis and classification of COPD based on CT scan images and multivariate pulmonary respiratory diseases. *Comput. Methods Prog. Biomed.* **243**, 107911 (2024).
31. Sharma, P., Pandey, A. K. & Bhattacharyya, D. K. Determining crucial genes associated with covid-19 based on COPD findings. *Comput. Biol. Med.* **128**, 104126 (2021).
32. Han, Z. et al. White blood cell count and chronic obstructive pulmonary disease: a Mendelian randomization study. *Comput. Biol. Med.* **151**, 106187 (2022).
33. Yang, Y. et al. Revealing exph5 as a potential diagnostic gene biomarker of the late stage of COPD based on machine learning analysis. *Comput. Biol. Med.* **154**, 106621 (2023).
34. Lee, Y. et al. Meta-analysis of single-cell RNA-sequencing data for depicting the transcriptomic landscape of chronic obstructive pulmonary disease. *Comput. Biol. Med.* **167**, 107685 (2023).
35. Hong, S. et al. Holmes: Health online model ensemble serving for deep learning models in intensive care units. In *Proc. 26th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining* 1614–1624. <https://doi.org/10.1145/3394486.3403212> (Association for Computing Machinery, New York, 2020).
36. Hochreiter, S. & Schmidhuber, J. Long short-term memory. *Neural Comput.* **9**, 1735–1780 (1997).
37. Wei, X., Rao, C., Xiao, X., Chen, L. & Goh, M. Risk assessment of cardiovascular disease based on SOLSSA-CatBoost model. *Expert Syst. Appl.* **219**, 119648 (2023).
38. Raihan, M. J., Khan, M. A.-M., Kee, S.-H. & Nahid, A.-A. Detection of the chronic kidney disease using XGBoost classifier and explaining the influence of the attributes on the model using SHAP. *Sci. Rep.* **13**, 6263 (2023).
39. Sun, Z. et al. An improved random forest based on the classification accuracy and correlation measurement of decision trees. *Expert Syst. Appl.* **237**, 121549 (2024).
40. Mirsadraee, M., Salarifar, E. & Attaran, D. Evaluation of superiority of FEV1/VC over FEV1/FVC for classification of pulmonary disorders. *J. Cardio-Thorac. Med.* **3**, 355–359 (2015).

Acknowledgements

The authors gratefully acknowledge the financial supports by the National Natural Science Foundation of China under Grant 62202332, Grant 62102008, Grant 62376197, Grant 62020106004 and Grant 92048301; Clinical Medicine Plus X—Young Scholars Project of Peking University, the Fundamental Research Funds for the Central Universities (PKU2024LCXQ030); PKU-OPPO Fund (B0202301); Beijing Natural Science Foundations (QY23040); Natural Science Foundation of Tianjin City (23JCYBJC00360); College Student Innovation and Entrepreneurship Training Program (202410060109).

Author contributions

S.M., Y. Zhou, and S.H. conceptualized the study. S.H., Y. Zhou, and S. Chen discussed the solution. S.M., Y. Zhou, and S.H. developed the methodology. S.M. developed the software. S.M., Y. Zhou, S.H., and S. Cao validated the results. S.M., S. Cao., X.L., and Y. Zhou performed the formal analysis. J. Xu, Y.W., S.G., and Q.Z. carried out the investigation. S.M., S.H., and J. Xie were responsible for data curation. S.M., Y. Zhou, X.L., and S.H. wrote the original draft. Y. Zhou, Y. Zhang, and S.H. reviewed and edited the manuscript. S.M. and Y.W. were responsible for visualization. All authors read and approved the final manuscript.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information The online version contains supplementary material available at <https://doi.org/10.1038/s41540-025-00489-y>.

Correspondence and requests for materials should be addressed to Yuxi Zhou, Yong Zhang, Shengyong Chen or Shenda Hong.

Reprints and permissions information is available at <http://www.nature.com/reprints>

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

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

© The Author(s) 2025