Check for updates

### **OPEN ACCESS**

EDITED BY Liang Zhao, Dalian University of Technology, China

REVIEWED BY Eleuterio A. Sánchez Romero, Hospital Universitario Puerta de Hierro Majadahonda, Spain Suixue Wang, Hainan University, China

\*CORRESPONDENCE Min Feng ⊠ fengfeng0716@163.com Yanxia Li ⊠ liyanxia001@163.com

 $^{\dagger}\mbox{These}$  authors have contributed equally to this work

RECEIVED 28 March 2025 ACCEPTED 09 May 2025 PUBLISHED 11 July 2025

### CITATION

Yi Y, Shi L, Liu H, Wang M, Feng M and Li Y (2025) COPD-MMDDxNet: a multimodal deep learning framework for accurate COPD diagnosis using electronic medical records. *Front. Med.* 12:1601736. doi: 10.3389/fmed.2025.1601736

### COPYRIGHT

© 2025 Yi, Shi, Liu, Wang, Feng and Li. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

# COPD-MMDDxNet: a multimodal deep learning framework for accurate COPD diagnosis using electronic medical records

Yuanyuan Yi<sup>1†</sup>, Lei Shi<sup>2†</sup>, Haoran Liu<sup>3</sup>, Mingyu Wang<sup>3</sup>, Min Feng<sup>1\*</sup> and Yanxia Li<sup>1\*</sup>

<sup>1</sup>Department of Respiratory Medicine, The First Affiliated Hospital of Dalian Medical University, Dalian, China, <sup>2</sup>Department of Emergency Medicine, The First Affiliated Hospital of Dalian Medical University, Dalian, China, <sup>3</sup>Dalian Medical University, Dalian, China

COPD affects approximately 391 million people globally. While spirometry is recognized as the gold standard for diagnosing COPD according to the GOLD guidelines, its availability is limited in primary healthcare settings, particularly in low- and middle-income countries. Furthermore, spirometry requires patient cooperation, which may be challenging for individuals with physical limitations or comorbidities, potentially impacting its accuracy. As a result, there is a need for alternative diagnostic methods, particularly those suited for resource-constrained environments. This study proposes a novel multimodal deep learning framework, COPD-MMDDxNet, which integrates structured pulmonary CT reports, blood gas analysis, and hematological analysis from electronic medical records (EMRs) to overcome the limitations of existing diagnostic methods. This framework develops the first multimodal diagnostic tool for COPD that does not rely on spirometry. It innovatively fuses cross-modal data, incorporating four key components: parametric numerical embedding, hierarchical interaction mechanisms, contrastive regularization strategies, and dynamic fusion coefficients. These innovations significantly enhance the model's ability to capture complex cross-modal relationships, thereby improving diagnostic accuracy. The dataset used in this study comprises 800 COPD patients, with a balanced age and sex distribution, and data were collected over a 24-month period. Experimental results demonstrate that COPD-MMDDxNet outperforms traditional single-modality models and other state-of-the-art multimodal models in terms of accuracy (81.76%), precision (78.87%), recall (77.78%), and F1 score (78.32%). Ablation studies confirm the critical importance of each model component, particularly the contrastive learning module and cross-modal attention mechanism, in enhancing model performance. This framework offers a robust solution for more accurate and accessible COPD diagnosis, particularly in resource-constrained environments, without the need for spirometry.

#### KEYWORDS

COPD, multimodal deep learning, diagnosis, cross-modal interaction, electronic medical records

## **1** Introduction

Chronic Obstructive Pulmonary Disease (COPD) is a serious chronic condition affecting approximately 391 million people globally (1). The primary risk factors for COPD include smoking, air pollution, and occupational exposure to harmful substances (2). Early diagnosis and management are crucial for improving patient outcomes, as timely identification of COPD plays a significant role in reducing acute exacerbations and slowing disease progression (3).

Although the Global Initiative for Chronic Obstructive Lung Disease (GOLD) guidelines recognize spirometry as the gold standard for diagnosing COPD, its availability in primary healthcare facilities is limited (4). Particularly in low- and middleincome countries, where the availability of spirometry equipment remains limited, failing to meet the screening needs of the population (5). Furthermore, spirometry requires patients to perform forced expiratory maneuvers, which may be challenging for patients with physical limitations or comorbidities (6). Therefore, despite its importance as a diagnostic tool, spirometry has inherent limitations, particularly in certain clinical and resource-constrained settings. In this context, multi-view dataowing to its rich complementary and consistent information across modalities-has garnered substantial interest for its potential to significantly enhance the performance of downstream diagnostic tasks (7).

Given these challenges, it is crucial to research disease diagnosis and treatment in resource-limited environments (8). Developing alternative diagnostic approaches that do not rely on spirometry, particularly non-invasive technologies, is essential. This study presents a solution in the form of a deep learning model specifically designed for diagnosing COPD in environments where spirometry may not be accessible.

In recent years, the rapid development of artificial intelligence (AI) technology has provided new opportunities for the diagnosis and treatment of COPD. Gompelmann et al. (9) investigate the application of AI software, ArtiQ.PFT, to assist pulmonologists in diagnosing interstitial lung disease by improving the interpretation of pulmonary function tests. Zhao et al. (10) proposed a Multi-Sentence Complementary Generative Adversarial Network to synthesize COPD-related diagnostic images by fusing semantic information from multiple clinical narrative sentences-leveraging BERT for semantic feature extraction, a multi-semantic fusion module, a pre-trained cross-modal contrast similarity model for fine-grained loss, and a multi-sentence joint discriminator to ensure that the generated images accurately reflect all input descriptions. Alves Pegoraro et al. (11) employed a Hidden Markov Model to analyze a range of breathing variables, including daily measures of oxygen therapy duration, mean breathing rate, and mean inspiratory amplitude, with the goal of detecting pre-exacerbation periods in COPD patients. However, existing studies primarily rely on single-modal data, such as pulmonary function tests or imaging, limiting the comprehensiveness of COPD diagnosis. At the same time, the widespread adoption of electronic medical record (EMR) systems globally has provided rich multimodal data. When effectively utilized, this information can enhance diagnostic accuracy.

Quennelle et al. (12) utilized EMR data to extract rare adverse events associated with pediatric cardiac catheterization, applying active learning (AL) methods to annotate the dataset, which was subsequently used to train a deep learning text classifier, incorporating clinical expert knowledge. Takeuchi et al. (13) applied natural language processing and machine learning techniques to unstructured electronic health records (EHRs) of lung cancer patients. They developed a method to predict tumor response evaluations, treatment durations, and survival curves, thereby aiding researchers and clinicians in analyzing and utilizing EHRs more efficiently.

These studies demonstrate that integrating AI with EMRs can uncover latent information within patient data, providing new approaches for early diagnosis and personalized treatment of COPD. However, current COPD diagnostic practices have yet to fully leverage the multimodal information contained in EMRs, limiting the effectiveness of AI applications in this field. Therefore, developing new methods to fully exploit the multimodal data within EMRs is crucial for improving the accuracy and efficiency of COPD diagnosis.

To address these challenges, this study proposes a novel multimodal deep learning framework, COPD-MMDDxNet, which integrates pulmonary imaging reports, medical history texts, and structured data such as blood gas analysis and hematological analysis from EMR systems to overcome the limitations of existing diagnostic methods. The proposed framework introduces four key innovations to address the challenges in COPD diagnosis:

- **Parametric numerical embedding:** A nonlinear representation mechanism utilizing learnable affine parameters replaces conventional linear transformations, enabling adaptive feature scaling across heterogeneous data modalities.
- Hierarchical interaction mechanism: Gated cross-attention layers facilitate multi-granular feature synergy between imaging biomarkers and physiological parameters, mitigating information degradation in shallow fusion architectures.
- **Contrastive regularization strategy:** Cross-modal semantic consistency constraints are implemented through prototype alignment in latent space, enhancing robustness against clinical data noise and missing values.
- **Dynamic fusion coefficients:** Attention-weighted fusion gates automatically calibrate hierarchical feature contributions based on pathological relevance, simultaneously optimizing prediction accuracy and model interpretability.

## 2 Method

# 2.1 Dataset construction and preprocessing

The data for this study were sourced from the electronic medical records (EMR) system of the First Affiliated Hospital of Dalian Medical University, with ethical review approval. A total of 800 COPD patients who met the inclusion criteria, based on



clinical manifestations such as chronic cough, excessive sputum production, and airflow limitation, were enrolled. The cohort consisted of 276 female patients and 524 male patients, with ages ranging from 35 to 91 years, following a normal distribution. Patients with incomplete records or those who did not undergo necessary examinations were excluded. The dataset comprises multimodal medical records, collected retrospectively over a 24month period. The raw dataset includes unstructured EMR texts and structured laboratory parameters. All personally identifiable information was anonymized using hash encryption.

For the unstructured features of the EMRs, medical natural language processing (NLP) techniques were employed for information extraction. Symptom duration, triggering factors, and other clinical elements in chief complaints were converted into standardized fields using regular expressions matched with a medical ontology database. Radiological features in pulmonary CT reports underwent entity recognition and terminology normalization.

The structured data underwent multidimensional quality control, including numerical range validation, logical relationship verification, and outlier correction. Specifically, for the systematic missingness of C-reactive protein (CRP) measurements, a clinically context-aware imputation strategy was implemented. For patients without fever symptoms and no CRP test results, a baseline value of 3.5 mg/L for CRP was assigned, based on standard clinical practice, while retaining the original testing status markers to ensure transparency and consistency in the imputation approach. For other missing values, different imputation methods were used based on the type of variable. Continuous variables were imputed using mean imputation, while categorical variables were imputed using the most frequent value.

Finally, discrete variables were encoded using binary representations, and continuous variables were Z-score standardized before model integration. The final multidimensional dataset consisted of 11 structured feature fields and 2 standardized textual description units, providing a reliable data foundation for subsequent analysis.

### 2.2 Model

The proposed COPD-MMDxNet multimodal deep learning framework, with its core mechanism centered on establishing a hierarchical progressive feature interaction process, achieves deep integration of multimodal EMR information through parametric numerical embedding, cross-modal attention routing, and contrastive regularization strategies. The model architecture is shown in Figure 1.

### 2.2.1 Multimodal feature encoding layer

In the feature extraction stage of this study, the data from three modalities achieve in-depth semantic representation through differentiated encoding strategies.

### 2.2.1.1 Structured feature encoding

For categorical features, a learnable embedding matrix  $E \in \mathbb{R}^{|C| \times d}$  is adopted, where  $|C| = \sum_{i=1}^{n} c_i$  represents the total number of categories (including special identifiers). Given the input  $x_{\text{catalog}} \in \mathbb{R}^{k \times m}$ , its embedding process is defined as:

$$\tilde{x}_i^{\text{catalog}} = E[x_i + \phi_i], \ \phi_i = \text{offset}(c_{i,j})$$
(1)

where  $\phi_i$  is the category offset to avoid ID conflicts.

For numerical features, non - linear embedding is achieved through parameterized affine transformation:

$$\tilde{x}_j^{\text{num}} = w_j \odot x_j^{\text{num}} + b_j, \ w_j, b_j \in \mathbb{R}^d$$
(2)

where  $\odot$  denotes element - wise multiplication, and  $w_j, b_j$  are learnable parameters.

### 2.2.1.2 Text feature extraction

In the text feature extraction process, we use a pre-trained BERT-Base model, which was trained on the English Wikipedia and the BookCorpus dataset (the original pre-training corpora for BERT). This model is widely used in natural language processing tasks and has demonstrated strong performance across a variety of benchmark tasks. To ensure efficient feature extraction, we project the final hidden states from BERT and normalize them to align with features from other modalities. Let *T* represent the input text and *l* the sequence length. The feature space alignment is achieved through a projection layer  $\mathcal{P}: \mathbb{R}^{l \times d_{bert}} \to \mathbb{R}^{l \times d}$ :

$$H^{\text{text}} = \text{BERT}(T) \in \mathbb{R}^{l \times d_{\text{bert}}}$$
(3)

$$\hat{H}^{\text{text}} = \text{LayerNorm}(\text{GELU}(W_p H^{\text{text}} + b_p))$$
(4)

**Cross-modal attention:** We design a bidirectional Gated Cross-Attention (GCA) mechanism to facilitate feature interaction between modalities:

$$Q = W_q \cdot \mathrm{LN}(X^{(l)}) \tag{5}$$

$$K, V = W_{kv} \cdot \mathrm{LN}(\hat{H}^{\mathrm{text}}) \tag{6}$$

$$GCA(X,H) = \text{Softmax}\left(\frac{QK^T}{\sqrt{d_k}}\right) V \odot \sigma(W_g[X^{(l)}; \hat{H}^{\text{text}}])$$
(7)

where  $\sigma$  represents the Sigmoid gating function, and  $W_g$  is the learnable gating parameter that controls the cross-modal interaction.

**Feature fusion:** We adopt a residual connection and hierarchical fusion strategy to combine the features from different modalities:

$$X_{\text{out}}^{(l)} = X^{(l)} + \text{FFN}(\text{LN}(X^{(l)})) + \lambda \cdot \text{GCA}(X^{(l)}, \hat{H}^{\text{text}})$$
(8)

where  $\lambda$  is an adjustable fusion coefficient, allowing the dynamic adjustment of the contributions from different modalities through learnable parameters.

# 2.2.2 Contrastive regularization and dynamic classifier

To enhance cross-modal semantic consistency, we introduce a contrastive loss based on InfoNCE: Feature projection: We establish a shared projection space  $\mathcal{Z} \in \mathbb{R}^{d_z}$  to project features from different modalities into a common space:

$$z^{\text{struct}} = \operatorname{Proj}_{\theta}(x_{\text{cls}}) \tag{9}$$

$$z^{\text{text}} = \operatorname{Proj}_{\theta}(h_{cls}^{\text{text}})$$
(10)

where  $\operatorname{Proj}_{\theta}$  denotes the projection function with learnable parameters  $\theta$ , mapping features from the structured data ( $x_{cls}$ ) and text ( $h_{cls}^{text}$ ) into the shared feature space  $\mathcal{Z}$ .

Loss function: We define the symmetric contrastive loss as follows:

$$\mathcal{L}_{\text{cont}} = -\frac{1}{2N} \sum_{i=1}^{N} \left[ \log \frac{e^{s(z_i^{\text{struct}}, z_i^{\text{struct}})/\tau}}{\sum_{j=1}^{N} e^{s(z_i^{\text{struct}}, z_j^{\text{struct}})/\tau}} + \log \frac{e^{s(z_i^{\text{struct}}, z_i^{\text{struct}})/\tau}}{\sum_{j=1}^{N} e^{s(z_i^{\text{struct}}, z_j^{\text{struct}})/\tau}} \right]$$
(11)

where  $s(\cdot)$  represents the cosine similarity function, and  $\tau$  is a learnable temperature parameter that controls the scale of the similarity between different feature pairs.

### 2.2.3 Joint optimization objective

The final objective function for multi-task joint optimization is given by:

$$\mathcal{L}_{\text{total}} = \alpha \cdot \mathcal{L}_{\text{task}} + \beta \cdot \mathcal{L}_{\text{cont}} + \gamma \cdot \mathcal{L}_{\text{reg}}$$
(12)

where  $\mathcal{L}_{task}$  represents the main task loss (such as cross-entropy),  $\mathcal{L}_{reg}$  is the weight regularization term, and  $\alpha$ ,  $\beta$ , and  $\gamma$  are the balance coefficients that control the contribution of each loss term.

## 3 Experiments

This section presents experiments designed to evaluate the effectiveness of the COPD-MMDxNet framework. Initially, we assess single-modality models to underscore the limitations inherent in relying on individual data sources. Subsequently, we compare our multimodal approach against other state-of-the-art methods to demonstrate its relative advantages. Finally, ablation studies are conducted to investigate the impact of critical components, including the contrastive learning module and the cross-modal attention mechanism. These experiments collectively highlight the superior performance of our model and emphasize the significance of each component in enhancing diagnostic accuracy and robustness.

# 3.1 Experiment setup and computational costs

In all experiments, we employed a 5-fold cross-validation approach for model training and evaluation. This method ensures robust performance assessment by splitting the dataset into five subsets, with each subset serving as the validation set once while the remaining four are used for training. The results from each fold were averaged to obtain the final performance metrics, mitigating the risk of overfitting and providing a comprehensive evaluation of the model's generalization ability.

For computational efficiency, all training was conducted on a single NVIDIA V100 GPU. Each fold of the 5-fold cross-validation took approximately 10 min, with the total training time for the entire process being 50 min on the V100 GPU. To optimize memory usage, we implemented dynamic batching and gradient accumulation during training.

# 3.2 Comparison with single-modality baseline models

To validate the effectiveness of multimodal diagnosis, this study first performed a comparative experiment with singlemodality baseline models. The single-modality models were based on structured clinical data and textual data analysis, respectively. The experimental results are shown in Table 1.

The experimental results indicate that while the traditional XGBoost (14) model and FT-Transformer (15) provide certain predictive capabilities using pulmonary function parameters and inflammatory biomarkers, they exhibit high missed diagnosis

Data	Method	Acc	Prec	Rec	Spec	F1 score
CSV	XGBoost	0.6706	0.6909	0.4935	0.8172	0.5758
CSV	FT-Transformer	0.6433	0.5652	0.3881	0.8077	0.4602
TEXT	DPCNN	0.6941	0.6769	0.5867	0.7789	0.6286
TEXT	BERT	0.7176	0.7077	0.6133	0.8000	0.6571
CSV&TEXT	COPD-MMDxNet	0.8176	0.7887	0.7778	0.8469	0.7832

#### TABLE 1 Comparison of single-modality baseline models.

The bolding indicates the parts where our model stands out.

TABLE 2 Comparison with advanced multimodal models.

Version	Acc	Prec	Rec	Spec	F1 score
RETAIN	0.7423	0.7198	0.7045	0.7812	0.7098
BRLTM	0.7834	0.7556	0.7412	0.8145	0.7483
ExBEHRT	0.7956	0.7645	0.7534	0.8212	0.7589
Ours	0.8176	0.7887	0.7778	0.8469	0.7832

The bolding indicates the parts where our model stands out.

rates and low recall when handling early-stage cases, highlighting the limitations of single-modality data in capturing complex pathological mechanisms. In textual data analysis, although the DPCNN (16) and BERT (17) models can extract useful information from radiological reports, and BERT improves accuracy, it still struggles to effectively capture dynamic imaging features and exhibits relatively low specificity. This reflects the shortcomings of single-modality text models in integrating and analyzing diverse features. The COPD-MMDxNet framework, by integrating multimodal information from both structured data and text, successfully overcomes the limitations of single-modality models in capturing cross-dimensional pathological relationships. This demonstrates that multimodal fusion can effectively reduce the information omissions and biases inherent in single data sources, providing a more comprehensive representation of the complex characteristics of the disease.

# 3.3 Comparison with advanced multimodal models

To evaluate the effectiveness of the proposed COPD-MMDxNet model, we conducted comparative experiments with three representative models: RETAIN (18), BRLTM (19), and ExBEHRT (20). The experimental results are shown in Table 2. The RETAIN model is based on a reverse-time attention mechanism, which improves accuracy and interpretability by processing EHRs data in reverse time order and applying a two-level attention mechanism. BRLTM, on the other hand, uses a Transformer architecture and self-attention mechanism to capture forward and backward dependencies in sequential data, enhancing the model's temporal learning capability. ExBEHRT is an extension of BEHRT, improving performance through stacking and pre-training strategies. The experimental results are shown in the Table 2.

Compared to the three baseline models, COPD-MMDxNet demonstrates significant advantages in multiple aspects, particularly in the fusion of multimodal features and cross-modal interaction for complex EHRs data. Experimental results show that our model outperforms RETAIN, BRLTM, and ExBEHRT in terms of accuracy, precision, recall, and AUC. COPD-MMDxNet effectively integrates information from structured data, text, and numerical data through a hierarchical cross-modal interaction mechanism. This stepwise feature interaction approach allows for deeper and more accurate cross-modal information complementarity and integration, which is superior to traditional methods like unimodal self-attention and simple cross-modal mapping. Additionally, COPD-MMDxNet adopts a dynamic feature fusion strategy that adjusts the fusion weights based on the contribution of each modality, thereby enhancing the model's performance across different tasks. In contrast, RETAIN and BRLTM use fixed-weight fusion strategies, and while ExBEHRT incorporates multimodal data, it lacks an effective cross-modal consistency training mechanism, limiting the full potential of multimodal information. Furthermore, by introducing InfoNCEbased contrastive regularization, we further enhance the semantic consistency across multimodal data. These innovations enable COPD-MMDxNet to exhibit stronger capabilities in multimodal data integration, semantic consistency, and feature processing, further demonstrating its superiority in EHRs data processing.

# 3.4 Ablation analysis of key model components

We also conducted systematic ablation studies to evaluate the effectiveness of our model. The experimental results are presented in Table 3. These experiments highlight the synergistic interactions between the key components within the multimodal framework and their significant influence on the model's robustness. The results indicate that the removal of the contrastive learning module significantly reduces the alignment of cross-modal semantics, which in turn leads to an increase in false-positive cases. This can be attributed to the weakened latent correlations between modalities, impairing the model's ability to establish distinct decision boundaries in the feature space. Further analysis shows that removing the cross-modal attention mechanism severely hinders the integration of heterogeneous features. The absence of dynamic weight allocation prevents the model from effectively capturing the complex pathological correlations across modalities, which results in a substantial increase in false-negative rates.

In addition, we performed a comparison using a direct fusion approach, where FT-Transformer or XGBoost was used to process structured data features, and these features were directly concatenated with the text features processed by BERT.

Version	Acc	Prec	Rec	Spec	F1 score
Original model	0.8176	0.7887	0.7778	0.8469	0.7832
Remove contrastive learning	0.8000	0.7639	0.7639	0.8265	0.7639
Remove cross-modal attention	0.7895	0.7049	0.7049	0.8364	0.7049
XGBoost+BERT direct fusion	0.7647	0.7864	0.6623	0.8495	0.7183
FTTransform+BERT fusion	0.7412	0.7656	0.6282	0.8370	0.6908

TABLE 3 Ablation analysis of key model components.

The bolding indicates the parts where our model stands out.

The experimental results show that, although this direct fusion method can extract certain structured and textual features, its performance is significantly lower than that of the proposed multimodal framework. The direct concatenation approach fails to capture the deep inter-modal correlations, preventing the model from effectively utilizing the complementary information between modalities.

Through these comparative experiments, we demonstrate the significant limitations of single-data fusion methods and cross-modal fusion strategies that lack dynamic weighting. Our framework, by leveraging more complex feature interactions and dynamic weight adjustments, is better able to capture pathological correlations between modalities, thus significantly enhancing the model's diagnostic ability and robustness.

## 4 Conclusion

This study presents COPD-MMDDxNet, a novel multimodal deep learning framework designed to enhance COPD diagnosis by integrating structured clinical data, radiological reports, and blood gas analysis from electronic medical records (EMR). The framework addresses the limitations of spirometry, making it particularly suitable for resource-constrained environments. Key innovations, including parametric numerical embedding, hierarchical interaction mechanisms, contrastive regularization, and dynamic fusion coefficients, enable effective multimodal integration and capture of complex cross-modal relationships.

Experimental results demonstrate that COPD-MMDDxNet outperforms traditional single-modality models and other state-ofthe-art multimodal models in terms of accuracy, precision, recall, and F1 score. Ablation studies confirm the importance of each component, particularly the contrastive learning module and crossmodal attention mechanism, in enhancing model performance.

COPD-MMDDxNet offers a robust solution for more accurate and accessible COPD diagnosis, particularly in settings where spirometry is unavailable or impractical. This framework has the potential to improve early diagnosis and disease management in COPD. However, it is important to note that the current study was conducted using a single-center dataset, which limits the generalizability of the results. Future work will focus on expanding clinical applicability and validating the model across multiple data centers to further assess its robustness and reliability in diverse clinical environments. The code for the model, including the architecture and training procedures, has been made publicly available on GitHub for transparency and reproducibility. The repository can be accessed at: https://github.com/yiyuanyuan827/COPD-MMDDxNet.

## Data availability statement

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

## Author contributions

YY: Data curation, Writing – review & editing, Formal analysis, Writing – original draft, Investigation. LS: Formal analysis, Data curation, Writing – original draft, Writing – review & editing, Investigation. HL: Investigation, Writing – review & editing, Formal analysis, Software, Writing – original draft, Data curation. MW: Writing – original draft, Formal analysis, Writing – review & editing, Software, Data curation, Investigation. MF: Formal analysis, Writing – original draft, Software, Data curation, Writing – review & editing, Investigation. YL: Formal analysis, Writing – original draft, Data curation, Software, Writing – review & editing, Investigation.

## Funding

The author(s) declare that financial support was received for the research and/or publication of this article. We would acknowledge the support of the Natural Science Foundation of Liaoning Province, China (No. 2024-MS-163).

# **Conflict of interest**

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

## **Generative AI statement**

The author(s) declare that no Gen AI was used in the creation of this manuscript.

## Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated

References

1. Adeloye D, Song P, Zhu Y, Campbell H, Sheikh A, Rudan I. Global, regional, and national prevalence of, and risk factors for, chronic obstructive pulmonary disease (COPD) in 2019: a systematic review and modelling analysis. *Lancet Respir Med.* (2022) 10:447–58. doi: 10.1016/S2213-2600(21)00511-7

2. Zhao L, Wang X, Liu Z, Wang Z, Chen Z. Learnable graph guided deep multiview representation learning via information bottleneck. *IEEE Trans Circ Syst Video Technol.* (2024) 35:3303–3314. doi: 10.1109/TCSVT.2024.3509892

3. Sun Y, Zhou J. New insights into early intervention of chronic obstructive pulmonary disease with mild airflow limitation. *Int J Chron Obstruct Pulmon Dis.* (2019) 14:1119–1125. doi: 10.2147/COPD.S205382

4. Agustí A, Celli BR, Criner GJ, Halpin D, Anzueto A, Barnes P, et al. Global initiative for chronic obstructive lung disease 2023 report: GOLD executive summary. *J Pan African Thoracic Soc.* (2022) 4:58–80.

5. Siddharthan T, Grealis K, Robertson NM, Lu M, Liu S, Pollard SL, et al. Assessing the prevalence and impact of preserved ratio impaired spirometry in low-income and middle-income countries: a post-hoc cross-sectional analysis. *Lancet Global Health.* (2024) 12:e1498–505. doi: 10.1016/S2214-109X(24)00233-X

6. Clark MV. Asthma: A Clinician's Guide. Burlington: Jones & Bartlett Learning. (2010).

7. Zhao L, Xie Q, Li Z, Wu S, Yang Y. Dynamic graph guided progressive partial view-aligned clustering. *IEEE Trans Neural Netw Learn Syst.* (2024) 36:9370–9382. doi: 10.1109/TNNLS.2024.3425457

8. Wang S, Zheng Z, Wang X, Zhang Q, Liu Z. A cloud-edge collaboration framework for cancer survival prediction to develop medical consumer electronic devices. *IEEE Trans Consum Electr.* (2024) 70:5251–5258. doi: 10.1109/TCE.2024.3413732

9. Gompelmann D, Gysan MR, Desbordes P, Maes J, Van Orshoven K, De Vos M, et al. AI-powered evaluation of lung function for diagnosis of interstitial lung disease. *Thorax.* (2025) 13:221537. doi: 10.1136/thorax-2024-221537

10. Zhao L, Huang P, Chen T, Fu C, Hu Q, Zhang Y. Multi-sentence complementarily generation for text-to-image synthesis. *IEEE Trans Multim.* (2023) 26:8323–8332. doi: 10.1109/TMM.2023.3297769

11. Alves Pegoraro J, Guerder A, Similowski T, Salamitou P, Gonzalez-Bermejo J, Birmelé E. Detection of COPD exacerbations with continuous monitoring of breathing

organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

rate and inspiratory amplitude under oxygen therapy. BMC Med Inform Dec Making. (2025) 25:101. doi: 10.1186/s12911-025-02939-3

12. Quennelle S, Malekzadeh-Milani S, Garcelon N, Faour H, Burgun A, Faviez C, et al. Active learning for extracting rare adverse events from electronic health records: a study in pediatric cardiology. *Int J Med Inform.* (2025) 195:105761. doi: 10.1016/j.ijmedinf.2024.105761

13. Takeuchi T, Horinouchi H, Takasawa K, Mukai M, Masuda K, Shinno Y, et al. A series of natural language processing for predicting tumor response evaluation and survival curve from electronic health records. *BMC Med Inform Decis Mak.* (2025) 25:1–11. doi: 10.1186/s12911-025-02928-6

14. Chen T, Guestrin C. Xgboost: a scalable tree boosting system. In: Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (2016). p. 785–794. doi: 10.1145/2939672.2939785

15. Gorishniy Y, Rubachev I, Khrulkov V, Babenko A. Revisiting deep learning models for tabular data. In: *Advances in Neural Information Processing Systems* (2021). p. 18932–43.

16. Zhang M, Pang J, Cai J, Huo Y, Yang C, Xiong H. DPCNN-based models for text classification. In: 2023 IEEE 10th International Conference on Cyber Security and Cloud Computing (CSCloud)/2023 IEEE 9th International Conference on Edge Computing and Scalable Cloud (EdgeCom). IEEE (2023). p. 363–368. doi: 10.1109/CSCloud-EdgeCom58631.2023.00068

17. Devlin J, Chang MW, Lee K, Toutanova K. Bert: pre-training of deep bidirectional transformers for language understanding. In: *Proceedings of the 2019 conference of the North American chapter of the association for computational linguistics: human language technologies, volume 1 (long and short papers)* (2019). p. 4171–4186.

18. Choi E, Bahadori MT, Sun J, Kulas J, Schuetz A, Stewart W. Retain: an interpretable predictive model for healthcare using reverse time attention mechanism. In: *Advances in Neural Information Processing Systems*. (2016). p. 29.

19. Meng Y, Speier W, Ong MK, Arnold CW. Bidirectional representation learning from transformers using multimodal electronic health record data to predict depression. *IEEE J Biomed Health Inform.* (2021) 25:3121–9. doi:10.1109/JBHL.2021.3063721

20. Rupp M, Peter O, Pattipaka T. Exbehrt: extended transformer for electronic health records. In: *International Workshop on Trustworthy Machine Learning for Healthcare*. Springer (2023). p. 73–84. doi: 10.1007/978-3-031-39539-0\_7