Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Multimodal Deep Learning Integrating Chest CT Lung-Component Segmentation and Clinical Variables for Prediction of 1-Year Composite Cardiovascular Events in COPD: A Single-Centre Retrospective Cohort Study

Kang Yu
Department of Radiology, The Fourth Affiliated Hospital of Harbin Medical University, Harbin,Heilongjiang, China.

Chunhui Qin
Department of Radiology, The Fourth Affiliated Hospital of Harbin Medical University, Harbin,Heilongjiang, China.

Yuxin Li
Department of Radiology, The Fourth Affiliated Hospital of Harbin Medical University, Harbin,Heilongjiang, China.

Chunxiao Wang
Department of Radiology, The Fourth Affiliated Hospital of Harbin Medical University, Harbin,Heilongjiang, China.

Yunjia Shuai
Department of Radiology, The Fourth Affiliated Hospital of Harbin Medical University, Harbin,Heilongjiang, China.

Ao Li
Department of Radiology, The Fourth Affiliated Hospital of Harbin Medical University, Harbin,Heilongjiang, China.

Jiahui Chen
Department of Radiology, The Fourth Affiliated Hospital of Harbin Medical University, Harbin,Heilongjiang, China.

tong zhang
yingxiang939@163.com

Department of Radiology, The Fourth Affiliated Hospital of Harbin Medical University, Harbin,Heilongjiang, China.

Research Article

Keywords: chronic obstructive pulmonary disease, chest computed tomography, lung segmentation, deep learning, multimodal fusion, cardiovascular events

Posted Date: February 27th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-8733857/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: No competing interests reported.

Abstract

Background

Patients with chronic obstructive pulmonary disease (COPD) are at increased risk of cardiovascular disease (CVD) events. We aimed to develop and internally validate a multimodal deep learning model that integrates chest CT–derived lung-component information with clinical variables to predict 1-year composite CVD events in COPD.

Methods

We enrolled COPD patients between January 2023 and January 2025 who underwent baseline thin-slice non-contrast chest CT and completed 12-month follow-up. The primary endpoint was the first occurrence within 12 months of a composite CVD event (acute myocardial infarction, ischaemic stroke, hospitalisation for unstable angina, or hospitalisation for heart failure). CT images were standardised and resampled to 1.0×1.0×1.0 mm voxels. Lung parenchyma was segmented using a U-Net–based model; emphysema regions were identified by thresholding (HU < − 950); pulmonary vessels were enhanced using a Frangi filter and extracted via connected component analysis. A three-branch 3D convolutional neural network (3D-CNN) was constructed to extract imaging features from normal lung tissue, emphysematous regions, and pulmonary vascular structures, which were fused with clinical features (including age, smoking pack-years, and FEV1%) using an attention mechanism. The dataset was split into training/validation/test sets (8:1:1) with stratified sampling. Discrimination and classification metrics were reported on the independent test set; calibration was assessed using a calibration curve, Brier score, and calibration intercept/slope.

Results

A total of 306 COPD patients were included (training n = 244, validation n = 30, test n = 32). Within 12 months, 182 composite CVD events occurred (59.5%). On the independent test set (events = 17/32), the model achieved an AUC of 0.8588 with an accuracy of 78.12%, precision of 81.25%, recall/sensitivity of 76.47%, specificity of 80.00%, and F1-score of 0.7879. The Brier score was 0.187, with a calibration slope of 0.435 and an intercept of − 0.606.

Conclusions

A multimodal deep learning framework combining lung-component CT features and clinical variables demonstrated good discrimination for predicting 1-year composite CVD events in COPD. Further external validation and calibration refinement are warranted before clinical deployment.

Introduction

Chronic obstructive pulmonary disease (COPD) is a common chronic respiratory condition characterized by persistent airflow limitation. Its core pathophysiological mechanisms involve multi-level and multi-system alterations, primarily including airway obstruction, destruction of lung parenchyma, and structural and functional remodeling of pulmonary vessels.[1] Long-term persistent airway inflammation leads to thickening of the small airway walls, luminal narrowing, and dynamic collapse, thereby exacerbating airflow limitation. Concurrently, alveolar wall destruction and decreased elastic recoil result in gas trapping and pulmonary hyperinflation, which significantly increase the work of breathing and reduce ventilation efficiency. Furthermore, chronic hypoxemia can induce pulmonary vasoconstriction, smooth muscle hyperplasia, and vascular wall remodeling. In some patients, this progresses to COPD-associated pulmonary hypertension, increasing right ventricular afterload and affecting systemic circulatory function. These pathological changes not only determine the respiratory manifestations of COPD but also lay the foundation for its multi-system complications.[2]

In recent years, increasing research has focused on the close association between COPD and cardiovascular diseases. Epidemiological studies indicate that COPD patients have a significantly higher risk of cardiovascular and cerebrovascular events, such as coronary artery disease, myocardial infarction, heart failure, and stroke, compared to the general population. Cardiovascular disease has become one of the leading causes of death in COPD patients. The risk of cardiovascular events is particularly prominent during or shortly after acute exacerbations, suggesting that respiratory inflammation, hypoxemia, and hemodynamic changes may act synergistically in the process of cardiovascular injury.[3] Therefore, early identification and assessment of cardiovascular disease risk in the COPD population hold significant clinical importance for optimizing individualized treatment strategies and reducing the incidence of adverse events.[4]

Currently, cardiovascular disease risk assessment primarily relies on clinical risk factors, hematological biomarkers, and cardiac imaging examinations such as cardiac CT or echocardiography. However, the routine application of these methods in COPD patients faces certain limitations, including high costs, complex procedures, and insufficient patient compliance. In contrast, chest CT has become the most commonly used imaging modality in the diagnosis, phenotyping, and treatment evaluation of COPD, offering relatively high accessibility and repeatability in clinical practice.[5]

In terms of imaging quantitative analysis, previous studies have mostly used whole-lung average-based metrics to assess COPD, such as the low attenuation area percentage (LAA%). While these methods can reflect the extent of parenchymal destruction to a certain degree, they overlook the spatial heterogeneity in the distribution of different lung tissue components, making it difficult to comprehensively characterize local structural abnormalities and their potential impact on cardiopulmonary interactions. In fact, variations in the proportion and spatial distribution of emphysematous regions, relatively normal lung tissue, and pulmonary vascular structures may be closely related to cardiovascular disease risk by

influencing pulmonary vascular resistance and cardiac load, among other pathways.[6] Therefore, in-depth exploration of chest CT imaging information at a finer tissue-structural level is of significant value.

With the rapid development of deep learning technology, segmentation and feature extraction methods based on convolutional neural networks provide new tools for analyzing complex medical imaging information. By performing fine segmentation of lung tissue components and extracting high-dimensional representative features from images, it is possible to more comprehensively reflect the structural and functional status of the lungs in COPD patients. Moreover, integrating imaging features with clinical data to construct multi-modal fusion models may further enhance the predictive capability for cardiovascular disease risk.[7]

Based on the above background, this study aims to use chest CT images from COPD patients as the research subject. It will involve automatic segmentation and quantitative analysis of normal lung tissue, emphysematous regions, and pulmonary vascular structures. Combined with deep learning methods to extract imaging features and integrated with clinical information, a multi-modal prediction model will be constructed to explore its application value in assessing cardiovascular disease risk in COPD patients. Through this research, we hope to provide new imaging-based evidence for the early risk identification of COPD combined with cardiovascular disease and offer references for formulating individualized clinical management strategies.

Methods

Study design and setting

This was a single-centre retrospective cohort study conducted at the Fourth Affiliated Hospital of Harbin Medical University, Harbin, China.

Ethics approval

The study protocol was approved by the institutional ethics committee and complied with the Declaration of Helsinki .The requirement for written informed consent was waived because of the retrospective design and the use of de-identified data.

Participants

We enrolled COPD patients between January 2023 and January 2025 who completed baseline thin-slice non-contrast chest CT and 12-month follow-up.

Inclusion criteria were: (1) spirometry-confirmed COPD (FEV1/FVC < 70%); and (2) availability of baseline thin-slice chest CT.

Exclusion criteria were: (1) severe CT artefacts affecting analysis; (2) missing key clinical or follow-up information; or (3) incomplete follow-up. (Figure 1)

Figure 1. Patient inclusion flow diagram.

Records were identified from the hospital database (Jan 2023–Jan 2025, **N = 1343**) and screened for eligibility. A total of **1037** records were excluded due to missing key clinical data (**n = 377**), incomplete 12-month follow-up (**n = 659**), or severe CT artifacts affecting analysis (**n = 1**). The final study cohort included **306** eligible COPD patients with baseline CT and was stratified into a training set (**n = 244**), validation set (**n = 30**), and test set (**n = 32**) at an 8:1:1 ratio.

Outcome definition

Outcome definition
The primary endpoint was the first occurrence within 12 months of a composite CVD event: acute myocardial infarction, ischaemic stroke, hospitalisation for unstable angina, or hospitalisation for heart failure. If multiple events occurred, only the first event was recorded.[8]

Outcomes were ascertained using electronic medical records, discharge summaries and follow-up telephone calls.[9]

CT acquisition and image preprocessing

All patients underwent thin-slice chest CT on multi-vendor spiral CT scanners (eg, Siemens and Philips). Typical parameters included 120 kVp, automatic tube current modulation, 1.0 mm slice thickness, 512×512 matrix, pitch approximately 1.0, with iterative reconstruction.

To reduce inter-scanner variability, CT intensities (Hounsfield units) were standardised and images were resampled to isotropic 1.0×1.0×1.0 mm voxels before analysis.

Lung-component segmentation

This study employed a deep learning network based on the U-Net architecture for the automatic segmentation of lung parenchyma. Emphysematous regions were identified using a threshold method (HU < − 950). The pulmonary vascular regions were first enhanced using a Frangi filter to highlight vascular structures, followed by extraction through connected component analysis.[10]

To ensure the reliability of the segmentation results, all segmented outputs were independently visually assessed by two radiologists, with a focus on the integrity of lung parenchyma boundaries, the accuracy of emphysematous region identification, and the continuity of vascular structures. The evaluation results showed that the segmentation outputs for most cases were highly consistent with the clinical anatomical structures. Minor discrepancies were observed only in some regions with severe

emphysema or complex vascular courses. As the primary objective of this study was the construction of a risk prediction model, and the segmentation results were primarily used for subsequent imaging feature extraction, a systematic quantitative evaluation of segmentation accuracy was not performed. Instead, visual consistency assessment by the imaging physicians was adopted as the quality control method.

Deep Learning-Based Feature Extraction and Fusion

Based on the segmentation results, this study constructed a three-branch three-dimensional convolutional neural network (3D-CNN) to extract imaging features from normal lung tissue, emphysematous regions, and pulmonary vascular structures, respectively. The design of the three-branch architecture aims to independently characterize the structural heterogeneity of different lung tissue components, preventing interference between features of distinct tissues within a single network.

Each imaging branch extracts multi-scale features through multiple layers of convolution, batch normalization, nonlinear activation, and pooling operations. The process culminates in a global average pooling layer to obtain fixed-dimensional imaging feature representations. The selection of clinical variables was based on previous research and clinical availability, taking into account sample size and variable stability. Variables such as age, smoking pack-years, and FEV1% were included as clinical input features.

During the feature fusion stage, imaging features and clinical features were concatenated. An attention mechanism was introduced to enhance the weighted representation of key features, followed by fully connected layers to perform the classification task. The final output is the probability of the patient developing a cardiovascular disease (CVD) event.(Figure 2)

Figure 2. Overview of data processing, model training, and evaluation.

Chest CT images were preprocessed (including intensity normalization and fixed-size resampling). Lung parenchyma was automatically segmented using a U-Net–based network and further decomposed into three components—relatively normal lung, emphysema (HU < − 950), and pulmonary vasculature (Frangi enhancement followed by connected component analysis). The three-component information was encoded into a single component mask and stacked with the corresponding 3D CT crop to form a two-channel input (CT + component-encoded mask) for a multimodal model. A 3D-CNN imaging encoder extracted imaging representations, which were fused with clinical features using an MLP-based clinical branch (with an attention-enhanced fusion module) to predict 1-year composite cardiovascular events. The dataset was split into training, validation, and independent test sets (8:1:1). Model performance was evaluated using ROC curves, confusion matrices, and standard classification metrics.

Model training and internal validation

The dataset was partitioned into training, validation, and test sets in an 8:1:1 ratio, using a stratified sampling method to ensure a relatively balanced proportion of CVD events across the different sets.

Training employed the cross-entropy loss function and the Adam optimizer. During the training process, the F1-score on the validation set served as the primary monitoring metric, and a learning rate scheduler was utilized to automatically adjust the learning rate when validation performance plateaued. Hyperparameters, including the number of training epochs, batch size, initial learning rate, and dropout rate, were uniformly set in the training script. A fixed random seed was used throughout the experiments to minimize randomness associated with data splitting and the training process.

Calibration assessment

In addition to discrimination metrics (AUC, accuracy, sensitivity, specificity, and F1-score), calibration was assessed using predicted probabilities of the 1-year composite cardiovascular event. A calibration curve was generated on the independent test set by grouping predicted probabilities into 10 uniform bins (n_bins = 10) and comparing the mean predicted probability with the observed event rate in each bin against the ideal line (y = x). Calibration error was quantified using the Brier score. Calibration intercept and slope were estimated by fitting a logistic regression model of the observed outcome on the logit-transformed predicted probability (intercept closer to 0 and slope closer to 1 indicate better calibration). [11]

Model Evaluation Methods

Model performance was evaluated in the training, validation, and test sets. The primary evaluation metrics included accuracy, precision, recall, F1-score, and the area under the receiver operating characteristic curve (AUC). ROC curves were plotted to assess the model's discriminative ability. A confusion matrix was also generated to display the distribution of classification error types, and from this matrix, clinically interpretable metrics such as sensitivity and specificity were further reported. All final performance results were reported on the independent test set.

To provide an interpretable baseline comparison relevant to real-world clinical scenarios, this study constructed a clinical baseline model. Common cardiovascular risk factors were selected as independent variables, including age, gender, smoking history (yes/no), hypertension, and diabetes. A multivariate logistic regression model was used to predict the occurrence of a composite cardiovascular event within one year. The model was fitted on the training set, and its discriminative ability (AUC) was evaluated on the independent test set. Additionally, accuracy, F1-score, sensitivity, and specificity at a threshold of 0.5 were reported. The 95% confidence interval for the AUC was obtained using bootstrap resampling.

Results

Participant characteristics

A total of 306 patients with COPD were included and randomly split (stratified by outcome) into a training set (n = 244), a validation set (n = 30) and an independent test set (n = 32) at an 8:1:1 ratio. The three datasets were mutually exclusive and derived from the same centre and enrolment framework. Baseline demographic and clinical characteristics are summarised in Table 1.

Table 1
Baseline characteristics of the study population (N = 306)

Characteristic	**Value**
Characteristic	Overall
Age, years	69.9 ± 8.9
Male sex, n (%)	183 (59.8)
Weight, kg	61.7 ± 14.7 (n = 263)
Height, cm	164.6 ± 8.8 (n = 99)
Diabetes, n (%)	52 (17.0)
Renal insufficiency, n (%)	23 (7.5)
Other vascular disease, n (%)	127 (41.5)
Hypertension, n (%)	134 (43.8)
Coronary artery disease, n (%)	84 (27.5)
Smoking history, n (%)	195 (63.7)
Alcohol history, n (%)	64 (20.9)
Hemoglobin, g/L	138.3 ± 19.9 (n = 269)
Hematocrit, %	42.3 ± 6.0 (n = 258)
C-reactive protein (CRP), mg/L	15.76 (4.01–57.70) (n = 275)
D-dimer, mg/L	0.51 (0.33–0.85) (n = 249)
Total cholesterol, mmol/L	4.8 ± 1.1 (n = 151)
LDL-C, mmol/L	2.9 ± 0.9 (n = 136)
HDL-C, mmol/L	1.2 ± 0.3 (n = 136)
Triglycerides, mmol/L	1.3 ± 1.0 (n = 151)
B-type natriuretic peptide (BNP), pg/mL	196 (58–850) (n = 187)
Interleukin-6 (IL-6), pg/mL	20.86 (8.02–36.89) (n = 30)

Follow-up outcomes

Within 12 months, 182 composite cardiovascular events occurred (59.5%, 182/306). Event rates were 61.5% (150/244) in the training set, 50.0% (15/30) in the validation set and 53.1% (17/32) in the test set. During follow-up, 219 patients (71.6%) experienced at least one acute exacerbation, while 87 (28.4%) had no acute exacerbations; among those with exacerbations, the mean number of exacerbations was 1.39. Detailed outcomes are provided in Table 2.

Table 2
Follow-up outcomes within 1 year (N = 306)

Outcome	Value
Composite CVD events within 1 year, n (%)	182 (59.5%)
Acute exacerbations during follow-up, n (%)	219 (71.6%)
No exacerbations, n (%)	87 (28.4%)
Mean exacerbations per patient (among those with any)	1.39

Performance of the multimodal deep learning model

Model discrimination and classification performance are presented in Table 3.

Training set (n = 244)

accuracy 85.25%; AUC 0.8794. The confusion matrix yielded sensitivity 83.33% (125/150) and specificity 88.30% (83/94) (TN = 83, FP = 11, FN = 25, TP = 125).(**Fig. 3**)

Figure 3. Model discrimination and classification performance in the training set (n = 244).

Receiver operating characteristic (ROC) curve and confusion matrix for the training set. The model achieved **AUC = 0.8794**. Confusion matrix counts: **TN = 83, FP = 11, FN = 25, TP = 125**.

Validation set (n = 30)

accuracy 96.67%; precision 93.75%; recall 100.00%; F1-score 0.9677; AUC 0.9822. Sensitivity was 100.00% (15/15) and specificity 93.33% (14/15) (TN = 14, FP = 1, FN = 0, TP = 15). (**Fig. 4**)

Figure 4. Model discrimination and classification performance in the validation set (n = 30).

ROC curve and confusion matrix for the validation set. The model achieved **AUC = 0.9822**. Confusion matrix counts: **TN = 14, FP = 1, FN = 0, TP = 15**.

Test set (n = 32)

accuracy 78.12%; precision 81.25%; recall 76.47%; F1-score 0.7879; AUC 0.8588. Sensitivity was 76.47% (13/17) and specificity 80.00% (12/15) (TN = 12, FP = 3, FN = 4, TP = 13). (**Fig. 5**)

Figure 5. Model discrimination and classification performance in the test set (n = 32).

ROC curve and confusion matrix for the independent test set. The model achieved **AUC = 0.8588**. Confusion matrix counts: **TN = 12, FP = 3, FN = 4, TP = 13**.

On the independent test set (n = 32, events = 17), the model demonstrated acceptable calibration (Fig. 4). The Brier score was 0.187, with a calibration slope of 0.435 and an intercept of − 0.606. The discrimination performance remained good (AUC = 0.859).

Table 3
Performance of the deep learning model on the training, validation, and test sets

Dataset	Accuracy	F1-score	Loss
Training	0.8525	0.8741	0.3940
Validation	0.9667	0.9677	0.2163
Test	0.7812	0.7879	0.5906

Clinical baseline model

To provide an interpretable clinical benchmark, we fitted a multivariable logistic regression model including age, sex, smoking history (yes/no), hypertension and diabetes (Table 4). In the full cohort, hypertension showed the strongest association with 1-year composite events (OR 17.38, 95% CI 8.65 to 34.92; p < 0.001). Age (per 1-year increase) was associated with higher risk (OR 1.06, 95% CI 1.03 to 1.10; p = 0.001), and smoking history was also associated with events (OR 1.94, 95% CI 1.01 to 3.70; p = 0.045). Sex and diabetes were not significantly associated with events (both p > 0.05).

Table 4
Multivariable logistic regression for 1-year composite CVD events (N = 306)

Predictor	OR (95% CI)	P value
Age (per 1-year increase)	1.06 (1.03–1.10)	0.001
Male sex	0.78 (0.42–1.47)	0.449
Smoking history (yes vs no)	1.94 (1.01–3.70)	0.045
Hypertension (yes vs no)	17.38 (8.65–34.92)	< 0.001
Diabetes (yes vs no)	1.58 (0.65–3.83)	0.310

Using the same stratified split (244/30/32), the clinical baseline model achieved: training set accuracy 0.6352 and F1-score 0.6764; validation set accuracy 0.6333 and F1-score 0.6207; and test set accuracy 0.4688 and F1-score 0.3704. On the test set, the confusion matrix was TN = 10, FP = 5, FN = 12 and TP = 5, corresponding to sensitivity 29.41% (5/17) and specificity 66.67% (10/15).(Table 5) On the

independent test set (n = 32, events = 17), the model demonstrated acceptable calibration (Fig. 6). The Brier score was 0.187, with a calibration slope of 0.435 and an intercept of − 0.606. The discrimination performance remained good (AUC = 0.859).

Figure 6. Calibration curve of the multimodal model in the independent test set.

Calibration plot comparing mean predicted probability and observed event rate in the test set (**n = 32**). The dashed line indicates perfect calibration. Model calibration statistics are reported in the panel (**Brier score = 0.187; calibration slope = 0.435; intercept = − 0.606; AUC = 0.859**).

Table 5
Performance of the clinical baseline logistic regression model

Dataset	Event / Non-event	Accuracy	F1-score
Training	150 / 94	0.6352	0.6764
Validation	15 / 15	0.6333	0.6207
Test	17 / 15	0.4688	0.3704

Comparison across datasets

Across the training, validation and test sets, the multimodal deep learning model showed higher discrimination and classification performance than the clinical baseline model.

Discussion

In this single-centre retrospective cohort study of patients with COPD, we developed a multimodal deep learning framework that integrates routine chest CT–derived lung-component representations with clinical variables to predict 1-year composite cardiovascular events. The model showed good discrimination in the training and validation sets and maintained performance in an independent held-out test set, suggesting predictive potential within the current sample size. A key contribution of this work is not simply the use of deep learning for classification but the component-level decomposition of lung structure. By partitioning the lung into three biologically interpretable regions—relatively normal lung, emphysema and pulmonary vasculature—and learning representations from each region separately before multimodal fusion, the framework aims to capture structural heterogeneity in COPD in a more organised and potentially interpretable manner for outcome-oriented risk stratification. [12]

COPD-related structural injury extends beyond airway and parenchymal changes and involves multi-pathway alterations, including reduction and remodelling of the pulmonary vascular bed. Previous studies have suggested that emphysema-related alveolar destruction may reduce the alveolar–capillary bed, thereby affecting pulmonary vascular resistance and right-heart load; pulmonary vascular pruning and perfusion abnormalities may also interact with systemic inflammation, oxidative stress and endothelial dysfunction, potentially contributing to elevated cardiovascular risk in COPD.[13] The present

study is designed for outcome prediction, and the observed associations between imaging phenotypes and subsequent events do not establish causality.[14] Nonetheless, a component-level representation may offer a more granular imaging perspective than conventional whole-lung or single-threshold metrics, which can be limited in capturing spatially heterogeneous structural patterns. Future work that combines interpretability approaches (eg, saliency or occlusion analyses) with comparisons against established quantitative CT measures may help examine whether the model's informative regions align with plausible physiological mechanisms.

We observed higher performance in the validation set than in the test set. This pattern is common in medical imaging deep learning studies and may be amplified under small-sample conditions. [15]The limited size of the validation and test sets means performance estimates may be sensitive to individual cases. In addition, although stratified sampling was used to balance event proportions, residual heterogeneity between subsets (eg, comorbidity profiles, CT acquisition parameters and reconstruction settings) may remain and influence generalisation. Moreover, some degree of adaptation to the training distribution is expected during optimisation, which can manifest as reduced performance on held-out data. Notably, the test-set AUC remained relatively high, indicating discrimination in the internal held-out evaluation. However, given the single split and the small test set, these estimates are uncertain and should not be overgeneralised as evidence of clinical generalisability. A more cautious interpretation is that the model demonstrated predictive potential in an internal test set, warranting further evaluation using repeated resampling/bootstrapping and, critically, external cohorts, together with calibration assessment and clinical utility analyses.

A practical advantage of the proposed approach is that the imaging input is derived from routine chest CT, which is commonly used in COPD assessment and follow-up and does not require additional tests. If the model proves stable in multi-centre external validation and undergoes appropriate calibration and clinical utility assessment, it could potentially be incorporated into imaging workflows to provide short-term cardiovascular risk estimates at the time of CT interpretation or follow-up review, supporting more refined risk stratification and targeted cardiovascular evaluation. [16]Importantly, the net benefit of any downstream clinical action based on model predictions must be demonstrated before implementation in real-world settings.

This study has limitations. First, it is a single-centre retrospective study with a limited sample size, particularly in the validation and test sets, which may affect the stability of performance estimates. Second, although multi-vendor CT scanners were included, differences in scanning protocols and reconstruction algorithms may introduce variability, and no domain adaptation or batch-effect correction was applied. Third, the work focused on predictive performance and did not include formal interpretability analyses, limiting transparency regarding model decision-making. Finally, follow-up duration was relatively short, and longer-term cardiovascular risk prediction was not assessed. [16]Future studies should validate the model in larger, multi-centre cohorts to better assess generalisability and clinical applicability, incorporate additional clinical indicators and longer follow-up,

and apply interpretability and robustness analyses to enhance transparency and support translation toward clinical use. [17]

Conclusion

In a cohort of 306 patients with COPD, we developed and internally evaluated a multimodal deep learning framework that integrates component-level lung structural representations from routine chest CT (relatively normal lung, emphysema and pulmonary vasculature) with clinical variables to predict 1-year composite cardiovascular events. The model achieved AUCs of 0.8794, 0.9822 and 0.8588 in the training, validation and test sets, respectively, and outperformed an interpretable clinical baseline model based on common cardiovascular risk factors, suggesting that CT-derived imaging representations may provide incremental information for short-term outcome prediction in this setting. Nevertheless, this was a single-centre retrospective study with limited validation/test sample sizes and internal testing only; external validation, calibration assessment and evaluation of clinical net benefit are required to determine generalisability and real-world utility.

Declarations

Ethics approval and consent to participate

The study protocol was approved by the Ethics Committee of the Fourth Affiliated Hospital of Harbin Medical University (approval No. 2025-115, 15 August 2025). The requirement for written informed consent was waived due to the retrospective design and the use of de-identified data.

Consent for publication

Not applicable.

Availability of data and materials

The data that support the findings of this study are not publicly available due to patient privacy and ethical restrictions. De-identified data may be made available from the corresponding author upon reasonable request and with approval from the Ethics Committee of the Fourth Affiliated Hospital of Harbin Medical University (or the institutional data access committee, as applicable).

Competing interests

The authors declare that they have no competing interests.

Funding

This research received no specific grant from any funding agency in the public, commercial or not-for-profit sectors.

Authors’ contributions

Yu Kang: Data curation; Methodology; Writing—original draft.

Chunhui Qin: Data acquisition; Validation.

Yuxin Li: Image processing; Software.

Chunxiao Wang: Formal analysis (statistical analysis).

Shuai Yunjia: Data curation.

Li Ao: Data curation.

Chen Jiahui: Data curation.

Tong Zhang: Conceptualization; Supervision; Project administration; Writing—review & editing; Guarantor.

All authors read and approved the final manuscript.

Patient and public involvement

Patients and/or the public were not involved in the design, conduct, reporting or dissemination plans of this research.

Acknowledgements

Not applicable.

Authors’ information

Tong Zhang, MD

Department of Radiology, The Fourth Affiliated Hospital of Harbin Medical University

No. 37, YiYuan Street, NanGang District, Harbin 150001, Heilongjiang, China

Tel: +86-451-82576831

E-mail: yingxiang939@163.com

References

1. Cornelius T. Clinical guideline highlights for the hospitalist: GOLD COPD update 2024. J Hosp Med. 2024;19(9):818–20.

2. Agustí AG, Noguera A, Sauleda J, Sala E, Pons J, Busquets X. Systemic effects of chronic obstructive pulmonary disease. Eur Respir J. 2003;21(2):347–60.
3. Chen W, Thomas J, Sadatsafavi M, FitzGerald JM. Risk of cardiovascular comorbidity in patients with chronic obstructive pulmonary disease: a systematic review and meta-analysis. Lancet Respir Med. 2015;3(8):631–9.
4. Donaldson GC, Hurst JR, Smith CJ, Hubbard RB, Wedzicha JA. Increased risk of myocardial infarction and stroke following exacerbation of COPD. Chest. 2010;137(5):1091–7.
5. Lynch DA, Al-Qaisi MA. Quantitative computed tomography in chronic obstructive pulmonary disease. J Thorac Imaging. 2013;28(5):284–90.
6. Washko GR, Nardelli P, Ash SY, Vegas Sanchez-Ferrero G, Rahaghi FN, Come CE, Dransfield MT, Kalhan R, Han MK, Bhatt SP, et al. Arterial Vascular Pruning, Right Ventricular Size, and Clinical Outcomes in Chronic Obstructive Pulmonary Disease. A Longitudinal Observational Study. Am J Respir Crit Care Med. 2019;200(4):454–61.
7. Falk T, Mai D, Bensch R, Çiçek Ö, Abdulkadir A, Marrakchi Y, Böhm A, Deubner J, Jäckel Z, Seiwald K, et al. U-Net: deep learning for cell counting, detection, and morphometry. Nat Methods. 2019;16(1):67–70.
8. Weintraub WS. Statistical Approaches to Composite Endpoints. JACC Cardiovasc Interv. 2016;9(22):2289–91.
9. Sinaiko AR, Jacobs DR Jr., Woo JG, Bazzano L, Burns T, Hu T, Juonala M, Prineas R, Raitakari O, Steinberger J, et al. The International Childhood Cardiovascular Cohort (i3C) consortium outcomes study of childhood cardiovascular risk factors and adult cardiovascular morbidity and mortality: Design and recruitment. Contemp Clin Trials. 2018;69:55–64.
10. Hasenstab KA, Tabalon J, Yuan N, Retson T, Hsiao A. CNN-based Deformable Registration Facilitates Fast and Accurate Air Trapping Measurements at Inspiratory and Expiratory CT. Radiol Artif Intell. 2022;4(1):e210211.
11. Van Calster B, McLernon DJ, van Smeden M, Wynants L, Steyerberg EW. Calibration: the Achilles heel of predictive analytics. BMC Med. 2019;17(1):230.
12. Pistenmaa CL, Nardelli P, Ash SY, Come CE, Diaz AA, Rahaghi FN, Barr RG, Young KA, Kinney GL, Simmons JP, et al. Pulmonary Arterial Pruning and Longitudinal Change in Percent Emphysema and Lung Function: The Genetic Epidemiology of COPD Study. Chest. 2021;160(2):470–80.
13. Barr RG, Bluemke DA, Ahmed FS, Carr JJ, Enright PL, Hoffman EA, Jiang R, Kawut SM, Kronmal RA, Lima JA, et al. Percent emphysema, airflow obstruction, and impaired left ventricular filling. N Engl J Med. 2010;362(3):217–27.
14. Polverino F, Celli BR, Owen CA. COPD as an endothelial disorder: endothelial injury linking lesions in the lungs and other organs? (2017 Grover Conference Series). *Pulm Circ* 2018, 8(1):2045894018758528.
15. Zantvoort K, Nacke B, Görlich D, Hornstein S, Jacobi C, Funk B. Estimation of minimal data sets sizes for machine learning predictions in digital mental health interventions. NPJ Digit Med.

2024;7(1):361.
16. Vickers AJ, Elkin EB. Decision curve analysis: a novel method for evaluating prediction models. Med Decis Mak. 2006;26(6):565–74.
17. Collins GS, Reitsma JB, Altman DG, Moons KG. Transparent Reporting of a multivariable prediction model for Individual Prognosis or Diagnosis (TRIPOD): the TRIPOD statement. Ann Intern Med. 2015;162(1):55–63.

Figures

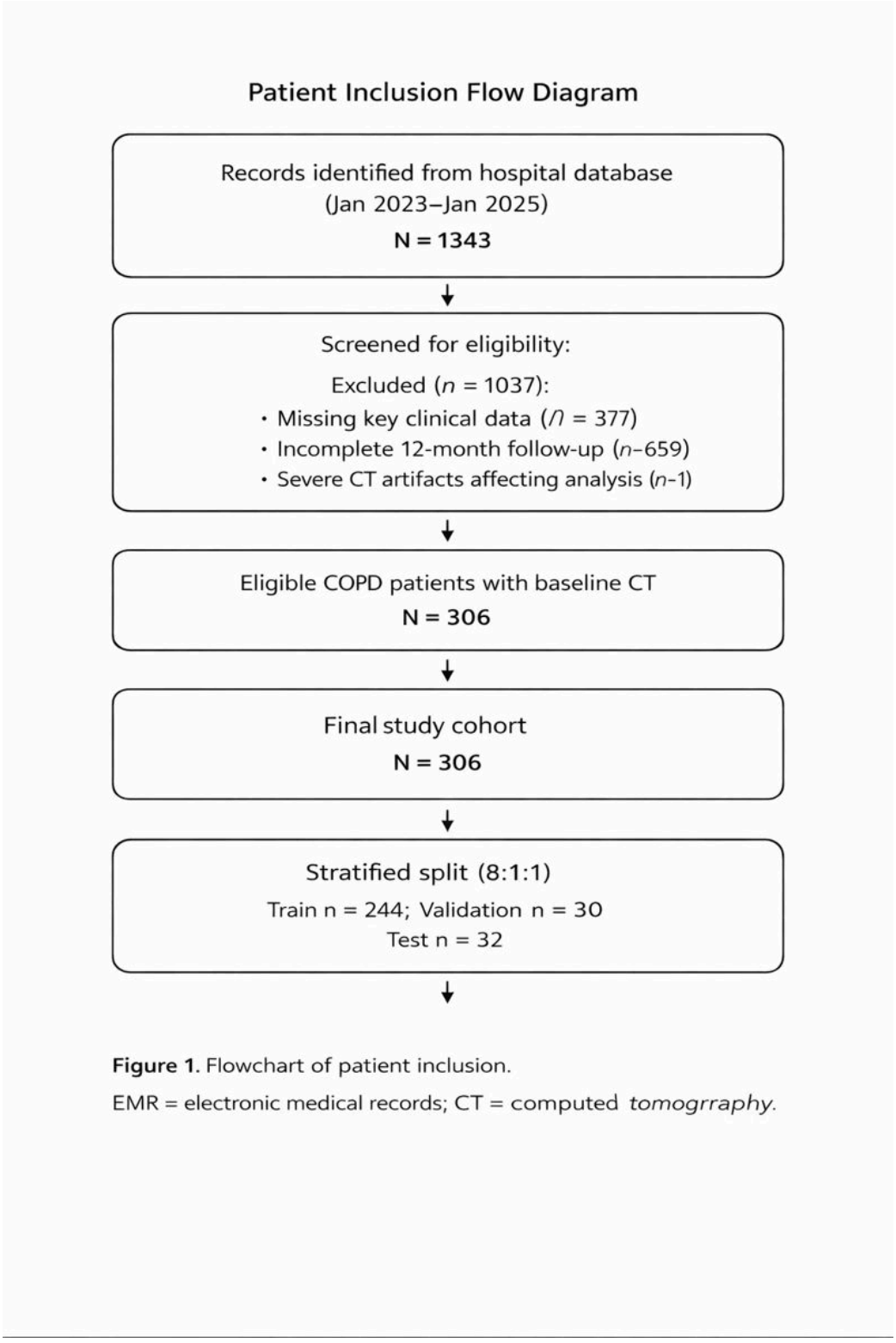


Figure 1. Flowchart of patient inclusion.

EMR = electronic medical records; CT = computed *tomogrraphy*.

Figure 1

Patient inclusion flow diagram.

Records were identified from the hospital database (Jan 2023–Jan 2025, **N = 1343**) and screened for eligibility. A total of **1037**records were excluded due to missing key clinical data (**n = 377**), incomplete 12-month follow-up (**n = 659**), or severe CT artifacts affecting analysis (**n = 1**). The final study cohort

included **306**eligible COPD patients with baseline CT and was stratified into a training set (**n = 244**), validation set (**n = 30**), and test set (**n = 32**) at an 8:1:1 ratio.

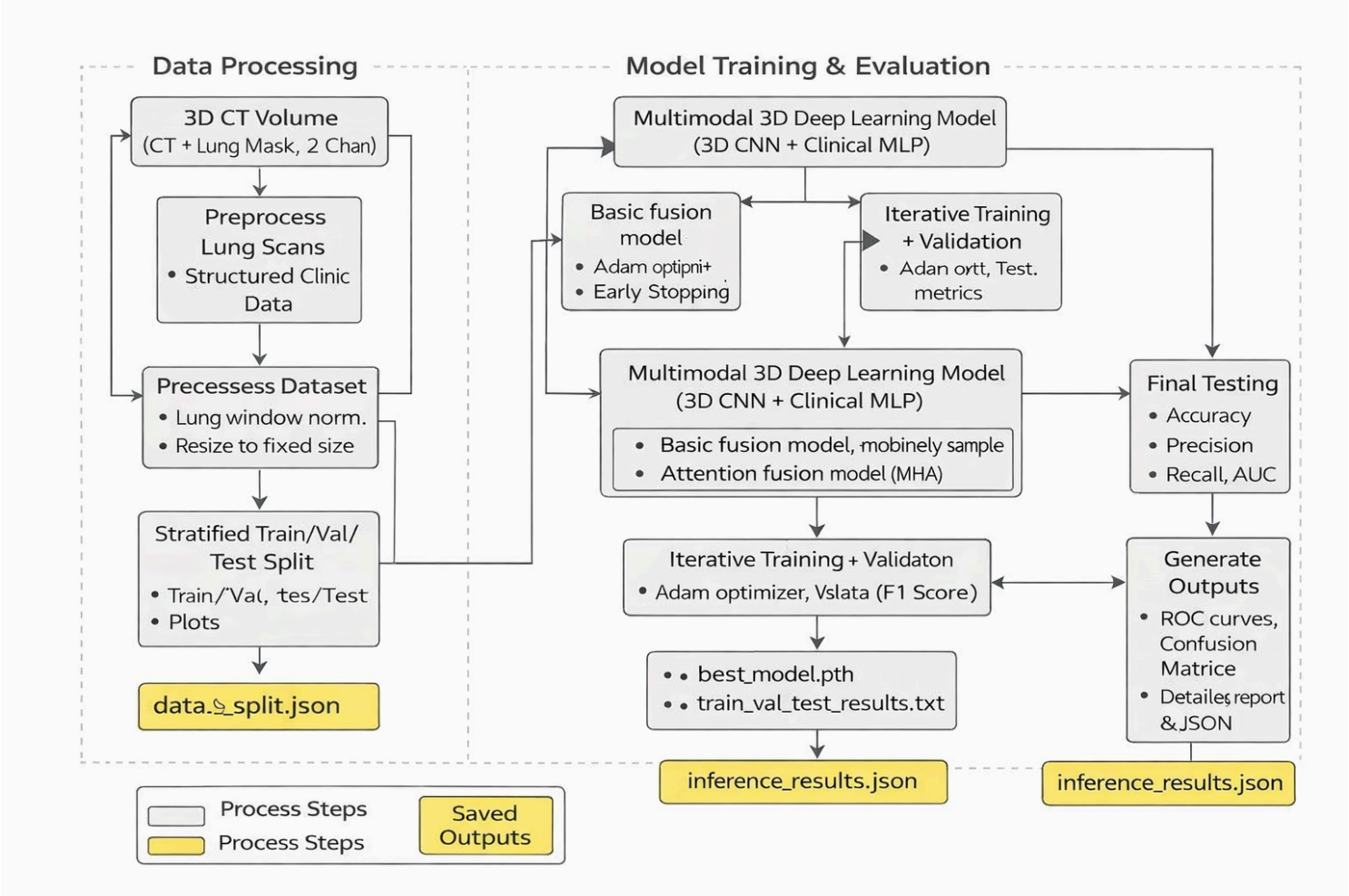


Figure 2

Overview of data processing, model training, and evaluation.

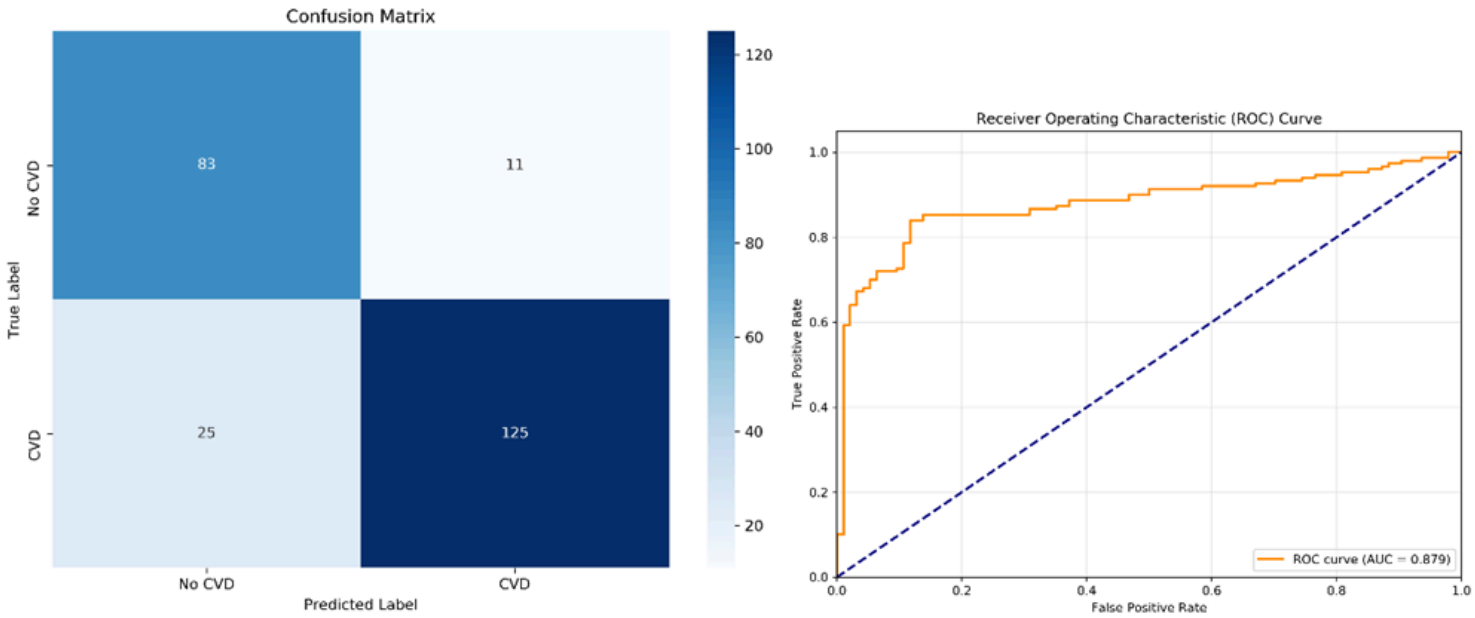


Figure 3

Model discrimination and classification performance in the training set (n = 244).

Receiver operating characteristic (ROC) curve and confusion matrix for the training set. The model achieved **AUC = 0.8794**. Confusion matrix counts: **TN = 83, FP = 11, FN = 25, TP = 125**.

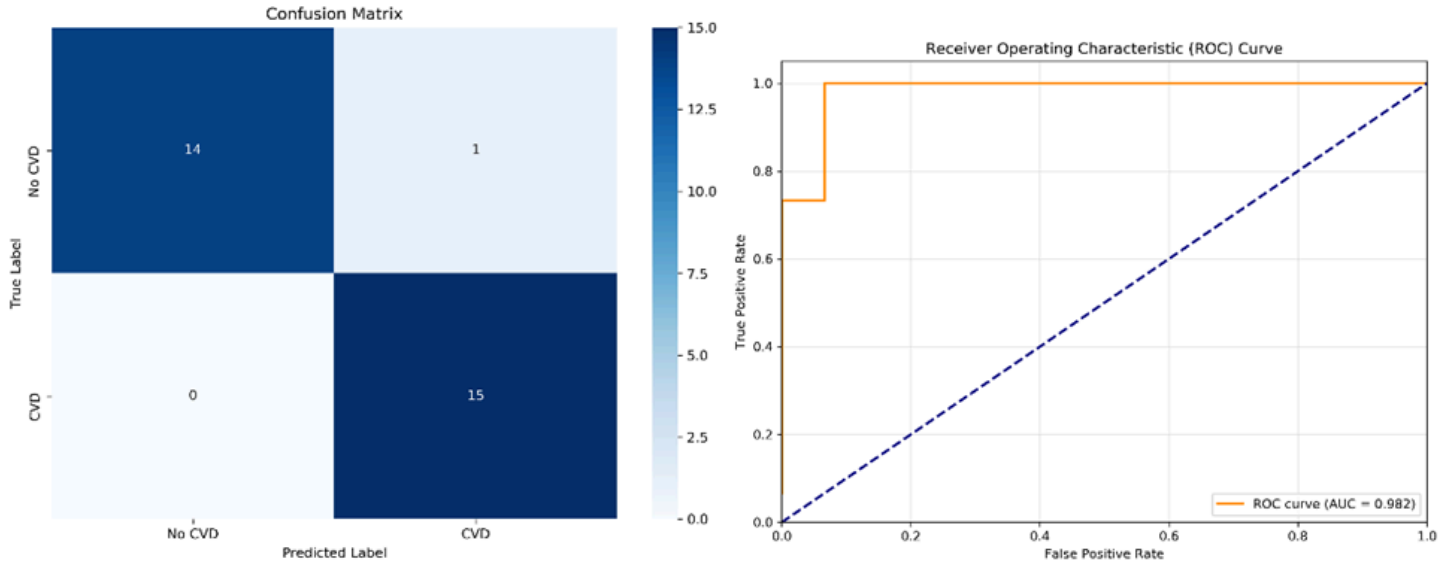


Figure 4

Model discrimination and classification performance in the validation set (n = 30).

ROC curve and confusion matrix for the validation set. The model achieved **AUC = 0.9822**. Confusion matrix counts: **TN = 14, FP = 1, FN = 0, TP = 15**.

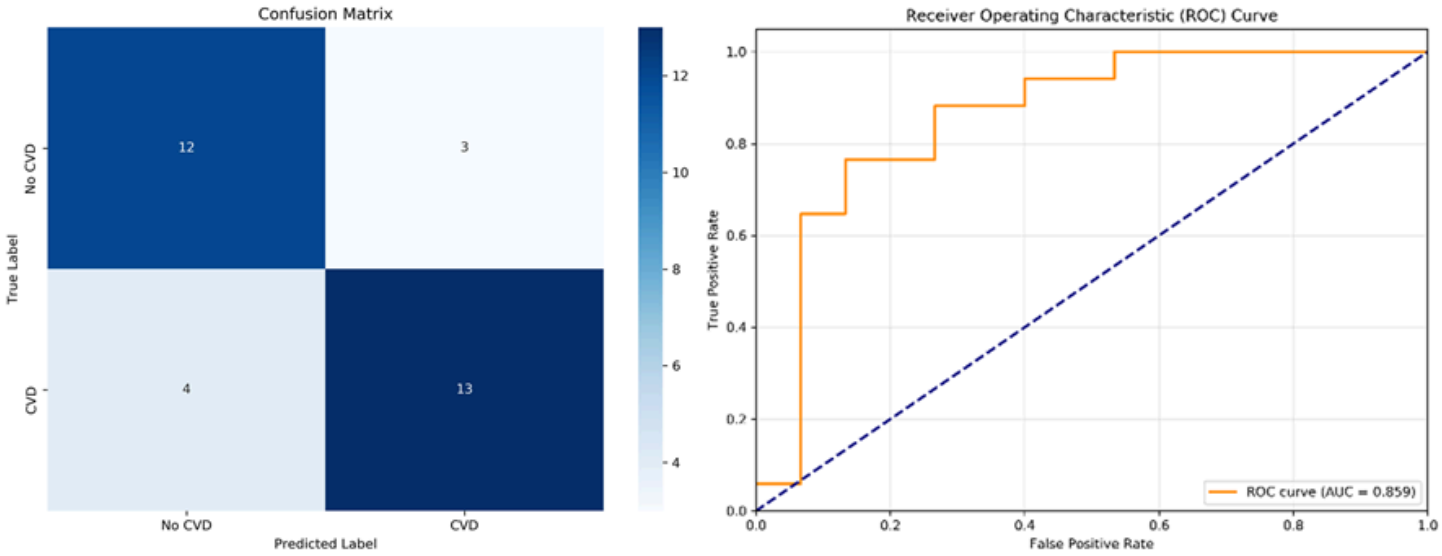


Figure 5

Model discrimination and classification performance in the test set (n = 32).

ROC curve and confusion matrix for the independent test set. The model achieved **AUC = 0.8588**. Confusion matrix counts: **TN = 12, FP = 3, FN = 4, TP = 13**.

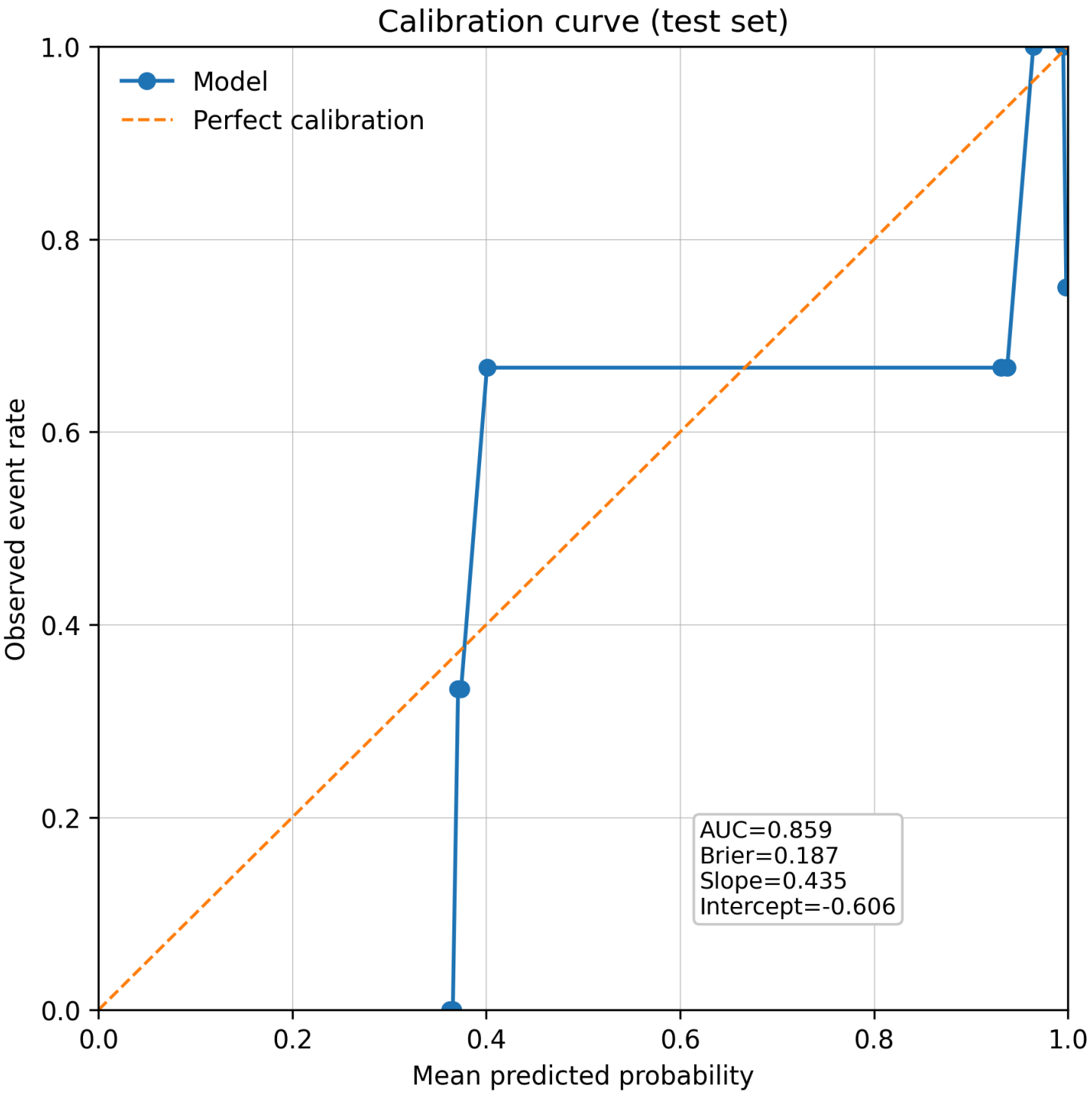


Figure 6

Calibration curve of the multimodal model in the independent test set.

Calibration plot comparing mean predicted probability and observed event rate in the test set (**n = 32**). The dashed line indicates perfect calibration. Model calibration statistics are reported in the panel (**Brier score = 0.187; calibration slope = 0.435; intercept = −0.606; AUC = 0.859**).