

Appendix 1

1. What is the name of your institution? [Fill in the blank]

2. What is the grade level of your institution? [Single-choice question]
 - A. Grade A tertiary hospital;
 - B. Grade B tertiary hospital;
 - C. Grade A secondary hospital;
 - D. Grade B secondary hospital;
 - E. Others.

3. In which department do you practice as a doctor? [Single-choice question]
 - A. Respiratory and Critical Care Medicine;
 - B. Infectious Diseases;
 - C. Critical Care Medicine;
 - D. Radiology;
 - E. Other subspecialties of Internal Medicine;
 - F. Surgery.

4. How long have you been engaged in clinical work? [Single-choice question]
 - A. 1-5 years;
 - B. 6-10 years;
 - C. 11-20 years;
 - D. 21-30 years;
 - E. >30 years.

5. How many cases of pulmonary aspergillosis (with definitive histopathological evidence) are diagnosed in your department each year? [Single-choice question]

- A. <20 cases;
- B. 20-50 cases;
- C. 50-100 cases;
- D. 100-200 cases;
- E. >200 cases

6. Can your institution independently and routinely conduct fungal antigen tests? [Single-choice question]

- A. Independently conduct fungal antigen tests (serum);
- B. Independently conduct fungal antigen tests (bronchoalveolar lavage fluid);
- C. Independently conduct fungal antigen tests (serum + bronchoalveolar lavage fluid);
- D. Do not independently conduct, but send to a third-party testing institution;
- E. Do not independently conduct and do not send to a third-party testing institution.

7. Does your institution routinely conduct special histopathological staining tests such as PAS (Periodic Acid-Schiff), AB (Alcian Blue), and GMS (Gomori Methenamine Silver)? [Single-choice question]

- A. Yes;
- B. No

8. Which type of pulmonary aspergillosis do you find the most challenging to diagnose? [Single-choice question]

- A. Invasive pulmonary aspergillosis;
- B. Invasive tracheobronchial aspergillosis;
- C. Chronic necrotizing pulmonary aspergillosis;
- D. Aspergilloma;
- E. Allergic bronchopulmonary aspergillosis

9. How familiar are you with the use of various antifungal medications?

[Single-choice question]

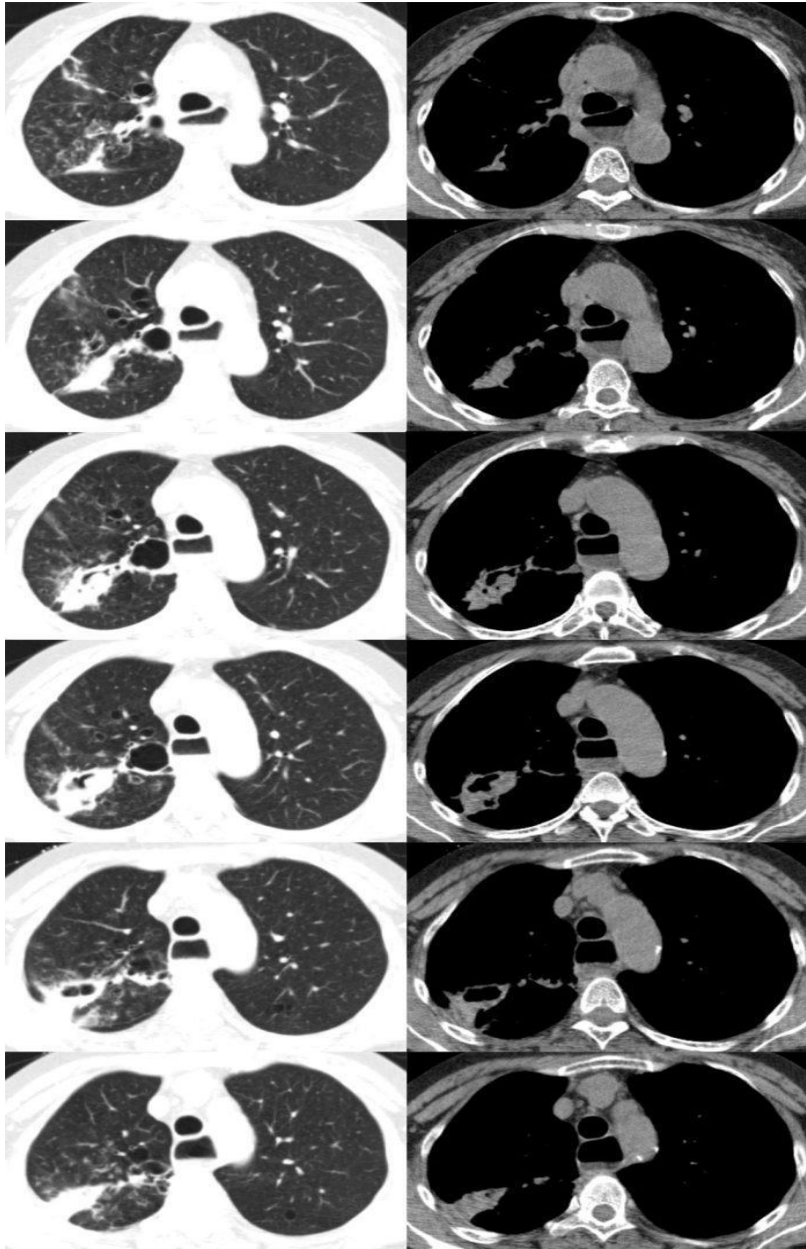
- A. Very unfamiliar (never used);
- B. Somewhat unfamiliar (used, but rarely);
- C. Average;
- D. Fairly familiar;
- E. Very familiar

10. What is your opinion on the application prospect of AI in CT imaging for differentiating pulmonary aspergillosis from bacterial pneumonia? [Single-choice question]

- A. Undetermined;
- B. Bad;
- C. So-so;
- D. Promising.

Appendix 2

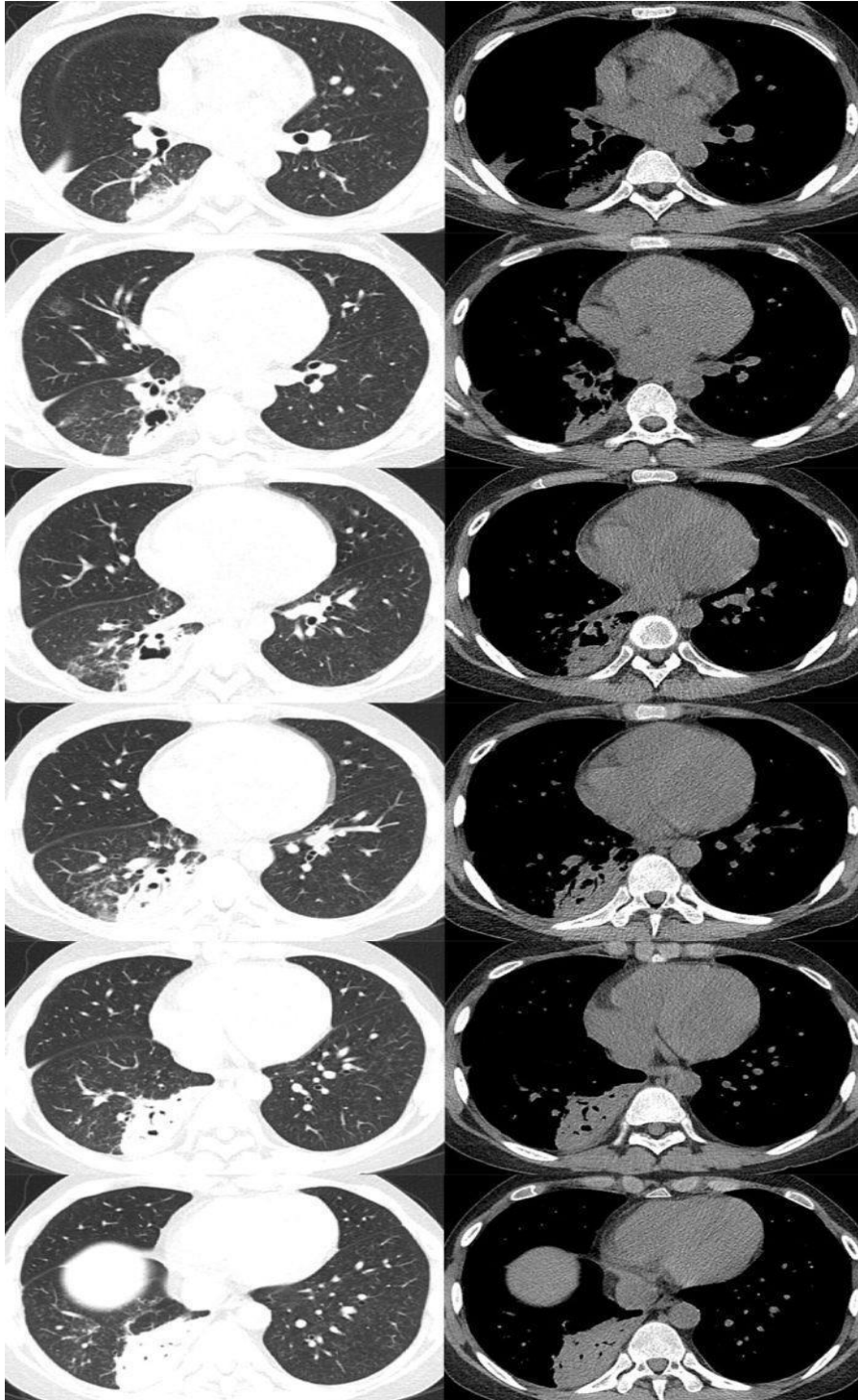
Case 1: A 56-year-old female with no underlying diseases was hospitalized for a chief complaint of chronic cough and hemoptysis persisting for a duration of 5 years, and episodes recurring for a period exceeding two weeks.



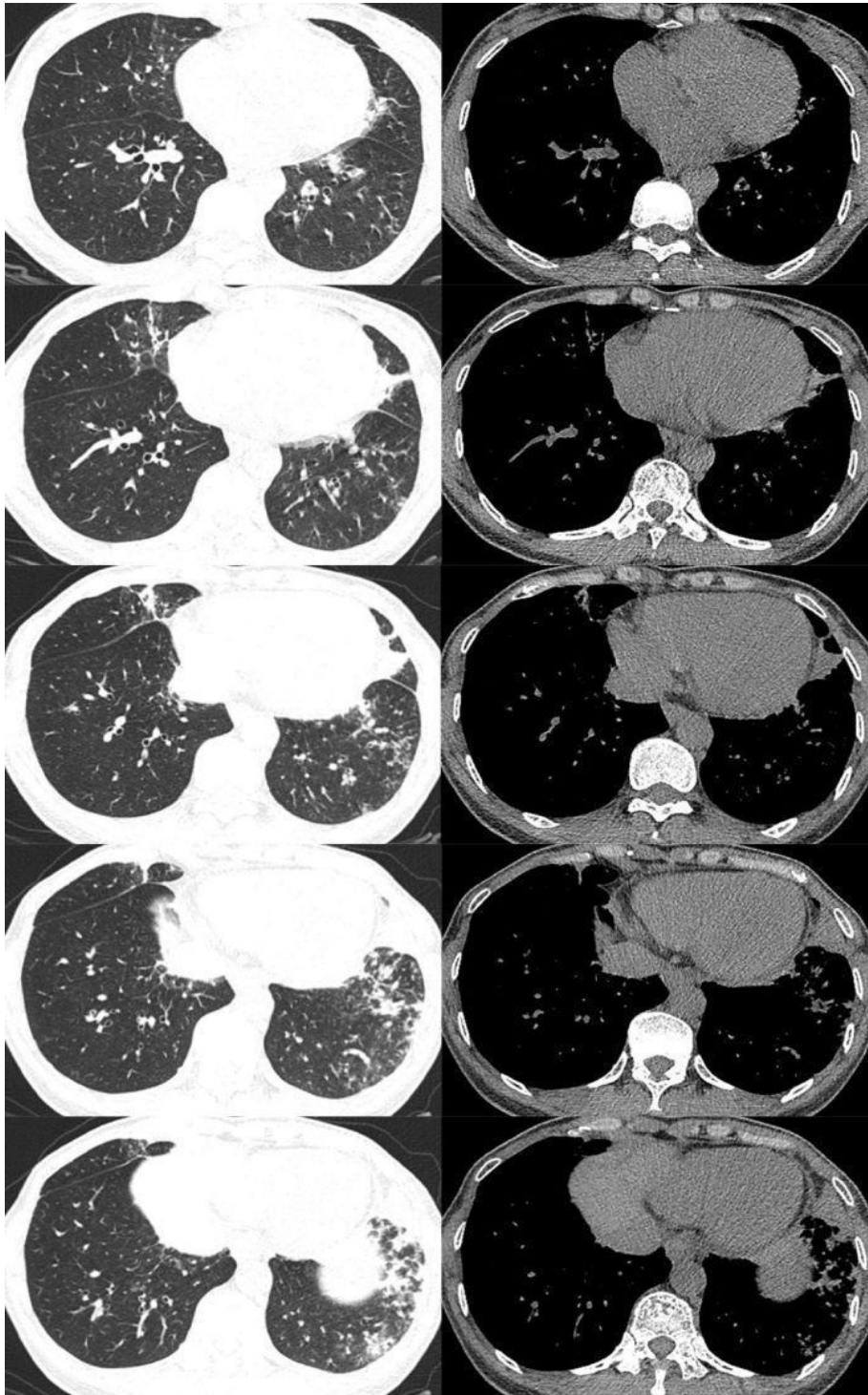
Case 2: A 65-year-old male, who has a medical history of hypertension and diabetes, was hospitalized for a chief complaint of persistent cough, expectoration, and hemoptysis for over a year, with worsening symptoms over a month.



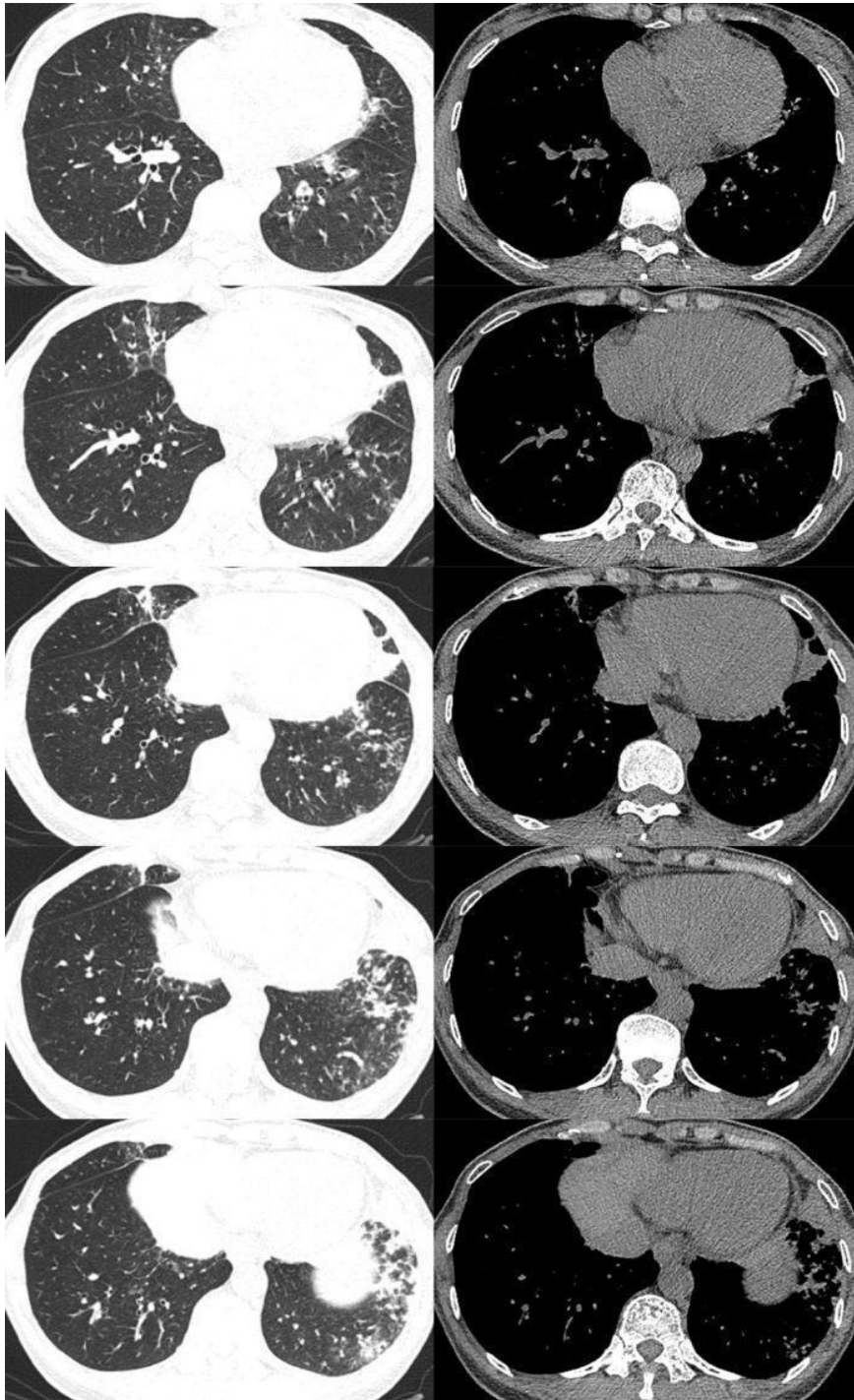
Case 3: A 20-year-old male, who has a medical history of intrahepatic bile duct calculus, chronic renal insufficiency, and chronic glomerulonephritis, was hospitalized for a chief complaint of cough, phlegm, fever, and chest pain persisting for a duration of 20 days.



Case 4: A 53-year-old female, who has a medical history of moderate-severe mitral valve insufficiency, was hospitalized for a chief complaint of recurrent fever and cough persisting for 2 years, with exacerbation over the past 3 days.



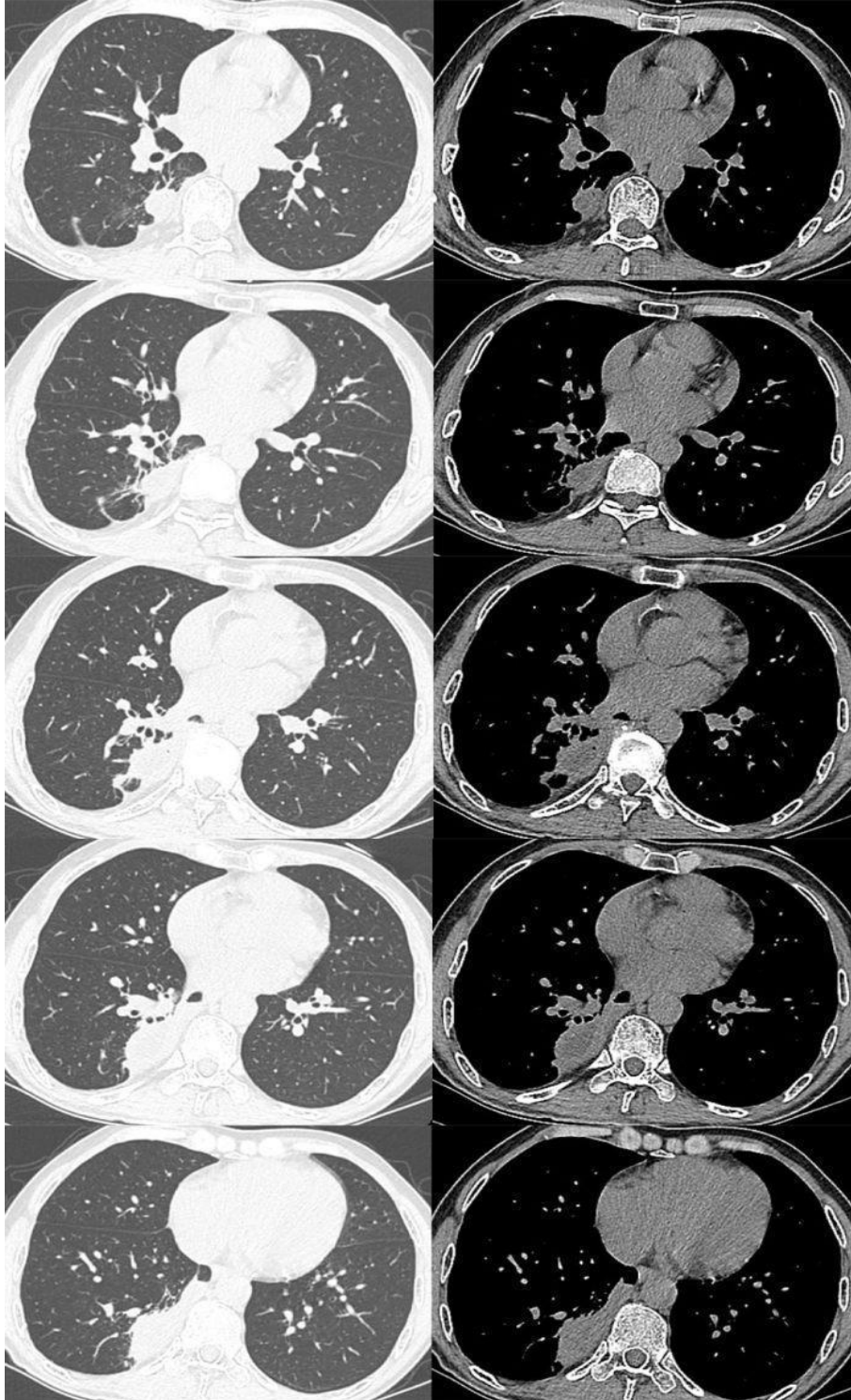
Case 5: A 62-year-old male, who has a medical history of cirrhosis in the decompensated stage, chronic viral hepatitis B in the active stage, right ureteral calculi, and renal effusion, was hospitalized for a chief complaint of fever, cough, and sputum persisting for a duration of two weeks.



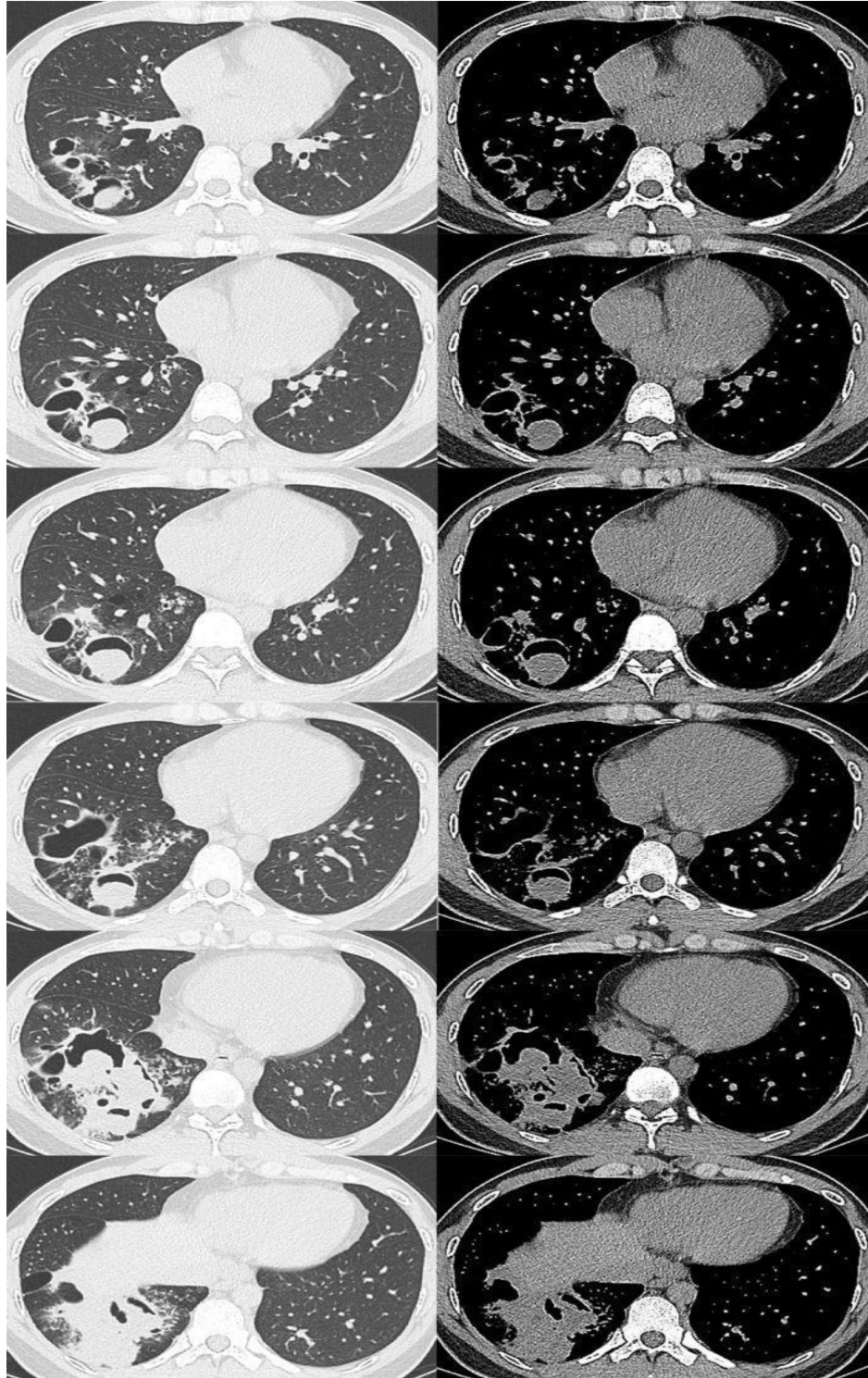
Case 6: A 38-year-old female, who has a medical history of pulmonary tuberculosis and bronchiectasis, was hospitalized for a chief complaint of hemoptysis for a duration exceeding four months.



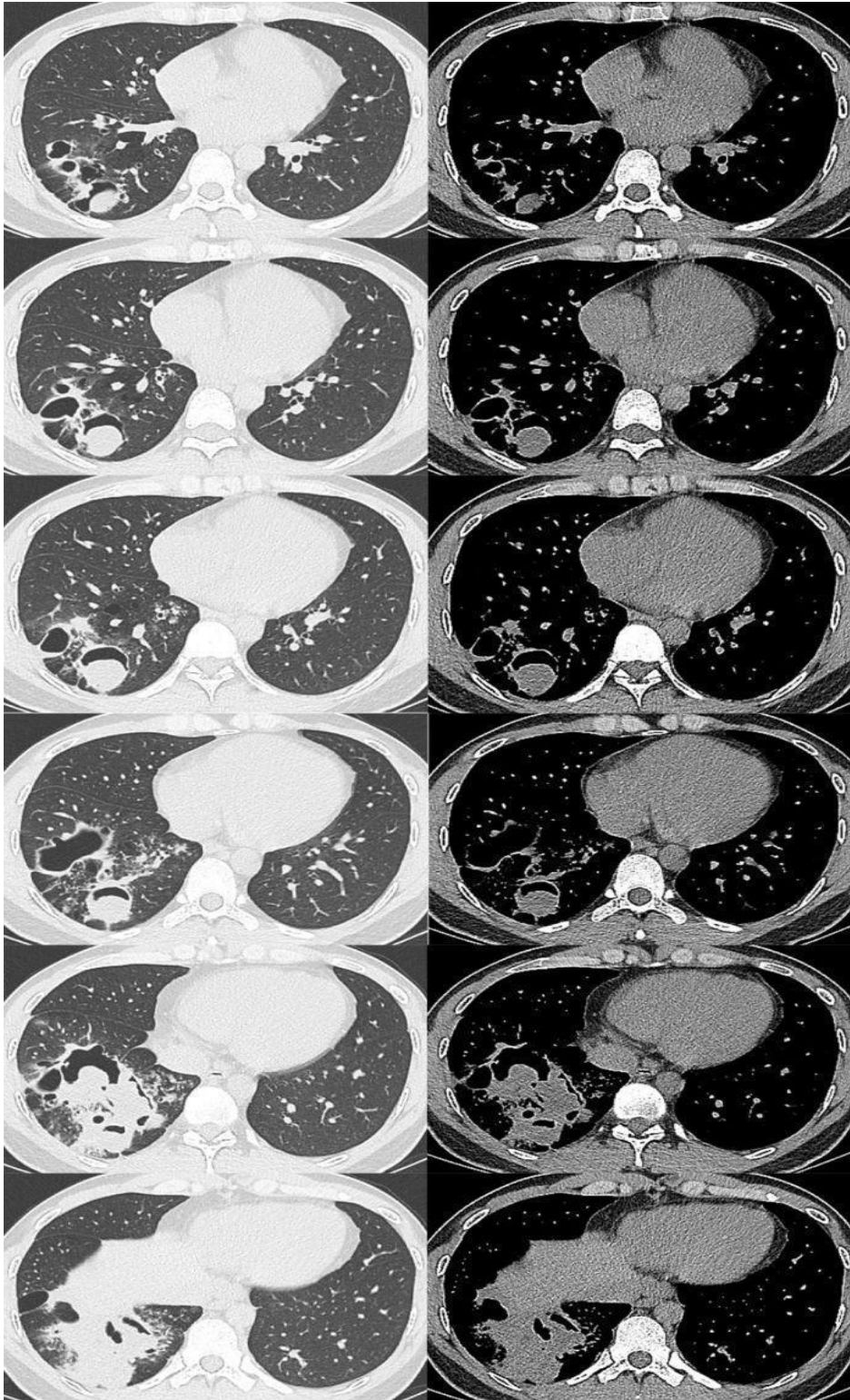
Case 7: A 53-year-old male, who has a medical history of pulmonary tuberculosis, was hospitalized for a chief complaint of hemoptysis for two months, and cough and expectoration for a month.



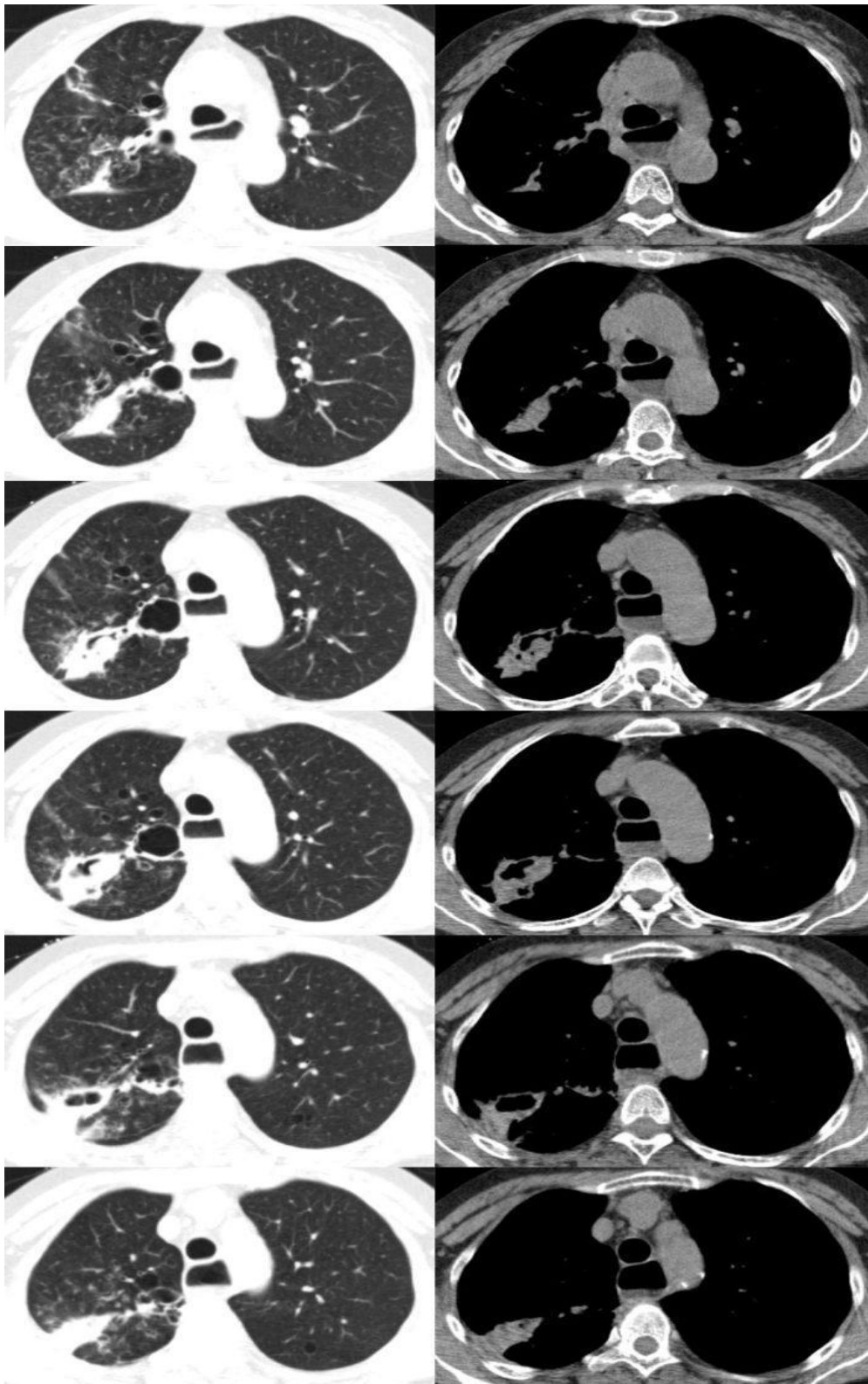
Case 8: A 25-year-old male with no underlying diseases was hospitalized for a chief complaint of cough, sputum production, hemoptysis, and chest tightness for 2 months, with fever for 1 day.



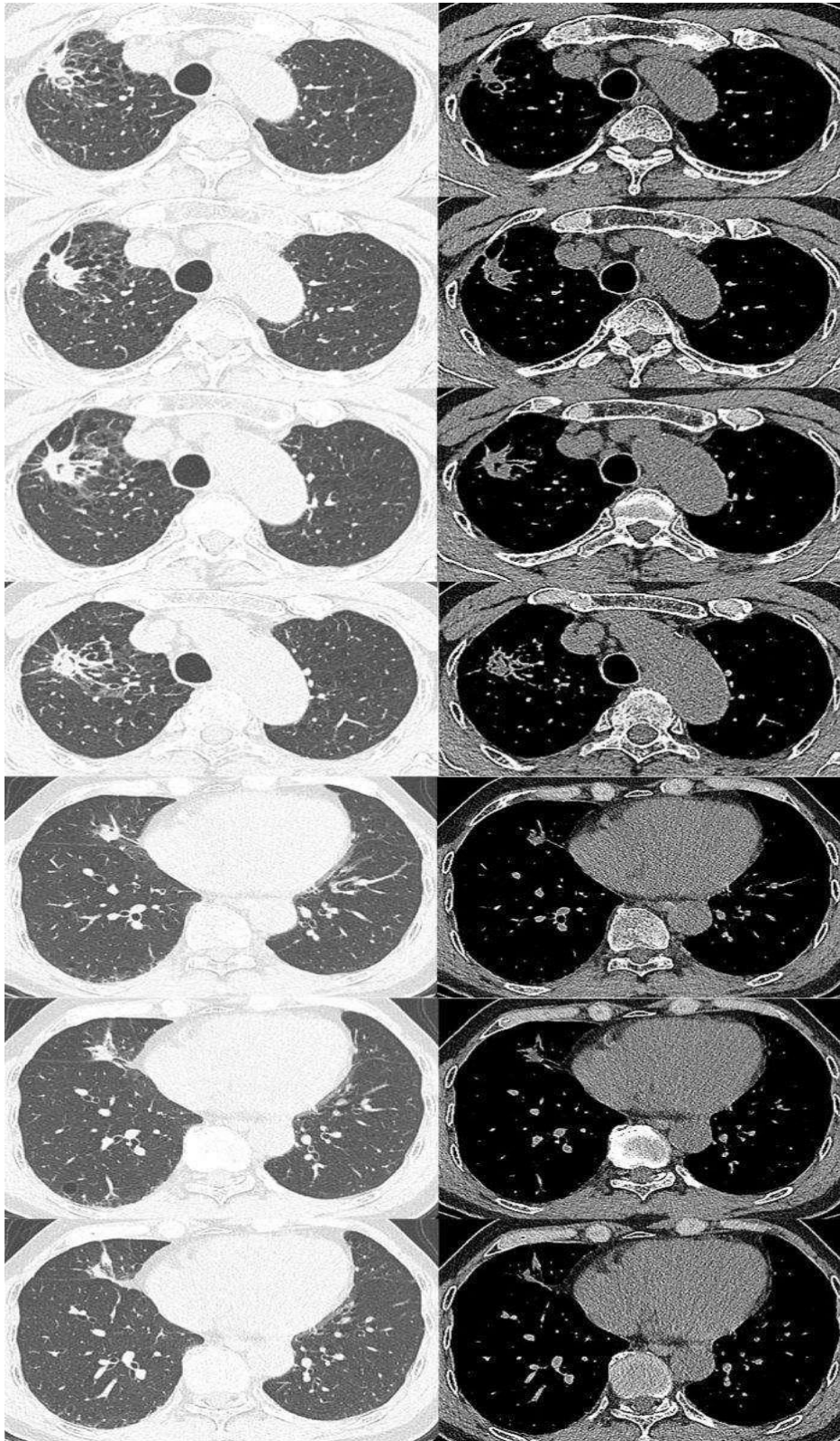
Case 9: A 80-year-old female, who had a history of chronic obstructive pulmonary disease, hypertension, and cerebral infarction, was hospitalized for a chief complaint of cough, fever, and shortness of breath for more than 10 days after a fall.



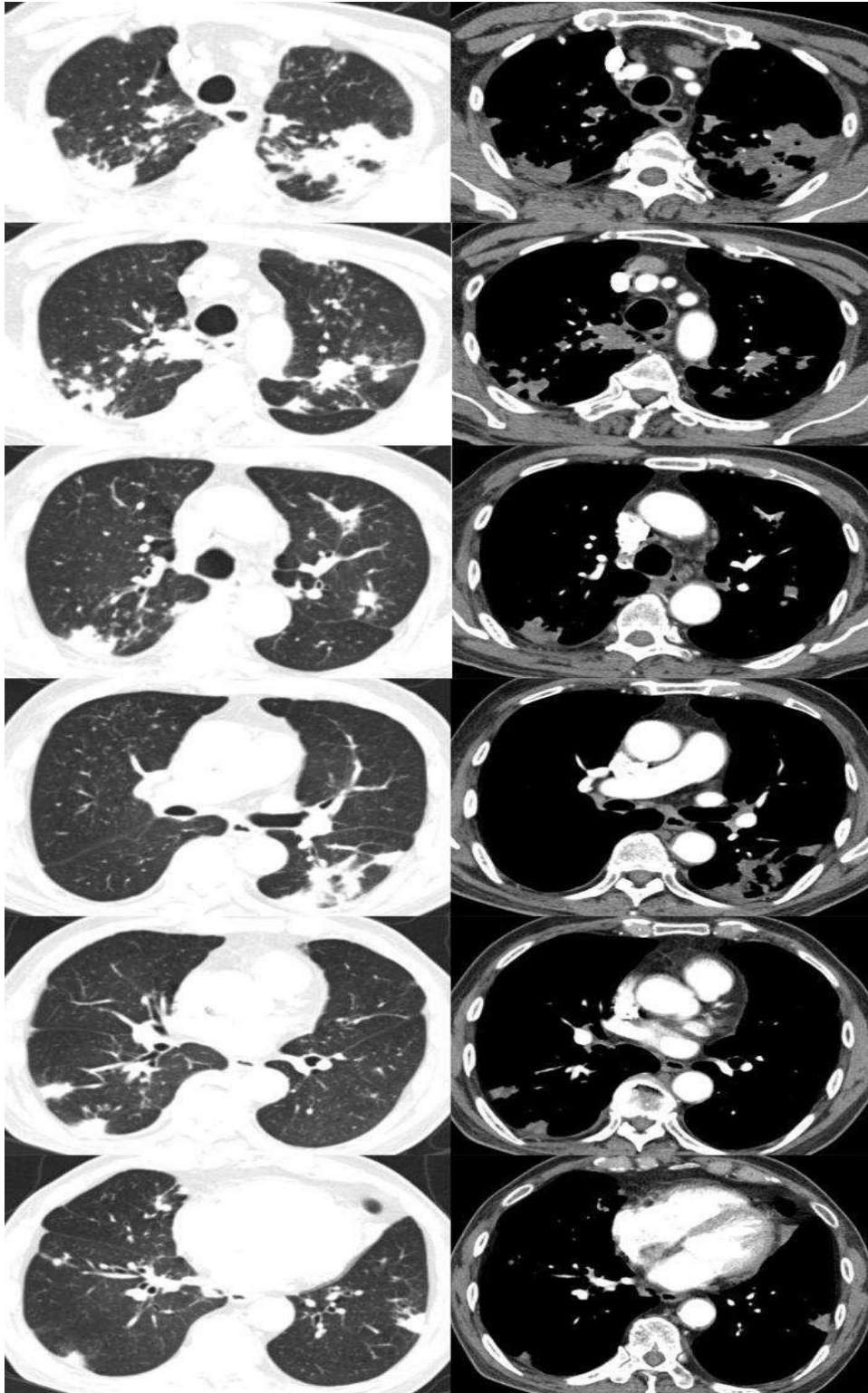
Case 10: A 58-year-old male, who had a history of chronic cholecystitis and multiple gallstones, was hospitalized for a chief complaint of hemoptysis for more than 20 days.



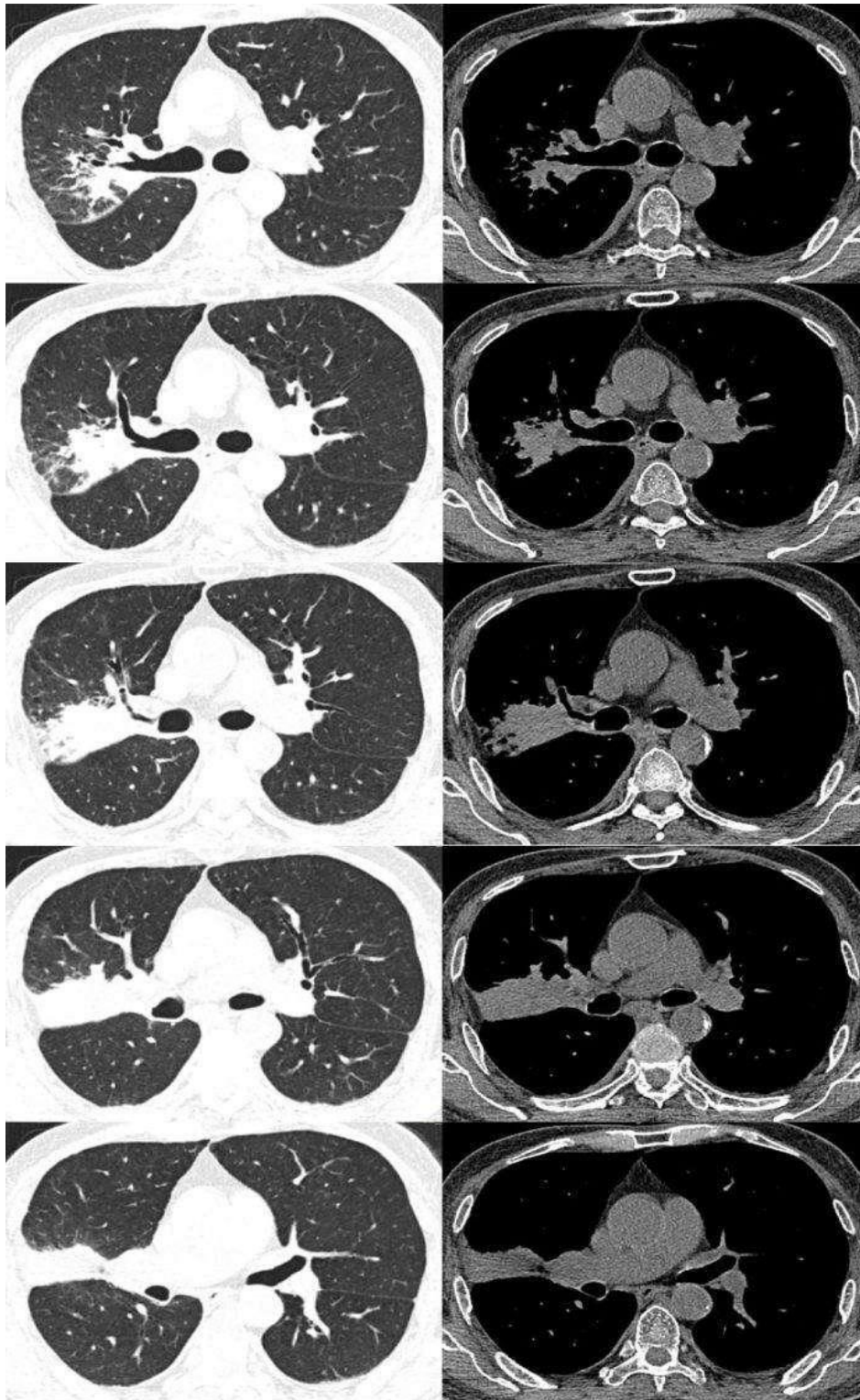
Case 11: A 70-year-old male, who had a history of pulmonary tuberculosis, was hospitalized for a chief complaint of intermittent hemoptysis for 4 days.



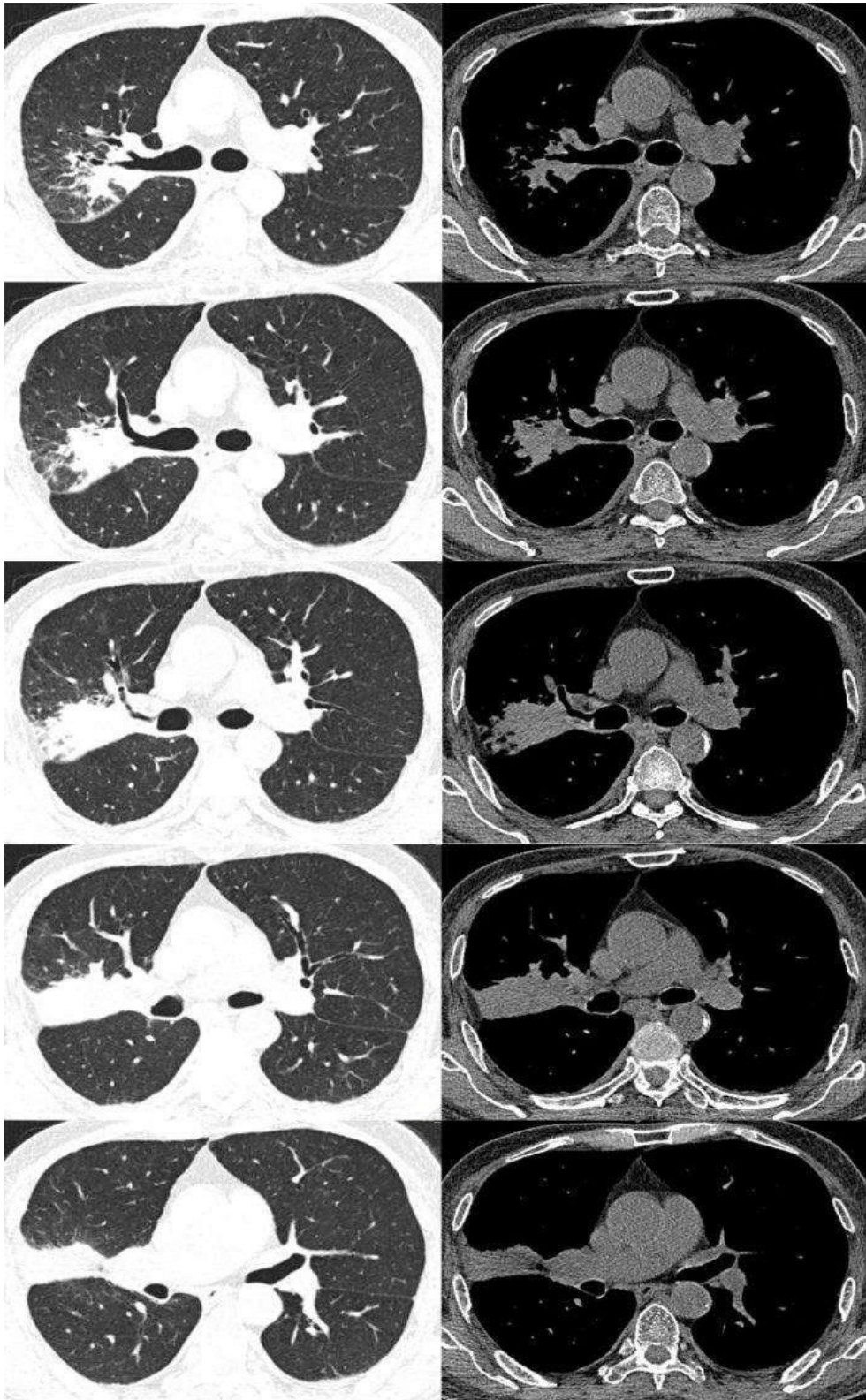
Case 12: A 77-year-old male, who had a history of chronic obstructive pulmonary disease, was hospitalized for a chief complaint of recurrent cough, sputum production, and shortness of breath for more than two months.



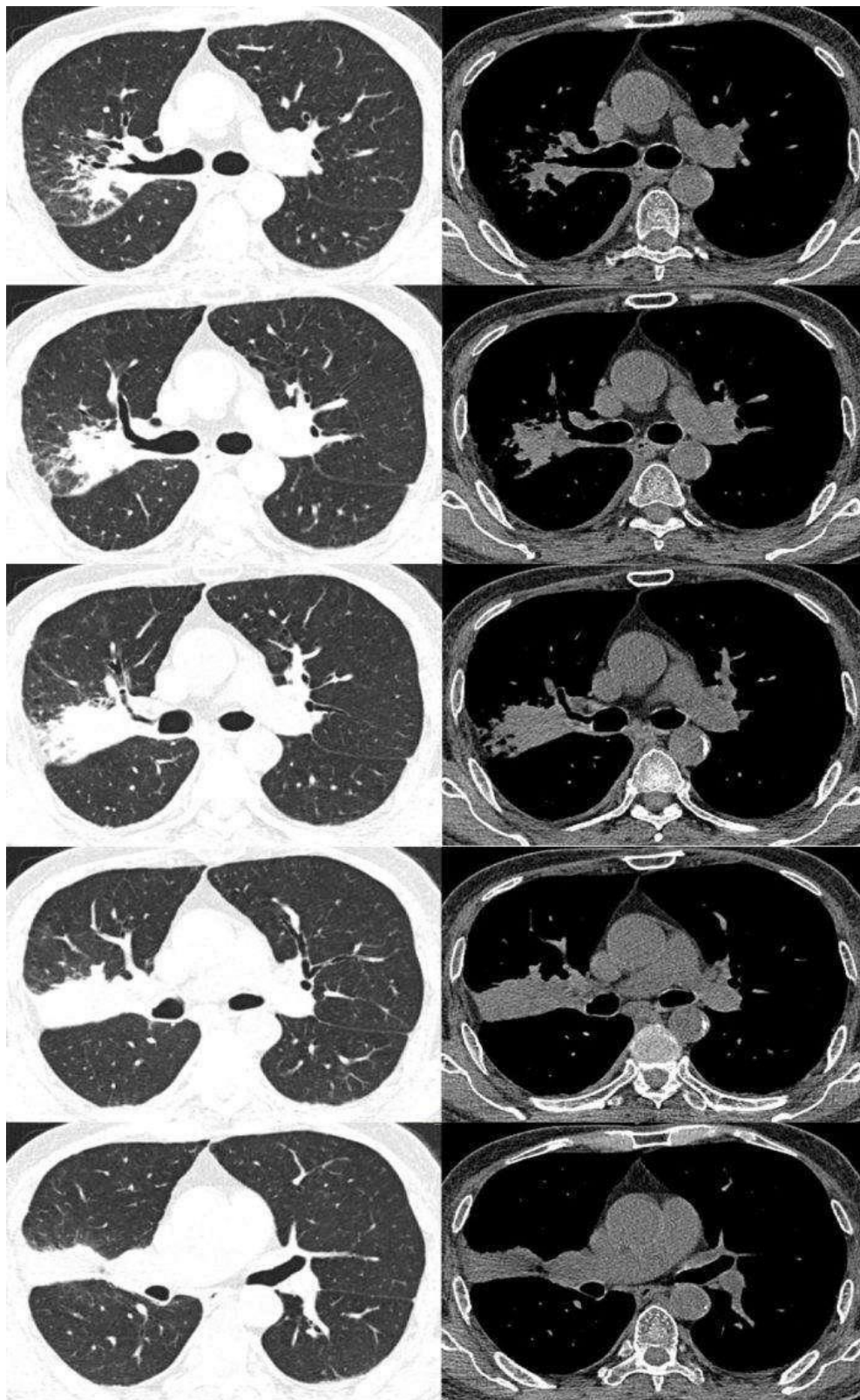
Case 13: A 70-year-old male, who had a history of chronic obstructive pulmonary disease and erythroderma, was hospitalized for a chief complaint of recurrent cough, sputum production, and shortness of breath for over two months.



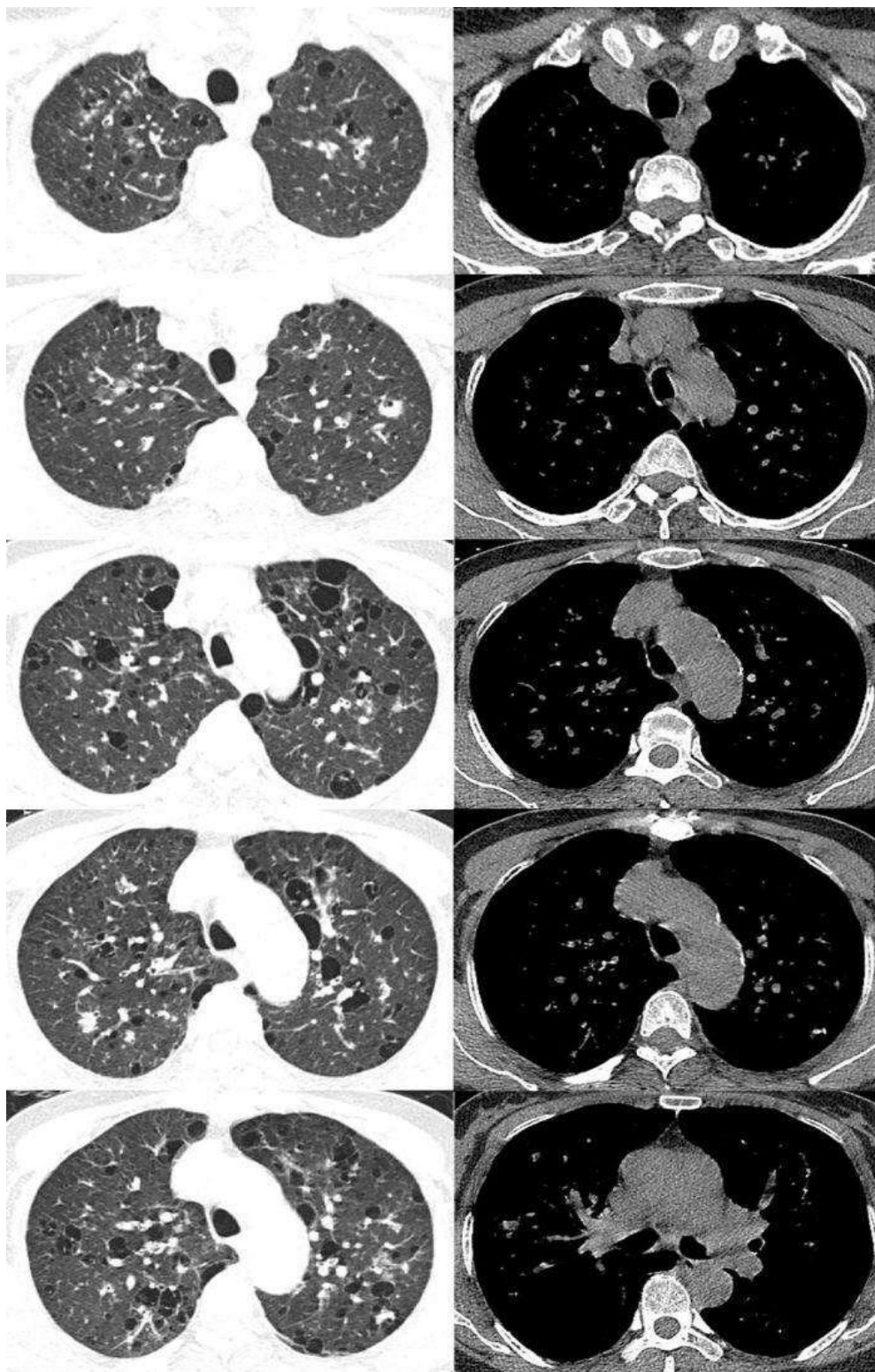
Case 14: A 57-year-old female, who had a history of gallbladder stones, was hospitalized for a chief complaint of fever and cough for one week.



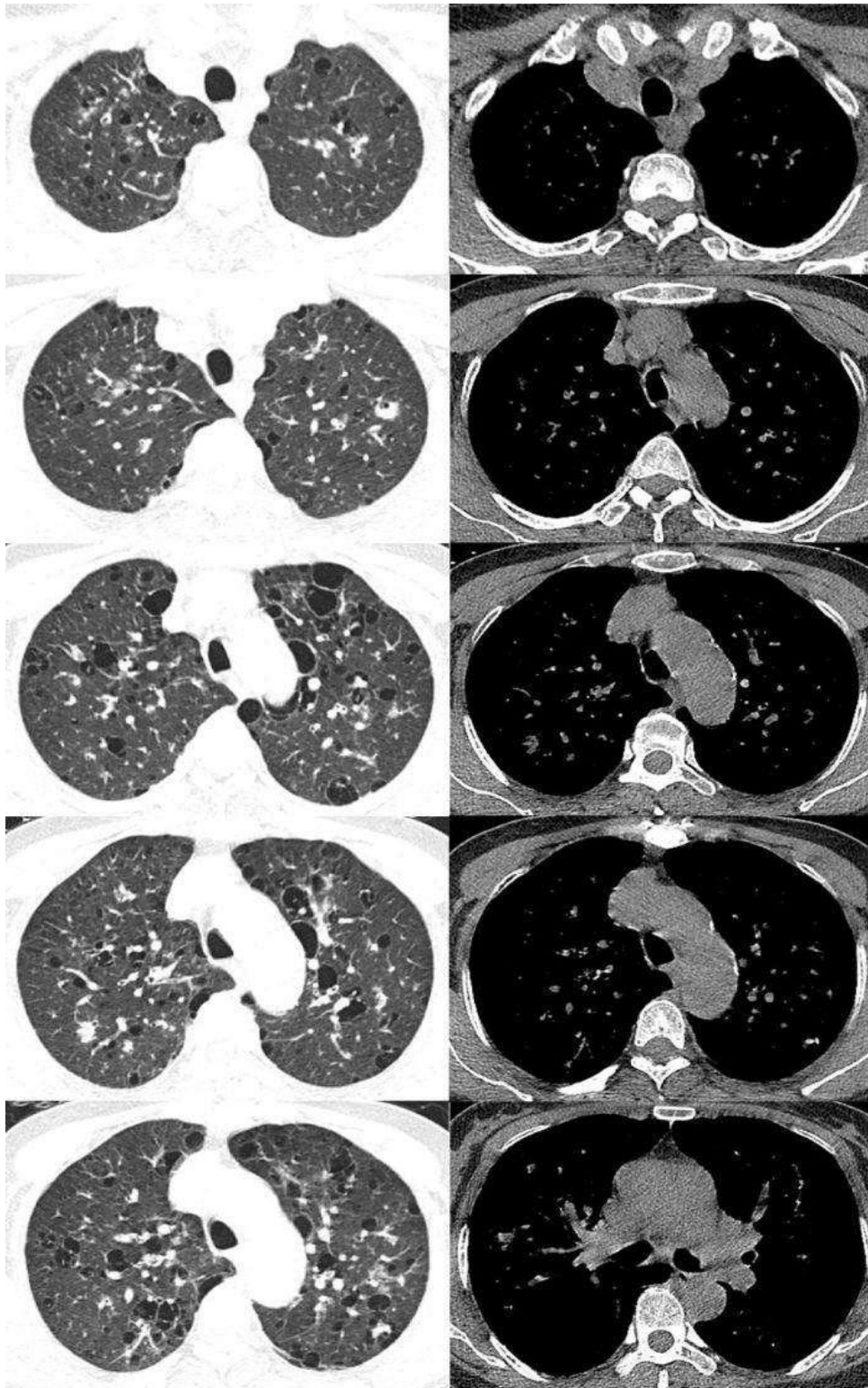
Case 15: A 55-year-old male, who had a history of diabetes, was hospitalized for a chief complaint of cough and sputum production for over two months.



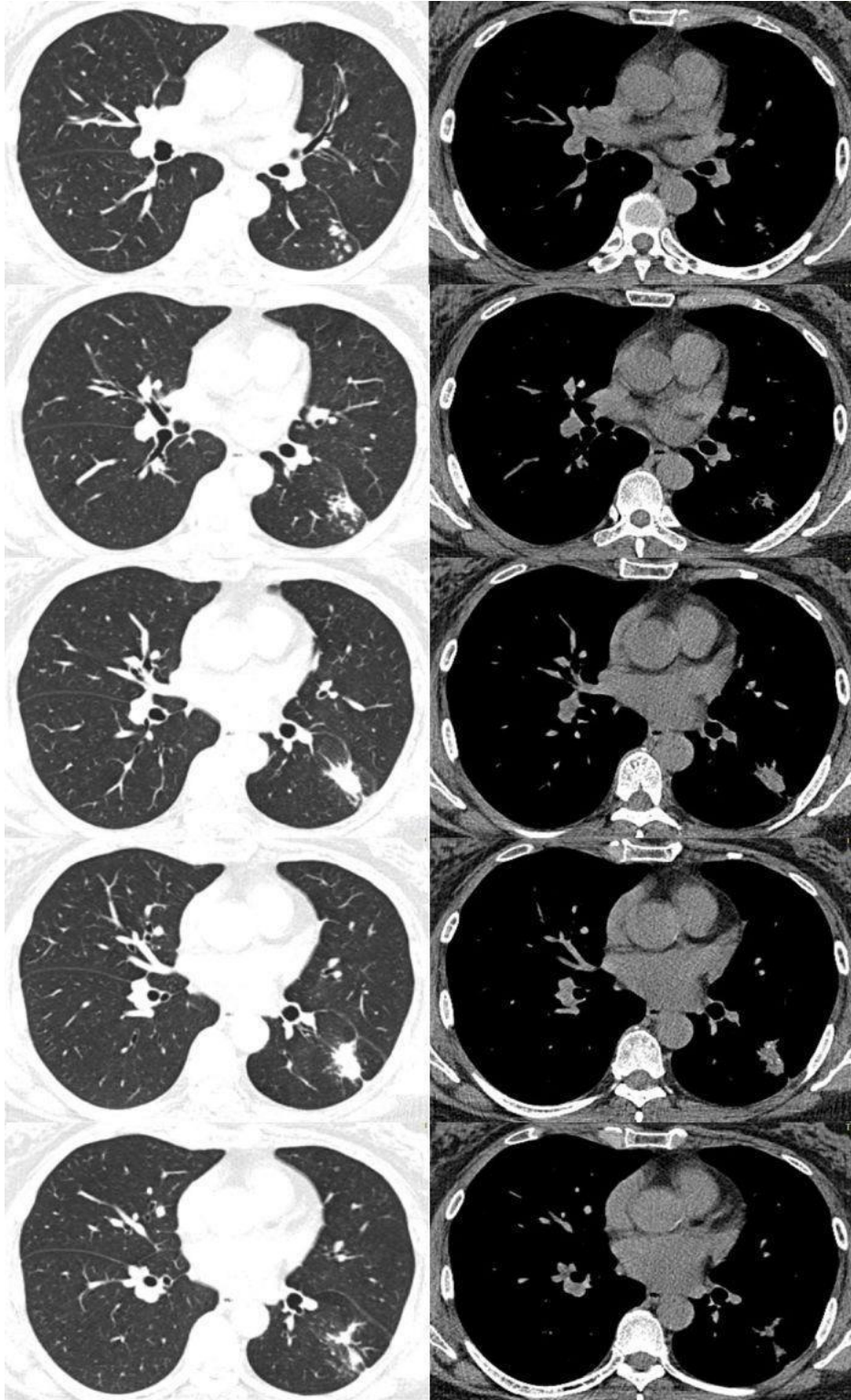
Case 16: A 56-year-old female, who had a history of Sjogren's syndrome, was hospitalized for a chief complaint of cough and sputum production for over a month, and shortness of breath for one week.



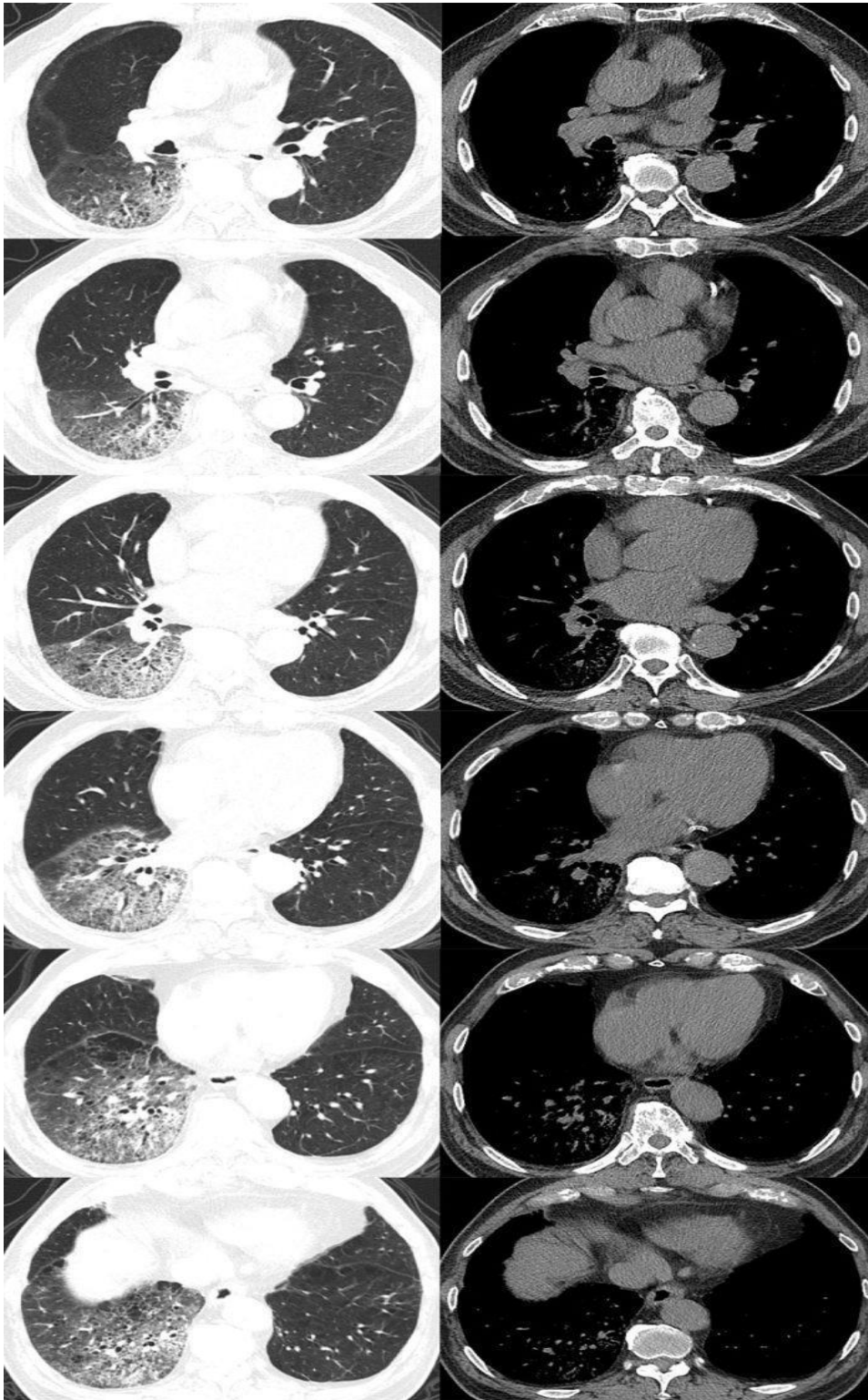
Case 17: A 53-year-old female, who had a history of asthma, chronic obstructive pulmonary disease, sinusitis, and diabetes, was hospitalized for a chief complaint of recurrent wheezing for over 10 years and symptoms worsening for the past 3 days.



Case 18: A 37-year-old male with no underlying diseases was hospitalized for a chief complaint of a left lung mass shadow identified during a physical examination three months prior.



Case 19: A 78-year-old male, who had a history of hypertension and coronary heart disease, was hospitalized for a chief complaint of cough, sputum production, and fever for seven days.



Case 20: A 37-year-old male with no underlying diseases was hospitalized for a chief complaint of fever, cough, and sputum production for 17 days.



Table S1 Architecture, image preprocessing, and training configurations of the unannotated supervised training and annotated supervised training models

Category	UST model	AST model
Manual lesion annotation	No	Yes, expert-reviewed lesion annotation
Foreground definition	Automatically generated lung mask	Lesion-containing bounding box defined from the expert-reviewed lesion mask
Lung segmentation	Pretrained 3D U-Net developed using 95 manually annotated CT examinations; used to generate the UST lung mask	Not used to define the AST foreground
Mask use	The automatically generated lung mask was used to define the pulmonary foreground	The expert-reviewed lesion mask was used to define the lesion-containing foreground and bounding box
Context margin	Not applicable	10 voxels around the lesion foreground
Image orientation	RAS	RAS
HU clipping range	−1,000 to 500 HU	−1,000 to 500 HU
Intensity normalization	Performed after HU clipping	Performed after HU clipping
Resampled voxel spacing	0.8×0.8×1.0 mm	0.8×0.8×1.0 mm
Foreground cropping	Pulmonary foreground defined by the automatically generated lung mask	Lesion bounding box with a 10-voxel context margin
Model-specific batch-processing dimensions	224×224×128 voxels	160×160×80 voxels
Data augmentation	Random flipping, rotation, affine transformation, mild contrast adjustment, and Gaussian noise; training data only	Random flipping, rotation, affine transformation, mild contrast adjustment, and Gaussian noise; training data only
Network dimensionality	3D	3D
Network architecture	3D residual convolutional neural network with an initial convolutional stem, two residual blocks, subsequent 3D convolutional and pooling layers, fully connected layers, and a two-node Softmax output	3D residual convolutional neural network with an initial convolutional stem, two residual blocks, subsequent 3D convolutional and pooling layers, fully connected layers, and a two-node Softmax output
Residual blocks	ResBlock1 and ResBlock2	ResBlock1 and ResBlock2
Output layer	Two-node Softmax output for PA/BP classification	Two-node Softmax output for PA/BP classification
Class labels	PA =1; BP =0	PA =1; BP =0
Optimizer	AdamW	AdamW
Initial learning rate	1×10^{-4}	1×10^{-4}
Weight decay	1×10^{-5}	1×10^{-5}
Learning-rate scheduler	CosineAnnealingLR	CosineAnnealingLR
Minimum learning rate	1×10^{-6}	1×10^{-6}
Loss function	Weighted cross-entropy loss	Weighted cross-entropy loss
Training-set class distribution	283 PA cases and 225 BP cases	283 PA cases and 225 BP cases
Class-weight calculation	$w_c = N/(K \times n_c)$, based on training-set class frequencies	$w_c = N/(K \times n_c)$, based on training-set class frequencies
Class weights	PA: 0.898; BP: 1.129	PA: 0.898; BP: 1.129

Training batch size	4	4
Validation batch size	1	1
Maximum epochs	200	200
Early-stopping criterion	Mean validation loss	Mean validation loss
Early-stopping patience	15 consecutive epochs	15 consecutive epochs
Checkpoint-selection criterion	Highest mean validation accuracy	Highest mean validation accuracy
Prediction rule	Class corresponding to the larger of the two Softmax output probabilities	Class corresponding to the larger of the two Softmax output probabilities
Hyperparameter selection	Empirical selection based on prior experiments and training stability; no formal grid, random, or Bayesian search	Empirical selection based on prior experiments and training stability; no formal grid, random, or Bayesian search
Random seed	42	42
Deterministic setting	MONAI set_determinism(seed=42); deterministic operations enabled where supported	MONAI set_determinism(seed=42); deterministic operations enabled where supported
Software environment	Python 3.9.16; PyTorch 1.12.1; MONAI 1.0.1; SimpleITK 2.1.1.1; scikit-learn 1.1.2; matplotlib 3.5.2	Python 3.9.16; PyTorch 1.12.1; MONAI 1.0.1; SimpleITK 2.1.1.1; scikit-learn 1.1.2; matplotlib 3.5.2
CUDA version	11.7	11.7
Hardware	One NVIDIA Tesla V100 GPU	One NVIDIA Tesla V100 GPU

Model-specific batch-processing dimensions are reported as width × height × depth. Class weights were calculated from the training-set distribution of 283 PA cases and 225 BP cases using $w_c = N/(K \times n_c)$. Data augmentation was applied only during training. AST, annotated supervised training; BP, bacterial pneumonia; CT, computed tomography; HU, Hounsfield unit; PA, pulmonary aspergillosis; RAS, right-anterior-superior; UST, unannotated supervised training.

Table S2 Comparison of demographic and clinical characteristics between the primary-center development cohort and the external-validation cohort

Characteristic	Primary-center development cohort (n=635)	External-validation cohort (n=63)	P value
Demographic and disease characteristics			
Age, years	56.3±16.4	54.8±14.6	0.459
Weight, kg	55.0 [48.0–62.0] (n=529)	56.0 [51.0–66.0] (n=38)	0.225
Male sex	393/635 (61.9%)	32/63 (50.8%)	0.085
Current or former smoker	216/635 (34.0%)	18/63 (28.6%)	0.383
Disease composition			0.608
Pulmonary aspergillosis	354/635 (55.7%)	33/63 (52.4%)	
Bacterial pneumonia	281/635 (44.3%)	30/63 (47.6%)	
PA subtype distribution			0.161
IPA	115/354 (32.5%)	16/33 (48.5%)	
CPA	214/354 (60.5%)	16/33 (48.5%)	
ABPA	25/354 (7.1%)	1/33 (3.0%)	
Host immune status			
Clearly documented major immunocompromising condition	22/635 (3.5%)	4/63 (6.3%)	0.282

Data are presented as mean ± standard deviation, median [interquartile range], or n/N (%), as appropriate. Percentages for IPA, CPA, and ABPA were calculated among patients with PA in the corresponding cohort. Clearly documented major immunocompromising conditions included confirmed neutropenia, hematologic malignancy, hematopoietic stem-cell transplantation, solid-organ transplantation, or explicitly documented severe immunodeficiency. Suspected or unconfirmed diagnoses were not included. Chronic underlying diseases were reported separately and were not considered equivalent to immunocompromised status. Available-case denominators are shown for variables with missing data, and missing values were not considered negative results. ABPA, allergic bronchopulmonary aspergillosis; CPA, chronic pulmonary aspergillosis; IPA, invasive pulmonary aspergillosis; PA, pulmonary aspergillosis.

Table S3 Patient-level confusion matrices of the UST and AST models in the internal and external validation cohorts

Validation cohort	Model	Actual diagnosis	Predicted PA	Predicted BP	Total
Internal validation	UST	PA	65	6	71
		BP	31	25	56
	AST	PA	58	13	71
		BP	15	41	56
External validation	UST	PA	25	8	33
		BP	17	13	30
	AST	PA	20	13	33
		BP	4	26	30

Data are presented as numbers of patients. Pulmonary aspergillosis was designated as the positive class and bacterial pneumonia as the negative class. AST, annotated supervised training; BP, bacterial pneumonia; PA, pulmonary aspergillosis; UST, unannotated supervised training.

Table S4 Calibration and probability-based performance of the UST and AST models

Cohort	Model	Brier score (95% CI)	Calibration intercept	Calibration slope	Average precision (95% CI)
Internal validation	UST	0.198 (0.174–0.224)	−0.303	1.423	0.804 (0.730–0.879)
	AST	0.180 (0.132–0.230)	0.025	0.467	0.839 (0.757–0.916)
External validation	UST	0.257 (0.228–0.291)	0.056	0.218	0.575 (0.483–0.723)
	AST	0.195 (0.133–0.265)	0.561	0.590	0.827 (0.716–0.928)

Lower Brier scores indicate better overall probabilistic accuracy. The ideal calibration intercept and slope are 0 and 1, respectively. Higher average precision indicates better precision-recall performance. Ninety-five percent confidence intervals for the Brier score and average precision were estimated using 2,000 patient-level stratified bootstrap resamples. PA was designated as the positive class. AP, average precision; AST, annotated supervised training; CI, confidence interval; PA, pulmonary aspergillosis; UST, unannotated supervised training.

Table S5 Distribution and exploratory subtype-specific performance of the UST and AST models across pulmonary aspergillosis subtypes

Panel A. Distribution of pulmonary aspergillosis subtypes

Cohort	IPA, n	CPA, n	ABPA, n	Total PA, n
Training cohort	88	174	21	283
Internal validation cohort	27	40	4	71
Primary-center total	115	214	25	354
External validation cohort	16	16	1	33
Overall study population	131	230	26	387

Panel B. Exploratory subtype-versus-BP performance

Cohort	PA subtype	Model	Correctly identified/total	Sensitivity (95% CI)	Specificity (95% CI)	AUC (95% CI)
Internal validation	IPA	UST	24/27	88.9% (71.9–96.1%)	44.6% (32.4–57.6%)	0.73 (0.62–0.85)
	IPA	AST	21/27	77.8% (59.2–89.4%)	73.2% (60.4–83.0%)	0.80 (0.70–0.90)
	CPA	UST	38/40	95.0% (83.5–98.6%)	44.6% (32.4–57.6%)	0.83 (0.74–0.91)
	CPA	AST	34/40	85.0% (70.9–92.9%)	73.2% (60.4–83.0%)	0.86 (0.79–0.93)
	ABPA	UST	3/4	75.0% (30.1–95.4%)	44.6% (32.4–57.6%)	0.56 (0.22–0.90)
	ABPA	AST	3/4	75.0% (30.1–95.4%)	73.2% (60.4–83.0%)	0.78 (0.58–0.98)
External validation	IPA	UST	12/16	75.0% (50.5–89.8%)	43.3% (27.4–60.8%)	0.56 (0.38–0.74)
	IPA	AST	9/16	56.3% (33.2–76.9%)	86.7% (70.3–94.7%)	0.76 (0.61–0.92)
	CPA	UST	12/16	75.0% (50.5–89.8%)	43.3% (27.4–60.8%)	0.53 (0.35–0.71)
	CPA	AST	10/16	62.5% (38.6–81.5%)	86.7% (70.3–94.7%)	0.85 (0.74–0.96)
	ABPA	UST	1/1	1/1 (100.0%; descriptive only)	43.3% (27.4–60.8%)	Not estimated
	ABPA	AST	1/1	1/1 (100.0%; descriptive only)	86.7% (70.3–94.7%)	Not estimated

Each subtype-versus-BP analysis included patients with the specified PA subtype as the positive group and all patients with BP from the same validation cohort as the negative group. The models were not retrained for individual PA subtypes. Sensitivity and specificity were calculated using the prespecified probability threshold of 0.5, with 95% confidence intervals estimated using the Wilson score method. AUCs and corresponding 95% confidence intervals were calculated from continuous PA prediction probabilities using the DeLong method. Specificity was identical across subtype analyses for a given model and cohort because the same BP comparator group was used. The single external ABPA case was reported descriptively, and an AUC was not estimated. ABPA, allergic bronchopulmonary aspergillosis; AST, annotated supervised training; AUC, area under the receiver operating characteristic curve; BP, bacterial pneumonia; CI, confidence interval; CPA, chronic pulmonary aspergillosis; IPA, invasive pulmonary aspergillosis; PA, pulmonary aspergillosis; UST, unannotated supervised training.

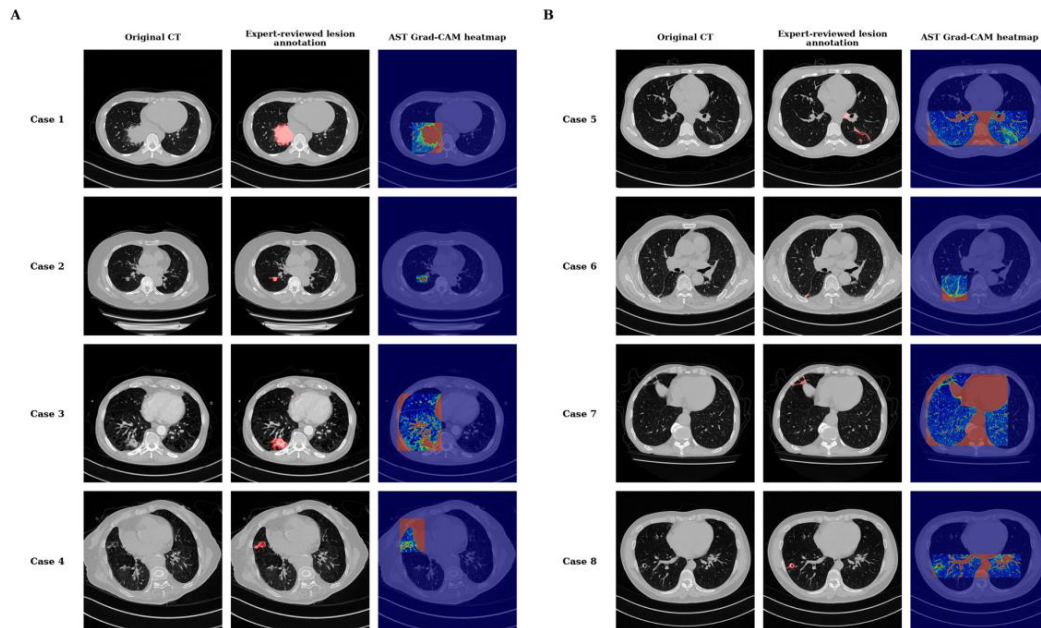


Figure S1 Representative Grad-CAM visualizations of correctly classified and misclassified cases generated by the AST model. Each row shows the original axial computed tomography image, the expert-reviewed lesion annotation, and the corresponding AST Grad-CAM heatmap from the same imaging level. Panel (A) shows four correctly classified cases: Cases 1 and 2 were patients with pulmonary aspergillosis correctly predicted as pulmonary aspergillosis, and Cases 3 and 4 were patients with bacterial pneumonia correctly predicted as bacterial pneumonia. Panel (B) shows four misclassified cases: Cases 5 and 6 were patients with pulmonary aspergillosis incorrectly predicted as bacterial pneumonia, representing false-negative pulmonary aspergillosis results, and Cases 7 and 8 were patients with bacterial pneumonia incorrectly predicted as pulmonary aspergillosis, representing false-positive pulmonary aspergillosis results. Grad-CAM provides a qualitative visualization of image regions contributing to individual model outputs and should not be interpreted as quantitative lesion segmentation, causal model reasoning, or evidence of clinical validity. AST, annotated supervised training; BP, bacterial pneumonia; CT, computed tomography; Grad-CAM, gradient-weighted class activation mapping; PA, pulmonary aspergillosis.

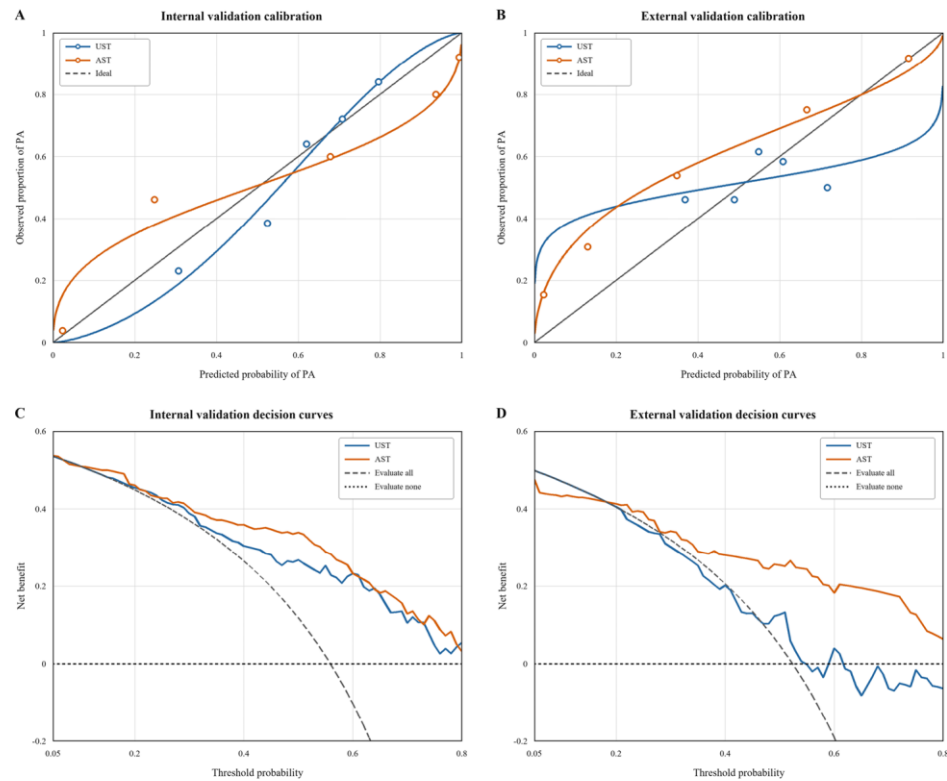


Figure S2 Calibration and decision curve analyses of the UST and AST models. Calibration plots are shown for the internal validation cohort (A) and external validation cohort (B). Grouped points represent the observed proportion of pulmonary aspergillosis within five approximately equal-sized groups of predicted probability, and the diagonal line represents ideal calibration. Decision curve analyses are shown for the internal validation cohort (C) and external validation cohort (D). Net benefit was evaluated over threshold probabilities from 0.05 to 0.80. A model-positive result was conceptualized as prompting additional diagnostic evaluation for pulmonary aspergillosis rather than directly initiating antifungal treatment. AST, annotated supervised training; PA, pulmonary aspergillosis; UST, unannotated supervised training.

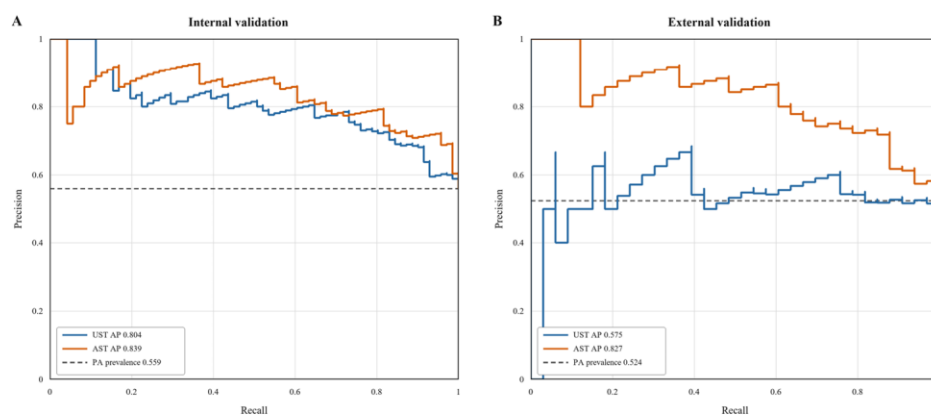


Figure S3 Precision-recall curves of the UST and AST models. Precision-recall curves are shown for the internal validation cohort (A) and external validation cohort (B), with pulmonary aspergillosis designated as the positive class. The horizontal reference lines indicate the prevalence of pulmonary aspergillosis in the corresponding cohort. Average precision was used as the summary measure of precision-recall performance. AP, average precision; AST, annotated supervised training; PA, pulmonary aspergillosis; UST, unannotated supervised training.

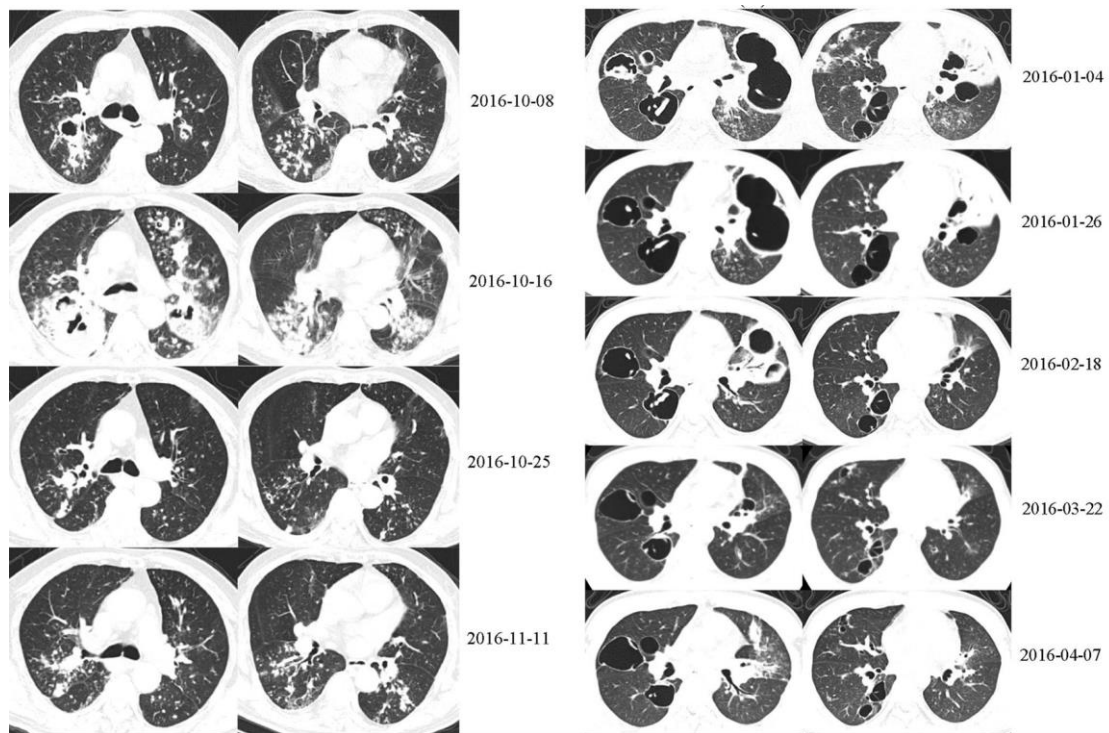


Figure S4 Serial CT images of pulmonary aspergillosis cases correctly identified by both models. Serial CT examinations from two representative patients with pulmonary aspergillosis are shown. All examinations were correctly classified by both the UST and AST models. AST, annotated supervised training; CT, computed tomography; PA, pulmonary aspergillosis; UST, unannotated supervised training.

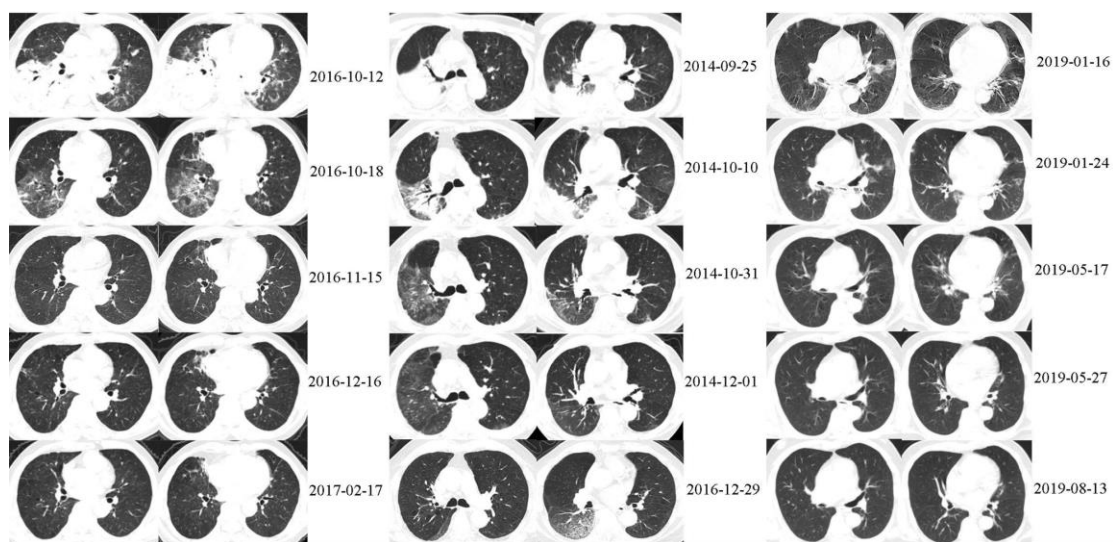


Figure S5 Serial CT images of bacterial pneumonia cases correctly identified by both models. Serial CT examinations from three representative patients with bacterial pneumonia are shown. All examinations were correctly classified by both the UST and AST models. AST, annotated supervised training; BP, bacterial pneumonia; CT, computed tomography; UST, unannotated supervised training.