

Differentiating Emphysema From Emphysema-Dominated COPD Patients with CT Imaging Feature and Machine Learning

Wanjin Guo¹, Mengqi Li¹, Ying Li², Xiaole Fan³, Lei Wu⁴

¹Department of Respiratory and Critical Care Medicine, Shanxi Provincial People's Hospital, Taiyuan, People's Republic of China; ²Department of Radiology, Shanxi Provincial People's Hospital, Taiyuan, People's Republic of China; ³Department of Information Management, Shanxi Provincial People's Hospital, Taiyuan, People's Republic of China; ⁴Department of Oncology, The Fifth Clinical Medical College of Shanxi Medical University, Taiyuan, People's Republic of China

Correspondence: Wanjin Guo, Email vipguowanjin@126.com

Background: Differentiating between emphysema and emphysema-dominant chronic obstructive pulmonary disease (COPD) remains challenging but crucial for appropriate management. Quantitative computed tomography (QCT) offers potential for improved characterization, yet its optimal application in conjunction with machine learning for this differentiation is not fully established.

Methods: This prospective study enrolled 476 participants (99 with emphysema, 377 with emphysema-dominant COPD) aged 34–88 years. All participants underwent spirometry and chest CT scans. QCT features including emphysema index, mean lung density, airway measurements, and vessel measurements were extracted. A random forest model was developed using these QCT features to differentiate between the two groups. The model's performance was assessed using area under the receiver operating characteristic curve (AUC-ROC). Correlations between QCT parameters and pulmonary function tests were analyzed.

Results: The model achieved an AUC-ROC of 0.97 (95% CI: 0.96–0.99) in differentiating emphysema from emphysema-dominant COPD. Emphysema index and airway wall thickness were the most important features for classification. QCT-derived emphysema index showed strong negative correlation with FEV1/FVC ($\rho = -0.54$, $p < 0.001$) in the emphysema-dominant COPD group, but no significant correlation in the emphysema group ($\rho = 0.001$, $p = 0.993$). Mean lung density was significantly lower in the emphysema-dominant COPD group compared to the isolated emphysema group ($p < 0.001$).

Conclusion: Machine learning analysis of QCT features can accurately differentiate emphysema from emphysema-dominant COPD. The differing relationships between QCT parameters and lung function in these two groups suggest distinct pathophysiological processes. These findings may contribute to improved diagnosis, phenotyping, and management strategies in emphysema and COPD.

Keywords: quantitative computed tomography, emphysema, emphysema-dominant COPD, chronic obstructive pulmonary disease, computed tomography

Background

Chronic Obstructive Pulmonary Disease (COPD) is a leading cause of morbidity and mortality worldwide, affecting an estimated 384 million individuals globally.¹ Characterized by persistent respiratory symptoms and airflow limitation, COPD encompasses a spectrum of pathological changes, including emphysema, small airway disease, and chronic bronchitis.^{2,3} Among these, emphysema, defined as the permanent enlargement of airspaces distal to terminal bronchioles, plays a crucial role in disease progression and patient outcomes.^{4,5}

The differentiation between emphysema and emphysema-dominant COPD represents a significant clinical challenge. While both conditions involve parenchymal destruction, emphysema-dominant COPD is characterized by more severe airflow limitation and often accompanies additional pathological changes in airways and pulmonary vasculature.^{6,7} Emphysema-dominant COPD is associated with faster lung function decline and higher mortality rates compared to emphysema alone.⁸ Patients with emphysema-dominant COPD may respond differently to bronchodilators and anti-

inflammatory therapies compared to those with predominant emphysema.⁹ Emphysema-dominant COPD is often accompanied by systemic manifestations and comorbidities, necessitating a more comprehensive management approach.¹⁰

Traditional diagnostic approaches, relying primarily on pulmonary function tests (PFTs) and visual assessment of computed tomography (CT) scans, are integral to the diagnosis and management of COPD. However, these methods have limitations in providing a comprehensive understanding of the disease. PFTs, while essential for diagnosing COPD and assessing airflow limitation, are unable to differentiate between specific phenotypes, such as emphysema, as they do not provide direct information about underlying structural changes or the heterogeneity of the disease.^{11,12} Visual CT assessment, on the other hand, is subject to inter-observer variability and may overlook subtle changes in lung parenchyma, airways, and vasculature.^{13–15}

Clinical guidelines, such as the Global Initiative for Chronic Obstructive Lung Disease (GOLD), recognize the importance of imaging in the assessment of COPD. Quantitative CT imaging, in particular, is recommended for evaluating structural abnormalities, such as emphysema and airway remodeling, that contribute to disease phenotyping and management.¹⁶ Similarly, the ATS/ERS joint statement emphasizes that imaging serves as a valuable adjunct to pulmonary function tests, providing insights into disease heterogeneity and severity.¹⁷ These recommendations underscore the need for advanced imaging-based methods, such as machine learning, to further enhance COPD diagnosis and phenotyping.

Recent advances in CT technology and image analysis techniques have opened new avenues for quantitative assessment of lung pathology. Quantitative CT (QCT) allows for objective measurement of lung density, airway dimensions, and vascular morphology, providing a comprehensive evaluation of COPD-related changes.^{18,19} Parameters such as the emphysema index (EI), mean lung density (MLD), airway wall thickness, and pulmonary vascular volumes have shown correlations with disease severity and progression.^{20–22} However, the complex interplay between these quantitative CT features and their relationship to pulmonary function remains incompletely understood. Moreover, the optimal combination of QCT parameters for differentiating emphysema from emphysema-dominant COPD has not been established.

The advent of machine learning techniques offers promising opportunities to leverage the rich data provided by QCT. Machine learning algorithms can identify complex patterns and relationships within high-dimensional data that may not be apparent through conventional statistical approaches.^{23,24} Previous studies have demonstrated the potential of machine learning in COPD for various applications, including disease detection, phenotyping, and outcome prediction.^{25–27} AI tools, such as GPT-4, have shown promise in aiding the interpretation of complex tests like CPET, improving diagnostic accuracy and patient stratification in COPD.²⁸

Despite these advancements, there remains a gap in our understanding of how machine learning can be optimally applied to QCT data for differentiating emphysema from emphysema-dominant COPD. Additionally, the relationship between QCT-derived features and pulmonary function across the spectrum of emphysema to emphysema-dominant COPD warrants further investigation. This study aims to: (1) Developing and validating a machine learning model based on QCT features to differentiate emphysema from emphysema-dominant COPD. (2) Investigating the relationships between QCT-derived parameters and pulmonary function tests. (3) Exploring the relative importance of various QCT features in characterizing the emphysema to emphysema-dominant COPD spectrum. By integrating advanced imaging analysis with machine learning, we seek to enhance our understanding of COPD pathophysiology and potentially improve diagnostic accuracy and phenotyping in clinical practice.

Materials and Methods

Participants

This prospective study enrolled 476 participants (99 emphysema cases; 377 emphysema-dominant COPD cases; mean age, 65 years; SD, 9.2 years; 381 men [80.0%]) (Table 1). Inclusion criteria for the Emphysema group were: post-bronchodilator Forced expiratory volume in 1 second (FEV1)/FVC ≥ 0.7 , The percentage ratio of low attenuation area to corresponding lung area (LAA%) $\geq 5\%$, and no history of chronic bronchitis. For the Emphysema-dominant COPD

Table 1 Clinical Characteristics, CT Parameters, and PFT Parameters of the Dataset

Characteristics	COPD	Emphysema
Number (%)	377	99
Age, yr, mean (SD)	66 (8.6)	58 (8.2)
Sex, male (%)	312 (82.8)	69 (69.7)
BMI, kg/m ² , mean (SD)	23.4 (4.1)	26.1 (5.5)
kVp, kV	113.28 (14.71)	114.25 (13.42)
Slice thickness, mm, mean (SD)	1.25 (0.18)	1.35 (0.14)
X-ray tube current (mA), mean ± SD	426.76 (237.41)	354.6 (178.54)
FEV1, mean (SD)	1.45 (0.72)	2.94 (0.65)
FVC, mean (SD)	2.68 (0.91)	3.65 (0.79)
FEV1/FVC, mean (SD)	0.52 (0.12)	0.81 (0.04)
FEV1%predicted, mean (SD)	56.37 (24.47)	105.90 (14.05)

group: post-bronchodilator FEV1/FVC <0.7, LAA% ≥10% of lung volume, and emphysema as the predominant feature on CT. Exclusion criteria included: active pulmonary infection, history of lung surgery or radiation therapy, diagnosis of asthma, bronchiectasis, or interstitial lung disease, unstable cardiovascular disease, pregnancy or breastfeeding, and inability to perform pulmonary function tests or CT scans.

Spirometry data were retrospectively extracted from patients' medical files. All patients included in the study had prior CT scans, and no additional scans were performed specifically for this study. Patients were selected based on the availability of spirometry and CT scans, as well as a confirmed diagnosis of emphysema-predominant COPD. A total of 47 patients were excluded due to not meeting the criteria for emphysema-predominant COPD.

The study was reviewed and approved by the Ethics Committee of Shanxi Provincial People's Hospital and conducted in accordance with the Declaration of Helsinki and Good Clinical Practice guidelines. Informed consent was obtained from all participants prior to the commencement of the study. Participants were fully informed about the purpose, procedures, potential risks, and benefits of the study before providing their written consent. All data were de-identified and securely stored with restricted access.

CT Acquisition and Quantitative Analysis

All participants underwent non-contrast high-resolution CT in supine position as recommended. The acquired CT datasets were subsequently transferred to a dedicated image analysis workstation for quantitative analysis. The nn-UNet algorithm was utilized for automatic lung lobe segmentation and applied a density threshold method (−950 HU) to calculate the emphysema index (EI) for each lobe.²⁹ Mean lung density (MLD) was obtained by averaging the CT values of all voxels within the lung parenchyma.

Airway analysis employed an advanced deep learning algorithm, measuring airways from the trachea to the 6th generation bronchi.³⁰ Extracted parameters included relative wall thickness (WT%), wall area percentage (WA%), and Pi10 (standardized airway wall thickness at an internal perimeter of 10 mm). Vascular analysis utilized a graph theory-based vessel segmentation algorithm to extract the 3D network of pulmonary vessels. Surface area, total blood volume (TBV), and the volume of small vessels (diameter <5mm) (BV5) along with its ratio to total blood vessel volume (BV5/TBV) were calculated. These parameters reflect the degree of pulmonary vascular remodeling.

All quantitative analysis results were independently reviewed by two experienced chest radiologists to ensure accuracy of segmentation and measurements. In cases of discrepancy, consensus was reached through discussion. Finally, all parameters were integrated into a structured database for subsequent statistical analysis and machine learning modeling.

Feature Selection and Machine-Learning

Feature selection was conducted using the Least Absolute Shrinkage and Selection Operator (LASSO) regression method. This technique was chosen for its ability to perform both variable selection and regularization, thereby

improving the prediction accuracy and interpretability of the resulting statistical model. The LASSO regression was implemented using the 'scikit-learn' package in Python (version 3.8, Python Software Foundation, Wilmington, DE, USA).

For machine learning model development, we employed a random forest algorithm, known for its robustness in handling high-dimensional data and its ability to capture complex, non-linear relationships between predictors. The random forest model was constructed using the 'scikit-learn' package in Python, with the number of trees set to 1000 and the number of variables randomly sampled as candidates at each split optimized through cross-validation. The dataset was randomly split into training (70%) and testing (30%) sets. The model was trained on the training set using 5-fold cross-validation to tune hyperparameters and prevent overfitting.

Statistical Analysis

Statistical analyses were performed using Python (version 3.8, Python Software Foundation, Wilmington, DE, USA) and SPSS (version 28.0, IBM Corp., Armonk, NY, USA). A two-tailed p value < 0.05 was considered statistically significant for all analyses.

For univariate analysis, comparisons between emphysema and emphysema-dominant COPD groups were conducted using Mann–Whitney U -tests. To account for multiple comparisons, p -values were adjusted using the Bonferroni correction method. The significance level was set at $\alpha = 0.05/n$, where n is the number of comparisons performed.

Correlations between QCT parameters and PFT variables were assessed using Spearman's rank correlation coefficient for non-normally distributed variables. The performance of the machine learning model was evaluated on the held-out test set using area under the receiver operating characteristic curve (AUC-ROC), accuracy, precision, recall, and F1-score. The performance of the machine learning model was compared to conventional statistical methods using DeLong's test for comparing AUC-ROC curves.

Results

Participant Demographics and Pulmonary Function

A total of 476 participants were included in the study, comprising 99 patients with emphysema and 377 with emphysema-dominant COPD (Table 1). The mean age of the cohort was 65 ± 9.2 years, with 80% male participants. Body Mass Index (BMI) was lower in the emphysema-dominant COPD group (23.4 ± 4.1 kg/m²) than in the emphysema group (26.1 ± 5.5 kg/m²).

Pulmonary function tests revealed significant differences between the groups. FEV1 was significantly lower in the emphysema-dominant COPD group ($56.37 \pm 24.47\%$ predicted) compared to the emphysema group ($105.90 \pm 14.05\%$ predicted; $p < 0.001$). The FEV1/FVC ratio was also significantly reduced in the emphysema-dominant COPD group (0.52 ± 0.12) compared to the emphysema group (0.81 ± 0.04 ; $p < 0.001$).

Analysis of Imaging Features

As shown in Table 2, quantitative CT analysis revealed distinct differences in imaging features between the two groups. The mean emphysema index (EI) was significantly higher in the emphysema-dominant COPD group compared to the emphysema group across all lung lobes. In the upper lobes, EI was $26.4 \pm 15.6\%$ versus $18.9 \pm 7.9\%$ ($p < 0.001$) in right upper lobes and $25.6 \pm 14.3\%$ versus $20.3 \pm 7.9\%$ ($p < 0.001$) in left upper lobes. In the lower lobes, EI was $21.3 \pm 13.2\%$ versus $15.8 \pm 7.4\%$ ($p < 0.001$). Mean lung density (MLD) was lower in the emphysema-dominant COPD group across all lobes. In the upper lobes, MLD was -848 ± 38 HU versus -815 ± 41 HU ($p < 0.001$) in right upper lobes and -850 ± 34 HU versus -826 ± 33 HU ($p < 0.001$) in left upper lobes. In the lower lobes, MLD was -824 ± 54 HU versus -803 ± 58 HU ($p < 0.001$) in right lower lobes and -850 ± 34 HU versus -797 ± 55 HU ($p < 0.001$) in left lower lobes.

Airway parameters showed significant differences, with the emphysema-dominant COPD group demonstrating increased mean values of airway wall thickness (WT) (1.38 ± 0.69 mm vs 0.98 ± 0.26 mm; $p < 0.001$) and decreased airway wall area percent (WA%) ($52.7 \pm 6.4\%$ vs $59.4 \pm 8.4\%$; $p < 0.001$). The Pi10 was also significantly higher in the emphysema-dominant COPD group (4.25 ± 1.81 mm vs 3.38 ± 0.69 mm; $p < 0.001$) (Table 2).

Table 2 Results for Quantitative Imaging Feature

Features		Emphysema	COPD	p-value*
EI%	Right upper lobe	18.93 (7.91)	26.40 (15.62)	<0.001
	Right middle lobe	21.11 (7.48)	24.81 (14.12)	0.057
	Right lower lobe	16.01 (7.69)	20.15 (13.12)	0.098
	Right lung	18.17 (7.47)	23.91 (13.38)	<0.001
	Left upper lobe	20.35 (7.86)	25.60 (14.29)	<0.001
	Left lower lobe	15.79 (7.43)	21.31 (13.22)	<0.001
	Left lung	18.32 (7.47)	23.82 (13.15)	<0.001
	Overall lung	18.25 (7.44)	23.88 (13.10)	<0.001
	Perc15	-945.38 (17.57)	-955.68 (22.48)	<0.001
MLD	Overall lung, HU	-814.45 (41.02)	-840.25 (38.21)	<0.001
	Right upper lobe, HU	-815.33 (40.58)	-847.82 (38.18)	<0.001
	Right middle lobe, HU	-834.07 (34.27)	-849.06 (37.73)	<0.001
	Right lower lobe, HU	-803.23 (57.57)	-823.97 (53.59)	<0.001
	Left upper lobe, HU	-826.01 (33.12)	-849.82 (34.14)	<0.001
	Left lower lobe, HU	-796.84 (54.58)	-824.10 (57.21)	<0.001
Airway	Pi10, mm	3.38 (0.69)	4.25 (1.81)	<0.001
	VVA%	0.59 (0.08)	0.53 (0.06)	<0.001
	WT, mm	0.98 (0.26)	1.38 (0.69)	<0.001
Vessel	Surface area, mm ²	4392.26 (1241.34)	4961.37 (1648.46)	0.058
	TBV, mm ³	241.51 (74.13)	287.16 (99.55)	<0.001
	BV5, mm ³	141.03 (39.31)	154.59 (54.86)	0.972
	BV5/TBV	0.59 (0.05)	0.55 (0.09)	<0.001

Note: *Mann-Whitney U-test.

Abbreviations: EI, emphysema; MLD, Mean lung density.

Vascular parameters revealed increased total pulmonary vascular volume in the emphysema-dominant COPD group ($287.2 \pm 99.6 \text{ cm}^3$ vs $231.5 \pm 74.1 \text{ cm}^3$; $p < 0.001$) (Table 2). The BV5/TBV ratio was significantly lower in the emphysema-dominant COPD group (0.55 ± 0.09 vs 0.59 ± 0.05 ; $p < 0.001$), indicating more severe small vessel loss.

Relationship of Imaging Features and Pulmonary Function

We performed a comprehensive analysis using Spearman's rank correlation to investigate the associations between quantitative CT imaging features and pulmonary function parameters in both emphysema and emphysema-dominant COPD groups (Tables 3 and 4, Figure 1). The EI exhibited a strong negative correlation with FEV₁ ($\rho = -0.39$, $p <$

Table 3 Correlation Coefficients (p-value) Between QCT Measurements and Lung Function Measurements in Emphysema Group

	FEV ₁	FVC	FEV1/FVC	FEV1%Predicted
EI of right upper lobe	0.092 (0.366)	0.075 (0.458)	0.032 (0.756)	0.197 (0.051)
EI of right middle lobe	0.053 (0.601)	0.046 (0.649)	0.001 (0.994)	0.181 (0.074)
EI of right lower lobe	0.076 (0.455)	0.065 (0.525)	-0.002 (0.988)	0.104 (0.307)
EI of right lung	0.077 (0.449)	0.063 (0.533)	0.015 (0.887)	0.149 (0.142)
EI of left upper lobe	0.102 (0.313)	0.090 (0.374)	0.011 (0.916)	0.169 (0.095)
EI of left lower lobe	0.100 (0.326)	0.092 (0.366)	-0.007 (0.948)	0.108 (0.285)
EI of left lung	0.092 (0.363)	0.082 (0.420)	0.002 (0.988)	0.132 (0.194)
Overall lung	0.083 (0.412)	0.073 (0.471)	0.001 (0.993)	0.142 (0.161)

(Continued)

Table 3 (Continued).

	FEV₁	FVC	FEV₁/FVC	FEV₁%Predicted
Perc15	-0.135 (0.181)	-0.131 (0.197)	0.018 (0.858)	-0.168 (0.096)
MLD of overall lung, HU	-0.242 (0.016)	-0.269 (0.007)	0.105 (0.303)	-0.209 (0.038)
MLD of right upper lobe, HU	-0.177 (0.080)	-0.211 (0.036)	0.117 (0.250)	-0.230 (0.022)
MLD of right middle lobe, HU	-0.166 (0.101)	-0.210 (0.037)	0.155 (0.126)	-0.185 (0.067)
MLD of right lower lobe, HU	-0.219 (0.030)	-0.236 (0.019)	0.078 (0.441)	-0.190 (0.060)
MLD of left upper lobe, HU	-0.275 (0.006)	-0.307 (0.002)	0.117 (0.250)	-0.186 (0.065)
MLD of left lower lobe, HU	-0.241 (0.016)	-0.261 (0.009)	0.064 (0.529)	-0.221 (0.028)
Pi10, mm	-0.044 (0.669)	-0.052 (0.612)	0.031 (0.764)	-0.067 (0.509)
WA%	-0.121 (0.231)	-0.156 (0.124)	0.133 (0.190)	-0.046 (0.650)
WT, mm	0.017 (0.870)	0.010 (0.922)	0.067 (0.507)	-0.128 (0.208)
Surface area, mm ²	0.494 (<0.001)	0.507 (<0.001)	-0.091 (0.371)	0.135 (0.181)
TBV, mm ³	0.486 (<0.001)	0.505 (<0.001)	-0.095 (0.350)	0.137 (0.178)
BV5, mm ³	0.473 (<0.001)	0.481 (<0.001)	-0.087 (0.394)	0.139 (0.169)
BV5/TBV	-0.251 (0.012)	-0.283 (0.005)	0.082 (0.418)	-0.012 (0.904)

Notes: a. Bold text indicates statistically significant results ($p < 0.05$).

Abbreviations: FEV₁, Forced Expiratory Volume in 1 second; FVC, Forced Vital Capacity; Perc15, The 15th percentile of lung; density MLD, Mean Lung Density, measured in Hounsfield Units (HU); Pi10, Square root of the wall area for a hypothetical airway with an internal perimeter of 10 mm; WA%, Airway Wall Area Percentage; WT, Airway Wall Thickness; Surface Area, Total surface area of the lungs; TBV, Total Blood Volume; BV5, Blood Volume in vessels smaller than 5 mm³; BV5/TBV, Ratio of blood volume in vessels smaller than 5 mm³ to Total Blood Volume.

Table 4 Correlation Coefficients (p-value) Between QCT Measurements and Lung Function Measurements in Emphysema-Dominated COPD Group

	FEV₁	FVC	FEV₁/FVC	FEV₁%Predicted
El of right upper lobe	-0.318 (<0.001)	-0.139 (0.007)	-0.475 (<0.001)	-0.418 (<0.001)
El of right middle lobe	-0.328 (<0.001)	-0.155 (0.003)	-0.461 (<0.001)	-0.379 (<0.001)
El of right lower lobe	-0.386 (<0.001)	-0.186 (<0.001)	-0.537 (<0.001)	-0.440 (<0.001)
El of right lung	-0.383 (<0.001)	-0.192 (<0.001)	-0.531 (<0.001)	-0.462 (<0.001)
El of left upper lobe	-0.308 (<0.001)	-0.129 (0.012)	-0.468 (<0.001)	-0.397 (<0.001)
El of left lower lobe	-0.395 (<0.001)	-0.181 (<0.001)	-0.563 (<0.001)	-0.470 (<0.001)
El of left lung	-0.376 (<0.001)	-0.176 (0.001)	-0.538 (<0.001)	-0.462 (<0.001)
Overall lung	-0.385 (<0.001)	-0.188 (<0.001)	-0.540 (<0.001)	-0.468 (<0.001)
Perc15	0.420 (<0.001)	0.196 (<0.001)	0.600 (<0.001)	0.521 (<0.001)
MLD of overall lung, HU	0.345 (<0.001)	0.124 (0.016)	0.553 (<0.001)	0.466 (<0.001)
MLD of right upper lobe, HU	0.285 (<0.001)	0.093 (0.073)	0.479 (<0.001)	0.406 (<0.001)
MLD of right middle lobe, HU	0.231 (<0.001)	0.062 (0.233)	0.399 (<0.001)	0.321 (<0.001)
MLD of right lower lobe, HU	0.342 (<0.001)	0.129 (0.012)	0.535 (<0.001)	0.428 (<0.001)
MLD of left upper lobe, HU	0.236 (<0.001)	0.042 (0.416)	0.453 (<0.001)	0.363 (<0.001)
MLD of left lower lobe, HU	0.336 (<0.001)	0.109 (0.035)	0.555 (<0.001)	0.454 (<0.001)
Pi10, mm	-0.140 (0.006)	-0.251 (<0.001)	0.080 (0.120)	0.000 (0.999)
WA%	-0.122 (0.018)	-0.198 (<0.001)	0.040 (0.435)	-0.039 (0.454)
WT, mm	-0.116 (0.025)	-0.234 (<0.001)	0.106 (0.040)	0.011 (0.839)

(Continued)

Table 4 (Continued).

	FEV ₁	FVC	FEV ₁ /FVC	FEV ₁ %Predicted
Surface area, mm ²	0.002 (0.967)	0.229 (<0.001)	-0.344 (<0.001)	-0.225 (<0.001)
TBV, mm ³	-0.019 (0.717)	0.205 (<0.001)	-0.356 (<0.001)	-0.249 (<0.001)
BV5, mm ³	0.010 (0.853)	0.223 (<0.001)	-0.315 (<0.001)	-0.190 (<0.001)
BV5/TBV	0.066 (0.200)	0.018 (0.725)	0.124 (0.016)	0.158 (0.002)

Abbreviations: FEV₁, Forced Expiratory Volume in 1 second; FVC, Forced Vital Capacity; Perc15, The 15th percentile of lung density; MLD, Mean Lung Density, measured in Hounsfield Units (HU); Pi10, Square root of the wall area for a hypothetical airway with an internal perimeter of 10 mm; WA%, Airway Wall Area Percentage; WT, Airway Wall Thickness; Surface Area, Total surface area of the lungs; TBV, Total Blood Volume; BV5, Blood Volume in vessels smaller than 5 mm³; BV5/TBV, Ratio of blood volume in vessels smaller than 5 mm³ to Total Blood Volume.

0.001), FVC ($\rho = -0.19$, $p < 0.001$), FEV₁ % predicted ($\rho = -0.47$, $p < 0.001$), and the FEV₁/FVC ratio ($\rho = -0.54$, $p < 0.001$) in the emphysema-dominant COPD group (Table 4), indicating a stronger relationship between structural lung changes and airflow limitation in more advanced disease states.

Additionally, MLD demonstrated a positive correlation with FEV₁ ($\rho = 0.34$, $p < 0.001$), FVC ($\rho = 0.12$, $p < 0.05$), FEV₁ % predicted ($\rho = 0.47$, $p < 0.001$), and the FEV₁/FVC ratio ($\rho = 0.55$, $p < 0.001$) in the emphysema-dominant COPD group (Table 4). These positive correlations were stronger in the lower lobes ($\rho = 0.54$, $p < 0.001$) than in the upper lobes ($\rho = 0.46$, $p < 0.001$), suggesting potential regional differences in the structure-function relationship within the lungs. Notably, MLD demonstrated a negative correlation with FEV₁ ($\rho = -0.24$, $p < 0.05$), FVC ($\rho = -0.27$, $p < 0.05$), and FEV₁ % predicted ($\rho = -0.21$, $p < 0.05$) in the emphysema group (Table 3).

Airway parameters exhibited varying degrees of correlation with pulmonary function in the emphysema-dominant COPD group, as shown in Table 4. Specifically, the WA% of segmental bronchi was negatively correlated with FEV₁ ($\rho = -0.12$, $p < 0.05$) and FVC ($\rho = -0.20$, $p < 0.001$). WT demonstrated a moderate negative correlation with FEV₁ ($\rho = -0.12$, $p < 0.05$) and positive correlation with the FEV₁/FVC ratio ($\rho = 0.11$, $p < 0.05$). Among the airway parameters, Pi10 showed the strongest negative correlation with FVC ($\rho = -0.25$, $p < 0.001$).

Vascular parameters also displayed significant associations with pulmonary function. In the emphysema-dominant COPD group (Table 4), the BV5/TBV ratio positively correlated with FEV₁ % predicted ($\rho = 0.16$, $p < 0.05$) and the

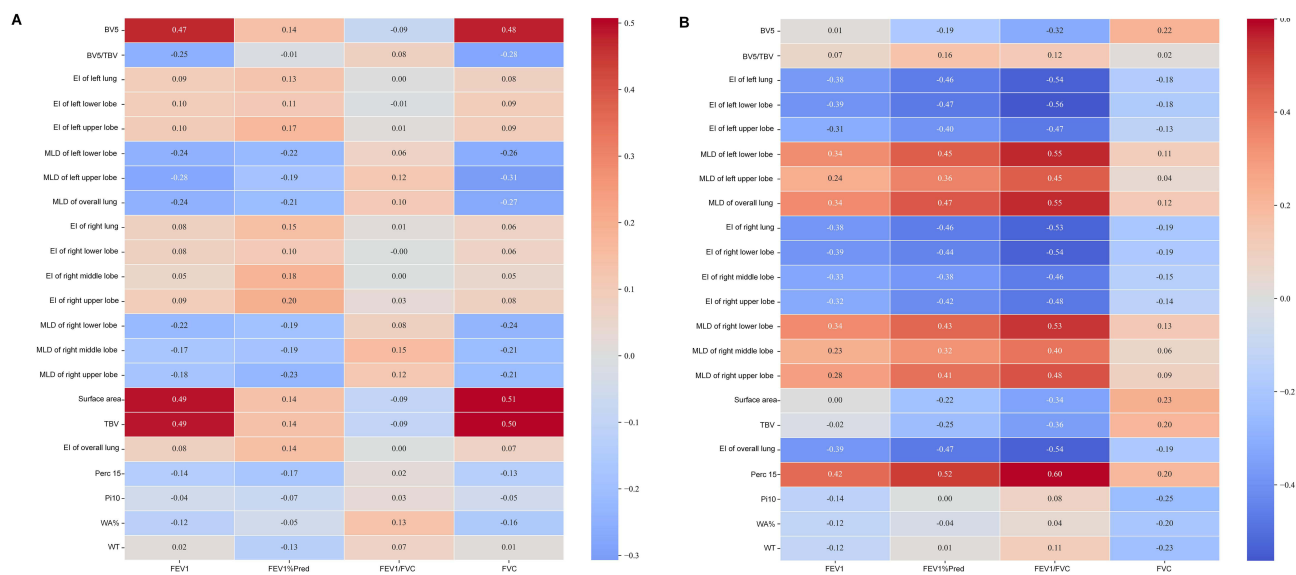


Figure 1 Correlation coefficients between QCT measurements and lung function measurements in (A) the emphysema group and (B) the emphysema-dominant COPD group. The figure highlights the relationships between quantitative imaging features (eg, low attenuation volume, mean lung density) and physiological measures (eg, FEV₁, FEV₁/FVC ratio). Strong positive or negative correlations suggest potential imaging biomarkers for disease assessment.

FEV₁/FVC ratio ($\rho = 0.12$, $p < 0.05$). Additionally, total pulmonary vascular volume exhibited a strong positive correlation with FVC ($\rho = 0.20$, $p < 0.001$) and a negative correlation with FEV₁ % predicted ($\rho = -0.25$, $p < 0.001$), and the FEV₁/FVC ratio ($\rho = -0.36$, $p < 0.001$). In the emphysema group, the TBV shows strongly positive correlation with FEV₁ ($\rho = 0.49$, $p < 0.001$) and FVC ($\rho = 0.50$, $p < 0.001$). Otherwise, the BV5/TBV ratio negatively correlated with FEV₁ ($\rho = -0.25$, $p = 0.012$) and FVC ($\rho = -0.28$, $p = 0.004$).

Machine-Learning Classification

The random forest model, trained on the selected features from the LASSO regression, demonstrated excellent performance in distinguishing between emphysema and emphysema-dominant COPD patients. The model's performance was evaluated using five-fold cross-validation, where each fold contained approximately 20% of the dataset as the validation set. A confusion matrix summarizing the number of correct and incorrect predictions, including the classification of emphysema without COPD in the validation cohort, is presented in Figure 2. As described in Figure 3, the model achieved an AUC-ROC of 0.97 (95% CI: 0.96–0.99) on the held-out test set. The optimal cut-off point, determined by the Youden index, yielded an accuracy of 0.91, precision of 0.89, recall of 0.82, and F1-score of 0.85 (Table 5).

As shown in Figures 4 and 5, using SHAP values for feature importance analysis, we identified the top five CT-derived parameters impacting model predictions. The features of WA%, WT, and Pi10 were crucial in shaping the model's output, with higher SHAP values indicating substantial contributions to prediction changes, while the EI of the left lower lobe and left lung also played important roles in prediction decisions.

The model outperformed traditional logistic regression (AUC-ROC: 0.85, 95% CI: 0.80–0.90) in classifying emphysema and emphysema-dominant COPD patients (DeLong's test, $p = 0.003$). To assess the model's stability, we performed bootstrap resampling with 1000 iterations. The bootstrapped 95% CI for AUC-ROC was 0.96–0.99, demonstrating the model's robustness.

A decision curve analysis in Figure 6 showed that the Random Forest model provided a higher net benefit compared to the "treat all" or "treat none" strategies across a wide range of threshold probabilities (10–90%), indicating potential clinical utility.

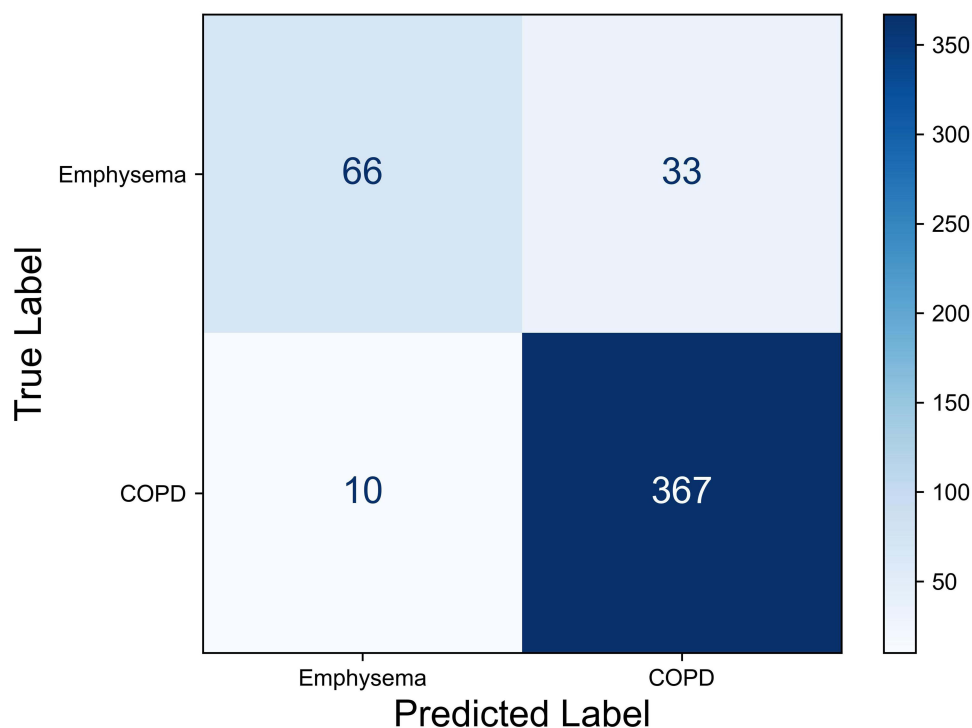


Figure 2 Confusion matrix of the random forest model for classifying emphysema and emphysema-dominant COPD. The matrix displays the number of correctly and incorrectly classified cases, with accuracy, sensitivity, and specificity highlighted.

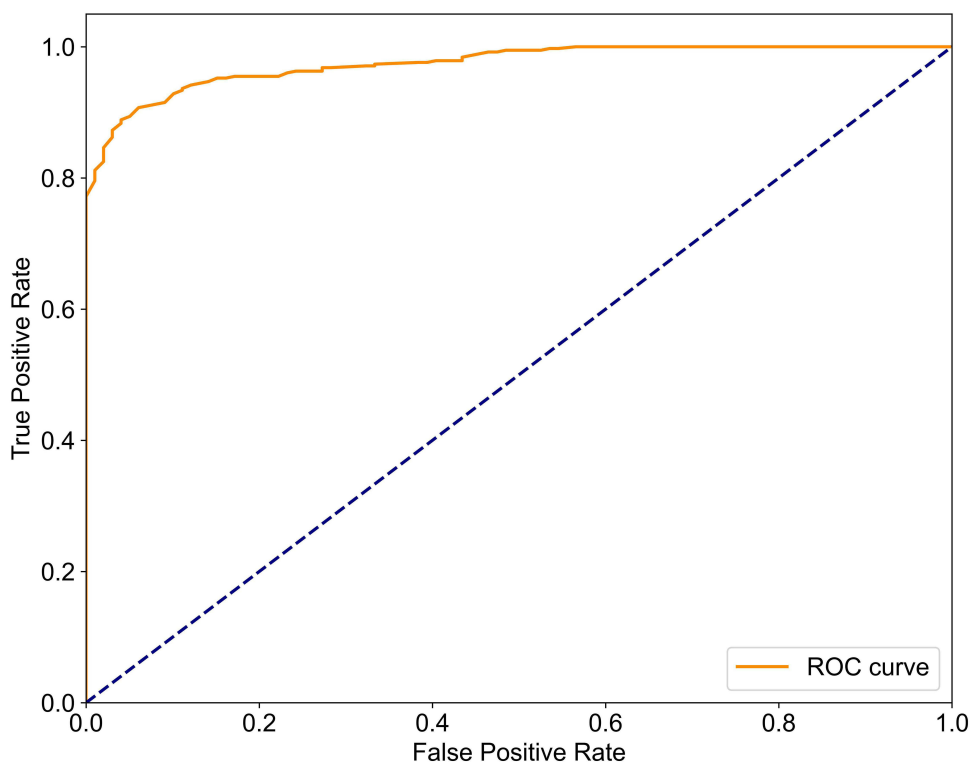


Figure 3 ROC curve of the random forest model in classifying emphysema and emphysema-dominant COPD, with an area under the curve (AUC) of 0.97 (95% CI: [0.96, 0.99]). The ROC curve demonstrates the model's excellent diagnostic accuracy, with a high true positive rate and low false positive rate across different thresholds.

Discussion

This study demonstrates the potential of quantitative CT analysis and machine learning in differentiating emphysema from emphysema-dominant COPD and elucidating the complex relationships between imaging features and pulmonary function. Our findings have important implications for disease characterization, progression monitoring, and personalized treatment strategies.

Machine Learning Classification

The high accuracy (91%) and AUC-ROC (0.97) of our random forest model in distinguishing emphysema from emphysema-dominant COPD highlight the power of combining quantitative CT features with advanced machine learning techniques. This performance surpasses previous studies using conventional CT visual scoring or individual quantitative parameters.^{1,2} The superior discriminative ability of our model compared to traditional logistic regression underscores the advantage of machine learning in capturing complex, non-linear relationships within high-dimensional data.

The identification of WA%, WT, Pi10, EI of left lower lobe, and EI of left lung as the top discriminative features aligns with the current understanding of COPD pathophysiology. These findings support the notion that COPD is a heterogeneous disease involving parenchymal destruction, airway remodeling, and vascular alterations.^{2,31}

Table 5 The Performance of Classification Using Random Forest

Model	Accuracy	Precision	Recall	F1-score	AUC
Random forest	0.91	0.89	0.82	0.85	0.97

Abbreviation: AUC, Area Under the Curve.

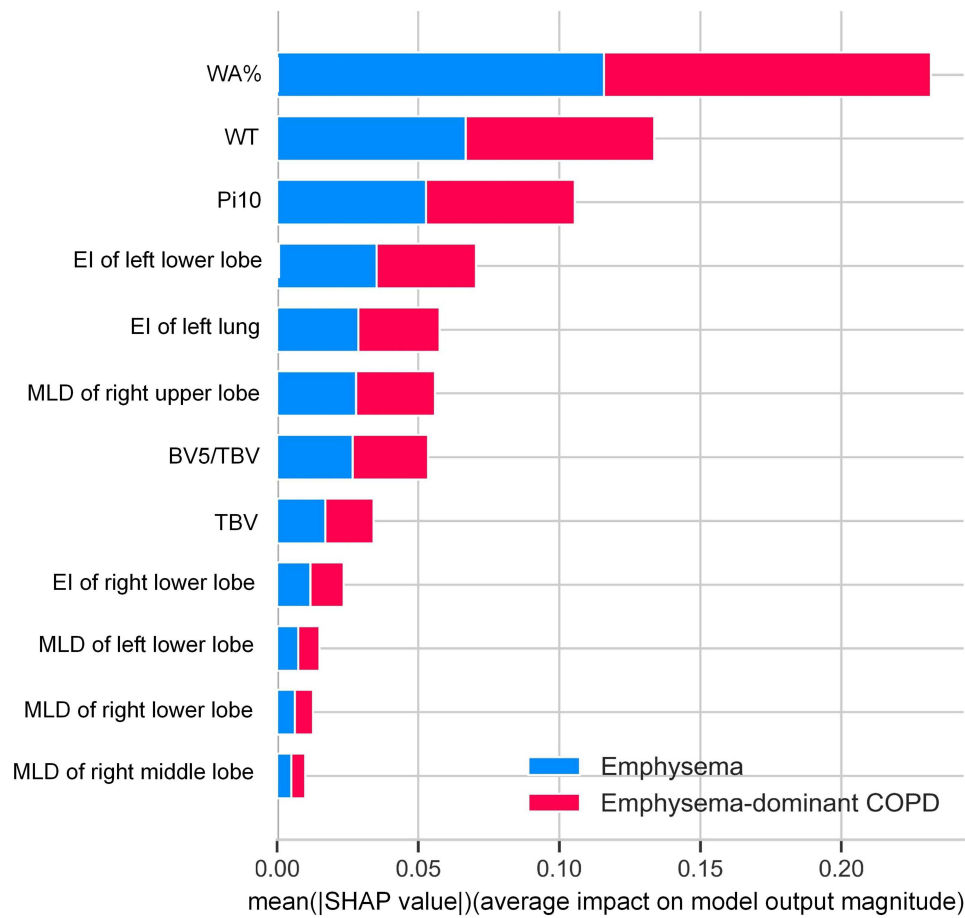


Figure 4 SHAP bar plot showing the average feature impact on the random forest model's classification of emphysema and emphysema-dominant COPD. Features such as WA% and airway wall thickness are shown to have the highest contribution to the classification decision, providing insights into the most influential imaging parameters.

Imaging-Function Relationships

The stronger correlations observed between quantitative CT parameters and pulmonary function tests provide insights into structure-function relationships in COPD compared with emphysema patients. The correlation coefficients in [Table 4](#) are generally stronger with more significant p-values. This suggests that in emphysema-dominated COPD patients, CT parameters are more closely linked to lung function. Our findings reveal that both the emphysema index and mean lung density are strongly correlated with pulmonary function parameters such as FEV1 and FVC in the emphysema-dominated COPD group. This underscores the significant influence of structural changes on lung function, aligning with previous studies that emphasize the role of emphysema in reducing airflow and lung capacity.^{32,33} While, in the emphysema group, EI showed weak correlations with FEV1, FVC, and other respiratory measures, with no statistically significant associations. MLD exhibited moderate, significant negative correlations with FEV1 and FVC, indicating that as emphysema worsens (ie, lower density), lung function declines. This reinforces the role of MLD as a potential marker for disease severity.

Compared to [Table 3](#), vascular parameters (eg, Surface area, TBV, BV5) in [Table 4](#) show weaker correlations with FEV1 but maintain positive correlations with FVC. This might reflect specific vascular remodeling patterns in emphysema patients. In [Table 4](#), airway wall parameters (eg, Pi10, WA%, WT) show relatively weak correlations with lung function. This aligns with the characteristics of emphysema-dominant COPD, where parenchymal destruction may be more significant than airway changes.

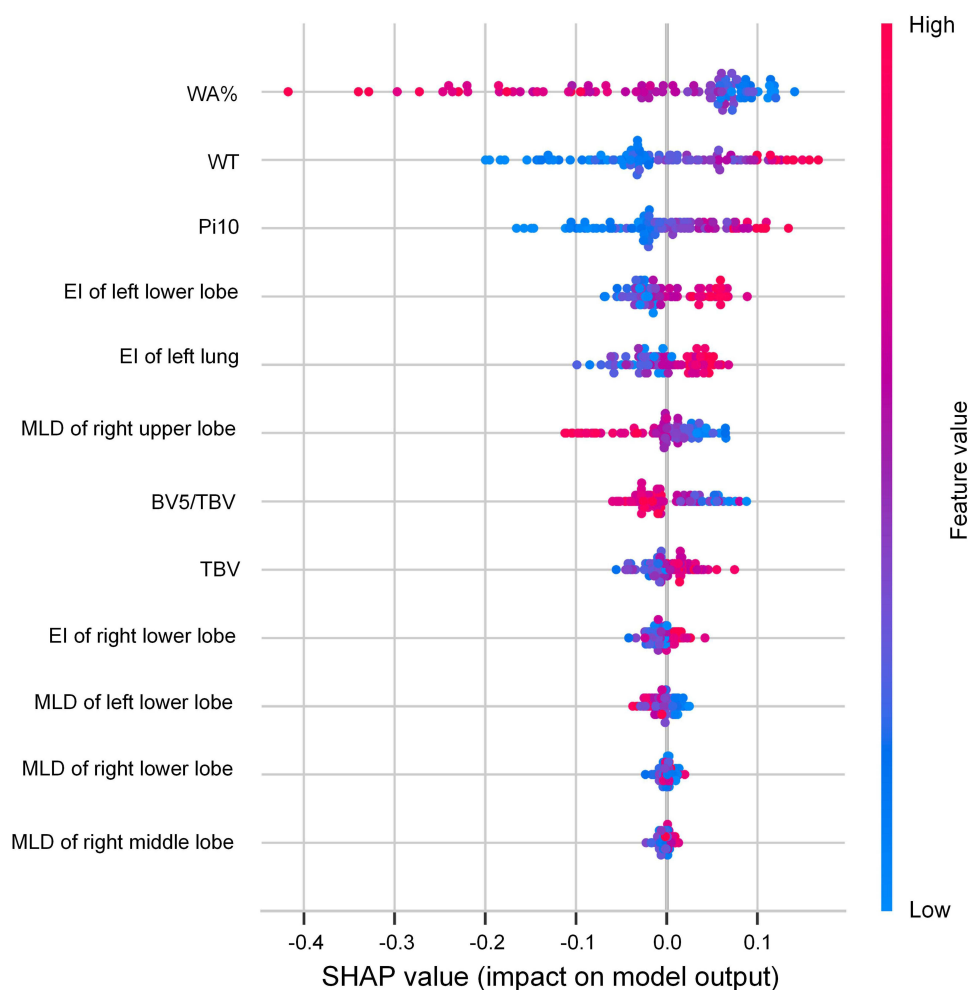


Figure 5 SHAP value plot illustrating individual feature contributions for specific cases in classifying emphysema and emphysema-dominant COPD. The plot highlights how key features, such as WA%, airway wall thickness, and Pi10, influence the model's predictions for individual patients, offering a more patient-specific interpretability.

Clinical Implications

Our machine learning model's ability to differentiate emphysema from emphysema-dominant COPD with high accuracy highlights its potential in early identification of COPD patients. While our findings are specific to emphysema-predominant COPD, the model could be applied in broader contexts, such as lung cancer screening programs, to identify undiagnosed COPD patients. Early detection in such settings could enable timely intervention and management, improving patient outcomes.^{34–36} The comprehensive quantitative CT analysis used in this study provides a multi-dimensional assessment of lung pathology, offering a more nuanced evaluation of disease severity and progression than pulmonary function tests alone.^{37,38} This approach could also be valuable in clinical trials by facilitating patient stratification and serving as an outcome measure for novel therapies targeting specific aspects of COPD pathology. By focusing on the model's ability to aid in diagnosis and phenotyping, our work lays the groundwork for integrating advanced imaging-based machine learning tools into clinical workflows, particularly for early detection and personalized management of respiratory diseases.

Limitations and Future Work

Despite the strengths of our study, several limitations should be acknowledged. First, the cross-sectional design precludes assessment of temporal changes in CT imaging features and their relationship with disease progression. Longitudinal studies are essential to evaluate how imaging features evolve over time and to assess the model's performance in

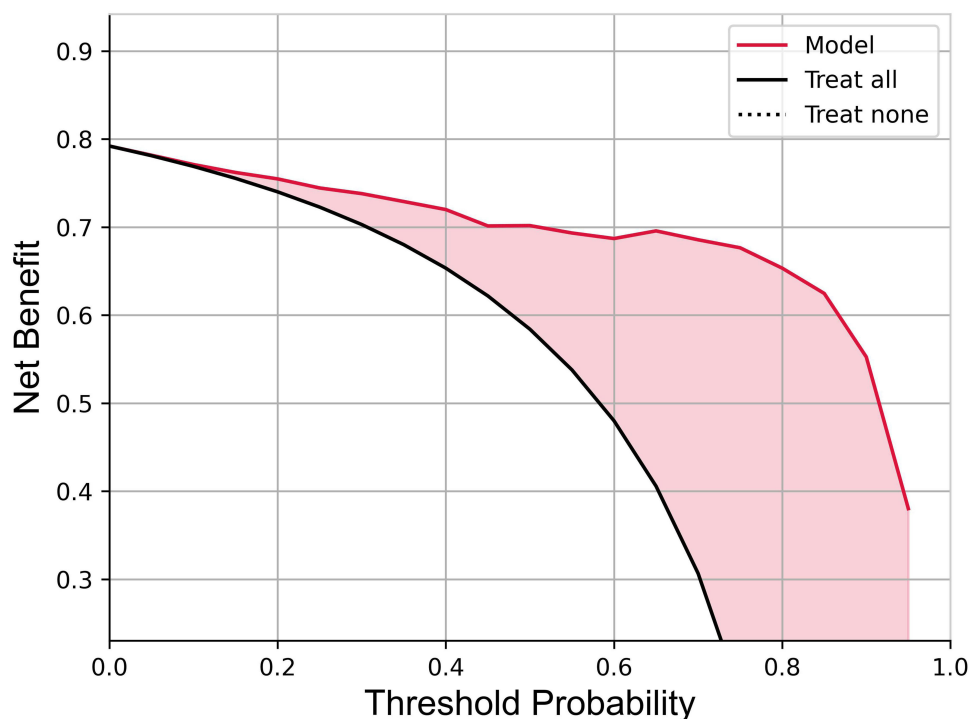


Figure 6 Decision curve analysis of the random forest model evaluating the clinical net benefit of classifying emphysema and emphysema-dominant COPD. The curve compares the net benefit of the model with default strategies (treat-all vs treat-none) across a range of threshold probabilities, demonstrating its potential utility in clinical decision-making.

predicting clinical outcomes, such as exacerbations and mortality. These studies would provide a more detailed understanding of disease dynamics and improve the clinical applicability of our findings. Second, while our sample size was adequate for the primary analysis, larger, multi-center studies are required to validate the generalizability of our findings across diverse populations, imaging protocols, and CT scanners. Third, the binary classification of emphysema versus emphysema-dominant COPD may oversimplify the spectrum of disease. Future research should explore multi-class classification approaches to capture the full range of COPD phenotypes, including airway-predominant and mixed phenotypes. Lastly, integrating genetic and molecular data with imaging features could provide a more comprehensive understanding of disease mechanisms. This multimodal approach may enhance classification performance and offer insights into personalized management strategies for COPD.

In conclusion, our study demonstrates the power of combining quantitative CT analysis with machine learning in differentiating emphysema from emphysema-dominant COPD and unraveling complex structure-function relationships. Future studies should focus on prospective designs to validate the model's predictive performance and evaluate its utility in monitoring disease progression. Additionally, integrating genetic and molecular biomarkers with imaging features may further improve diagnostic accuracy and provide a more comprehensive understanding of COPD pathogenesis, paving the way for novel therapeutic strategies.

Abbreviations

COPD, chronic obstructive pulmonary disease; QCT, Quantitative computed tomography; FEV1, forced expiratory volume in one second; FVC, forced vital capacity; EI, emphysema index; MLD, mean lung density; LAA, low attenuation area; BMI, body mass index; HU, Hounsfield Unit; WT, wall thickness; WA%, wall area percentage; Pi10, standardized airway wall thickness at an internal perimeter of 10 mm; TBV, total blood volume; BV5, total volume of vessels with a cross-sectional area smaller than 5 mm².

Data Sharing Statement

The data will be supplied upon request to corresponding authors.

Author Contributions

All authors made a significant contribution to the work reported, whether that is in the conception, study design, execution, acquisition of data, analysis and interpretation, or in all these areas; took part in drafting, revising or critically reviewing the article; gave final approval of the version to be published; have agreed on the journal to which the article has been submitted; and agree to be accountable for all aspects of the work.

Funding

There is no funding to report.

Disclosure

The authors declare no competing interests.

References

1. Adeloje 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(5):447–458. doi:10.1016/S2213-2600(21)00511-7
2. Brandsma CA, Van den Berge M, Hackett TL, Brusselle G, Timens W. Recent advances in chronic obstructive pulmonary disease pathogenesis: from disease mechanisms to precision medicine. *J Pathol.* 2020;250(5):624–635. doi:10.1002/path.5364
3. Ritchie AI, Wedzicha JA. Definition, causes, pathogenesis, and consequences of chronic obstructive pulmonary disease exacerbations. *Clinics Chest Med.* 2020;41(3):421–438. doi:10.1016/j.ccm.2020.06.007
4. Bodduluri S, Reinhardt JM, Hoffman EA, Newell JD, Bhatt SP. Recent advances in computed tomography imaging in chronic obstructive pulmonary disease. *Ann Am Thoracic Soc.* 2018;15(3):281–289. doi:10.1513/AnnalsATS.201705-377FR
5. Bhatt SP, Washko GR, Hoffman EA, et al. Imaging advances in chronic obstructive pulmonary disease: insights from COPDGene. *Am J Respir Crit Care Med.* 2018;199(3):286–301. doi:10.1164/rccm.201807-1351SO
6. Walters EH, Shukla SD, Mahmood MQ, Ward C. Fully integrating pathophysiological insights in COPD: an updated working disease model to broaden therapeutic vision. *Eur Respir Rev.* 2021;30(160):200364. doi:10.1183/16000617.0364-2020
7. Lange P, Ahmed E, Lahmar ZM, Martinez FJ, Bourdin A. Natural history and mechanisms of COPD. *Respirology.* 2021;26(4):298–321. doi:10.1111/resp.14007
8. Han MK, Agusti A, Celli BR, et al. From GOLD 0 to pre-COPD. *Am J Respir Crit Care Med.* 2021;203(4):414–423. doi:10.1164/rccm.202008-3328PP
9. Segal LN, Martinez FJ. Chronic obstructive pulmonary disease subpopulations and phenotyping. *J Allergy Clin Immunol.* 2018;141(6):1961–1971. doi:10.1016/j.jaci.2018.02.035
10. Cazzola M, Rogliani P, Puxeddu E, Ora J, Matera MG. An overview of the current management of chronic obstructive pulmonary disease: can we go beyond the GOLD recommendations? *Expert Rev Respir Med.* 2018;12(1):43–54. doi:10.1080/17476348.2018.1398086
11. Neder JA, de-Torres JP, Milne KM, O'Donnell DE. Lung function testing in chronic obstructive pulmonary disease. *Clinics Chest Med.* 2020;41(3):347–366. doi:10.1016/j.ccm.2020.06.004
12. Petousi N, Talbot NP, Pavord I, Robbins PA. Measuring lung function in airways diseases: current and emerging techniques. *Thorax.* 2019;74(8):797–805. doi:10.1136/thoraxjnl-2018-212441
13. Amaza IP, O'Shea AM, Fortis S, Comellas AP. Discordant quantitative and visual CT assessments in the diagnosis of emphysema. *Int J Chronic Obstr.* 2021;16:1231–1242. doi:10.2147/COPD.S284477
14. Kim SS, Seo JB, Lee HY, et al. Chronic obstructive pulmonary disease: lobe-based visual assessment of volumetric CT by using standard images—comparison with quantitative CT and pulmonary function test in the COPDGene study. *Radiology.* 2013;266(2):626–635. doi:10.1148/radiol.12120385
15. Lynch DA, Moore CM, Wilson C, et al. CT-based visual classification of emphysema: association with mortality in the COPDGene study. *Radiology.* 2018;288(3):859–866. doi:10.1148/radiol.2018172294
16. Agustí A, Celli BR, Criner GJ, et al. Global initiative for chronic obstructive lung disease 2023 report: GOLD executive summary. *Am J Respir Crit Care Med.* 2023;207(7):819–837. doi:10.1164/rccm.202301-0106PP
17. Celli BR, MacNee W, Agusti A, et al. Standards for the diagnosis and treatment of patients with COPD: a summary of the ATS/ERS position paper. *Eur Respir J.* 2004;23(6):932–946. doi:10.1183/09031936.04.00014304
18. Motahari A, Barr RG, Han MK, et al. Repeatability of pulmonary quantitative computed tomography measurements in chronic obstructive pulmonary disease. *Am J Respir Crit Care Med.* 2023;208(6):657–665. doi:10.1164/rccm.202209-1698PP
19. Koo MC, Tan WC, Hogg JC, et al. Quantitative computed tomography and visual emphysema scores: association with lung function decline. *ERJ Open Res.* 2023;9(2):00523–2022. doi:10.1183/23120541.00523-2022
20. Chaudhary MF, Hoffman EA, Guo J, et al. Predicting severe chronic obstructive pulmonary disease exacerbations using quantitative CT: a retrospective model development and external validation study. *Lancet Digital Health.* 2023;5(2):e83–e92. doi:10.1016/S2589-7500(22)00232-1

21. Baraghoshi D, Strand M, Humphries SM, et al. Quantitative CT evaluation of emphysema progression over 10 years in the COPDGene study. *Radiology*. 2023;307(4):e222786. doi:10.1148/radiol.222786
22. Kirby M, Smith BM. Quantitative CT scan imaging of the airways for diagnosis and management of lung disease. *Chest*. 2023;164(5):1150–1158. doi:10.1016/j.chest.2023.02.044
23. Jordan MI, Mitchell TM. Machine learning: trends, perspectives, and prospects. *Science*. 2015;349(6245):255–260. doi:10.1126/science.aaa8415
24. Wu Y, Xia S, Liang Z, Chen R, Qi S. Artificial intelligence in COPD CT images: identification, staging, and quantitation. *Respir Res*. 2024;25(1):319. doi:10.1186/s12931-024-02913-z
25. Zhao M, Wu Y, Li Y, et al. Learning and depicting lobe-based radiomics feature for COPD severity staging in low-dose CT images. *BMC Pulm Med*. 2024;24(1):294. doi:10.1186/s12890-024-03109-3
26. Castaldi J, Boueiz A, Yun J, et al. Machine learning characterization of COPD subtypes: insights from the COPDGene study. *Chest*. 2020;157(5):1147–1157. doi:10.1016/j.chest.2019.11.039
27. Verstraete K, Gyselink I, Huts H, et al. Estimating individual treatment effects on COPD exacerbations by causal machine learning on randomised controlled trials. *Thorax*. 2023;78(10):983–989. doi:10.1136/thorax-2022-219382
28. Kleinhendler E, Pinkhasov A, Hayek S, et al. Interpretation of cardiopulmonary exercise test by GPT—promising tool as a first step to identify normal results. *Expert Rev Respiratory Med*. 2025;19(4):371–378. doi:10.1080/17476348.2025.2474138
29. Pang H, Wu Y, Qi S, et al. A fully automatic segmentation pipeline of pulmonary lobes before and after lobectomy from computed tomography images. *Comput Biol Med*. 2022;147:105792. doi:10.1016/j.compbiomed.2022.105792
30. Wu Y, Zhao S, Qi S, et al. Two-stage contextual transformer-based convolutional neural network for airway extraction from CT images. *Artif Intell Med*. 2023;143:102637. doi:10.1016/j.artmed.2023.102637
31. Polosukhin VV, Gutor SS, Du R-H, et al. Small airway determinants of airflow limitation in chronic obstructive pulmonary disease. *Thorax*. 2021;76(11):1079–1088. doi:10.1136/thoraxjnl-2020-216037
32. Hoesein FAM, de Hoop B, Zanen P, et al. CT-quantified emphysema in male heavy smokers: association with lung function decline. *Thorax*. 2011;66(9):782–787. doi:10.1136/thx.2010.145995
33. Han MK, Agusti A, Calverley PM, et al. Chronic obstructive pulmonary disease phenotypes: the future of COPD. *Am J Respir Crit Care Med*. 2010;182(5):598–604. doi:10.1164/rccm.200912-1843CC
34. Young RP, Hopkins RJ. Chronic obstructive pulmonary disease (COPD) and lung cancer screening. *Transl Lung Cancer Res*. 2018;7(3):347. doi:10.21037/tlcr.2018.05.04
35. Mulshine JL, Aldigé CR, Ambrose LF, et al. Emphysema detection in the course of lung cancer screening: optimizing a rare opportunity to impact population health. *Ann Am Thoracic Soc*. 2023;20(4):499–503. doi:10.1513/AnnalsATS.202207-631PS
36. Gendarme S, Maitre B, Hanash S, Pairen J-C, Canoui-Poitrine F, Chouaïd C. Beyond lung cancer screening, an opportunity for early detection of chronic obstructive pulmonary disease and cardiovascular diseases. *JNCI Cancer Spectrum*. 2024;8(5):kae082. doi:10.1093/jncics/pkae082
37. Mets O, De Jong P, Van Ginneken B, Gietema H, Lammers J. Quantitative computed tomography in COPD: possibilities and limitations. *Lung*. 2012;190:133–145. doi:10.1007/s00408-011-9353-9
38. Coxson HO, Leipsic J, Parraga G, Sin DD. Using pulmonary imaging to move chronic obstructive pulmonary disease beyond FEV1. *Am J Respir Crit Care Med*. 2014;190(2):135–144. doi:10.1164/rccm.201402-0256PP

International Journal of Chronic Obstructive Pulmonary Disease

Publish your work in this journal

The International Journal of COPD is an international, peer-reviewed journal of therapeutics and pharmacology focusing on concise rapid reporting of clinical studies and reviews in COPD. Special focus is given to the pathophysiological processes underlying the disease, intervention programs, patient focused education, and self management protocols. This journal is indexed on PubMed Central, MedLine and CAS. The manuscript management system is completely online and includes a very quick and fair peer-review system, which is all easy to use. Visit <http://www.dovepress.com/testimonials.php> to read real quotes from published authors.

Submit your manuscript here: <https://www.dovepress.com/international-journal-of-chronic-obstructive-pulmonary-disease-journal>

Dovepress
Taylor & Francis Group