machine learning & knowledge extraction

MDPI

Article

Manifold Integration of Lung Emphysema Signatures (MILES): A Radiomic-Based Study

Marek Socha [1], Agata Durawa [2], Małgorzata Jelito [2], Katarzyna Dziadziuszko [2], Witold Rzyman [3], Edyta Szurowska [2] and Joanna Polanska [1,*]

[1] Department of Data Science and Engineering, Silesian University of Technology, 44-100 Gliwice, Poland; marek.socha@polsl.pl
[2] 2nd Department of Radiology, Medical University of Gdansk, 80-210 Gdansk, Poland
[3] Department of Thoracic Surgery, Medical University of Gdansk, 80-210 Gdansk, Poland
* Correspondence: joanna.polanska@polsl.pl

Abstract

Chronic obstructive pulmonary disease (COPD) is the third leading cause of death worldwide, and emphysema is present in the majority of affected patients and can be identified on computed tomography (CT). This study investigated whether radiomic features derived from automatically and adaptively segmented low-attenuation lung regions can capture distinct imaging characteristics of COPD beyond conventional emphysema measures. Radiomic features were extracted from 6078 chest CT scans of 2243 participants from the COPDGene cohort. Emphysematous regions were segmented using the MimSeg method based on Gaussian mixture modelling with patient-adjusted thresholding, and radiomic features were computed for individual lesion clusters and aggregated per patient using summary statistics, yielding 780 features per subject. Uniform Manifold Approximation and Projection (UMAP) was used to generate a low-dimensional embedding, and feature contributions were evaluated using SHAP analysis and statistical testing. The resulting embedding demonstrated structured patterns broadly aligned with Global Initiative for Chronic Obstructive Lung Disease (GOLD) stages, with greater overlap among GOLD 0–2 and more consolidated groupings for GOLD 3 and 4, reflecting differences in disease severity. The most influential features were predominantly derived from Grey Level Run Length Matrix measures, capturing textural heterogeneity and spatial organisation of emphysematous changes that are not directly described by standard density-based metrics. These findings suggest that radiomic analysis of adaptively segmented CT data may provide complementary and structurally distinct information relative to conventional emphysema measures, supporting a more nuanced characterisation of emphysema patterns in COPD.

Keywords: unsupervised learning; radiomic features; computer-aided diagnosis; thorax computed tomography; emphysema

Check for updates

Academic Editor: Guoqing Chao

Received: 18 December 2025
Revised: 26 January 2026
Accepted: 27 January 2026
Published: 30 January 2026

Copyright: © 2026 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license.

1. Introduction

COPD (Chronic Obstructive Pulmonary Disease) is a progressive respiratory condition characterised by a chronic and irreversible limitation of airflow through the airways. This airflow obstruction arises primarily from two underlying pathological processes: chronic bronchitis—marked by persistent inflammation of the airways and excessive mucus production, which narrows the bronchial passages and pulmonary emphysema—defined by permanent damage to the alveoli and a loss of their elasticity, leading to reduced surface area for

gas exchange [1]. Identifying the predominant underlying mechanism of COPD—whether emphysema or chronic bronchitis—allows for more tailored treatment decisions, a more accurate prognosis, and a better understanding of the patient's symptoms. Phenotyping COPD (for example, "emphysema-dominant" vs. "chronic bronchitis-dominant") is now a key component of modern clinical management of the disease. The key diagnostic tool is spirometry, performed after administration of a bronchodilator. A post-bronchodilator FEV_1/FVC (Forced Expiratory Volume in 1 s/Forced Vital Capacity) ratio below 0.7 confirms persistent airflow limitation, which is the hallmark of COPD and the basis for the GOLD standard [1,2]. The severity is then graded based on the FEV_1 percentage of the predicted value. Additional assessments, such as symptom questionnaires (e.g., mMRC or CAT), exacerbation history, CT imaging, and arterial blood gases, may help guide treatment decisions and rule out other conditions. The presented work explores radiomic features derived from the CT images of the COPDGene dataset to capture a broader range of imaging biomarkers [3,4]. By visualising case similarities through this expanded feature space, the proposed method aims to support more accurate, individualised decision-making and improve responsiveness to patient-specific disease characteristics.

For the assessment of emphysematous changes on chest CT, the widely accepted standard is the measurement of the percentage of low-attenuation areas relative to total lung volume using a fixed threshold of −950 HU (%LAA−950). This metric shows partial correlation with GOLD staging and is one of the most commonly used thresholds in emphysema research. This threshold performs well in identifying large, discrete low-attenuation clusters, but it does not fully capture the diversity of emphysema patterns observed in COPD [5]. Before the dominance of clearly delineated low-attenuation regions, early disease stages, so-called incipient emphysema, may be characterised by more subtle and spatially diffuse structural changes affecting broader areas of the lung [6]. Recent studies have shown that increased parenchymal heterogeneity in lung regions that appear visually normal in COPD patients is associated with disease progression and mortality [7]. These findings suggest that COPD-related structural alterations may exist in forms that are not fully captured by traditional fixed–threshold–based metrics and that complementary textural and spatial descriptors may provide additional insight into disease presence and progression.

These limitations motivated the use of other descriptors and metrics measuring heterogeneity and textural features of patterns related to COPD. Radiomic features enable the extraction of such descriptors from defined regions of interest [8,9]. They comprise a collection of values characterising a specific region within a region-of-interest (ROI) of a medical image. Various methodologies have been applied to test their effectiveness. Cheplygina et al. suggested a logistic classifier to categorise patients as healthy or diseased [10]. In pursuing similar objectives, Li et al. employed the SVM algorithm, leveraging computed lung radiomic characteristics to additionally identify COPD subtypes [11]. Yang et al. introduced two methodologies: the first utilises the Lasso method to pinpoint significant lung radiomic features, followed by categorisation via a multi-layer perceptron [12], and the second also employs Lasso for feature selection but features a graph neural network for the final classification [13]. Amudala et al. concentrated on lung and bronchial tree features, examining their texture and shape, and adopted a Gradient Boosting-based algorithm to classify lungs as either healthy or exhibiting COPD [14]. While radiomics has shown strong performance in detecting COPD on both standard-dose and low-dose CT scans, most approaches frame the problem as binary classification, providing limited insight into disease heterogeneity, phenotypic structure, and the relationships among radiomic features concerning COPD.

Neural networks offer an alternative to radiomic features by learning to extract meaningful image representations correlated with clinical targets. González et al. trained a CNN using canonical CT views to classify COPD and GOLD stage [15]. Singla et al. used a generative network with attention to predict spirometric indices (FEV1, FEV1/FVC) [16]. Xu et al. employed a CNN for feature extraction, followed by PCA and classical classifiers to stage COPD [17]. Du et al. focused on the bronchial tree, using CNN and a voting classifier for COPD detection [18]. Tang et al. applied a deep residual network to regional CT inputs for spirometry-based COPD classification [19]. Ho et al. used parametric response mapping with CNN to identify emphysema and classify COPD [20]. Chen et al. applied graph convolutional networks on CT cubes to detect COPD [21]. Zhang et al. combined DenseNet models trained on lung volumes and bronchial walls via logistic regression [22]. Wu et al. used 3D lung and airway snapshots for CNN-based COPD detection [23]. Sun et al. utilised ResNet18 with attention on axial CT slices [24]. Li et al. modelled spatial-temporal ROI graphs for COPD classification using graph neural networks [25]. Another approach by Wu et al. fused CT slices, lung, and bronchial masks using an attention-based CNN [23]. Xue et al. adopted a multiple-instance learning framework with attention to detect COPD and localise affected slices [26]. Yu et al. introduced a self-supervised contrastive learning method for emphysema detection [27]. Deep learning–based methods have shown strong performance in COPD classification; however, many approaches focus on end-to-end prediction and provide limited information regarding disease staging or regional patterns relevant for clinical interpretation.

Clinicians remain cautious about AI-driven tools due to the need for transparency in high-stakes medical decisions [28,29]. Simply outputting a diagnostic label is insufficient; explainability is essential. Some studies use explainable AI (xAI) techniques to visualise network attention on CT scans, highlighting regions influencing the decision [20,23,26]. However, these methods can lead to ambiguity—activation maps may extend beyond relevant regions, potentially reflecting biases from training data or imaging centre artefacts. As such, neural networks are not yet suitable for diagnosis but may be useful as screening tools if validated on diverse datasets with strong predictive performance [30,31]. In contrast, radiomic approaches offer greater interpretability but may miss subtle, localised abnormalities when applied to entire lung volumes.

The present work focuses on characterising COPD using both quantitative and qualitative analysis of radiomic features extracted from chest CT images. Unlike prior studies that either focus on the whole lung or rely on fixed-threshold segmentations that may miss subtle changes, our approach uses adaptive, patient-specific thresholding to delineate emphysema-related tissue. This allows the method to capture large, clearly defined emphysematous regions in severe cases, as well as diffuse, incipient structural changes in early or mild cases, which are often overlooked by standard global fixed thresholds. By projecting the resulting radiomic descriptors into a low-dimensional space using UMAP, the pipeline (MILES) highlights both distinct emphysema patterns and the features that drive them, providing interpretable structure in the data. While the analysis also reflects GOLD stage trends, the key contribution is the identification of subtle, patient-specific tissue changes across the spectrum of disease, offering a more nuanced and individualised assessment of emphysema phenotypes in COPD.

2. Materials and Methods

The MILES pipeline is presented in Figure 1, CT series are processed through lung and lobe segmentation, as well as the removal of bronchi from the lung parenchyma. This is followed by MiMSeg based adaptive thresholding, lesion clustering, and per-cluster radiomic feature extraction. Features are then aggregated at the patient-level and

passed to the UMAP method. The created embedding is explained using SHAP, capturing meaningful features.

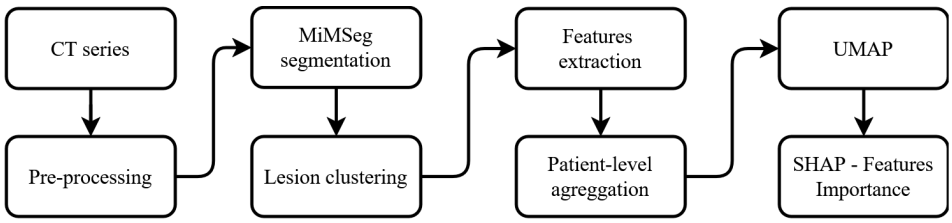


Figure 1. Schematic pipeline showing key points of presented approach.

2.1. Dataset

The COPDGene dataset was developed through a multi-phase study involving 21 U.S. institutions [2]. It includes genetic data, inspiratory and expiratory CT scans, and patient metadata such as age, sex, ancestry, scanner type, convolutional kernel, and GOLD stage derived from the spirometry results.

The GOLD stage is a spirometry-based metric that classifies the severity of airflow limitation in COPD patients. GOLD stages range from 0 to 4, where GOLD 0 represents individuals with symptoms suggestive of COPD but without spirometric confirmation of airflow obstruction, when $FEV1/FVC > 0.7$. GOLD stages 1 through 4 correspond to progressively increasing severity of obstruction based on post-bronchodilator FEV1 as a percentage of the predicted value. A patient is classified as GOLD 1 when $FEV1 \geq 80\%$, GOLD 2 when $50\% \leq FEV1 < 80\%$, GOLD 3 when $30\% \leq FEV1 < 50\%$ and GOLD 4 when $FEV1 < 30\%$ of predicted [1].

From the entire dataset, for each patient, the series with a soft or standard convolution kernel was selected. Only the scans conducted during inspiration were considered in this study. This work analyses 1107 CT series of unique patients from the COPDGene population. The resulting dataset consisted of studies conducted using 12 different scanners from 3 popular CT scanner manufacturers, as shown in Table 1. The distribution of the GOLD standard values across the dataset is shown in Table 2.

Table 1. Scanner Manufacturer DICOM attribute distribution.

Manufacturer	Count
GE Medical Systems	570
SIEMENS	519
Philips	18

Table 2. GOLD stage distribution.

GOLD Stage	Count
0	234
1	76
2	374
3	263
4	160

2.2. Region of Interest (ROI) Definition

The region of interest was emphysematous lesions. The following sequence of operations was used to segment them: lung segmentation, lobe segmentation, bronchial tree segmentation and removal, and adaptive thresholding to localise emphysematous lesions.

For both lungs segmentation and lobes segmentation, the UNet proposed and trained by Hofmanninger et al. [32] was used. A bronchial bundle was removed from the lung area using the dual-velocity growing method [33]. This way, reducing the number of voxels similar in intensity to the emphysematous lesion.

The emphysematous lesion was segmented using adaptive thresholding based on the Gaussian Mixture Models employed by the MimSeg algorithm [34]. The algorithm iterates over the slices of the given CT series and performs decomposition of the Gaussian components, where the number of components was selected using the Bayesian Information Criterion (BIC). The parameters of components (mean, standard deviation, and alpha) are stored in each iteration. After all slices were analysed, the grouping using K-Means was done in the space of mean, standard deviation, and alpha; the number of clusters was optimised using the elbow curve. For each cluster, a membership curve was created. The curve characterised by the lowest overall mean is assumed to contain emphysema. The segmentation threshold was defined at the intersection of the membership curves.

Emphysema segmentation was split based on the lobes. Additionally, sub-regions were created in each lobe based on the voxel coordinates using K-Means. As a result, voxels located next to each other and in close proximity were incorporated into a single ROI.

2.3. Features Extraction

Three feature sets were extracted per region of interest (ROI). The first set included 110 radiomic features [8,9] from the ROI, using the pyRadiomics package. The second set was computed from contours drawn within a 3-voxel margin around the ROI. The third set described the spatial position of the ROI, based on a vector from the lung's weighted centre to the ROI centre, capturing its magnitude and angles in x, y, and z directions. In total, 156 features were calculated per ROI.

Radiomic features came from seven groups: (1) First-Order (intensity distribution), (2) Shape (geometry; not used for contour features), (3) GLCM (texture via voxel pair relationships), (4) GLRLM (run lengths of equal intensity), (5) GLSZM (size of same-intensity zones), (6) GLDM (intensity dependence in neighbourhoods), and (7) NGTDM (contrast with surrounding voxels).

For each patient, features from all ROIs were aggregated using the mean, median, min, max, and standard deviation, resulting in 780 features per patient.

2.4. Features Space Reduction

As a first step in the pipeline, features with zero variance were removed. The remaining features were then standardised using their mean and standard deviation. Principal Component Analysis (PCA) was applied for initial dimensionality reduction, retaining components that explained 90% of the total variance.

2.5. UMAP Embedding

Uniform Manifold Approximation and Projection (UMAP) [35] was used to create the final 2D embedding. UMAP models high-dimensional data as a weighted graph and finds a low-dimensional embedding by minimising differences between the original and reduced graphs. For key hyperparameters, the distance metric $d(\dots)$ was set to cosine distance, the number of neighbours k was set to 70, and the minimum distance between points d_{min} was set to 0.3.

2.6. Features Significance

To evaluate the importance of individual features in the UMAP embedding process, SHAP values were applied [36]. A separate neural network model was trained, where the input consisted of standardised features and the output corresponded to the UMAP

coordinates (U1 and U2). The architecture consisted of five fully connected (dense) layers with ELU activation functions [37], and two dropout layers (with a dropout rate of 0.3) placed between the dense layers to prevent overfitting. The network's backbone contained between 200 and 50 neurons, with 2 neurons at the input layer. Training was carried out using the Adam optimiser with a learning rate of 0.001, and mean absolute error (MAE) was used as the loss function. The model was trained for 100 epochs.

To assess how radiomic features identified as influential by SHAP relate to conventional emphysema metrics, Spearman correlation coefficients were calculated between each feature and both %LAA−950 and mean lung density. Correlation magnitudes were qualitatively classified as negligible (<0.1), small (<0.3), medium (<0.5) or large (<0.7) according to [38]. This approach allows the evaluation of how SHAP-selected features reflect conventional emphysema measures while identifying additional structural information captured by radiomic descriptors.

3. Results

The feature set was first preprocessed and then passed through Principal Component Analysis (PCA). Components explaining 90% of the total variance were retained, resulting in 29 principal components, each representing a linear combination of the original features. UMAP was then applied to these PCA components. The resulting embedding is presented in Figure 2 as a density plot, with separate subplots illustrating the distribution of patients from each GOLD group. In each subplot, clusters—areas with a high concentration of points—were labelled as C(GOLD) (cluster number, ordered left to right). Specifically, three clusters were identified for GOLD 0 patients (C00, C01, C02); two similarly sized clusters for GOLD 1 (C10, C11); two differently sized clusters for GOLD 2 (C20, C21); and a single cluster each for GOLD 3 (C30) and GOLD 4 (C40) patients.

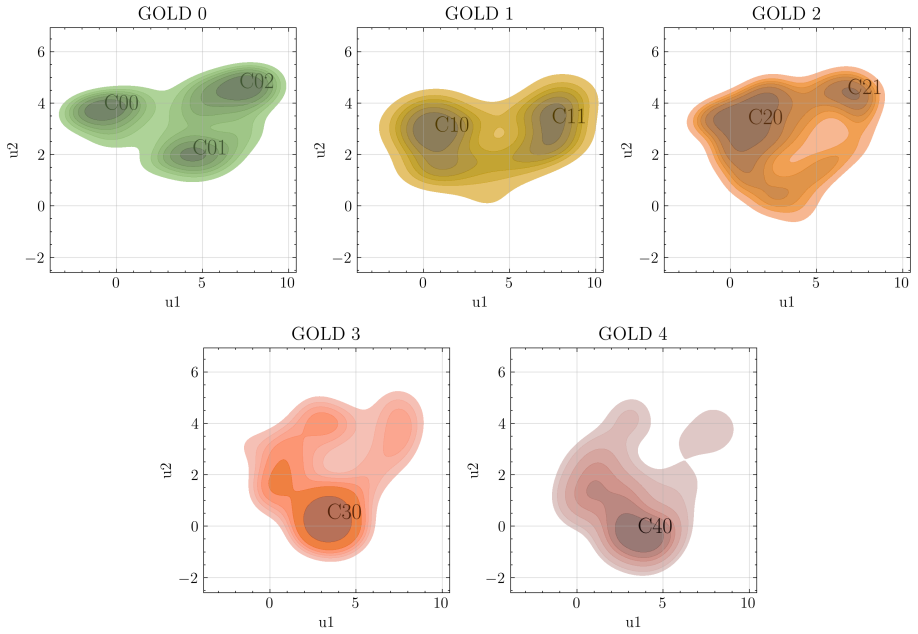


Figure 2. Density plots of resulting UMAP embedding shown on different subplots. For each GOLD group, areas of the largest densities have been found and named following the convention C(GOLD) (cluster number left–right).

Some overlap is observed among clusters corresponding to GOLD groups 0, 1, and 2. Specifically, clusters C00 and C02 are positioned slightly above clusters C10, C11, C20, and C21. Cluster C10 lies between C00 and C01, while C11 is located between C01 and C02. Cluster C20 overlaps with C10 but does not significantly intersect with C00 or C01. Cluster C21 overlaps with C11 and occupies the lower portion of C02. Clusters C30 and C40 are positioned near each other, with the highest density area of C40 located below the C30 cluster.

The number of clusters and the degree of overlap within the GOLD groups indicate a level of non-homogeneity. However, this variation is not primarily driven by the disease itself, but rather by individual patient differences. A trend toward increased cluster homogeneity is observed with higher GOLD stages. Each identified cluster was visualised and further analysed to support this observation.

An example image from the C00 cluster is shown in the second subplot of Figure 3. These scans show low attenuation values (−982 HU), with abnormalities near the lung edges or blood vessels. Thin alveolar walls are visible, indicating high-resolution imaging. The third subplot presents an example from the C01 cluster, which includes both normal scans and those with emphysematous changes. Segmented regions are large, with densities around −910 HU, mostly located away from lung borders and minimally overlapping blood vessels. The C02 cluster example, shown in the fourth subplot, displays unusually high threshold values (−788 HU). Despite this, lung boundaries are not fully outlined, possibly due to scanning issues or high-attenuation patterns. Some images are noisy, obscuring alveolar detail. Segmentations often extend to the chest wall, suggesting contour features may influence this cluster's placement in the UMAP space.

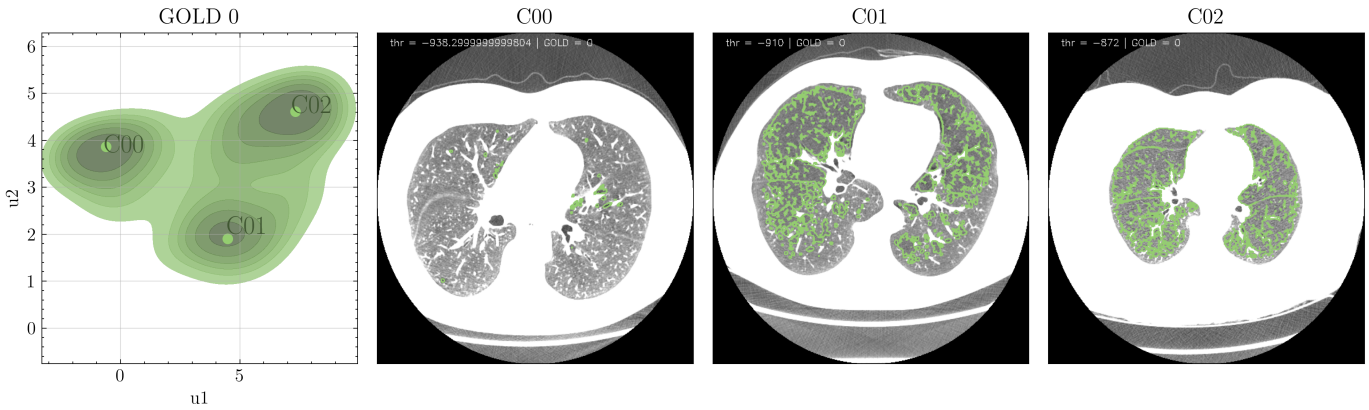


Figure 3. UMAP embedding density plot for GOLD 0 group with example images taken from the centre of each cluster.

An example image from the C10 cluster is shown in the second subplot of Figure 4. These scans exhibit small-volume but highly diffuse emphysematous lesions, often in contact with lung walls and vessels. Threshold values vary between −900 and −940 HU, though voxel values within segmentations appear relatively uniform. The C11 cluster example, shown in the third subplot, includes scans with larger volumes of abnormal tissue and lower diffusion; lesions are more compact. Thresholds range from −850 to −910 HU, and the segmented regions show greater heterogeneity in voxel values.

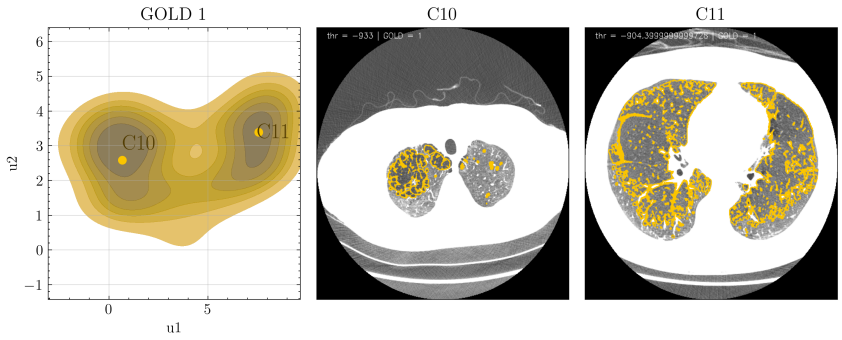


Figure 4. UMAP embedding density plot for GOLD 1 group with example images taken from the centre of each cluster.

An example from the larger, more diffuse C20 cluster is shown in the second subplot of Figure 5. Threshold values in this group range widely from −915 to −990 HU, reflecting the cluster's broad coverage. These scans show moderate structural changes with increasing intensity lower on the UMAP y-axis. The lesions are highly diffuse, and voxel values appear homogeneous. The C21 cluster example, shown in the third subplot, represents a smaller group with higher threshold values, indicating larger lesion volumes. The segmented areas are compact, often adjacent to the lung boundary, and show heterogeneous voxel intensities.

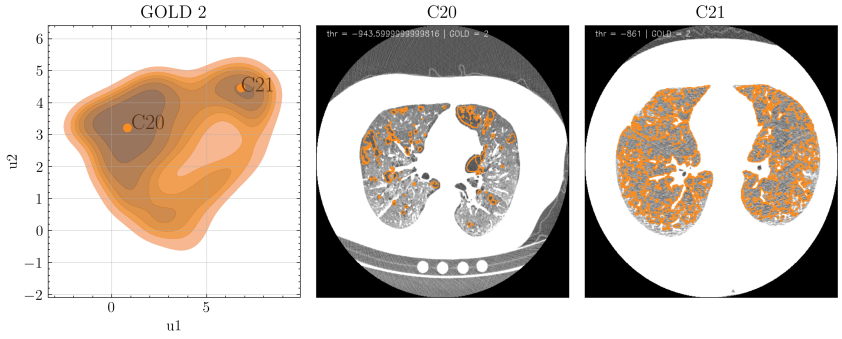


Figure 5. UMAP embedding density plot for GOLD 2 group with example images taken from the centre of each cluster.

An example from the high-density C30 cluster is shown in the second subplot of Figure 6. These scans contain a large volume of emphysematous lesions with no consistent distribution pattern. The HU values within the segmented areas are heterogeneous, featuring distinct patches of lesions separated by prominent white lines.

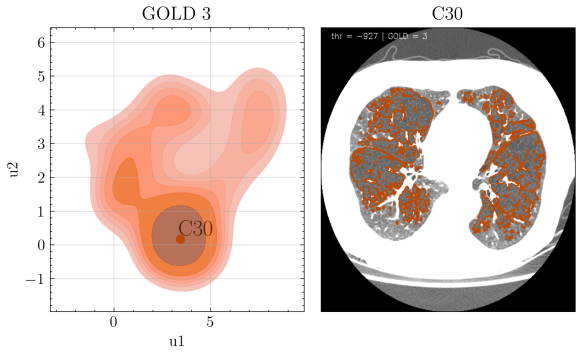


Figure 6. UMAP embedding density plot for GOLD 3 group with example image taken from the centre of the cluster.

An example from the main C40 cluster is shown in the second subplot of Figure 7. These images feature widespread, large, and homogeneous emphysematous patches, re-

flecting extensive alveolar wall destruction. Segmented areas often border vessels and lung walls or are divided by fissures and reticulations.

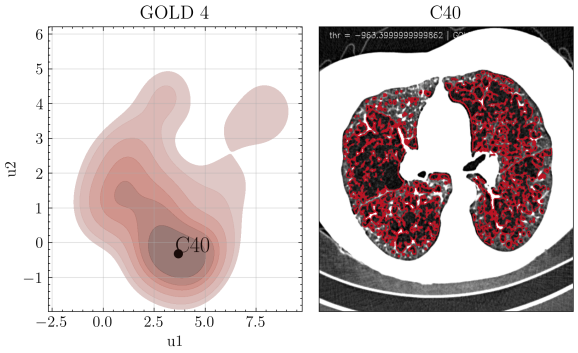


Figure 7. UMAP embedding density plot for GOLD 4 group with example image taken from the centre of the cluster.

A visual analysis of the UMAP embedding reveals a progression in emphysematous changes. In the upper left region, images show small segmented areas with homogeneous grey-level intensities. Moving rightward, the extent of abnormalities increases—initially with high spatial dispersion, later transitioning to more uniform regions. Further to the right, segmentations become more heterogeneous in terms of grey levels and often border lung walls. As one moves downward, the segmentations gradually shift away from lung boundaries. The lower part of the UMAP becomes increasingly narrow, with segmentations displaying growing homogeneity. On the left and right, segmentations with highly scattered regions (ROIs) are observed, while at the bottom, segmentations form large, uniform patches, suggesting advanced alveolar wall destruction.

To better understand which features influenced the UMAP embedding, a backwards feature analysis was performed. SHAP values, computed from the trained neural network model, identified the most impactful features. The mean absolute SHAP values for the U1 dimension were as follows: min Run Entropy shap = 0.0245, std Contour Small Area Emphasis shap = 0.0245, mean Surface Area shap = 0.017943, max Dependence Non-Uniformity shap = 0.017147 and std Run Entropy shap = 0.01698. Following that, the mean absolute SHAP values for the U2 dimension were: std Minor Axis Length shap = 0.017045, min Run Percentage shap = 0.01312, min Run Length Non-Uniformity Normalised shap = 0.01251, max Long Run Emphasis shap = 0.01219 and min Short Run Emphasis shap = 0.01127. The minimum run entropy is the most influential feature along the U1 (X) axis, while the standard deviation of the minor axis length is most significant along the U2 (Y) axis. Other features show relatively uniform SHAP value distributions.

The overall structure reveals a clear separation between the GOLD 0–1–2 classes and the GOLD 3–4 classes. GOLD 0, 1, and 2 occupy the upper part of the UMAP embedding, while GOLD 3 and 4 occupy the lower part of the embedding. Separation ability of the joint groups is shown in the Figure 8.

Based on the SHAP features indicated, the Kruskal-Wallis test was conducted to assess the influence of these features on the groups GOLD 0–1–2 and GOLD 3–4. The resulting *p*-values and boxplot of the features values are presented in Figure 9.

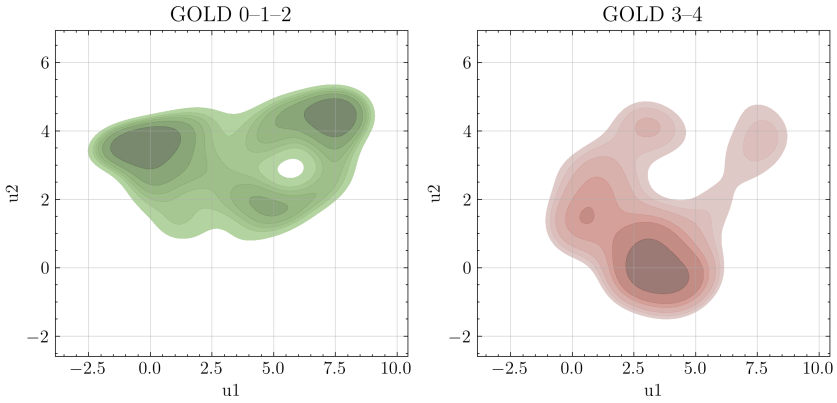


Figure 8. UMAP embedding density plot presenting separation ability of joint densities of GOLD groups 0–1–2 and 3–4.

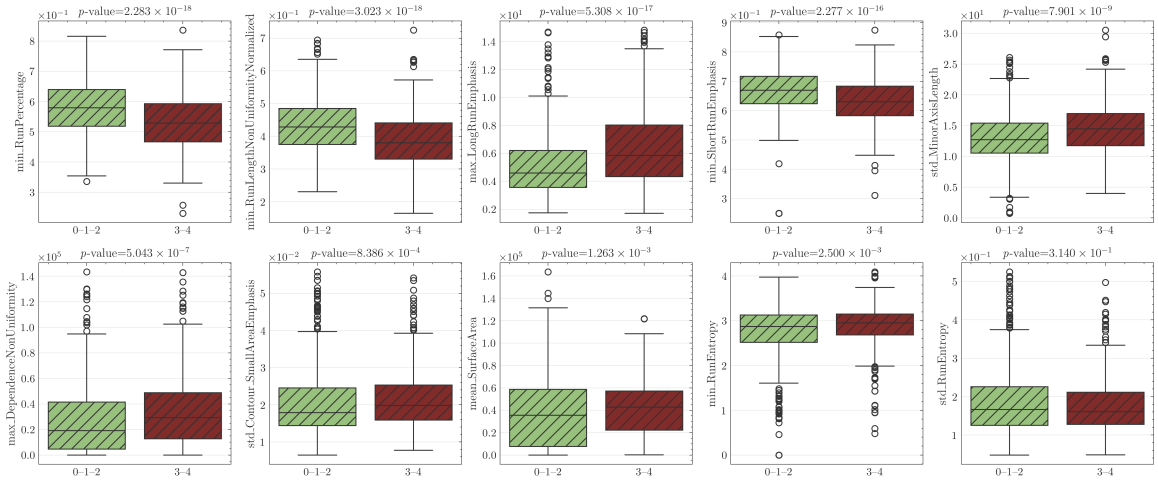


Figure 9. Boxplots with Kruskal-Wallis *p*-values in the title showing the distribution of the features indicated by the SHAP analysis and their strength regarding differentiating between groups of joint GOLD standards 0–1–2 and 3–4.

Assuming a statistical significance level of $\alpha = 0.05$, each of the analysed features significantly differentiates the GOLD 0–1–2 group from the GOLD 3–4 group. The most distinguishing features belong to dimension U2; for example, features such as min Run Percentage and min Run Length Non-Uniformity Normalised, as seen in the first row of Figure 9. The features selected by SHAP analysis for U1 are also statistically significant, but their *p*-values are higher than those for U2 features. This result was expected, considering the density plots shown in Figure 8.

Spearman correlation analysis showed that U1 features (min Run Entropy, std Run Entropy, std Contour Small Area Emphasis, mean Surface Area, std Minor Axis Length) generally exhibited negligible to weak correlations with %LAA−950 and mean lung density. In contrast, U2 features (min Run Percentage, min Run Length Non-Uniformity Normalised, max Long Run Emphasis, min Short Run Emphasis) demonstrated moderate correlations with both metrics, particularly with mean lung density. These findings indicate that U1 features capture subtle or incipient emphysematous changes that are not fully represented by global threshold-based metrics, while U2 features reflect larger-scale emphysema patterns associated with conventional measures (Table 3).

Table 3. Spearman correlation of SHAP-indicated radiomic features with standard emphysema metrics (%LAA−950 HU and mean lung density). Effect sizes are categorised as negligible, weak, moderate, or strong based on the correlation magnitude.

UMAP	Feature	Metric	Spearman's ρ	Effect
U1	min Run Entropy	%LAA−950 HU	0.03	negligible
U1	std Run Entropy	%LAA−950 HU	−0.06	negligible
U1	std Contour Small Area Emphasis	%LAA−950 HU	0.15	small
U1	mean Surface Area	%LAA−950 HU	0.13	small
U1	std Minor Axis Length	%LAA−950 HU	0.31	moderate
U1	min Run Entropy	Mean Lungs Density	0.04	negligible
U1	std Run Entropy	Mean Lungs Density	0.00	negligible
U1	std Contour Small Area Emphasis	Mean Lungs Density	−0.22	small
U1	mean Surface Area	Mean Lungs Density	−0.15	small
U1	std Minor Axis Length	Mean Lungs Density	−0.24	small
U2	min Run Percentage	%LAA−950 HU	−0.43	moderate
U2	min Run Length Non-Uniformity Normalized	%LAA−950 HU	−0.43	moderate
U2	max Long Run Emphasis	%LAA−950 HU	0.44	moderate
U2	min Short Run Emphasis	%LAA−950 HU	−0.41	moderate
U2	min Run Percentage	Mean Lungs Density	0.57	large
U2	min Run Length Non-Uniformity Normalized	Mean Lungs Density	0.58	large
U2	max Long Run Emphasis	Mean Lungs Density	−0.58	large
U2	min Short Run Emphasis	Mean Lungs Density	0.56	large

4. Features Impact

This section explores the impact of individual features on the UMAP embedding. UMAP visualisations include example CT slices representing low and high feature values, with corresponding voxel-based radiomic features displayed below each image. Each low-attenuation pattern segmentation was divided into regions of interest (ROIs) based on lung lobes. Radiomic features were calculated per ROI and then aggregated using a difference-based method.

The most influential feature along the U1 axis is the minimum of Run Entropy, a texture metric derived from the Grey Level Run Length Matrix (GLRLM). It measures the randomness of run-length distributions across intensities—higher values indicate complex textures, while lower values reflect more uniform patterns. Figure 10 shows examples with high and low values. In GOLD 0 and 1 cases, emphysema appeared in small, uniform regions; otherwise, masks covered non-uniform lung areas. This suggests the feature helps differentiate emphysematous from non-emphysematous cases in early GOLD stages.

Run Entropy appears again in the importance ranking—its standard deviation is the fifth most significant feature. While both measures derive from the same origin, their UMAP distributions differ (Figure 11). The gradient of feature variation is reversed, but the overall spatial trend remains consistent.

The standard deviation of contour Small Area Emphasis is the second most important U1 feature. This texture feature, derived from the Grey Level Size Zone Matrix (GLSZM), quantifies the frequency of small homogeneous zones along the segmentation contour. High values indicate varied contour sizes and textures, often seen in centrally located lesions; low values reflect more uniform contours, typically along lung borders (Figure 12).

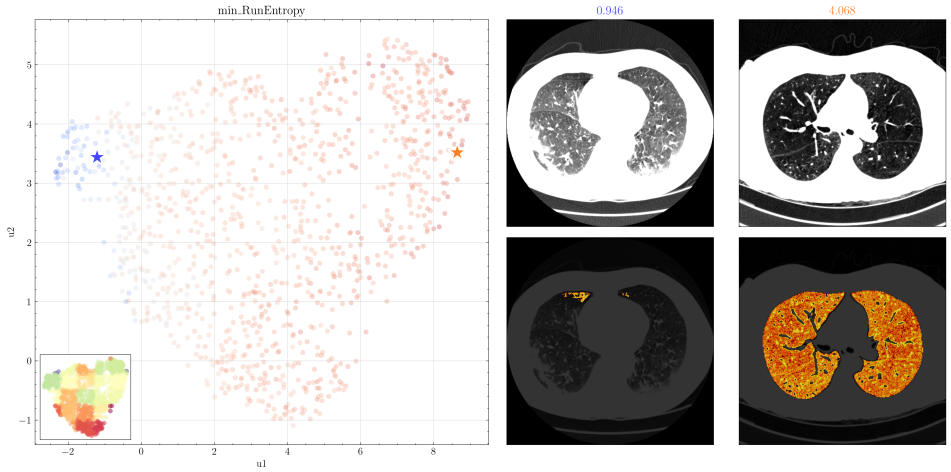


Figure 10. Minimum of Run Entropy distribution across the UMAP embedding (reference GOLD distribution in bottom-left corner). Example images (indicated by stars on the scatter plot) with low and high feature values with their voxel-based feature heatmaps.

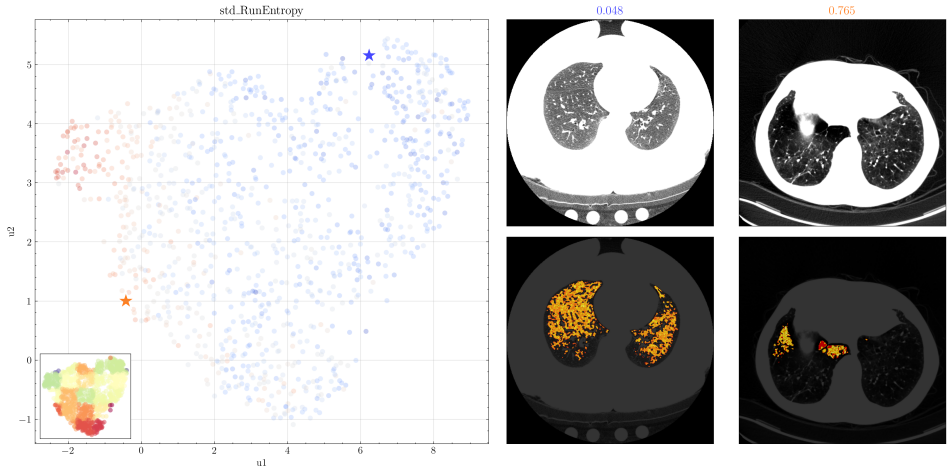


Figure 11. Standard deviation of Run Entropy distribution across the UMAP embedding (reference GOLD distribution in bottom-left corner). Example images (indicated by stars on the scatter plot) with low and high feature values with their voxel-based feature heatmaps.

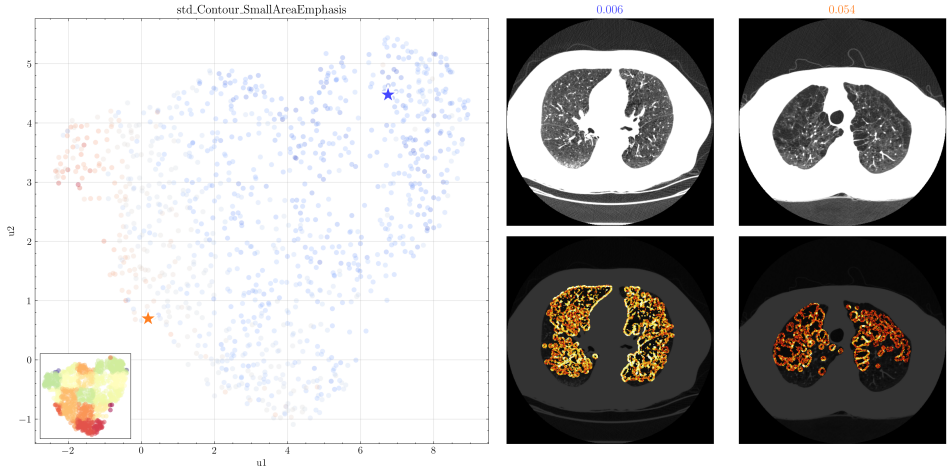


Figure 12. Standard deviation of Contours' Small Area Emphasis distribution across the UMAP embedding (reference GOLD distribution in the bottom-left corner). Example images (indicated by stars on the scatter plot) with low and high feature values with their voxel-based feature heatmaps.

https://doi.org/10.3390/make8020032

The mean Surface Area ranks third in U1 significance. As a shape-based feature, it represents the average outer surface area of segmented 3D ROIs and correlates directly with mask size. Examples with high and low values are shown in Figure 13.

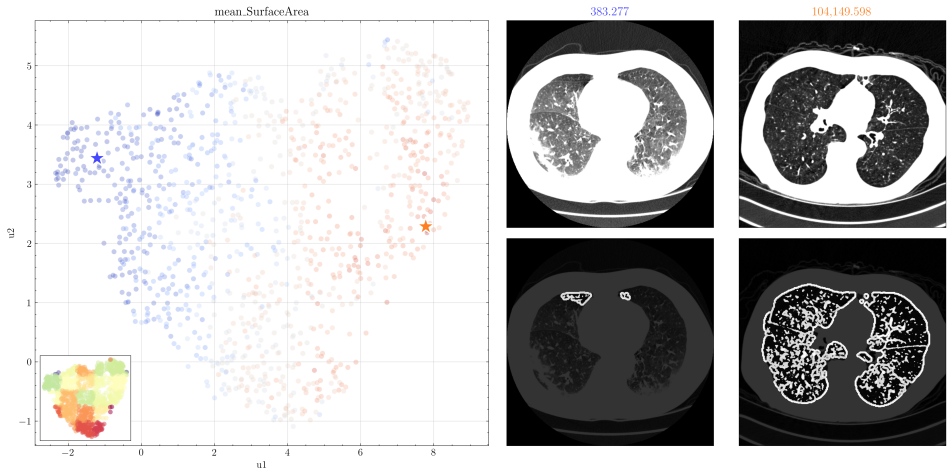


Figure 13. Mean of ROIs' Surface Area distribution across the UMAP embedding (reference GOLD distribution in bottom-left corner). Example images (indicated by stars on the scatter plot) with low and high feature values, with and without overlaid segmentations' contours.

The standard deviation of Minor Axis Length is the most influential feature along the U2 axis. This shape-based feature reflects the variation in the shortest axis of ellipses fitted to segmented ROIs. High values suggest many lesions of varying sizes scattered across the lungs, while low values indicate fewer or similarly sized regions. Example images are shown in Figure 14.

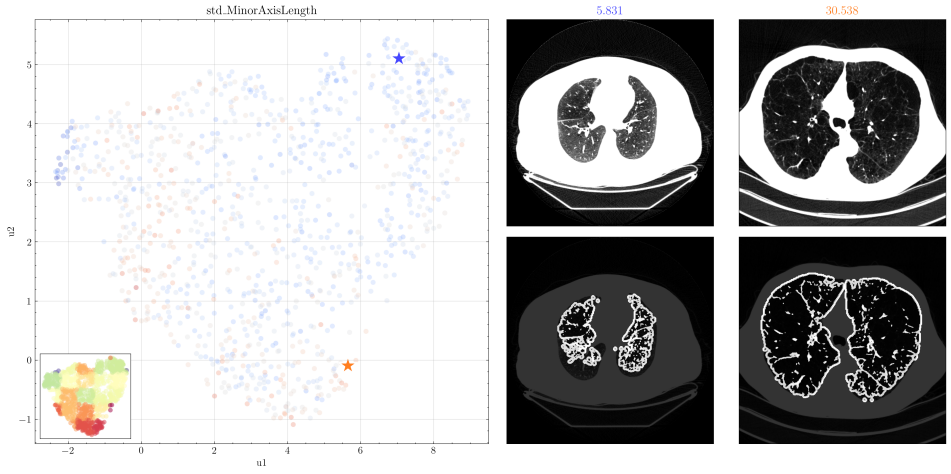


Figure 14. Standard deviation of Minor Axis Length distribution across the UMAP embedding (reference GOLD distribution in the bottom-left corner). Example images (indicated by stars on the scatter plot) with low and high feature values, with and without overlaid segmentations' contours.

The minimum of Run Percentage ranks second. This texture feature from the GLRLM measures the density of homogeneous grey-level runs. Low values indicate more uniform regions, typical of emphysematous lesions. Examples are in Figure 15.

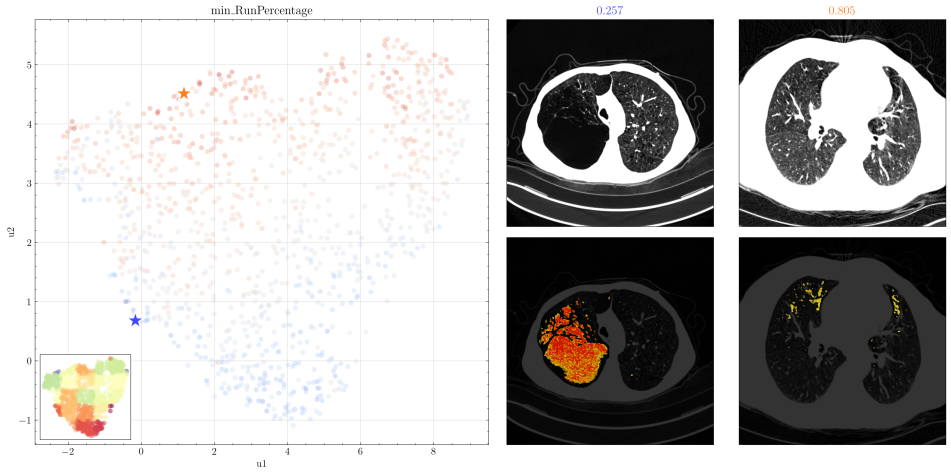


Figure 15. Minimum of Run Percentage distribution across the UMAP embedding (reference GOLD distribution in bottom-left corner). Example images (indicated by stars on the scatter plot) with low and high feature values with their voxel-based feature heatmaps.

Third in importance is the minimum of Run Length Non-Uniformity Normalised, also from GLRLM. It captures variation in run lengths—low values indicate that some lengths dominate, suggesting a repetitive texture, as seen in emphysematous areas (Figure 16).

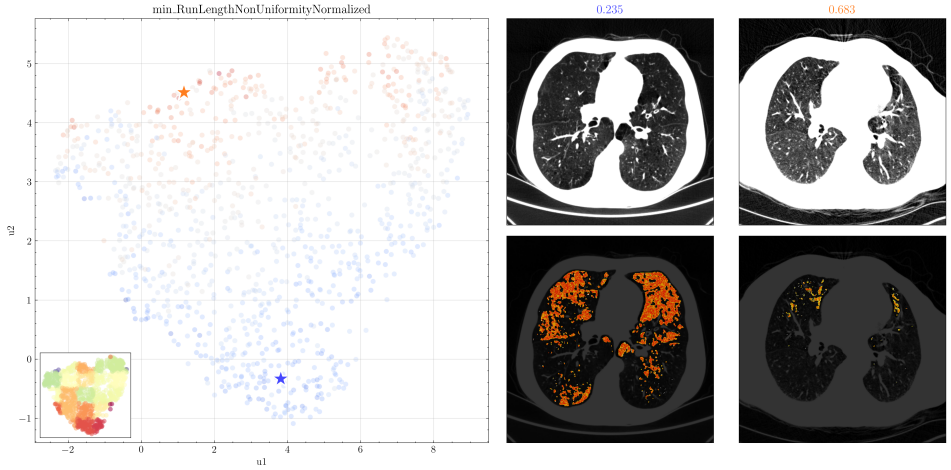


Figure 16. Minimum of Run Length Non-Uniformity Normalised distribution across the UMAP embedding (reference GOLD distribution in the bottom-left corner). Example images (indicated by stars on the scatter plot) with low and high feature values with their voxel-based feature heatmaps.

The last two key features are the maximum of Long Run Emphasis (LRE) and minimum of Short Run Emphasis (SRE). Both reflect texture homogeneity, but at different scales. High LRE and low SRE values both indicate more uniform textures, typical in emphysematous tissue. LRE captures large-scale structure, while SRE reflects finer patterns. Example slices are in Figures 17 and 18.

The first UMAP component (U1) appears to reflect both the location and size of ROIs, as well as the extent of low grey-level regions (emphysema).

1. Run Entropy distinguishes patients by texture randomness—larger, non-emphysematous ROIs show higher randomness, capturing both emphysema presence and extent.
2. The standard deviation of contour Small Area Emphasis reflects whether ROIs are near the lung borders and their spread. Small, peripheral ROIs with low volume yield high variation; large ROIs extending to fissures show lower variation.
3. Mean Surface Area directly measures the extent of emphysematous changes.

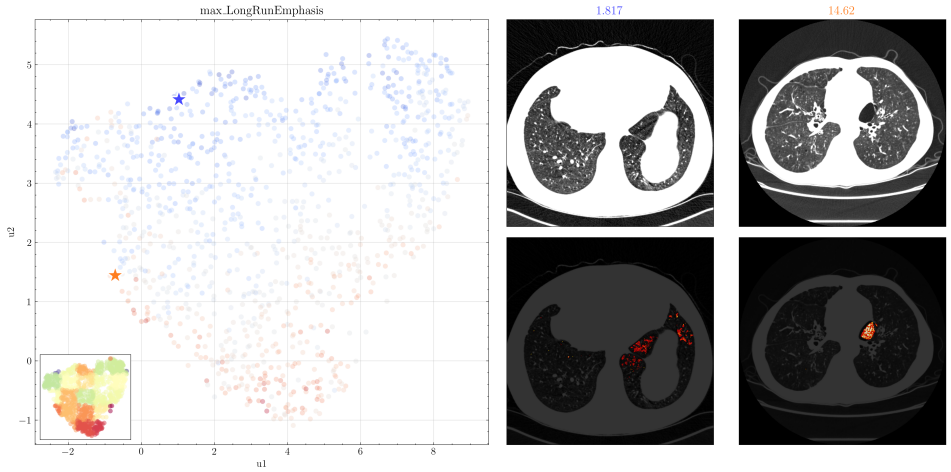


Figure 17. Maximum of Long Run Emphasis (LRE) distribution across the UMAP embedding (reference GOLD distribution in the bottom-left corner). Example images (indicated by stars on the scatter plot) with low and high feature values with their voxel-based feature heatmaps.

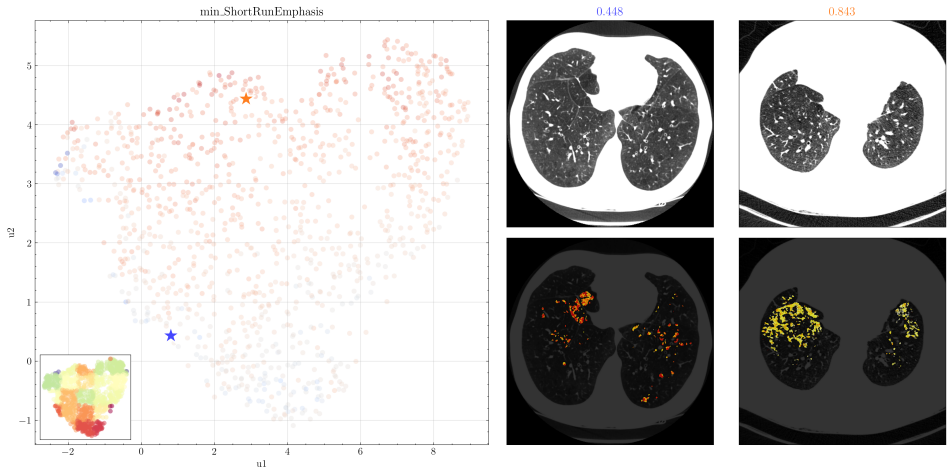


Figure 18. Minimum of Short Run Emphasis (SRE) distribution across the UMAP embedding (reference GOLD distribution in the bottom-left corner). Example images (indicated by stars on the scatter plot) with low and high feature values with their voxel-based feature heatmaps.

The second UMAP component (U2) mainly reflects the extent and homogeneity of emphysema within ROIs.

1. Standard deviation of Minor Axis Length indicates the number and size variability of ROIs—more varied ROIs suggest more widespread emphysema.
2. Run Percentage and Run Length Non-Uniformity Normalised both describe homogeneity; lower values signal more uniform emphysematous texture.
3. Long Run Emphasis (LRE) and Short Run Emphasis (SRE) also measure homogeneity but at different scales—LRE captures global patterns, SRE local ones.

Component U2, and the features driving it, primarily capture differences between GOLD severity groups, both for the full GOLD classification and for the combined GOLD 0–1–2 versus 3–4 grouping. In contrast, component U1 mainly reflects variability within individual GOLD groups, highlighting heterogeneity among patients of similar spirometric severity.

5. Discussion

The present study adopts an unsupervised, data-driven analysis strategy based on adaptive emphysema segmentation and region-level feature aggregation. By including

incipient emphysema through patient-specific thresholding, the resulting manifold representation captures structural variability that is not directly optimised for the prediction of standard metrics. This enables the identification of radiomic features shaping the embedding beyond conventional density-based measures, positioning the proposed framework as complementary to existing supervised radiomics approaches.

In contrast, most radiomics-based COPD studies adapt supervised learning frameworks that aim to classify disease presence, GOLD stage, or predict clinical or densitometric outcomes using CT-derived features extracted from predefined anatomical regions, most commonly the whole lung [11–13]. These approaches typically rely on feature selection methods such as LASSO or generalised linear models, followed by conventional machine-learning classifiers or neural networks, and demonstrate strong associations between radiomic features, emphysema burden, and established clinical metrics. More recent studies have gone a step further by analysing how radiomic features change over time and by including additional anatomical regions such as the airways and bronchi. Using features extracted from whole-lung and airway masks, these works have shown that radiomics can predict changes in lung density and the rate of emphysema progression [39], and can improve COPD detection in both low-dose and standard-dose CT scans [14]. While these approaches expand the clinical applications of radiomics, they still rely on predefined anatomical regions and supervised models trained to predict specific outcome measures.

While the GOLD stage is used as the standard for COPD severity, it has important limitations. It is based on spirometry and reflects functional airflow limitation, which can be influenced by both airway disease and parenchymal destruction. As a result, the GOLD stage does not translate directly into the actual amount of emphysematous changes. This is especially true in early disease when structural changes may be present before significant airflow limitation occurs. Imaging-based features, such as those presented in the article, can provide complementary information about the extent, distribution, and heterogeneity of emphysematous changes, resulting in a clearer interpretation of the COPD background.

In the created embedding, areas corresponding to clusters C02, C11, and C21 are particularly noteworthy. These clusters likely represent regions with incipient or diffuse emphysematous changes, characterised by subtle structural alterations that are faint or difficult to detect using conventional fixed-threshold approaches. Because the segmentation threshold is automatically adjusted on a per-patient basis, such regions may be included in the analysis even when clearly delineated emphysematous lesions are not present. Assigning these regions to distinct clusters allows them to be systematically separated from the core emphysema patterns while still preserving information about early or mild structural changes. This may be beneficial when analysing patients with early-stage disease, as it enables the identification of subtle emphysema-related alterations without confounding the characterisation of more advanced phenotypes.

The overlap in the GOLD 0, GOLD 1, and GOLD 2 clusters, GOLD 0 represents individuals who do not meet spirometric criteria for COPD but exhibit respiratory symptoms consistent with early disease. As such, this group is expected to overlap structurally with early GOLD stages rather than form a clearly distinct category. Quantitative imaging studies support this view. Konietzke et al. [40] reported only modest differences in emphysema burden between GOLD 0 and GOLD 1 patients, with emphysema index (EI) values of approximately $1.0 \pm 1.8\%$ and $4.5 \pm 9.9\%$, respectively. In contrast, GOLD 2 patients showed a marked increase in EI ($\sim 19.4 \pm 15.8\%$), accompanied by the largest variability among all stages, indicating substantial heterogeneity in structural lung changes.

Similar patterns have been reported in other large cohorts. Young et al. [41] observed overlapping emphysema phenotypes across early GOLD stages, although in their cohort, variability was more pronounced in GOLD 3 and GOLD 4. Importantly, GOLD 2 consis-

tently emerges as a heterogeneous transitional stage, reflecting the coexistence of relatively preserved lung tissue and regions of established emphysematous destruction. In addition, GOLD 2 typically represents the largest subgroup in clinical and observational studies, often substantially outnumbering GOLD 1 patients, as shown in studies by Konietzke et al. [40], Young et al. [41], and Li et al. [42]. This broad and heterogeneous patient composition further contributes to overlap with adjacent GOLD stages. Those findings support the interpretation that the overlap among GOLD 0, 1, and 2 in the radiomic embedding space reflects the biological continuum of early COPD rather than methodological limitations.

Sensitivity to UMAP parameter selection was assessed by testing multiple configurations, varying the minimum distance and the number of nearest neighbours while keeping the distance metric constant. Although changes in global rotation and compactness of the embedding were observed, the overall structure and relative relationships between patients were maintained, as shown in Figure 19. Neighbourhood consistency was further quantified using the Jaccard index, calculated for 70 nearest neighbours to match the reference embedding. A high degree of neighbourhood overlap was observed across parameter settings, indicating that the grouping patterns are not dependent on a specific configuration but reflect the underlying structure in the radiomic feature space.

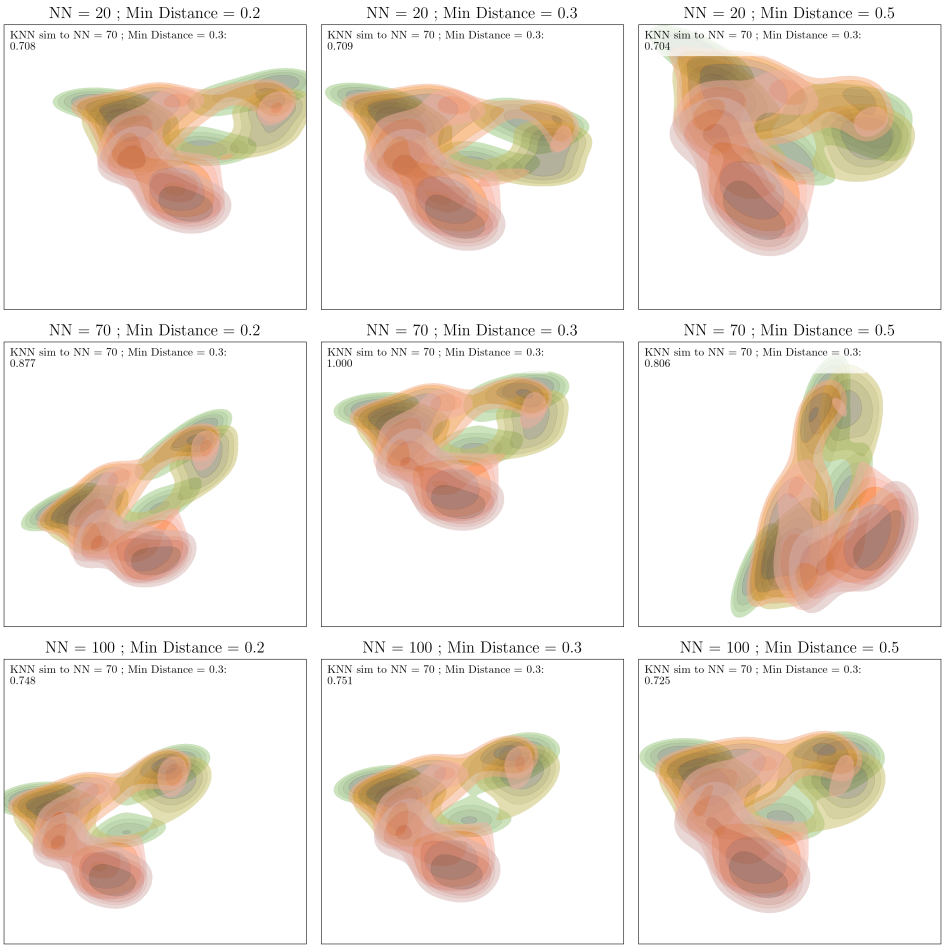


Figure 19. UMAP embedding using a cosine distance metric generated with different minimum distance and nearest-neighbour parameters, demonstrating consistency across embeddings obtained with varying parameter settings. Density plot colors represents GOLD 0 (green), GOLD 1 (yellow), GOLD 2 (orange), GOLD 3 (light red) and GOLD 4 (dark red) groups respectively. The mean Jaccard index was calculated to compare the neighbourhood structure of each embedding with the reference embedding. Overall, Jaccard index values ranged from 0.70 to 0.87, indicating strong overlap of point neighbourhoods.

6. Conclusions

The presented method enables the stratification of patients with chronic obstructive pulmonary disease (COPD) within a 2D embedding space, providing a visually interpretable representation of structural lung changes derived from chest CT. By applying patient-adaptive thresholds for emphysema segmentation, the method captures not only well-defined emphysematous regions in advanced disease but also diffuse and incipient structural changes in earlier stages that may not be detected by fixed-threshold approaches.

The extracted radiomic features differentiate patients across GOLD stages, separating GOLD 0–2 from GOLD 3–4, while also revealing substantial intra-stage heterogeneity, reflecting the complex and continuous nature of emphysema progression. Several features were identified as influential to the embedding structure, including the standard deviation of Minor Axis Length, the minimum of Run Percentage, the minimum of Run Length Non-Uniformity Normalised, the maximum of Long Run Emphasis, the minimum of Short Run Emphasis, the minimum and standard deviation of Run Entropy, the standard deviation of contour Small Area Emphasis, and the mean Surface Area.

Rather than being driven by a single dominant descriptor, the learnt embedding and SHAP-based analysis indicate that emphysema represents a multidimensional structural phenotype best characterised by the combined behaviour of multiple radiomic features. By incorporating information from both established emphysema and subtle early changes, this approach provides complementary information to standard emphysema metrics, supporting a more nuanced and interpretable assessment of COPD-related lung pathology.

Author Contributions: Conceptualisation, J.P. and M.S.; methodology, M.S.; software, M.S.; validation, J.P. and M.S.; formal analysis, E.S., W.R. and A.D.; investigation, M.S.; resources, M.J., A.D. and K.D.; data curation, M.S., M.J., A.D. and K.D.; writing—original draft preparation, M.S.; writing—review and editing, J.P. and E.S.; visualisation, M.S.; supervision, J.P.; project administration, J.P.; funding acquisition, J.P. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by Silesian University of Technology, no.02/070/BK26/0079 subsidy, and preliminary studies were funded by IP-CMC Pomeranian Interdisciplinary Centre of Digital Medicine, ABM, grant no.2023/ABM/02/00018.

Data Availability Statement: All analyses were performed on de-identified, publicly available data from the COPDGene study.

Acknowledgments: During the preparation of this manuscript/study, the authors used Grammarly v1.2.227.1811 and Writefull Premium 2025.68.0 for the purposes of grammar, spelling, punctuations checks and style correction of the sentences written by the authors of the article. Moreover, the DeepL version 26.1.1.19426 Translate was used for translation from Polish to English. The authors have reviewed and edited the output and take full responsibility for the content of this publication.

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Global Initiative for Chronic Obstructive Lung Disease. Global Strategy for the Diagnosis, Management, and Prevention of Chronic Obstructive Pulmonary Disease. 2025 Report. Available online: https://goldcopd.org/2025-gold-report/ (accessed on 12 January 2025).
2. Regan, E.A.; Hokanson, J.E.; Murphy, J.R.; Make, B.; Lynch, D.A.; Beaty, T.H.; Curran-Everett, D.; Silverman, E.K.; Crapo, J.D. Genetic Epidemiology of COPD (COPDGene) Study Design. *COPD J. Chronic Obstr. Pulm. Dis.* **2010**, *7*, 32–43. [CrossRef]
3. Lynch, D.A.; Austin, J.H.M.; Hogg, J.C.; Grenier, P.A.; Kauczor, H.U.; Bankier, A.A.; Barr, R.G.; Colby, T.V.; Galvin, J.R.; Gevenois, P.A.; et al. CT-Definable Subtypes of Chronic Obstructive Pulmonary Disease: A Statement of the Fleischner Society. *Radiology* **2015**, *277*, 192–205. [CrossRef]

4. Galbán, C.J.; Han, M.K.; Boes, J.L.; Chughtai, K.A.; Meyer, C.R.; Johnson, T.D.; Galbán, S.; Rehemtulla, A.; Kazerooni, E.A.; Martinez, F.J.; et al. Computed tomography–based biomarker provides unique signature for diagnosis of COPD phenotypes and disease progression. *Nat. Med.* **2012**, *18*, 1711–1715. [CrossRef]
5. Castaldi, P.J.; San José Estépar, R.; Mendoza, C.S.; Hersh, C.P.; Laird, N.; Crapo, J.D.; Lynch, D.A.; Silverman, E.K.; Washko, G.R. Distinct Quantitative Computed Tomography Emphysema Patterns Are Associated with Physiology and Function in Smokers. *Am. J. Respir. Crit. Care Med.* **2013**, *188*, 1083–1090. [CrossRef]
6. Nishio, M.; Tanaka, Y. Heterogeneity in pulmonary emphysema: Analysis of CT attenuation using Gaussian mixture model. *PLoS ONE* **2018**, *13*, e0192892. [CrossRef] [PubMed]
7. John, J.; Burrowes, K.S.; Bhatt, S.P.; Kumar, H.; O'Sullivan, J.M.; Golovina, E.; Jin, T.; Tawhai, M.H. Heterogeneity in Normal-Appearing Lung Tissue is Similar Between COPD and Smokers, and Predicts Morbidity and Progression in COPD. *Acad. Radiol.* **2025**, *32*, 7573–7580. [CrossRef] [PubMed]
8. Lambin, P.; Rios-Velazquez, E.; Leijenaar, R.; Carvalho, S.; Van Stiphout, R.G.; Granton, P.; Zegers, C.M.; Gillies, R.; Boellard, R.; Dekker, A.; et al. Radiomics: Extracting more information from medical images using advanced feature analysis. *Eur. J. Cancer* **2012**, *48*, 441–446. [CrossRef] [PubMed]
9. Lambin, P.; Leijenaar, R.T.; Deist, T.M.; Peerlings, J.; De Jong, E.E.; Van Timmeren, J.; Sanduleanu, S.; Larue, R.T.; Even, A.J.; Jochems, A.; et al. Radiomics: The bridge between medical imaging and personalized medicine. *Nat. Rev. Clin. Oncol.* **2017**, *14*, 749–762. [CrossRef]
10. Cheplygina, V.; Pena, I.P.; Pedersen, J.H.; Lynch, D.A.; Sørensen, L.; De Bruijne, M. Transfer learning for multicenter classification of chronic obstructive pulmonary disease. *IEEE J. Biomed. Health Inform.* **2017**, *22*, 1486–1496. [CrossRef]
11. Li, Z.; Liu, L.; Zhang, Z.; Yang, X.; Li, X.; Gao, Y.; Huang, K. A novel CT-based radiomics features analysis for identification and severity staging of COPD. *Acad. Radiol.* **2022**, *29*, 663–673. [CrossRef]
12. Yang, Y.; Li, W.; Guo, Y.; Zeng, N.; Wang, S.; Chen, Z.; Liu, Y.; Chen, H.; Duan, W.; Li, X.; et al. Lung radiomics features for characterizing and classifying COPD stage based on feature combination strategy and multi-layer perceptron classifier. *Math. Biosci. Eng.* **2022**, *19*, 7826–7855. [CrossRef]
13. Yang, Y.; Wang, S.; Zeng, N.; Duan, W.; Chen, Z.; Liu, Y.; Li, W.; Guo, Y.; Chen, H.; Li, X.; et al. Lung radiomics features selection for COPD stage classification based on auto-metric graph neural network. *Diagnostics* **2022**, *12*, 2274. [CrossRef] [PubMed]
14. Amudala Puchakayala, P.R.; Sthanam, V.L.; Nakhmani, A.; Chaudhary, M.F.; Kizhakke Puliyakote, A.; Reinhardt, J.M.; Zhang, C.; Bhatt, S.P.; Bodduluri, S. Radiomics for improved detection of chronic obstructive pulmonary disease in low-dose and standard-dose chest CT scans. *Radiology* **2023**, *307*, e222998. [CrossRef]
15. Gonzalez, G.; Ash, S.Y.; Vegas-Sánchez-Ferrero, G.; Onieva Onieva, J.; Rahaghi, F.N.; Ross, J.C.; Díaz, A.; San José Estépar, R.; Washko, G.R. Disease staging and prognosis in smokers using deep learning in chest computed tomography. *Am. J. Respir. Crit. Care Med.* **2018**, *197*, 193–203. [CrossRef]
16. Singla, S.; Gong, M.; Riley, C.; Sciurba, F.; Batmanghelich, K. Improving clinical disease subtyping and future events prediction through a chest CT-based deep learning approach. *Med. Phys.* **2021**, *48*, 1168–1181. [CrossRef] [PubMed]
17. Xu, C.; Qi, S.; Feng, J.; Xia, S.; Kang, Y.; Yao, Y.; Qian, W. DCT-MIL: Deep CNN transferred multiple instance learning for COPD identification using CT images. *Phys. Med. Biol.* **2020**, *65*, 145011. [CrossRef]
18. Du, R.; Qi, S.; Feng, J.; Xia, S.; Kang, Y.; Qian, W.; Yao, Y.D. Identification of COPD from multi-view snapshots of 3D lung airway tree via deep CNN. *IEEE Access* **2020**, *8*, 38907–38919. [CrossRef]
19. Tang, L.Y.; Coxson, H.O.; Lam, S.; Leipsic, J.; Tam, R.C.; Sin, D.D. Towards large-scale case-finding: Training and validation of residual networks for detection of chronic obstructive pulmonary disease using low-dose CT. *Lancet Digit. Health* **2020**, *2*, e259–e267. [CrossRef]
20. Ho, T.T.; Kim, T.; Kim, W.J.; Lee, C.H.; Chae, K.J.; Bak, S.H.; Kwon, S.O.; Jin, G.Y.; Park, E.K.; Choi, S. A 3D-CNN model with CT-based parametric response mapping for classifying COPD subjects. *Sci. Rep.* **2021**, *11*, 34. [CrossRef]
21. Chen, L.; Feng, Q.; Yin, X.; Min, X.; Shi, L.; Yang, D.; Chen, Y.W.; Zhang, D.; Zhu, W. A graph convolutional multiple instance learning on a hypersphere manifold approach for diagnosing chronic obstructive pulmonary disease in CT images. *IEEE J. Biomed. Health Inform.* **2022**, *26*, 6058–6069. [CrossRef]
22. Zhang, L.; Jiang, B.; Wisselink, H.J.; Vliegenthart, R.; Xie, X. COPD identification and grading based on deep learning of lung parenchyma and bronchial wall in chest CT images. *Br. J. Radiol.* **2022**, *95*, 20210637. [CrossRef]
23. Wu, Y.; Du, R.; Feng, J.; Qi, S.; Pang, H.; Xia, S.; Qian, W. Deep CNN for COPD identification by Multi-View snapshot integration of 3D airway tree and lung field. *Biomed. Signal Process. Control* **2023**, *79*, 104162. [CrossRef]
24. Sun, J.; Liao, X.; Yan, Y.; Zhang, X.; Sun, J.; Tan, W.; Liu, B.; Wu, J.; Guo, Q.; Gao, S.; et al. Detection and staging of chronic obstructive pulmonary disease using a computed tomography–based weakly supervised deep learning approach. *Eur. Radiol.* **2022**, *32*, 5319–5329. [CrossRef]
25. Li, Z.; Huang, K.; Liu, L.; Zhang, Z. Early detection of COPD based on graph convolutional network and small and weakly labeled data. *Med. Biol. Eng. Comput.* **2022**, *60*, 2321–2333. [CrossRef]

26. Xue, M.; Jia, S.; Chen, L.; Huang, H.; Yu, L.; Zhu, W. CT-based COPD identification using multiple instance learning with two-stage attention. *Comput. Methods Programs Biomed.* **2023**, *230*, 107356. [CrossRef]
27. Yu, K.; Sun, L.; Chen, J.; Reynolds, M.; Chaudhary, T.; Batmanghelich, K. DrasCLR: A self-supervised framework of learning disease-related and anatomy-specific representation for 3D lung CT images. *Med. Image Anal.* **2024**, *92*, 103062. [CrossRef]
28. DeCamp, M.; Tilburt, J.C. Why we cannot trust artificial intelligence in medicine. *Lancet Digit. Health* **2019**, *1*, e390. [CrossRef] [PubMed]
29. Hatherley, J.J. Limits of trust in medical AI. *J. Med. Ethics* **2020**, *46*, 478–481. [CrossRef]
30. Beck, J.T.; Rammage, M.; Jackson, G.P.; Preininger, A.M.; Dankwa-Mullan, I.; Roebuck, M.C.; Torres, A.; Holtzen, H.; Coverdill, S.E.; Williamson, M.P.; et al. Artificial Intelligence Tool for Optimizing Eligibility Screening for Clinical Trials in a Large Community Cancer Center. *JCO Clin. Cancer Inform.* **2020**, *4*, 50–59. [CrossRef] [PubMed]
31. McKinney, S.M.; Sieniek, M.; Godbole, V.; Godwin, J.; Antropova, N.; Ashrafian, H.; Back, T.; Chesus, M.; Corrado, G.S.; Darzi, A.; et al. International evaluation of an AI system for breast cancer screening. *Nature* **2020**, *577*, 89–94. [CrossRef]
32. Hofmanninger, J.; Prayer, F.; Pan, J.; Röhrich, S.; Prosch, H.; Langs, G. Automatic lung segmentation in routine imaging is primarily a data diversity problem, not a methodology problem. *Eur. Radiol. Exp.* **2020**, *4*, 50. [CrossRef]
33. Socha, M. Unsupervised Airway Segmentation for Enhanced Analysis of COPD in Lung Cancer Patients Using Dual-Velocity Growing. 2025. Available online: https://epos.myesr.org/poster/esr/ecr2025/C-21943 (accessed on 12 December 2025).
34. Binczyk, F.; Stjelties, B.; Weber, C.; Goetz, M.; Meier-Hein, K.; Meinzer, H.P.; Bobek-Billewicz, B.; Tarnawski, R.; Polanska, J. MiMSeg—An algorithm for automated detection of tumor tissue on NMR apparent diffusion coefficient maps. *Inf. Sci.* **2017**, *384*, 235–248. [CrossRef]
35. McInnes, L.; Healy, J.; Saul, N.; Großberger, L. UMAP: Uniform Manifold Approximation and Projection. *J. Open Source Softw.* **2018**, *3*, 861. [CrossRef]
36. Lundberg, S.M.; Lee, S.I. A Unified Approach to Interpreting Model Predictions. In *Advances in Neural Information Processing Systems 30*; Guyon, I., Luxburg, U.V., Bengio, S., Wallach, H., Fergus, R., Vishwanathan, S., Garnett, R., Eds.; Curran Associates, Inc.: Red Hook, NY, USA, 2017; pp. 4765–4774.
37. Clevert, D.A.; Unterthiner, T.; Hochreiter, S. Fast and Accurate Deep Network Learning by Exponential Linear Units (ELUs). *arXiv* **2016**, arXiv:1511.07289. [CrossRef]
38. Hopkins, W.; Marshall, S.; Batterham, A.; Hanin, J. Progressive statistics for studies in sports medicine and exercise science. *Med. Sci. Sports Exerc.* **2009**, *41*, 3–13. [CrossRef]
39. Saha, P.; Bodduluri, S.; Nakhmani, A.; Chaudhary, M.F.A.; Amudala Puchakayala, P.R.; Sthanam, V.; San Jose Estepar, R.; Reinhardt, J.M.; Zhang, C.; Bhatt, S.P. Computed Tomography Radiomics Features Predict Change in Lung Density and Rate of Emphysema Progression. *Ann. Am. Thorac. Soc.* **2025**, *22*, 83–92. [CrossRef] [PubMed]
40. Konietzke, P.; Brunner, C.; Konietzke, M.; Wagner, W.L.; Weinheimer, O.; Heußel, C.P.; Herth, F.J.F.; Trudzinski, F.; Kauczor, H.U.; Wielpütz, M.O. GOLD stage-specific phenotyping of emphysema and airway disease using quantitative computed tomography. *Front. Med.* **2023**, *10*, 1184784. [CrossRef] [PubMed]
41. Young, K.A.; Strand, M.J.; Ragland, M.F.; Kinney, G.L.; Austin, E.E.; Regan, E.A.; Lowe, K.E.; Make, B.J.; Silverman, E.K.; Crapo, J.D.; et al. Pulmonary Subtypes Exhibit Differential Global Initiative for Chronic Obstructive Lung Disease Spirometry Stage Progression: The COPDGene® Study. *Chronic Obstr. Pulm. Dis. J. COPD Found.* **2019**, *6*, 414–429. [CrossRef]
42. Li, C.L.; Chang, H.C.; Tseng, C.W.; Tsai, Y.C.; Liu, J.F.; Chan, C.C.; Tsai, M.L.; Liu, S.F. The DOSE index in chronic obstructive pulmonary disease: Evaluating healthcare costs. *BMC Pulm. Med.* **2024**, *24*, 560. [CrossRef] [PubMed]

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.