Hierarchical multiple instance learning for COPD grading with relatively specific similarity

Hao Zhang[1,3], Mingyue Zhao[2,3], Mingzhu Liu[6], Jiejun Luo[6], Yu Guan[7], Jin Zhang[7], Yi Xia[7], Di Zhang[7], Xiuxiu Zhou[7], Li Fan[7], Shiyuan Liu[7], and S. Kevin Zhou[2,3,4,5]

[1] School of Computer Science and Technology, University of Science and Technology of China

[2] School of Biomedical Engineering, Division of Life Sciences and Medicine, University of Science and Technology of China (USTC), Hefei Anhui, 230026, China

[3] Center for Medical Imaging, Robotics, Analytic Computing & Learning (MIRACLE), Suzhou Institute for Advance Research, USTC, Suzhou Jiangsu, 215123, China

[4] Key Laboratory of Precision and Intelligent Chemistry, USTC, Hefei Anhui,230026, China

[5] Key Laboratory of Intelligent Information Processing of Chinese Academy of Sciences (CAS), Institute of Computing Technology, CAS

[6] Department of Radiology, Tongji Hospital of Tongji University

[7] Department of Radiology, the Second Affiliated Hospital of PLA Naval Medical University

Abstract. Chronic obstructive pulmonary disease (COPD) is a type of obstructive lung disease characterized by persistent airflow limitation and ranks as the third leading cause of death globally. As a heterogeneous lung disorder, the diversity of COPD phenotypes and the complexity of its pathology pose significant challenges for recognizing its grade. Many existing deep learning models based on 3D CT scans overlook the spatial position information of lesion regions and the correlation within different lesion grades. To this, we define the COPD grading task as a multiple instance learning (MIL) task and propose a hierarchical multiple instance learning (H-MIL) model. Unlike previous MIL models, our H-MIL model pays more attention to the spatial position information of patches and achieves a fine-grained classification of COPD by extracting patch features in a multi-level and granularity-oriented manner. Furthermore, we recognize the significant correlations within lesions of different grades and propose a Relatively Specific Similarity (RSS) function to capture such relative correlations. We demonstrate that H-MIL achieves better performances than comparative methods on an internal dataset comprising 2,142 CT scans. Additionally, we validate the effectiveness of the model architecture and loss design through an ablation study. and the robustness of our model on different central datasets. Code is available at https://github.com/Mars-Zhang123/H-MIL.git.

Keywords: COPD Grading · Hierachical Mutiple Instance Learning · Relatively Specific Similarity Loss.

1 Introduction

Chronic obstructive pulmonary disease (COPD) is a prevalent chronic airflow obstruction disorder, characterized by a high incidence, mortality, and disability rate [17]. Numerous studies [6, 8, 15, 18] have indicated that early diagnosis of COPD can effectively prevent irreversible damage to the lungs. However, existing clinical diagnostic guidelines [22], which rely on pulmonary function tests, have been shown in many recent studies [2, 13] to significantly lag behind early structural changes in lung function(such as early remodeling of small airways and small blood vessels). Image-based pulmonary function assessments, such as CT scans, offer more intuitive and comprehensive pulmonary pathological structural changes, crucial for the diagnosis, and intervention, particularly early warning of COPD [4, 14, 19].

In recent years, deep learning methods have demonstrated excellent performance in various medical image tasks due to their superior ability to analyze complex data and perceive fine structures, showing a great potential for image-based COPD diagnosis [7,11,16,21]. Nevertheless, as the original lung CT volume is large, the objective of this study, which aims to achieve case-level disease interpretation through comprehensive and detailed image understanding, evidently contradicts the limitation of computational resources.

To handle this problem, most researchers endeavor to enhance model inputs to achieve information sparsification or information refinement. Ahmed *et al.* [1] utilize the resized CT volume as input to predict positive cases using a voxelResNet variant model. Hatt *et al.* [9] divide the lung CT into multiple regions, randomly sample slices from different regions, downsample the slices, and input the samples into a 2D CNN model. These methods attempt to directly compress CT information or utilize networks to self-distill from a vast array of complex data, aiming to achieve refined disease interpretation with a limited computational burden, which directly results in the loss of fine structures or limited perception capability of the network towards fine structures. *The challenge of extracting meaningful information from vast and complex data to achieve fine-grained disease interpretation remains.* Moreover, existing methods mostly focus on achieving early warning of COPD, which is essentially a binary classification problem. Studies related to refined grading diagnosis lag behind, and some directly treat the COPD grading task as a multi-class classification problem [20, 24]. *This fundamentally ignores the correlation between labels, namely, the similarity and diversity of disease manifestations.*

To address the aforementioned challenges, tailored to fine-grained COPD classification, we present a hierarchical multi-instance learning approach with a relatively specific similarity. Specifically, this involves the following two main contributions. **1) Hierarchical multi-instance learning (H-MIL).** To tackle the challenges posed by limited computational resources and constrained model information extraction capacity, we propose a Hierarchical multi-instance learning framework for a step-wise attention fusion and information extraction at three levels: pixel-level, slice-level, and sub-bag level, through spatial attention fusion and sub-bag design, enabling progressive information refinement. **2) Rel-**

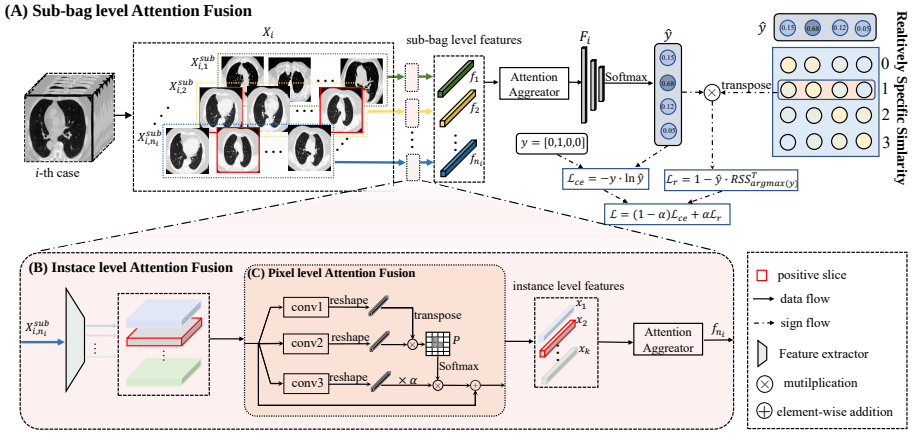

Fig. 1. Overview of our proposed Hierarchical Multi-Instance Learning.

atively Specific Similarity (RSS). To learn the correlation between labels, we innovatively introduce the Relatively Specific Similarity (RSS) into a loss to drive network learning. It enables the network to achieve progressive label association learning on top of the original cross-entropy loss, enhancing the model's perception of multi-class correlations and classification performance.

Experiments are conducted for the two tasks of COPD binary classification and COPD grading based on an in-house dataset as well as an external dataset, demonstrating that the H-MIL successively encourages the attention on crucial regions for classification, the correlation between the labels at the feature level, modeled by RSS, has a great impact, and our proposed method outperforms comparative method by a pronounced margin.

2 Method

In this task, the training dataset of labeled CT volumes is denoted as $\mathcal{D} = \{X_i, Y_i\}_{i=1}^{|\mathcal{D}|}$, where $X_i = \{x_{i,j}\}_{j=1}^{n_i}$ represents the i-th bag (i.e., CT volume). Our objective is to learn the mapping : $X \rightarrow Y$ to realize the grading prediction of COPD of each CT image. As shown in Fig. 1, there are two parts of our method: 1) Hierarchical Multi-Instance Learning, including pixel-level, slice-level, and sub-bag level attention fusion, and 2) Relatively Specific Similarity.

2.1 Hierarchical Multiple Instance Learning

To alleviate computational burden while preserving comprehensive fine-grained structural information as much as possible, we treat a 3D CT as a stack of 2D slices, that is, **a slice is an instance**. Then, we introduce the concept of sub-bags, each of which is regarded as a subset of a bag including a group of instances.

Let $X_{i,j}^{sub}$ denote the j-th sub-bag split from the i-th bag X_i, then $X_{i,j}^{sub} = \{x_{i,j,k}\}_{k=1}^{n_j}$, where $x_{i,j,k} \in \mathbb{R}^d$ is the k-th instance in this sub-bag after feature extraction. It is noteworthy that, to achieve sub-bag level attention perception, the sub-bag labels here are unknown, which differs from the "pseudo bag" design inherited from the bag label in [23].

In this paper, we employ attention mechanisms to fusion sub-bag level feature [12]. The bag level features F_i can be formulated as follows:

$$F_i = \sum_{j=1}^{n_i} a_{i,j} f_{i,j}, \tag{1}$$

where $f_{i,j} \in \mathbb{R}^d$ represents $X_{i,j}^{sub}$ feature, and $a_{i,j} \in \mathbb{R}$ represents its attention scores in the sub-bag level, which can be derived by the attention gate:

$$a_{i,j} = \frac{\exp\{w^T(\tanh(V f_{i,j}^T) \odot \mathrm{sigm}(U f_{i,j}^T))\}}{\sum_{k=1}^{n_i} \exp\{w^T(\tanh(V f_{i,k}^T) \odot \mathrm{sigm}(U f_{i,k}^T))\}}, \tag{2}$$

where $w \in \mathbb{R}^{l\times 1}$,$U \in \mathbb{R}^{l\times d}$,$V \in \mathbb{R}^{l\times d}$ are parameters and $\odot$ is an element-wise multiplication. H-MIL allows assigning different attention scores to various sub-bags, enabling the network to identify crucial sub-bags more effectively. In sub-bag level attention fusion, the model is unable to locate the specific positive instance within the crucial sub-bag. Therefore, we require slice level attention fusion to obtain $f_{i,j}$. Slice level attention fusion employs the same method as sub-bag attention fusion. After the slice level attention fusion , F_i can be formulated as follows:

$$F_i = \sum_{j=1}^{n_i} a_{i,j} \sum_{k=1}^{n_j} b_{i,j,k} x_{i,j,k}, \tag{3}$$

where $b_{i,j,k} \in \mathbb{R}$ represents the instance level scores of $x_{i,j,k}$. Through slice level attention fusion, the model can accurately locate those crucial instances.

We observe that lesions are unevenly distributed across different regions even within a single instance. The main reason for this is that a single instance typically contains various types of lung structures, and these structures are dispersed in different regions. As a result, the contribution of each region to the task varies, which limits the model's ability to accurately locate the fine structure of diseased lesion areas. Inspired by [5], we denote the feature after preliminary encoding as x', then we convolve the x' with three different convolution kernels to obtain three features $conv_1$, $conv_2$, and $conv_3$ of the same shape. The final instance-level feature map x with pixel level attention can be obtained by the following formula:

$$x = \alpha \times conv_3 \cdot P + x' = \alpha \times conv_3 \cdot \mathrm{softmax}({conv_1}^T \cdot conv_2) + x'. \tag{4}$$

where $x', conv_1, conv_2, conv_3, x \in \mathbb{R}^{C\times(H\times W)}$ and $P \in \mathbb{R}^{(H\times W)\times(H\times W)}$. P is the pixel-level position attention matrix, and each value of the matrix represents the relationship strength between two location pixels. Through the scale coefficient

α, the allocated attention weight is dynamically adjusted. Combined with the slice level attention map, H-MIL has a global contextual view at the instance level, which greatly enhance its attention to the crucial pixels of regions inside the instance.

2.2 Relatively Specific Similarity

In the task of COPD grading, there are four different GOLD grades based on the FEV_1 [22]: $\{\text{GOLD}_i;\ i = 1 : 4\}$, with the severity of COPD increasing from 1 to 4. From the perspective of similarity, we construct a corresponding model to acquire and articulate these challenging similarities. In the real world, the number of classes is infinite as FEV_1 is continuously-valued, but we can create an abstract **base class** that maintains the same feature similarity with all classes. We define the base class as β, the set of all classes as C, and the feature similarity between classes as γ. Then, we can express this relationship using the following formula:

$$\forall \eta \in C \longrightarrow \gamma(\beta, \eta) = \mu. \tag{5}$$

where η denotes an arbitrary class and μ is a positive constant. Considering the similarity of features, we can represent the feature similarity between all classes in the real world using a large adjacency matrix $S \in \mathbb{R}^{n \times n}$, whose entry $s(i, j) = \gamma(i, j)$ represents the similarity scores between the classes i and j. The notion of **absolutely abstract similarity (AAS)** is defined as:

$$AAS_{i,j} = \lim_{n \to \infty} \frac{\gamma(i, j)}{z} = \lim_{n \to \infty} \frac{\gamma(i, j)}{\sum_{c=1}^{n} \gamma(\beta, c)}. \tag{6}$$

where z is the normalization factor that represents the sum of similarities between all classes and the base class. The AAS satisfies a list of interesting properties that are provided in the supplementary materials.

The concept of AAS is extremely helpful in understanding the similarity between classes, but calculating the corresponding AAS for solving a specific problem becomes nearly impossible due to the enormity of computing the similarity between a large number of classes. In order to solve a multi-class classification problem, we usually only focus on the classes that are relevant to the specific problem. Therefore, there is no need to calculate all the AAS in real-world scenarios. As a result, we introduce the concept of **relatively specific similarity (RSS)** based on AAS. RSS solely considers the classes involved in a specific problem and can be calculated as opposed to ASS. Similar to ASS, we define $RSS_{i,j}$ as the similarity between class i and class j in a specific problem, which can be obtained by the following formula:

$$RSS_{i,j} = \frac{AAS_{i,j}}{\sum_{k=1}^{C} AAS_{i,k}}, \tag{7}$$

where C is the number of classes involved in this classification problem. The RRS also satisfies a list of interesting properties that are provided in the supplementary materials.

Based on RSS, we design a loss function as the following:

$$loss_r = 1 - \frac{\hat{y} \cdot RSS^T_{\text{argmax}(y)}}{Z_{\text{argmax}(y)}}, \quad Z_i = \sum_{k=1}^{C} RSS_{i,k}. \tag{8}$$

where $\hat{y}, y \in \mathbb{R}^{1\times C}$, and $Z_i \in \mathbb{R}$ is a normalization factor for the i-th class.

The supervision of CE loss provides us with the opportunity to learn the correlation between classes. In other words, the prediction distribution of the model acts as a type of soft label and implies the relative similarity between classes. Based on this view, we construct RSS to capture this similarity and utilize it as an additional form of supervision. As we increase the weight of RSS loss, the model undergoes a transition from hard to soft labels, thereby continuously enhancing its understanding of inter-class similarity through self-distillation:

$$Loss = (1-\alpha)loss_{ce} + \alpha \cdot loss_r. \tag{9}$$

where α is parameter to balance $loss_{ce}$ and $loss_r$, which will increase from very small values when the model is initially trained.

3 Experiments

Dataset Experiments are conducted using an in-house dataset of 2,142 normal-dose CT scans acquired at end-inspiration. The tasks include binary classification and GOLD grading of COPD. For binary classification, the labels are relatively balanced, with 43% positive and 57% negative. However, within the positive cases, the distribution of gold grades is extremely unbalanced, with proportions of GOLD_1 to GOLD_4 being 18.5%, 54.1%, 20.1%, and 7.3%, respectively.

Implementation details To mitigate the instability caused by random sampling, we randomly select samples from each label in equal proportions to form the training set, validation set, and test set for each training session. This approach enhances the consistency of distribution among these three sets. The ratio of the training set, validation set, and test set is 7:1:2. In each training session, the model with the best performance on the validation set is selected as the inference for the test set. All CT cases undergo the processing before being inputted into the prediction model, which leverages DenseNet121 [10] as the backbone network in our experiments.

Main results To verify the effectiveness of our proposed method, it is compared with the mainstream MIL methods [3, 20] and the 3D CNN-based method [1] both on the binary classification task and fine-grained grading task of COPD. From Table 1 that reports the results of our approach and the comparative methods, we observe the following. **1)** our approach demonstrates superior performance on both tasks. In particular, in the task of COPD grading, this superiority becomes even more apparent (+15.9% in precision, +16.4% in recall

and +16.4% in F-score) in contrast with the second-best results. **2)** Our method showcases the highest recall scores both in binary classification and grading tasks, which is very crucial in clinical practice, particularly for the early detection of diseases. **3)** With our proposed method, utilizing stacked 2D slices as input outperforms using 3D patches, indicating that for diseases with diffuse lesions like COPD, 2D slices can better preserve pulmonary structure and spatial information, thereby enhancing the model's ability to perceive abnormal regions.

Table 1. Quantitative results (%) of different methods on binary classification and grading of COPD (Mean±Standard deviattion).

I : COPD binary classification					
Method	Accuracy	Precision	Recall	F-score	AUC
3D CNN [1]	$70.1_{\pm 1.3}$	$68.3_{\pm 1.7}$	69.1±1.2	$68.5_{\pm 1.4}$	$70.2_{\pm 1.5}$
MIL+Max-pooling	$78.1_{\pm 0.6}$	$76.5_{\pm 0.9}$	$77.2_{\pm 1.3}$	$76.8_{\pm 1.0}$	$79.1_{\pm 0.3}$
MIL+Avg-pooling	$72.1_{\pm 1.6}$	$70.8_{\pm 1.3}$	$70.9_{\pm 1.5}$	$70.8_{\pm 1.5}$	$73.6_{\pm 0.8}$
MIL+RNN [3]	$79.5_{\pm 0.3}$	$78.1_{\pm 0.2}$	$77.7_{\pm 0.3}$	$77.9_{\pm 0.2}$	$80.4_{\pm 0.1}$
MIL+Attn with 3D patches [20]	$78.3_{\pm 0.3}$	$78.0_{\pm 0.1}$	$77.6_{\pm 0.2}$	$77.8_{\pm 0.1}$	$79.5_{\pm 0.2}$
MIL+Attn with 2D slices [20]	$79.4_{\pm 0.3}$	$79.6_{\pm 0.8}$	$78.9_{\pm 0.3}$	$79.3_{\pm 0.6}$	$81.0_{\pm 0.4}$
H-MIL with 3D patches	$85.4_{\pm 0.4}$	$84.8_{\pm 0.5}$	$85.3_{\pm 0.4}$	$85.1_{\pm 0.4}$	$86.7_{\pm 0.3}$
H-MIL with 2D slices(Ours)	$\mathbf{88.5}_{\pm 0.2}$	$\mathbf{88.0}_{\pm 0.4}$	$\mathbf{88.7}_{\pm 0.1}$	$\mathbf{88.5}_{\pm 0.3}$	$\mathbf{89.6}_{\pm 0.3}$
II : COPD grading					
3D CNN [1]	$40.6_{\pm 3.9}$	$37.4_{\pm 2.7}$	$39.6_{\pm 4.5}$	$38.1_{\pm 3.3}$	$44.1_{\pm 2.3}$
MIL+Max-pooling	$48.3_{\pm 1.3}$	$44.6_{\pm 2.2}$	$45.3_{\pm 1.0}$	$45.0_{\pm 1.8}$	$49.4_{\pm 1.1}$
MIL+Avg-pooling	$45.0_{\pm 2.2}$	$43.7_{\pm 1.3}$	$44.0_{\pm 1.5}$	$43.8_{\pm 1.5}$	$45.1_{\pm 1.9}$
MIL+RNN [3]	$54.3_{\pm 1.2}$	$53.6_{\pm 0.7}$	$52.5_{\pm 0.8}$	$53.2_{\pm 0.7}$	$55.9_{\pm 0.9}$
MIL+Attn with 3D patches [20]	$51.9_{\pm 1.1}$	$50.9_{\pm 0.9}$	$51.6_{\pm 0.7}$	$51.2_{\pm 0.8}$	$53.0_{\pm 0.9}$
MIL+Attn with 2D slices [20]	$50.8_{\pm 0.7}$	$48.9_{\pm 0.6}$	$49.2_{\pm 0.5}$	$49.0_{\pm 0.5}$	$51.4_{\pm 0.6}$
H-MIL with 3D patches	$65.7_{\pm 0.8}$	$64.6_{\pm 1.2}$	$65.0_{\pm 1.0}$	$64.9_{\pm 1.1}$	$66.9_{\pm 0.6}$
H-MIL with 2D slices(Ours)	$\mathbf{69.0}_{\pm 0.9}$	$\mathbf{68.5}_{\pm 0.6}$	$\mathbf{68.9}_{\pm 0.7}$	$\mathbf{68.6}_{\pm 0.6}$	$\mathbf{69.6}_{\pm 0.5}$

Ablation study To demonstrate the effectiveness of sub-bag, pixel level attention fusion (PLAF) in H-MIL and RSS-based loss, several comparative studies are carried out in Table 2. **1)** The significant performance improvement delivered by the pixel level attention fusion confirms that the importance of pixel attention inside instance. **2)** We show the superiority of sub-bag by introducing sub-bag partitioning, which improves the AUC by 2.8%. **3)** By introducing RSS-based loss, all the metrics of the model have been significantly improved, which strongly demonstrates that the correlation between labels at the feature level has a significant impact on COPD grading. **4)** By halving the number of

instances, the AUC of the model only shows a small decrease, which demonstrates that our model does not depend on a substantial number of instances as input and can achieve exceptional performance with only a moderate number of instances. **5)** To verify the influence of the number of sub-bags on the model, we reduce the number of sub-bags to either the minimum or add to the maximum while keeping the total number of instances unchanged. The results indicate that, there is a certain extent of reduction in model performance, which confirms that the number of sub-bags significantly affects feature attention aggregation in the model.

Table 2. Results(%) of Ablation study of our proposed method (Mean±Standard deviation).

Ablation for COPD grading							
Method	Params		Metrics				
Method	n_i	n_j	Accuracy	Precision	Recall	F-score	AUC
baseline	-	24	$50.8_{\pm 0.7}$	$48.9_{\pm 0.6}$	$49.2_{\pm 0.5}$	$49.0_{\pm 0.5}$	$51.4_{\pm 0.6}$
w/ PLAF	-	24	$55.2_{\pm 1.6}$	$56.1_{\pm 1.3}$	$54.9_{\pm 1.0}$	$55.4_{\pm 1.1}$	$56.3_{\pm 0.9}$
w/ PLAF & sub-bag	4	6	$58.4_{\pm 1.1}$	$57.3_{\pm 1.4}$	$58.0_{\pm 0.9}$	$57.7_{\pm 1.2}$	$59.1_{\pm 1.5}$
w/ PLAF & sub-bag & $Loss_r$	4	6	$69.0_{\pm 0.9}$	$68.5_{\pm 0.6}$	$68.9_{\pm 0.7}$	$68.6_{\pm 0.6}$	$69.6_{\pm 0.5}$
w/ PLAF & sub-bag & $Loss_r$	4	3	$67.7_{\pm 0.5}$	$67.3_{\pm 0.6}$	$66.7_{\pm 0.9}$	$66.9_{\pm 0.7}$	$68.4_{\pm 0.8}$
w/ PLAF & sub-bag & $Loss_r$	2	6	$66.8_{\pm 0.5}$	$66.4_{\pm 0.5}$	$66.5_{\pm 0.4}$	$66.4_{\pm 0.4}$	$67.3_{\pm 0.5}$
w/ PLAF & sub-bag & $Loss_r$	6	2	$66.1_{\pm 0.8}$	$66.0_{\pm 0.3}$	$65.7_{\pm 0.5}$	$65.8_{\pm 0.4}$	$66.5_{\pm 0.5}$

External validation To further validate the generalization ability of our model on low-dose CT datasets, we additionally employ a dataset curated from another clinical site for testing purposes. This dataset comprises 118 negatives and 220 positives. We compare our method with MIL+Attn and 3D CNN using this dataset. The AUC of our method is 87.3%, which indicates a reduction of 2.3% compared to its performance on the previous test set. Notably, this reduction is significantly smaller than the respective reduction of 5.4% for MIL+Attn citesun2022detection and 4.7% for 3D CNN, thereby further substantiating the robustness of our model across different data centers.

4 Conclusion

In this paper, we propose a hierarchical multiple-instance learning method for COPD grading with a relatively specific similarity. The hierarchical multi-instance learning strategy achieves a progressive attention fusion and an effective information refinement by introducing pixel-level fusion, slice-level fusion, and sub-bag level fusion, enabling resource-friendly and fine-grained interpretation of lung

lesions. Furthermore, to fully exploit the inter-label correlations in the task of COPD grading, we introduce a Relatively Specific Similarity in the loss function that constrains the model to continuously learn the correlations between disease severity levels, thereby assisting the model in achieving better and more robust disease grading. Extensive experiments illustrate the superior performance of our method over comparative approaches both in binary and multi-class grading tasks for COPD, thus highlighting the algorithm's great potential in clinical disease diagnosis.

Disclosure of Interests. The authors have no competing interests to declare that are relevant to the content of this article.

References

1. Ahmed, J., Vesal, S., Durlak, F., Kaergel, R., Ravikumar, N., Rémy-Jardin, M., Maier, A.: Copd classification in ct images using a 3d convolutional neural network. In: Bildverarbeitung für die Medizin 2020: Algorithmen–Systeme–Anwendungen. Proceedings des Workshops vom 15. bis 17. März 2020 in Berlin. pp. 39–45. Springer (2020)
2. Bhatt, S.P., Balte, P.P., Schwartz, J.E., Cassano, P.A., Couper, D., Jacobs, D.R., Kalhan, R., O'Connor, G.T., Yende, S., Sanders, J.L., et al.: Discriminative accuracy of fev1: Fvc thresholds for copd-related hospitalization and mortality. Jama **321**(24), 2438–2447 (2019)
3. Campanella, G., Hanna, M.G., Geneslaw, L., Miraflor, A., Werneck Krauss Silva, V., Busam, K.J., Brogi, E., Reuter, V.E., Klimstra, D.S., Fuchs, T.J.: Clinical-grade computational pathology using weakly supervised deep learning on whole slide images. Nature medicine **25**(8), 1301–1309 (2019)
4. El Kaddouri, B., Strand, M.J., Baraghoshi, D., Humphries, S.M., Charbonnier, J.P., van Rikxoort, E.M., Lynch, D.A.: Fleischner society visual emphysema ct patterns help predict progression of emphysema in current and former smokers: results from the copdgene study. Radiology **298**(2), 441–449 (2021)
5. Fu, J., Liu, J., Tian, H., Li, Y., Bao, Y., Fang, Z., Lu, H.: Dual attention network for scene segmentation. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. pp. 3146–3154 (2019)
6. Gershon, A.S., Thiruchelvam, D., Chapman, K.R., Aaron, S.D., Stanbrook, M.B., Bourbeau, J., Tan, W., To, T., Network, C.R.R., et al.: Health services burden of undiagnosed and overdiagnosed copd. Chest **153**(6), 1336–1346 (2018)
7. 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. American journal of respiratory and critical care medicine **197**(2), 193–203 (2018)
8. Guthrie, A.: Chronic obstructive pulmonary disease series part 4: Identifying, managing, and preventing exacerbations. Sr. Care Pharm. **38**(9), 361–369 (Sep 2023)
9. Hatt, C., Galban, C., Labaki, W., Kazerooni, E., Lynch, D., Han, M.: Convolutional neural network based copd and emphysema classifications are predictive of lung cancer diagnosis. In: International Workshop on Reconstruction and Analysis of Moving Body Organs. pp. 302–309. Springer (2018)
10. Huang, G., Liu, Z., Van Der Maaten, L., Weinberger, K.Q.: Densely connected convolutional networks. In: Proceedings of the IEEE conference on computer vision and pattern recognition. pp. 4700–4708 (2017)
11. Humphries, S.M., Notary, A.M., Centeno, J.P., Strand, M.J., Crapo, J.D., Silverman, E.K., Lynch, D.A., of COPD (COPDGene) Investigators, G.E.: Deep learning enables automatic classification of emphysema pattern at ct. Radiology **294**(2), 434–444 (2020)
12. Ilse, M., Tomczak, J., Welling, M.: Attention-based deep multiple instance learning. In: International conference on machine learning. pp. 2127–2136. PMLR (2018)
13. Oelsner, E.C., Hoffman, E.A., Folsom, A.R., Carr, J.J., Enright, P.L., Kawut, S.M., Kronmal, R., Lederer, D., Lima, J.A., Lovasi, G.S., et al.: Association between emphysema-like lung on cardiac computed tomography and mortality in persons without airflow obstruction: a cohort study. Annals of internal medicine **161**(12), 863–873 (2014)

14. Park, J., Hobbs, B.D., Crapo, J.D., Make, B.J., Regan, E.A., Humphries, S., Carey, V.J., Lynch, D.A., Silverman, E.K., Investigators, C., et al.: Subtyping copd by using visual and quantitative ct imaging features. Chest **157**(1), 47–60 (2020)
15. Riley, C.M., Sciurba, F.C.: Diagnosis and outpatient management of chronic obstructive pulmonary disease: a review. Jama **321**(8), 786–797 (2019)
16. Schroeder, J.D., Bigolin Lanfredi, R., Li, T., Chan, J., Vachet, C., Paine III, R., Srikumar, V., Tasdizen, T.: Prediction of obstructive lung disease from chest radiographs via deep learning trained on pulmonary function data. International Journal of Chronic Obstructive Pulmonary Disease pp. 3455–3466 (2020)
17. Singh, D., Agusti, A., Anzueto, A., Barnes, P.J., Bourbeau, J., Celli, B.R., Criner, G.J., Frith, P., Halpin, D.M., Han, M., et al.: Global strategy for the diagnosis, management, and prevention of chronic obstructive lung disease: the gold science committee report 2019. European Respiratory Journal **53**(5) (2019)
18. Soriano, J.B., Polverino, F., Cosio, B.G.: What is early copd and why is it important? European Respiratory Journal **52**(6) (2018)
19. Steiger, D., Siddiqi, M.F., Yip, R., Yankelevitz, D.F., Henschke, C.I., Jirapatnakul, A., Flores, R., Wolf, A., Libby, D.M., Smith, J.P., et al.: The importance of low-dose ct screening to identify emphysema in asymptomatic participants with and without a prior diagnosis of copd. Clinical imaging **78**, 136–141 (2021)
20. 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. European Radiology **32**(8), 5319–5329 (2022)
21. 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. The Lancet Digital Health **2**(5), e259–e267 (2020)
22. Venkatesan, P.: Gold copd report: 2024 update. The Lancet Respiratory Medicine **12**(1), 15–16 (2024)
23. Zhang, H., Meng, Y., Zhao, Y., Qiao, Y., Yang, X., Coupland, S.E., Zheng, Y.: Dtfd-mil: Double-tier feature distillation multiple instance learning for histopathology whole slide image classification. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 18802–18812 (2022)
24. 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. The British Journal of Radiology **95**(1133), 20210637 (2022)