

# Radiology: Artificial Intelligence

## **Improved Detection of Chronic Obstructive Pulmonary Disease on Chest CT using the Mean Curvature of Isophotes**

Journal:	<i>Radiology: Artificial Intelligence</i>
Manuscript ID	RYAI-21-0105.R3
Manuscript Type:	Technical Development
Manuscript Categorization Terms:	Chronic Obstructive Pulmonary Disease < 6. TOPICS, Quantification < Vision < Application Domain < 8. MACHINE LEARNING, Lung < 5. STRUCTURES, CT < 2. MODALITIES/TECHNIQUES

SCHOLARONE™  
Manuscripts

CLAIM: Checklist for Artificial Intelligence in Medical Imaging

Section / Topic	No.	Item	
TITLE / ABSTRACT			
	1	Identification as a study of AI methodology, specifying the category of technology used (e.g., deep learning)	✓
	2	Structured summary of study design, methods, results, and conclusions	✓
INTRODUCTION			
	3	Scientific and clinical background, including the intended use and clinical role of the AI approach	✓
	4	Study objectives and hypotheses	✓
METHODS			
Study Design	5	Prospective or retrospective study	✓
	6	Study goal, such as model creation, exploratory study, feasibility study, non-inferiority trial	✓
Data	7	Data sources	✓
	8	Eligibility criteria: how, where, and when potentially eligible participants or studies were identified (e.g., symptoms, results from previous tests, inclusion in registry, patient-care setting, location, dates)	✓
	9	Data pre-processing steps	✓
	10	Selection of data subsets, if applicable	N/A
	11	Definitions of data elements, with references to Common Data Elements	N/A
	12	De-identification methods	✓
	13	How missing data were handled	N/A
Ground Truth	14	Definition of ground truth reference standard, in sufficient detail to allow replication	✓
	15	Rationale for choosing the reference standard (if alternatives exist)	✓
	16	Source of ground-truth annotations; qualifications and preparation of annotators	N/A
	17	Annotation tools	N/A
	18	Measurement of inter- and intrarater variability; methods to mitigate variability and/or resolve discrepancies	N/A
Data Partitions	19	Intended sample size and how it was determined	✓
	20	How data were assigned to partitions; specify proportions	✓

	<b>21</b>	Level at which partitions are disjoint (e.g., image, study, patient, institution)	✓
<b>Model</b>	<b>22</b>	Detailed description of model, including inputs, outputs, all intermediate layers and connections	✓
	<b>23</b>	Software libraries, frameworks, and packages	✓
	<b>24</b>	Initialization of model parameters (e.g., randomization, transfer learning)	✓
<b>Training</b>	<b>25</b>	Details of training approach, including data augmentation, hyperparameters, number of models trained	✓
	<b>26</b>	Method of selecting the final model	✓
	<b>27</b>	Ensembling techniques, if applicable	N/A
<b>Evaluation</b>	<b>28</b>	Metrics of model performance	✓
	<b>29</b>	Statistical measures of significance and uncertainty (e.g., confidence intervals)	✓
	<b>30</b>	Robustness or sensitivity analysis	N/A
	<b>31</b>	Methods for explainability or interpretability (e.g., saliency maps), and how they were validated	N/A
	<b>32</b>	Validation or testing on external data	✓
<b>RESULTS</b>			
<b>Data</b>	<b>33</b>	Flow of participants or cases, using a diagram to indicate inclusion and exclusion	✓
	<b>34</b>	Demographic and clinical characteristics of cases in each partition	✓
<b>Model performance</b>	<b>35</b>	Performance metrics for optimal model(s) on all data partitions	✓
	<b>36</b>	Estimates of diagnostic accuracy and their precision (such as 95% confidence intervals)	✓
	<b>37</b>	Failure analysis of incorrectly classified cases	N/A
<b>DISCUSSION</b>			
	<b>38</b>	Study limitations, including potential bias, statistical uncertainty, and generalizability	✓
	<b>39</b>	Implications for practice, including the intended use and/or clinical role	✓
<b>OTHER INFORMATION</b>			
	<b>40</b>	Registration number and name of registry	N/A
	<b>41</b>	Where the full study protocol can be accessed	N/A
	<b>42</b>	Sources of funding and other support; role of funders	N/A

1  
2  
3  
4  
5  
6  
7  
8  
9  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

Mongan J, Moy L, Kahn CE Jr. Checklist for Artificial Intelligence in Medical Imaging (CLAIM): a guide for authors and reviewers. Radiol Artif Intell 2020; 2(2):e200029. <https://doi.org/10.1148/ryai.2020200029>

Title:

**Improved Detection of Chronic Obstructive Pulmonary Disease on Chest CT using the Mean Curvature of Isophotes**

Manuscript type: Technical Development

Summary statement:

**The mean curvature of isophotes is a classical computer vision technique that, when incorporated into a typical convolutional neural network pipeline or used on its own, may improve discrimination of chronic obstructive pulmonary disease in chest CT.**

Key points:

- The mean curvature of isophotes (MCI) computes a geometrical transformation on the original CT images.
- When trained on MCI images, a typical convolutional neural network (CNN) architecture sees improvement in precision and recall relative to the same architecture trained on the original CT images (McNemar test,  $p < .001$ ).
- Simple threshold-based classification with a single histogram-derived feature from the MCI images matched the improved recall value of the CNN model.

**Abstract**

**Purpose**

To determine if the mean curvature of isophotes (MCI), a standard computer vision technique, can be used to improve detection of chronic obstructive pulmonary disease (COPD) on chest CT.

**Materials and Methods**

In this retrospective study, chest CT scans were obtained from 243 patients with COPD and 31 controls (among all 274: 151 women [mean age, 70 years; range, 44-90 years] and 123 men [mean age, 71 years; range, 29-90 years]) from two community practices between 2006 and 2019. A convolutional neural network (CNN) architecture was trained on either CT images or CT images transformed through the MCI algorithm. Separately, a linear classification based on a single feature derived from the MCI computation (called hMCI1) was also evaluated. All three models were evaluated with cross-validation, using Precision-macro and Recall-macro metrics, i.e, the mean of per-class precision and recall values, respectively (the latter being equivalent to balanced accuracy).

**Results**

Linear classification based on hMCI1 resulted in a higher Recall-macro relative to the CNN trained and applied on CT images (0.85 [95% confidence interval: 0.84,0.86] vs 0.77 [95% CI: 0.75,0.79]) but with a similar reduction in Precision-macro (0.66 [95% CI: 0.65,0.67] vs 0.77 [95% CI: 0.75,0.79]). The CNN model trained and applied on MCI-transformed images had a higher Recall-macro (0.85 [95% CI: 0.83,0.87] vs 0.77 [95% CI: 0.75,0.79]) and Precision-macro (0.85 [95% CI: 0.83,0.87] vs 0.77 [95% CI: 0.75,0.79]) relative to CNN trained and applied on CT images.

**Conclusion**

The MCI algorithm may be valuable towards the automated detection and diagnosis of COPD in chest CT scans, as part of a CNN-based pipeline or with stand-alone features.

## Introduction

Chronic obstructive pulmonary disease (COPD) is clinically defined as airflow obstruction with respiratory symptoms as measured with spirometry (1,2). Traditionally, radiological evaluation has not played a central role in the diagnosis of COPD. However, due to known limitations of spirometry in sensitivity to early disease (3,4), and given that objectively measured imaging markers of disease are important elements of optimal decision-making in clinical practice, there is an increasing interest in quantitative CT assessment of the lungs for disease detection, stratification, and risk prediction (5-9).

In medical imaging, convolutional neural networks (CNNs) have gained popularity and widespread use in recent years, and in chest CT analysis, they have been applied to the detection of COPD and other lung diseases (8-11). While CNNs have been shown to outperform other image analysis algorithms in a large variety of applications, recent computational developments have demonstrated that CNN performance can be improved if certain types of image geometry information are computed explicitly and input into the CNN, as this type of information is not readily available by standard convolution operations (12,13).

In this work, we introduce an image transformation that computes geometrical information from CT images, specifically, the mean curvature of isophotes (MCI). Image isophotes (contours of equal luminance) and their properties have long been used in computer vision to capture geometric information important for image analysis and visual perception (14-16). To our knowledge, the application of MCI to chest CT scans has not previously been studied. Therefore, we investigate here the hypothesis that COPD changes in the lung parenchyma can be reflected in the geometrical properties of the image. Specifically, in this exploratory study, we test the hypothesis that training a CNN architecture on MCI images would result in a higher diagnostic performance for discriminating patients with COPD from controls, as compared to training the same architecture on the original CT images. We also test in a linear classification

framework the discriminatory power of a single histogram-derived feature computed on the MCI images and compare its performance to that of the CNN.

**Materials and Methods**

*Clinical Methods*

This study was conducted with retrospective data acquired between 2006 and 2019 from patients with COPD and controls from two community-based pulmonology practices. The study had institutional review board approval and patient informed consent was waived given the observational nature of the study. Using an initial database of 1160 patients, patients were included if they a non-contrast CT scan, and in subsets biometric, pulmonary function test, and respiratory oscillometry data available.

All patients with COPD had symptoms consistent with the disease, more than 10 pack-year tobacco use history, and satisfied the following Global Initiative for Chronic Obstructive Lung Disease (GOLD) spirometry criteria: either a FEV<sub>1</sub>/FVC ratio (forced expiratory volume in one second over full vital capacity) of less than 0.7 (GOLD COPD I-IV), or a maximal mid-expiratory flow of less than 65% predicted or a residual volume of greater than 130% predicted (Pre-COPD) (2,4,17). Controls were drawn from the same practices, had less than 10 pack-years of tobacco use history, and had been determined to be free of respiratory disease following clinical and physiology assessment.

In total, 274 individuals (31 controls and 247 patients with COPD) with CT scans acquired at eight different institutions were included. This sample included 151 women (mean age: 70 years, range 44-90) and 123 men (mean age: 71, range 29-90). CT scan data were imported into a radiology software (OsiriX MD v.2.6, Primeo, Bernex, Switzerland) for deidentification. Figure 1 summarizes our patient selection procedure.

*CT Acquisition*



A total of five different scanner models were used to image patients. Scanners and acquisition details are shown in Table 1.

### *Computational Methods Overview*

An overview of the historical and mathematical background underlying the MCI algorithm is provided in Appendices E1-E3 (supplement). To assess the benefit of the MCI transformation, we compared the diagnostic performance of a CNN architecture trained on CT images to the performance of the same CNN architecture trained on MCI images. Separately, we also defined a simple quantitative feature from the distribution of voxel values in the MCI images (called hMCI1), and we used it to perform linear (threshold-based) classification. Details on image preprocessing are described in Appendix E4 (supplement). A description of the linear classification based on hMCI1 is given in Appendix E5. Finally, a description on the development of the two CNN models trained with CT or MCI images is given in Appendix E6 (supplement), and a schematic illustration is provided in Supplementary Figure S1.

### *Statistical Analysis*

We report the performance of all our models in terms of the mean of per-class recall values, called the macro average of recall (Recall-macro), as well as the per-class average of precision (Precision-macro). Use of these metrics allows for comparison to previous studies (8,9), and also mitigates the imbalance in sample size between our two classes. In a binary classification, recall is equivalent to sensitivity, and precision is equivalent to positive predictive value. Furthermore, the mean of per-class recall (Recall-macro) is equivalent to the balanced accuracy (i.e., to the mean of sensitivity and specificity).

The difference in classification performance between the CNN model trained and applied on CT images, and the CNN model trained and applied on MCI images, is assessed statistically via McNemar's test (18). The diagnostic performance of the two models was compared in four different contexts, each of which involves a binary discrimination: (A) the overall discrimination between controls and all COPD patients; (B) between controls and Pre-COPD patients only; (C)

between controls and GOLD I-II patients only; and (D) between controls and GOLD III-IV patients only. In each of these four discrimination scenarios, McNemar’s test was performed separately for each cross-validation fold, and we retained only the largest *P* value obtained over all folds. Given the four reported p-values (one for each discrimination scenario), a Bonferroni-corrected threshold of significance is set to  $0.05/4=0.0125$ . McNemar’s test was computed with the “mcnemar” function implemented in Python in the statsmodels package v0.12.2 (<https://www.statsmodels.org/stable/index.html>). Further details are provided in Appendix E6 (supplement).

*Code*

The code for our experiments is available at <http://www.github.com/>

**Results**

*Clinical Characteristics*

Demographic and clinical characteristics, as well as spirometry data are provided in Table 2, with patients with COPD separated according to GOLD stage. In terms of racial characteristics, all patients are Caucasian with the exception of 1 Black and 2 Asian male patients.

*Discrimination Between the Control Group and Patients with COPD*

For the simple univariate classification based on hMCI1 between controls and all patients with COPD, we report a receiver operating characteristic (ROC) curve analysis of classifier performance (Figure 2). The mean (standard deviation) of AUC across cross-validation folds was 0.87 (0.01). The diagnostic performance of the linear classification based on hMCI1 was 0.85 (95% CI: [0.84,0.86]) for the Recall-macro and 0.66 (95% CI: [0.65,0.67]) for Precision-macro over the 10 folds of the stratified cross validation.

Table 3 reports the discrimination performance of the two CNN models trained and applied on CT and MCI images, as well as the statistical significance of the difference in their classification performance. All four p-values obtained via McNemar's test were lower than the Bonferonni-corrected significance threshold of 0.0125.

Additional results are presented in Appendix E7 (supplement) and Supplementary Figures S2-S4, including visualization of MCI images and comparisons to quantitative CT metrics commonly used in the COPD literature.

## Discussion

In standard clinical practice, COPD (and the subdivision of its severity into GOLD stages) is a diagnosis determined from clinical and physiological tests, such as spirometry. The lack of a well-defined relationship between spirometry-based ground-truth staging and radiological appearance makes the automated image-based detection and staging of COPD challenging, even for powerful algorithms such as CNNs. In this study, we approached this challenge through the use of an image transformation that computes geometrical information from CT images, specifically, the mean curvature of isophotes (MCI). Our results show that a CNN architecture trained on MCI images had higher diagnostic performance compared to the same CNN architecture trained on the original CT images; this trend was observed across all disease severity groups, including patients with pre-COPD. These findings support that CNNs do not always extract, on their own, all of the important information from an image, and may benefit from being provided geometric information relevant to tissue characteristics in a precomputed fashion, as shown in a small but growing body of literature(12,13).

It is likely that optimization of the CNN architecture and parameter tuning may have led to improved results with the CNN trained on CT images. In fact, one might argue that because of this lack of optimization, our CNN experiment on the original CT images may not be a good point

of comparison to the CNN trained on the MCI images. To counter this argument, we note that the objective of this study was not to determine the most optimal CNN model for this specific clinical problem. Rather, it was to showcase the relative improvement in performance for the same CNN architecture and the same training algorithm when provided MCI images as input, as opposed to original CT images. In addition, our results are similar to those reported in previous studies. To our knowledge, there are only two recent studies on COPD detection using CNN-based analysis of CT images (8,9). Tang et al (9) used the ResNet CNN architecture. We used ResNet with 50 layers, whereas they experimented with 50, 101 and 152 layers. Gonzalez et al (8) used a shallower network. Both studies used more complex preprocessing than ours, involving a selection of canonical sections or spatially corresponding regions to be compared across patients. Despite larger training sets, more advanced preprocessing steps, controlled acquisition protocols in all of the datasets involved, and a paucity of patients with early disease in some datasets (which should make discrimination easier), our results with a CNN on CT images were similar to their results. Depending on parameters (such as the number of layers in the ResNet architecture or a specific region of interest) the Precision-macro and Recall-macro values obtained by Tang et al (9) ranged from 0.69 to 0.76, and from 0.70 to 0.80, respectively. We obtained a value of 0.77 for both Precision-macro and Recall-macro (controls vs all COPD), towards the upper end of the range reported in (9). The approach from Gonzalez et al (8) results in Recall-macro and Precision-macro values of 0.77 for both measures (deduced from the confusion matrix in their Fig. 2A), which was very similar to our own results. We conclude that the diagnostic performance of our CNN architecture on CT images appears comparable to current state of the art in the literature.

Our experiment with univariate classification based on the hMCI1 measure achieved the same Recall-macro value of 0.85 as the CNN applied to MCI images, in a much simpler fashion. However, the Precision-macro was lower at a value of 0.66. Nevertheless, this high recall value comparable to that of a CNN suggests this approach might be a viable and simple alternative to a CNN model in the context of screening for COPD. In such an application where recall is more important than precision, suspected COPD cases could be flagged and those deemed uncertain can be flagged for further review by an expert to exclude potential false positives. Screening for

1  
2  
3 COPD is important for a variety of clinical reasons, including the initiation of tobacco use  
4 cessation, vaccination, and the prescription of medication to both relieve symptoms and treat  
5 exacerbations. Such strategies prolong lives and reduce healthcare costs (2).  
6  
7  
8  
9

10 Our study had limitations. As this was an exploratory study, our methodology was not fully  
11 optimized, neither at the deep learning level, nor at the univariate classification level. For  
12 instance, we chose a threshold in the definition of hMCI1 in a somewhat arbitrary fashion, and  
13 we picked a fixed value of  $\sigma$ . In future work, instead of picking a threshold we will work with the  
14 entire histogram data in combination with statistical techniques tailored to probability density  
15 estimates. We will also extend the approach to a range of spatial scales akin to classical scale-  
16 space approaches (14-16). Another limitation consists in our small dataset, compared to other  
17 studies that worked with large prospectively acquired datasets (19-21). In future work, we will  
18 assess the performance of the MCI technique on large public prospective datasets (19-21). In the  
19 present study, we felt it was informative to present results on a retrospective sample derived  
20 from a community practice, as it emulates real-world clinical practice where data and conditions  
21 are less standardized compared to research cohorts. Another limitation stems from the lack of  
22 racial diversity among our patient population, with only 3 patients (1%) being non-Caucasian. In  
23 future work, this limitation will be addressed with validation on the above mentioned prospective  
24 datasets (19-21), which are not only larger but also more racially and ethnically diverse.  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39

40 As MCI has not previously been used in the context of disease prediction on chest CT, we  
41 performed a comparison between training the CNN on CT images alone versus MCI images alone,  
42 in order to evaluate the performance gain achieved by using MCI images compared to original CT  
43 images. In future work, a CNN could also be trained on both types of images simultaneously.  
44  
45  
46  
47  
48

49 Whereas we have showcased the use of MCI in a discrimination context, an important next step  
50 is to identify the endophenotype(s) to which it is sensitive (22). Applying the technique to imaging  
51 data from explanted lung tissue samples with known characteristics could provide insight in that  
52 regard (23). In Appendix E7 (supplement) we discuss the spatial information inherent in MCI  
53  
54  
55  
56  
57  
58  
59  
60

values and how it relates to specific structures in the lungs (see also Supplementary Figures S2 and S3).

A potential application of using MCI images could be in detecting other diffuse lung disease, such as early interstitial lung disease (ILD). Currently, there is a lack of sensitive biomarkers with which to follow interstitial lung disease progression and monitor response to therapy (24). Preliminary data suggests that MCI potentially might also fill this unmet clinical need (25). In future work we will investigate the ability of the MCI technique to differentiate between COPD, interstitial lung disease, and other types of lung disease. Finally, as the proposed MCI methodology is organ and image modality agnostic, we plan to explore applications to other organs, diseases, and image modalities.

In summary, we evaluated the use of an image transformation, called the mean curvature of isophotes (MCI), to discriminate COPD on CT images. Incorporated into a CNN analysis, this transformation allows an improvement in classification performance. Used instead in the context of a basic univariate classification, it still performs at a level that may be useful in certain types of applications. Further investigation of MCI images for classifying diseases on CT is warranted.

**References**

1. Celli BR, MacNee W, 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):932-46.
2. Vogelmeier CF, Criner GJ, Martinez FJ, Anzueto A, Barnes PJ, Bourbeau J, et al. Global Strategy for the Diagnosis, Management, and Prevention of Chronic Obstructive Lung Disease 2017 Report. GOLD Executive Summary. Am J Respir Crit Care Med. 2017;195(5):557-82
3. Woodruff PG, Barr RG, Bleecker E, Christenson SA, Couper D, Curtis JL, et al. Clinical Significance of Symptoms in Smokers with Preserved Pulmonary Function. N Engl J Med. 2016;374(19):1811-21.

- 1  
2  
3 4. Regan EA, Lynch DA, Curran-Everett D, Curtis JL, Austin JH, Grenier PA, et al. Clinical and  
4 Radiologic Disease in Smokers With Normal Spirometry. *JAMA Intern Med.* 2015;175(9):1539-49.  
5  
6
- 7 5. Shaker SB, Dirksen A, Lo P, Skovgaard LT, de Bruijne M, Pedersen JH. Factors influencing  
8 the decline in lung density in a Danish lung cancer screening cohort. *Eur Respir J.*  
9 2012;40(5):1142-8.  
10  
11
- 12 6. Kirby M, Tanabe N, Tan WC, Zhou G, Obeidat M, Hague CJ, et al. Total Airway Count on  
13 Computed Tomography and the Risk of Chronic Obstructive Pulmonary Disease Progression.  
14 Findings from a Population-based Study. *Am J Respir Crit Care Med.* 2018;197(1):56-65.  
15  
16
- 17 7. San Jose Estepar R, Kinney GL, Black-Shinn JL, Bowler RP, Kindlmann GL, Ross JC, et al.  
18 Computed tomographic measures of pulmonary vascular morphology in smokers and their  
19 clinical implications. *Am J Respir Crit Care Med.* 2013;188(2):231-9.  
20  
21
- 22 8. Gonzalez G, Ash SY, Vegas-Sanchez-Ferrero G, Onieva Onieva J, Rahaghi FN, Ross JC, et al.  
23 Disease Staging and Prognosis in Smokers Using Deep Learning in Chest Computed Tomography.  
24 *Am J Respir Crit Care Med.* 2018;197(2):193-203.  
25  
26
- 27 9. Tang LYW, Coxson HO, Lam S, Leipsic J, Tam RC, Sin DD. Towards large-scale case-finding:  
28 training and validation of residual networks for detection of chronic obstructive pulmonary  
29 disease using low-dose CT. *The Lancet Digital Health.* 2020;2(5):e259-e67.  
30  
31
- 32 10. Walsh SLF, Calandriello L, Silva M, Sverzellati N. Deep learning for classifying fibrotic lung  
33 disease on high-resolution computed tomography: a case-cohort study. *The Lancet Respiratory*  
34 *Medicine.* 2018;6(11):837-45.  
35  
36
- 37 11. Ciompi F, Chung K, van Riel SJ, Setio AAA, Gerke PK, Jacobs C, et al. Towards automatic  
38 pulmonary nodule management in lung cancer screening with deep learning. *Sci Rep.*  
39 2017;7:46479.  
40  
41
- 42 12. Clough JR, Oksuz I, Byrne N, Schnabel JA, King AP. Explicit Topological Priors for Deep-  
43 Learning Based Image Segmentation Using Persistent Homology. *Information Processing in*  
44 *Medical Imaging. Lecture Notes in Computer Science* 2019. p. 16-28.  
45  
46
- 47 13. Rezanejad M, Downs G, Wilder J, Walther DB, Jepson A, Dickinson A, Siddiqi K. Scene  
48 categorization from contours: medial axis based salience measures. *2019 IEEE Conference on*  
49 *Computer Vision and Pattern Recognition (CVPR)2019*, pp. 4111-4119.  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

14. Koenderink JJ. The structure of images. *Biological cybernetics*. 1984;50(5):363-70.

15. Griffin LD, Colchester AC. Superficial and deep structure in linear diffusion scale space: Isophotes, critical points and separatrices. *Image and Vision Computing*. 1995;13(7):543-57.

16. Kuijper A, Florack LM. The relevance of non-generic events in scale space models. *International Journal of Computer Vision*. 2004;57(1):67-84.

17. Han MK, Agusti A, Celli BR, Criner GJ, Halpin DMG, Roche N, Papi A, Stockley RA, Wedzicha J, Vogelmeier CF. From GOLD 0 to Pre-COPD. *Am J Respir Crit Care Med*. 2021;203(4):414-423.

18. Everitt BS. *The analysis of contingency tables*. London: Chapman and Hall, 1977.

19. Agusti A, Calverley PM, Celli B, Coxson HO, Edwards LD, Lomas DA, et al. Characterisation of COPD heterogeneity in the ECLIPSE cohort. *Respir Res*. 2010;11:122.

20. Bourbeau J, Tan WC, Benedetti A, Aaron SD, Chapman KR, Coxson HO, et al. Canadian Cohort Obstructive Lung Disease (CanCOLD): Fulfilling the need for longitudinal observational studies in COPD. *COPD*. 2014;11(2):125-32.

21. Regan EA, Hokanson JE, Murphy JR, Make B, Lynch DA, Beaty TH, et al. Genetic epidemiology of COPD (COPDGene) study design. *COPD*. 2010;7(1):32-43.

22. Barnes PJ. Endo-phenotyping of COPD patients. *Expert Review of Respiratory Medicine*. 2020;1-11.

23. Vasilescu DM, Martinez FJ, Marchetti N, Galban CJ, Hatt C, Meldrum CA, et al. Noninvasive Imaging Biomarker Identifies Small Airway Damage in Severe Chronic Obstructive Pulmonary Disease. *Am J Respir Crit Care Med*. 2019;200(5):575-81.

24. Richeldi L, du Bois RM, Raghu G, Azuma A, Brown KK, Costabel U, et al. Efficacy and safety of nintedanib in idiopathic pulmonary fibrosis. *N Engl J Med*. 2014;370(22):2071-82.

25. Anonymized.

26. Quanjer PH, Tammeling GJ, Cotes JE, Pedersen OF, Peslin R, Yernault JC. Lung volumes and forced ventilatory flows. Report Working Party Standardization of Lung Function Tests, European Community for Steel and Coal. Official Statement of the European Respiratory Society. *Eur Respir J Suppl*. 1993;16:5-40.



Table 1. Technical Scan Parameters.

Model	Aquilion	Discovery CT750	LightSpeed VCT	Sensation 64	Somatom Definition Edge
Manufacturer	Toshiba	GE Medical Systems	GE Medical Systems	Siemens	Siemens
Number of scans	1 (0.4%)	25 (9%)	232 (85%)	14 (5%)	2 (1%)
Institution	A	B	C, D, E, F	B	G, H
Convolution Kernel (number of scans)	FC08 (1)	Lung (25)	Lung (1) Standard (231)	B70f (14)	B70f (1) I41f\3 (1)
X-ray Voltage (kV)	120	120	120	120	110* [100, 120]
X-ray Current (mA)	234	103 [100, 0.64]	266 [154, 420]	248 [205, 352]	125 [124, 126]
Pixel Spacing (mm)	0.65	0.70 [0.64, 0.74]	0.69 [0.64, 0.74]	0.67 [0.64, 0.71]	0.80 [0.72, 0.87]
Section Thickness (mm)	1	2.5	1.25	3	2.50 [2.00, 3.00]
Revolution Time (s)	N/A†	0.7	0.7	0.8	0.40 [0.30, 0.50]
Spiral Pitch Factor	N/A	0.98	1.38	1.38	0.78 [0.60, 0.95]
Single Collimation Width (mm)	N/A	0.63	0.63	0.63	0.6
Total Collimation Width (mm)	N/A	40	40	40	38

\*Data are presented as median [IRQ] where values vary between CT scans

†N/A = data not available

1  
2  
3  
4  
5  
6  
7  
8  
9  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

Table 2. Patient Characteristics and Spirometry

	Controls	Pre-COPD	GOLD I	GOLD II	GOLD III	GOLD IV	P-Value
A. Patient characteristics							
n	31	29	48	110	44	12	
Age (median years [IRQ])	61 [51,71]	70 [65,79]	73 [67,78]	74 [69,78]	70 [64,78]	64 [58,73]	< .001
No. women	18 (58%)	15 (52%)	32 (67%)	60 (54.5%)	22 (50%)	4 (33%)	.35
No. men	13 (42%)	14 (48%)	16 (33%)	50 (45.4%)	22 (50%)	8 (67%)	
Tobacco use History (pack-years)	4.0 [0.0,13.5]	38.0 [29.0,44.0]	39.0 [27.8,47.5]	44.0 [33.2,53.0]	45.0 [34.0,54.2]	41.0 [36.2,66.5]	< .001
Height (cm)	166 [160,170]	164 [160,172]	159 [155,168]	164 [158,170]	164 [159,170]	166 [164,170]	.075
Weight (kg)	79 [70,88]	83 [66,99]	76 [65,86]	76 [67,88]	75 [58,87]	70 [64,87]	.66
BMI (kg/m <sup>2</sup> )	28 [25,34]	30 [24,34]	28 [25,33]	28 [26,32]	26 [23,32]	26 [23,29]	.34
B. Spirometry							
n	29	29	48	110	44	12	
FEV <sub>1</sub> (L)	2.55 [2.19,3.47]	2.14 [1.58,2.49]	1.79 [1.54,2.41]	1.38 [1.16,1.73]	0.89 [0.76,1.10]	0.64 [0.51,0.68]	< .001
FEV <sub>1</sub> (% predicted)*	109 [97,115]	87 [80,106]	88 [85,95]	66 [59,73]	41 [37,46]	25 [20,27]	< .001
FVC (L)	3.19 [2.79,4.25]	2.81 [2.20,3.48]	2.88 [2.38,3.89]	2.61 [2.20,3.22]	2.38 [1.75,2.79]	1.96 [1.61,2.12]	< .001
FVC (% predicted)	107 [97,117]	100 [83,115]	116 [109,126]	96 [87,106]	78 [72,89]	59 [52,71]	< .001
FEV <sub>1</sub> /FVC (%)	80 [78,84]	72 [71,73]	63 [61,66]	54 [49,59]	40 [36,46]	35 [30,38]	< .001
MMEF (L/s)	2.78 [1.71,3.64]	1.36 [1.01,1.67]	0.87 [0.66,1.10]	0.51 [0.38,0.65]	0.29 [0.23,0.37]	0.23 [0.20,0.27]	< .001
MMEF (% predicted)	84 [65,107]	49 [43,60]	34 [27,39]	20 [15,25]	10 [9,13]	8 [7,9]	< .001

**Note.**— Data are shown as median (interquartile range). In each row, a p-value resulting from a Kruskal-Wallis test indicates whether the groups differ in terms of the specified characteristic or spirometry variable. While the groups did not differ in terms of sex distribution, height, weight and BMI, the individuals in the control group were younger than the Pre-COPD to GOLD III patients (post-hoc Dunn test  $P < .005$ ), and GOLD IV were younger than GOLD II (post-hoc Dunn test  $p < 0.01$ ). By definition, forced spirometry differed between groups. BMI = body mass index, FEV<sub>1</sub> = forced expiratory volume in one second, FVC forced vital capacity, GOLD = Global Initiative for Chronic Obstructive Lung Disease, MMEF = maximal mid-expiratory flow

\*European Community for Steel and Coal (ECSC) predicted values (26).

Table 3. Diagnostic Performance of the CNN Models Trained and Applied on CT or MCI Images

	CT images	MCI images	P value*
A. Control group vs All COPD			< .001
Recall-macro	0.77 (0.75,0.79)	0.85 (0.83,0.87)	
Precision-macro	0.77 (0.75,0.79)	0.85 (0.83,0.87)	
B. Control group vs Pre-COPD			< .001
Recall-macro	0.75 (0.72, 0.78)	0.84 (0.80,0.88)	
Precision-macro	0.62 (0.60,0.64)	0.68 (0.65,0.71)	
C. Control group vs GOLD I-II			< .001
Recall-macro	0.77 (0.74,0.80)	0.85 (0.82,0.88)	
Precision-macro	0.76 (0.73,0.79)	0.84 (0.81,0.87)	
D. Control group vs GOLD III-IV			< .001
Recall-macro	0.77 (0.75,0.79)	0.85 (0.82,0.88)	
Precision-macro	0.69 (0.66,0.72)	0.76 (0.73,0.79)	

Note.— Diagnostic metrics (95% confidence intervals), computed over the cross-validation folds. For each of the four comparisons reported in the Table, a p-value was computed with McNemar’s test to assess the significance of the difference in classification performance between the CNN trained and applied to CT images, and the CNN trained and applied to MCI images. For each comparison, McNemar’s test was performed separately for each cross-validation fold, and we report the largest *P* value obtained over all folds. Given the four reported p-values, a Bonferroni-corrected threshold of significance is set to 0.05/4=0.0125. All the reported *P* values were below this threshold.

\**P* value shown for comparison between model trained with CT or MCI images.

Figure Legends

**Figure 1.** Flowchart summarizing patient selection and analysis. FEV<sub>1</sub> = forced expiratory volume in one second, FVC = forced vital capacity, GOLD = Global Initiative for Chronic Obstructive Lung Disease, MMEF = maximal mid-expiratory flow, RV = residual volume.

**Figure 2.** Classification between controls and all patients with chronic obstructive pulmonary disease using hMCI1 as a single feature in a linear classification model. Mean receiver operating characteristic curve (green) computed over the 10 folds of stratified cross-validation Black: ± 1 standard deviation from the mean receiver operating characteristic curve. AUC: area under the curve.

## Supplementary material

### Improved Detection of Chronic Obstructive Pulmonary Disease on Chest CT using Mean Curvature of Isophotes

**Appendix E1. Background: Isophotes and image measures invariant to gray level intensity transformations**

In image analysis, it has long been known that important information is carried by geometric entities known as isophotes, or isocontours, i.e. curves or surfaces of equiluminance points (1-3). These geometric entities reflect intrinsic properties of object structure and are therefore important for the semantic description of image content. In the field of computer vision, isophotes and their geometry have been studied for their role in the perception of object shape by the human visual system in 2D images (4). In computer vision, they also play an important role in the analysis of image content at multiple scales of resolution, in what is known as scale space analysis (1-3). Because of their important geometric properties, isophotes have been used in the detection and tracking of points and structures of interest, either in medical images for applications such as image matching and registration (5) or in a wide variety of more general computer vision applications such eye-tracking in video (6) or face detection (7). Outside the field of computer vision, isophotes have been used in surface visualization for a variety of applications in computer-aided design and manufacturing (8, 9), or for visualization of astronomy data (10).

In addition to carrying important information about image geometry, it has also been shown that isophotes underlie several image *invariants*, i.e. image properties that are inherent to the underlying visual scene, and that are invariant to general transformations of the image gray level intensity. Examples of such transformations include changes in contrast or brightness (11). Such transformations do not alter the iso-intensity, or isophote structure of the image. Several image invariants, i.e., image features invariant to such transformations, can be computed from image isophotes, as discussed at length in the computer vision literature (11-13). Here, we focus on one such invariant, the mean curvature of the isophotes (MCI). Given that the MCI image transformation measures known aspects of image geometry, that have been shown to be fundamental in other areas of image analysis and computer vision, and given that emphysema/COPD attacks the structure of lung parenchyma, we hypothesized that the changes that occur in lung parenchyma can be captured by a robust geometrical method such as MCI.

## Appendix E2. Computing the MCI image transformation

For any given input image, the MCI computation results in a transformed image of the same size as the input. In the resulting MCI image, each voxel represents the MCI value for the corresponding voxel in the original input image. This is achieved by computing the divergence of the unit gradient of the input image  $I$ :

$$MCI(I) = \text{div}\left(\frac{\nabla I}{\|\nabla I\|}\right) \quad [1]$$

Other equivalent definitions can be formulated as well. For instance, MCI can alternatively be defined based on the trace of the Hessian matrix, or via equations involving the second derivatives of the image (11-14). Regardless of the particular definition used, however, the isophote curvature at a given point in the image is computed via local information only, i.e. using only information from the local image neighborhood.

## Appendix E3. Scale choice in computing the MCI image

Computing equation [1] requires a convolution with a 3D Gaussian filter. In addition to smoothing the image and thus facilitating numerical image gradient calculations involved in computing equation [1], this filtering step also serves to capture structures of a particular spatial scale in the image - a scale defined by the standard deviation  $\sigma$  of the Gaussian filter. The importance of MCI and of Gaussian filters for scale space analysis in computer vision has been discussed at length elsewhere (1-3). In the experiments reported in this paper, we worked with  $\sigma = 1$  mm. While this number may be reminiscent of the scanner's voxel resolution, it is important to note that filtering is implemented with Gaussian kernels covering a spatial extent of  $\pm 3\sigma$ , thus integrating information over a neighborhood that spans several voxels. In that sense, differences in smoothing due to variations in voxel sizes between different acquisitions will be small relative to the Gaussian smoothing implemented with our method. While small, such differences may nevertheless be present, and are part of the signal variability that is inherent in the data.

**Appendix E4. Image preprocessing**

The preprocessing used in our pipeline is very limited. We first interpolated the original CT image to an isotropic image grid with voxel size of 0.9x0.9x0.9mm<sup>3</sup>. This has the effect of removing the variation of voxel sizes across scans in the original dataset. Image interpolation to an isotropic grid is standard practice with the International Biomarker Standardization Initiative (IBSI)(15). We then performed a 3D segmentation of the lungs from the resampled CT image using the Chest Imaging Platform (<http://chestimagingplatform.org/>), an open-source software.

**Appendix E5. Classification based on a single histogram-derived MCI feature**

Since the MCI image is by construction in the same space as the CT image, the lung segmentation can be directly transferred onto the MCI image. We then analyse the segmented MCI images with histogram-based quantification, due to its simplicity, and with analogy to simple radiomics features and other quantitative methods popular in the literature on lung CT densitometry (15-19). We first compute a histogram  $H$  of MCI values over the lung mask with kernel density estimation as implemented in Matlab with the `ksdensity()` function (20), using a normal kernel and 100 equally spaced bins covering the MCI value range from -2.5mm<sup>-1</sup> to 2.5mm<sup>-1</sup>. The histogram  $H$  is then normalized so that its values sum to 1, which makes it equivalent to a probability density estimate.

Based on this, we propose the hMCI1 measure defined in equation [2] as the fraction of image voxels within the lung mask with an MCI value larger than 1 mm<sup>-1</sup>:

$$\text{hMCI1} = \int_1^{\infty} H(x)dx \tag{2}$$

Using this measure as a single feature, we performed a linear classification between control and patients with COPD. Given the large imbalance in our data and the small number of individuals in our control class, we did not create one random partition between training and validation datasets. Rather, we performed 10-fold stratified cross-validation (21). For each training partition, we computed a receiver operating characteristic (ROC) curve and its area under the



curve (AUC). We then used Youden's  $j$  on the ROC curve to select an optimal threshold for the hMCI1 value, which defines a univariate linear classification model. We then computed the diagnostic performance of this threshold on the validation dataset, and report diagnostic performance measures averaged over the 10 folds. 95% confidence intervals of each measure over the cross-validation folds were also computed. Finally, we also report the mean ROC curve with  $\pm 1$  standard deviation.

## Appendix E6. Classification based on deep learning

We also investigated the performance of a CNN for discriminating controls vs. COPD subjects, using either CT or MCI images as input. We chose the Deep Residual Network with 50 layers (ResNet50) architecture (22), as it is commonly used in the literature. We used the architecture implementation available in the Pytorch library (<http://pytorch.org/>) with pre-training on the popular ImageNet database (<http://www.image-net.org/>) and additional training as described below. Supplementary Figure S1 presents a visual summary of our experimental setup.

Deep learning models used in image analysis have been developed and optimized for 2D image data. Adapting them to 3D images is challenging – this is why analyses of volumetric images are typically done on a set of 2D slices extracted from the 3D image (23-25). In this work, to represent a given 3D scan, we extracted eleven axial slices equally spaced through the lungs. Each such slice is then given the same label as the originating 3D scan. Additionally, voxels outside the lungs in each slice are masked, i.e., set to a value outside of the allowed range. In the training phase, the network predicts labels for individual slices, which are then combined via majority voting to assign a predicted label to the corresponding 3D scan. The back-propagation process is then optimized based on whether the prediction error of scan labels decreases.

Prior to training, we applied data augmentation to increase the number of data samples in each class, and to balance the number of data samples in each class. Data augmentation was done in 3D, that is, it was applied on the 3D scans and not on individual slices. We applied two types of

augmentation: 1) non-uniform scaling of scans along each axis, and 2) small degree rotations about the z-axis. For non-uniform scaling, we chose randomly a scaling factor that could increase or decrease lung size up to 15% independently along each of the three axes. In addition, a rotation angle was chosen randomly for a maximum of 5-degree clockwise/counter-clockwise rotation about the z-axis. Using this method, we created 1000 scans for each of the two categories of "COPD" and "Control". We then performed 5-fold cross-validation, by randomly splitting this dataset into 80%-20% training and testing folds, performed 5 times without repetitions. On each fold, we trained the network initialized with pre-trained weights, and then we applied it to the testing dataset. We report diagnostic performance measures averaged over the 5 folds, as well as their 95% confidence intervals.

We first trained a CNN model on the anatomical CT images. Separately, we trained the same architecture on the MCI images, thus obtaining a second CNN model specific to MCI images. We ensured that the 5 folds of cross-validation were the same between both experiments. In this fashion, it was possible to perform McNemar's statistical test for each fold separately, in order to test whether the difference in performance is statistically significant for each fold. While we averaged the Precision-macro and Recall-macro metrics across folds, p-values cannot be averaged. Instead, we chose a maximally conservative approach by retaining and reporting the largest p-value obtained over the 5 folds, as a worst-case scenario.

We report results over the entire cohort comparing controls to all of the COPD subjects. Separately, we also report diagnostic performance results (and McNemar testing results) for comparing controls to three separate subgroups of COPD patients: pre-COPD only, GOLD I and GOLD II only, and GOLD III and GOLD IV only. In this fashion, we observe the diagnostic performance relative to different severity levels. We achieve this sub-group comparison by first recording the model predictions for each subject in our entire population, and then creating confusion matrices for each sub-group comparison separately, by accounting only for subjects with ground truth labeling among the targeted GOLD stages.

## Appendix E7. Supplementary Results

### *Illustrative examples*

In Supplementary Figure S2 (first and second row from top to bottom), we present an example section through a chest CT scan, with the corresponding MCI image. To illustrate the invariance properties of MCI to non-linear image transformations, we then artificially change the contrast of the chest CT by taking the original Hounsfield unit (HU) values to the power of 3 (Suppl. Figure S2, third row). We also present the corresponding MCI image calculated on the modified HU values. Finally, we perform another modification, where we take the cube root of the original HU values after shifting them to the positive half-line (Suppl. Figure S2, fourth row). We also present the corresponding MCI image calculated on the modified HU values. As expected, the MCI images are identical to each other, despite the obvious changes to the corresponding intensity images.

### *Threshold choice in the definition of the hMCI1 measure*

We defined the hMCI1 measure as the fraction of lung voxels with an MCI value larger than  $1 \text{ mm}^{-1}$ . In doing so, we focus on the positive extreme of the MCI histogram. Note the largest negative MCI values are associated with the brightest image structure (blood vessels in the case of lung CT), whereas the largest positive MCI values are associated with the dark spaces between these bright structures, i.e., lung parenchyma in the case of lung CT. This is illustrated in Supplementary Figure S2 and Figure S3. In fact, it has been shown that the local maxima of MCI values act as ridge detectors in images (5, 12, 13). Thus, by measuring the proportion of a MCI histogram allocated to large positive values ( $> 1 \text{ mm}^{-1}$  in the case of the hMCI1 measure), we measure the proportion of voxels that are the furthest in a certain geometric sense from blood vessels, for instance, the red voxels in Supplementary Figure S3 (right). Note that these voxels cannot be easily defined by a simple thresholding operation on HU values as done with other standard quantitative CT measures (18,19,26).

*Quantitative CT measures across GOLD stages*

In Supplementary Figure S4, we use swarmbox plots to show the distribution across GOLD stages of the hMCI1 measure, as well as of two standard quantitative CT measures popular in the literature:  $LAA_{<-950HU}$ , the fraction of lung voxels with a value less than -950 Hounsfield units (HU); and  $PD_{15}$ , the 15th percentile of the lung density histogram (18,19,26). The latter two measures have commonly been used in the respiratory literature to quantify emphysema in chest CT scans (18,19,26).

**Appendix E8. Supplementary Discussion**

Our control group should ideally be a random sample from the community. However, exposing individuals to unnecessary radiation would be unethical, and given the retrospective nature of our study, beyond the scope of our work. Thus, as a control sample, we chose a reasonable compromise of individuals who, following extensive clinical examinations including chest CT scans, were found to be free of respiratory disease. These patients had either incidental micronodules or transient dyspnea without detectable cardiopulmonary disease. Arguably, this group might be more “normal” in terms of lung disease than the local general population which is known to have a combined background incidence of asthma and COPD approaching 17%. (27)

In terms of methodological development, the current MCI-based pipeline can benefit from several improvements. One such improvement lies in the optimization of the histogram threshold parameter (currently set to  $1\text{ mm}^{-1}$ ), the use of the entire histogram (as opposed to choosing a single threshold), or the investigation of different, more sophisticated non-histogram based image quantification techniques borrowed from the world of radiomics (15-17). An improved analysis will also consider a range of spatial scales, as opposed to picking fixed scale values ( $\sigma=1\text{mm}$  in the present work). Finally, we will also optimize the use of the entire MCI image in conjunction with machine learning techniques such as deep learning.

## References

1. Griffin LD, Colchester AC. Superficial and deep structure in linear diffusion scale space: Isophotes, critical points and separatrices. *Image and Vision Computing*. 1995;13(7):543-57.
2. Koenderink JJ. The structure of images. *Biological cybernetics*. 1984;50(5):363-70.
3. Kuijper A, Florack LM. The relevance of non-generic events in scale space models. *International Journal of Computer Vision*. 2004;57(1):67-84.
4. Koenderink JJ, van Doorn AJ. Photometric invariants related to solid shape. *Optica Acta: International Journal of Optics*. 1980;27(7):981-96.
5. Maintz JBA, van den Elsen PA, Viergever MA. Evaluation of ridge seeking operators for multimodality medical image matching. *IEEE Transactions on pattern analysis and machine intelligence*. 1996;18(4):353-65.
6. Valenti R, Gevers T, editors. Accurate eye center location and tracking using isophote curvature. 2008 IEEE Conference on Computer Vision and Pattern Recognition; 2008: IEEE.
7. Lichtenauer J, Hendriks E, Reinders M, editors. Isophote properties as features for object detection. 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05); 2005: IEEE.
8. Poeschl T. Detecting surface irregularities using isophotes. *Computer Aided Geometric Design*. 1984;1(2):163-8.
9. Theisel H, Farin G. The curvature of characteristic curves on surfaces. *IEEE Computer Graphics and Applications*. 1997;17(6):88-96.
10. Chaware L, Cannon R, Kembhavi AK, Mahabal A, Pandey S. Isophotal shapes of early-type galaxies to very faint levels. *The Astrophysical Journal*. 2014;787(2):102.
11. Florack LM, Romeny BTH, Koenderink JJ, Viergever MA. General intensity transformations and differential invariants. *Journal of Mathematical Imaging and Vision*. 1994;4(2):171-87.
12. Eberly D, Gardner R, Morse B, Pizer S, Scharlach C. Ridges for image analysis. *Journal of Mathematical Imaging and Vision*. 1994;4(4):353-73.
13. López AM, Lloret D, Serrat J, Villanueva JJ. Multilocal creaseness based on the level-set extrinsic curvature. *Computer Vision and Image Understanding*. 2000;77(2):111-44.

14. Goldman R. Curvature formulas for implicit curves and surfaces. Computer Aided Geometric Design. 2005;22(7):632-58.

15. Zwanenburg A, Vallieres M, Abdalah MA, Aerts H, Andrearczyk V, Apte A, et al. The Image Biomarker Standardization Initiative: Standardized Quantitative Radiomics for High-Throughput Image-based Phenotyping. Radiology. 2020;295(2):328-38.

16. Aerts HJ, Velazquez ER, Leijenaar RT, Parmar C, Grossmann P, Carvalho S, et al. Decoding tumour phenotype by noninvasive imaging using a quantitative radiomics approach. Nat Commun. 2014;5:4006.

17. van Griethuysen JJM, Fedorov A, Parmar C, Hosny A, Aucoin N, Narayan V, Beets-Tan RGH, Fillion-Robin JC, Pieper S, Aerts HJWL. Computational Radiomics System to Decode the Radiographic Phenotype. Cancer Res. 2017;77(21):e104-e107.

18. Gevenois P, deMaertelaer V, de Vuyst, P, et al. Comparison of Computed Density and Macroscopic Morphometry In Pulmonary Emphysema. Am J Respir Crit Care Med. 1995;152:653-7.

19. Gevenois P, de Vuyst, P, de Maertelaer V, et al. Comparison of Computed Density and Microscopic Morphometry in Pulmonary Emphysema. Am J Respir Crit Care Med. 1996;154:187-92.

20. Silverman BW. Density estimation for statistics and data analysis. New York, N.Y., U.S.A.: Chapman & Hall/CRC; 1986.

21. Kohavi R. A Study of Cross-Validation and Bootstrap for Accuracy Estimation and Model Selection. International Joint Conference on Artificial Intelligence (IJCAI), volume 2, pp 1137–1143. 1995.

22. He K, Zhang X, Ren S, Sun J. Deep Residual Learning for Image Recognition. 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)2016. p. 770-8

23. Gonzalez G, Ash SY, Vegas-Sanchez-Ferrero G, Onieva Onieva J, Rahaghi FN, Ross JC, et al. Disease Staging and Prognosis in Smokers Using Deep Learning in Chest Computed Tomography. Am J Respir Crit Care Med. 2018;197(2):193-203.

- 1  
2  
3 24. Tang LYW, Coxson HO, Lam S, Leipsic J, Tam RC, Sin DD. Towards large-scale case-finding:  
4 training and validation of residual networks for detection of chronic obstructive pulmonary  
5 disease using low-dose CT. *The Lancet Digital Health*. 2020;2(5):e259-e67.  
6  
7  
8 25. Walsh SLF, Calandriello L, Silva M, Sverzellati N. Deep learning for classifying fibrotic lung  
9 disease on high-resolution computed tomography: a case-cohort study. *The Lancet Respiratory*  
10 *Medicine*. 2018;6(11):837-45.  
11  
12  
13 26. Shaker SB, Dirksen A, Lo P, Skovgaard LT, de Bruijne M, Pedersen JH. Factors influencing  
14 the decline in lung density in a Danish lung cancer screening cohort. *Eur Respir J*.  
15 2012;40(5):1142-8.  
16  
17  
18 27. Gershon AS, Thiruchelvam D, Chapman KR, Aaron SD, Stanbrook MB, Bourbeau J, Tan W,  
19 To T; Canadian Respiratory Research Network. Health Services Burden of Undiagnosed and  
20 Overdiagnosed COPD. *Chest* 2018;153(6):1336-1346.  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

## Supplementary Figure Legends

**Supplementary Figure S1.** A visual summary of our deep learning experiments. A,B: CNN-on-CT experiment. C,D: CNN-on-MCI experiment. In the training phase of both experiments, a pre-trained ResNet50 model (orange box) is further trained with 11 axial sections extracted from each 3D CT scan (A), or from each 3D MCI image volume (C) in our training dataset. In the inference phase, 11 axial sections are extracted from each 3D CT scan (B) or from each 3D MCI image volume (D) in the testing dataset. Each of the extracted test sections is analysed separately with the appropriately trained ResNet50 model, displayed as a gold box for ResNet50 trained on CT images (B), or as a blue box for ResNet50 trained on MCI (D). Predictions thus obtained for individual sections are then combined via majority voting to obtain a final prediction for the 3D CT scan (B) or the 3D MCI volume (D). See text for further details.

**Supplementary Figure S2.** Illustration of the invariance of MCI to invertible image transformations. From top to bottom, first row: axial section of CT image (left), corresponding MCI image (right). Second row: masked original image (left), masked MCI image (right). Third row: Masked original image to the power of 3 (left), masked MCI image (right). Fourth row: Cube root of masked original image (left), masked MCI image (right).

**Supplementary Figure S3.** Left: A zoom-in on a region of the chest CT section previously shown in the top row of Supplementary Figure S1. Right: a zoom-in on the corresponding MCI image. We remind the reader that bright structures in the CT image reflect blood vessels, and dark areas consist of air space and low-density lung tissue such as alveoles. Note that the MCI computation is a 3D operation. While we show only a 2D section, the MCI values reflect structures within a 3D neighborhood around each section pixel, including sections above and below the one displayed.

**Supplementary Figure S4.** Swarmbox plots for different measures for patients separated by GOLD stage. Top: hMCI1. Middle:  $LAA_{<-950HU}$ . Bottom:  $PD_{15}$ .



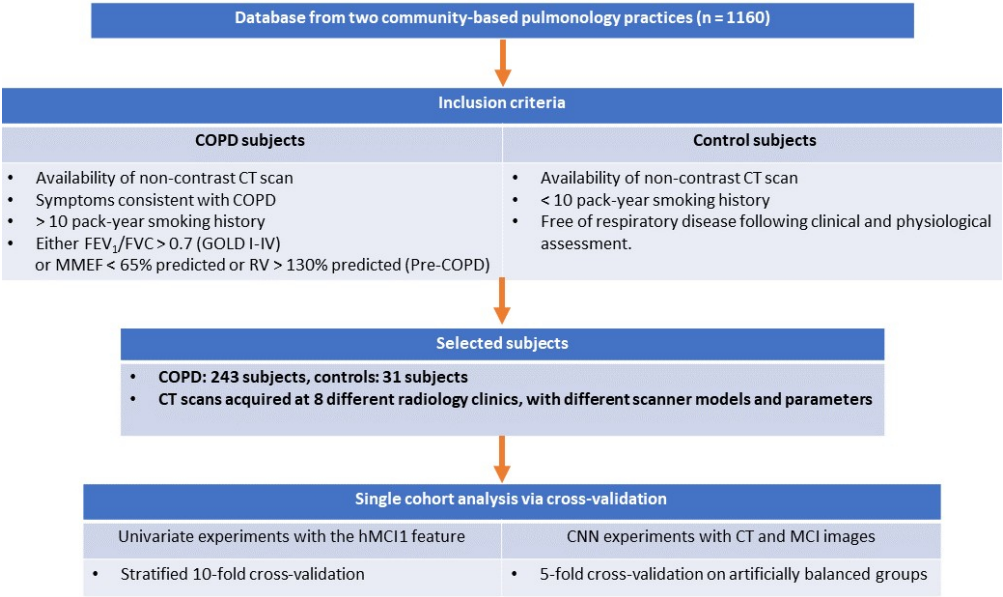


Figure 1. Flowchart summarizing subject selection and analysis. FEV1: forced expiratory volume in one second. FVC: forced vital capacity. GOLD: Global Initiative for Chronic Obstructive Lung Disease. MMEF: mid maximal expiratory flow. RV: residual volume.

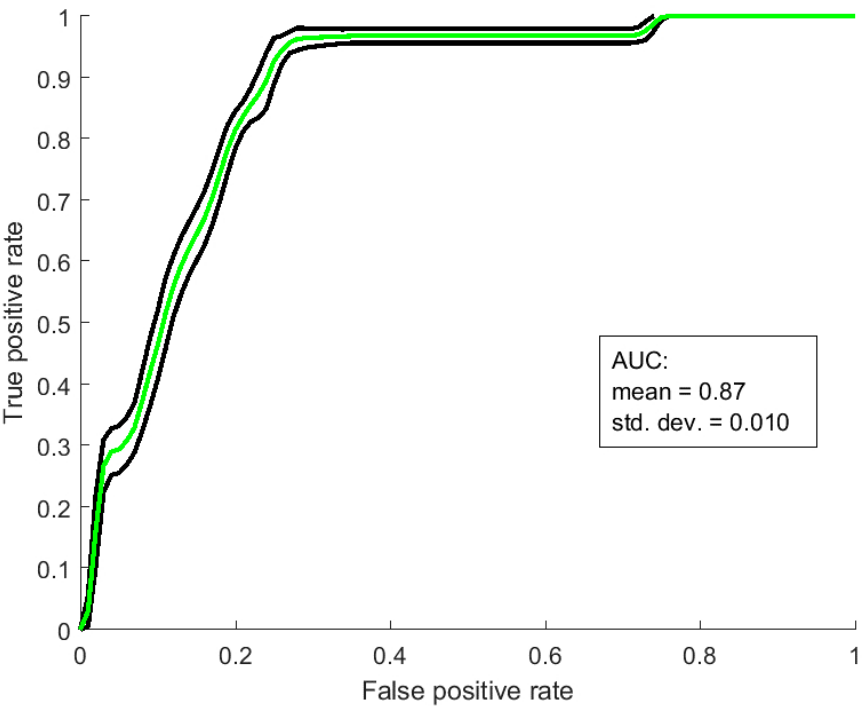
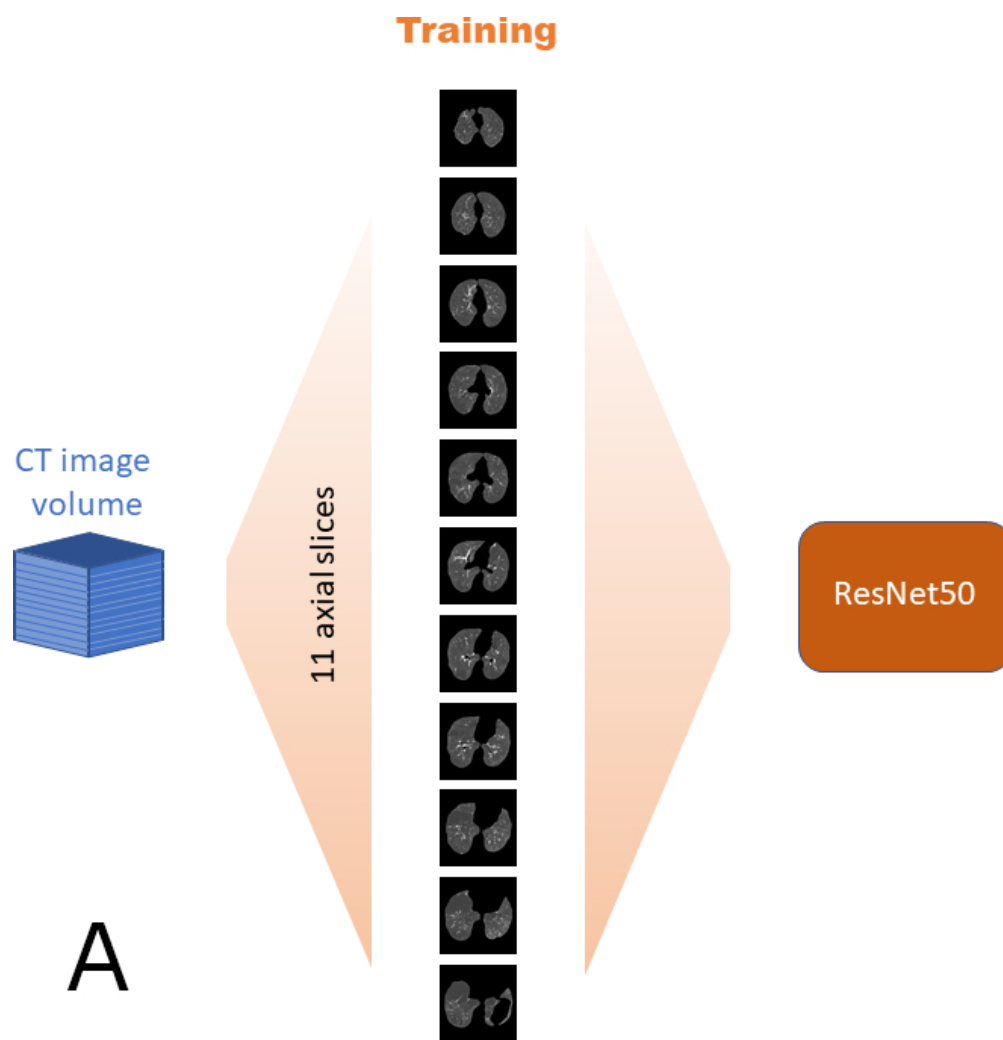


Figure 2. Classification between controls and all COPD subjects using hMCI1 as a single feature in a linear classification model. Mean ROC curve (green) computed over the 10 folds of stratified cross-validation Black:  $\pm 1$  standard deviation from the mean ROC curve. AUC: area under the curve. Std. dev: standard deviation.

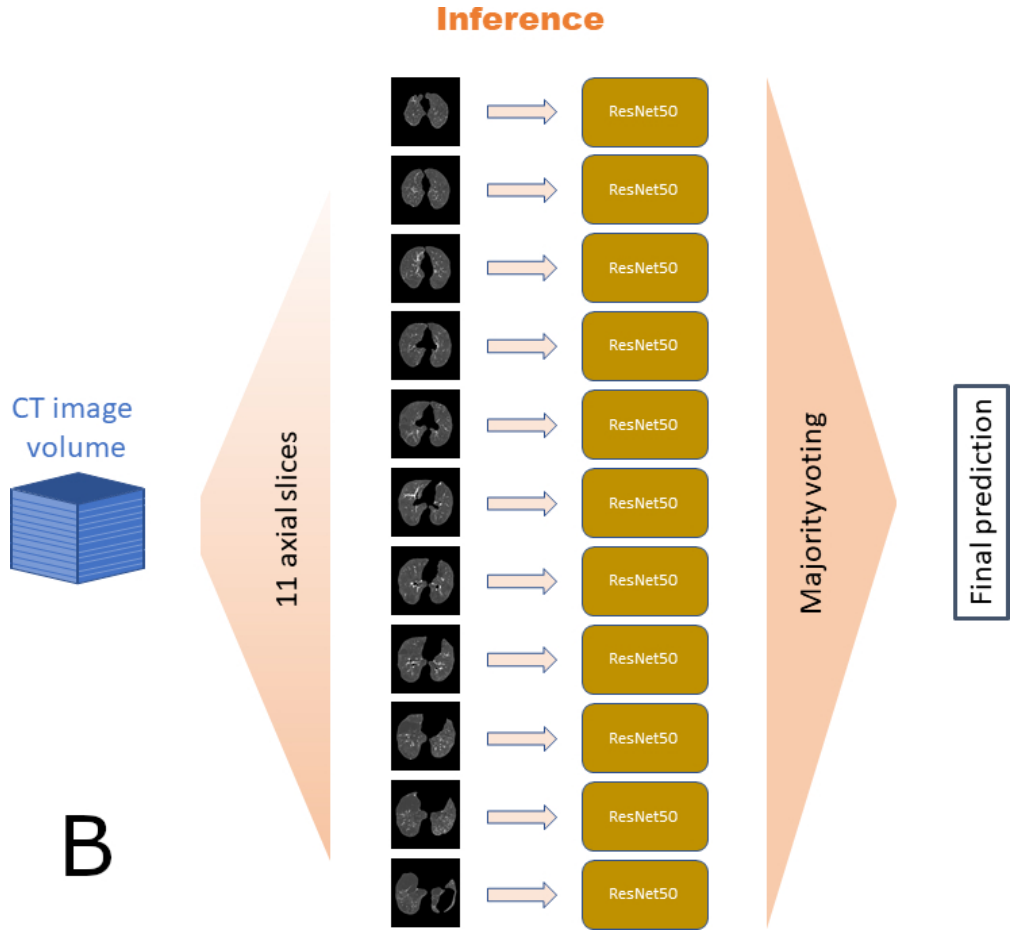
74x55mm (300 x 300 DPI)



39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

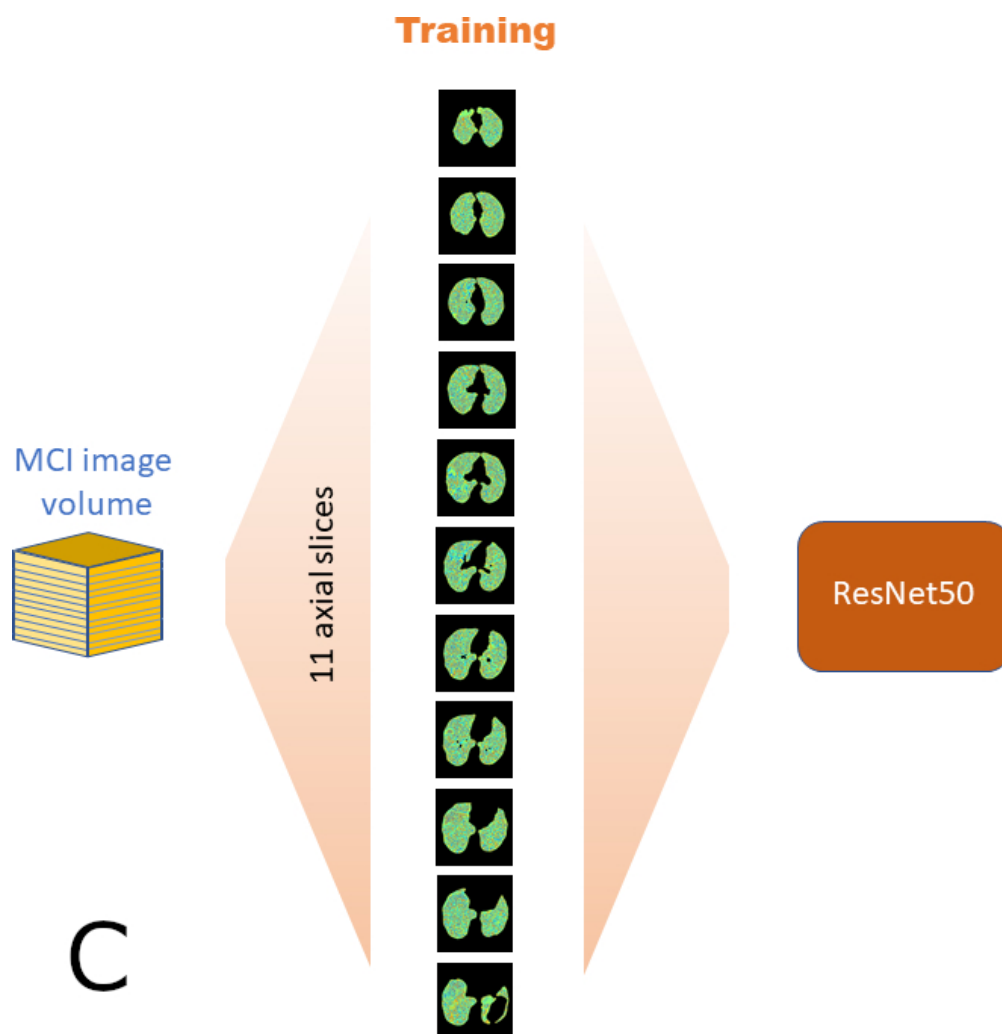
Supplementary Figure S1 (A).A visual summary of the CNN-on-CT experiment. In the training phase, a pre-trained ResNet50 model (orange box) is further trained with 11 axial slices extracted from each 3D CT scan in the training dataset

55x57mm (300 x 300 DPI)



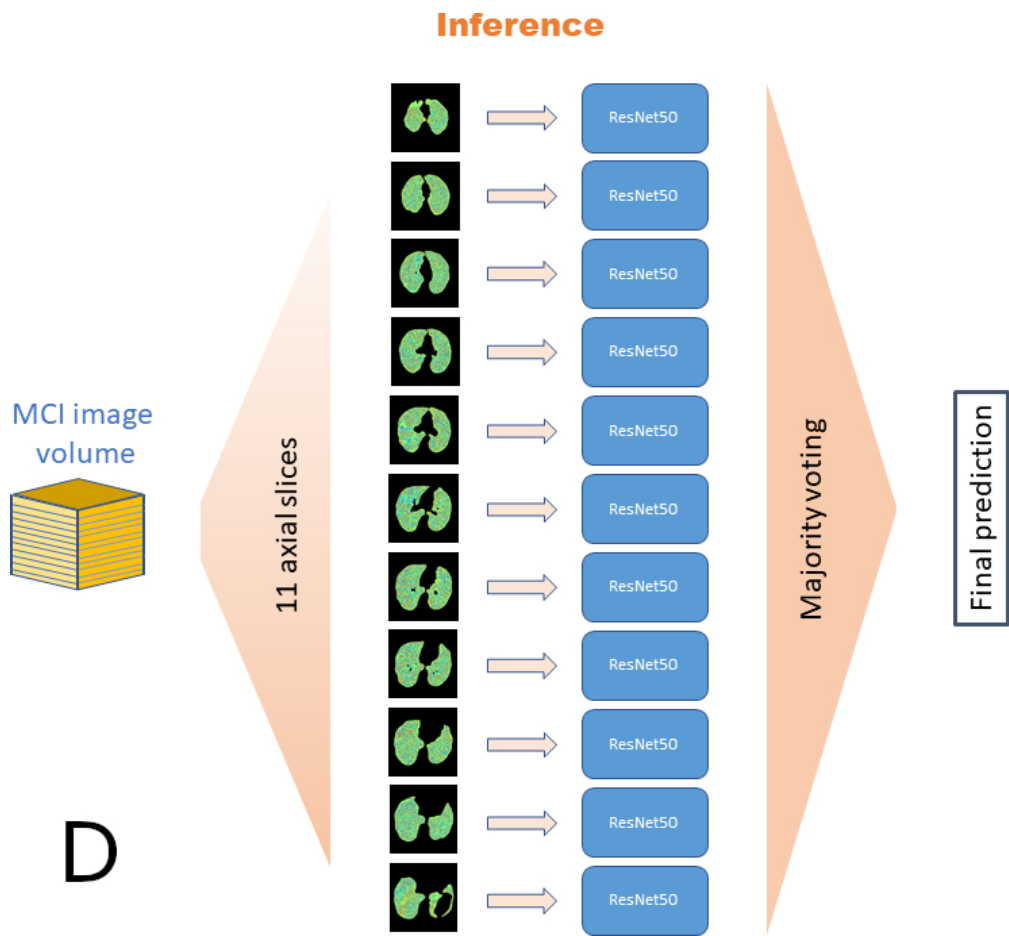
Supplementary Figure S1 (B).A visual summary of the CNN-on-CT experiment. In the inference phase, 11 axial slices are extracted from each 3D CT scan in the testing dataset. Each of the extracted test slices is analysed separately with the appropriately trained ResNet50 model, displayed as a gold box for ResNet50 trained on CT images. Predictions thus obtained for individual slices are then combined via majority voting to obtain a final prediction for the 3D CT scan.

62x57mm (300 x 300 DPI)



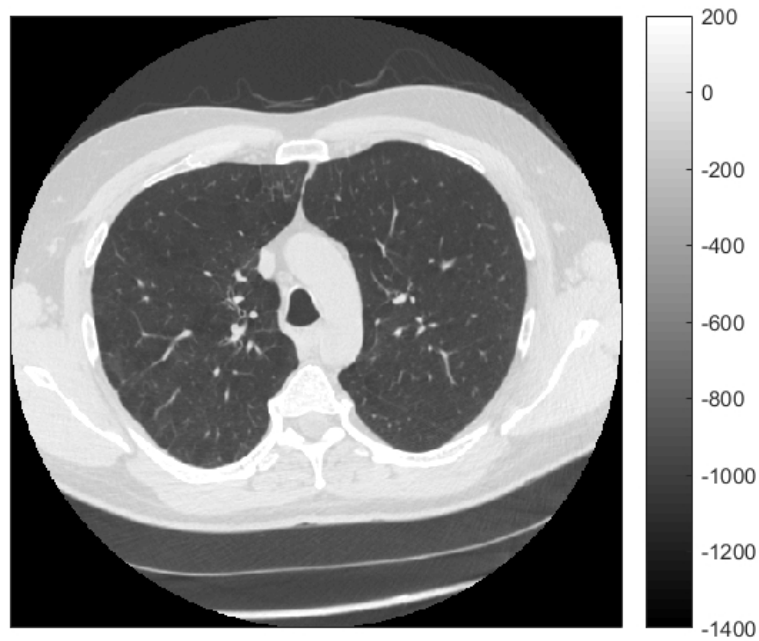
Supplementary Figure S1 (C).A visual summary of the CNN-on-MCI experiment. In the training phase, a pre-trained ResNet50 model (orange box) is further trained with 11 axial slices extracted from each 3D MCI image volume in the training dataset

55x57mm (300 x 300 DPI)



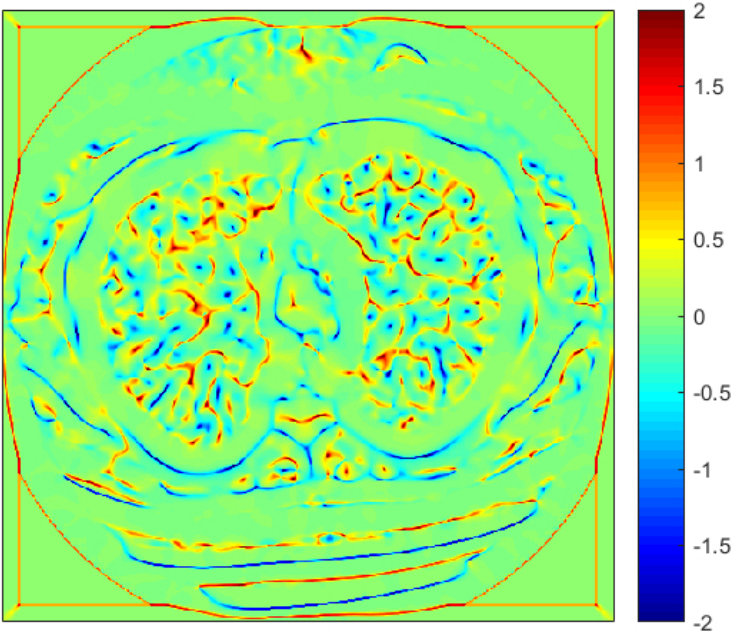
Supplementary Figure S1 (D).A visual summary of the CNN-on-MCI experiment. In the inference phase, 11 axial slices are extracted from each 3D MCI image volume in the testing dataset. Each of the extracted test slices is analysed separately with the appropriately trained ResNet50 model, displayed as a blue box for ResNet50 trained on MCI images. Predictions thus obtained for individual slices are then combined via majority voting to obtain a final prediction for the 3D MCI image volume.

62x57mm (300 x 300 DPI)



Supplementary Figure S2 Top row Left. Illustration of the invariance of MCI to invertible image transformations. Axial slice of CT scan.

59x44mm (300 x 300 DPI)



Supplementary Figure S2 Top row Right. Illustration of the invariance of MCI to invertible image transformations. MCI image corresponding to original axial CT slice.

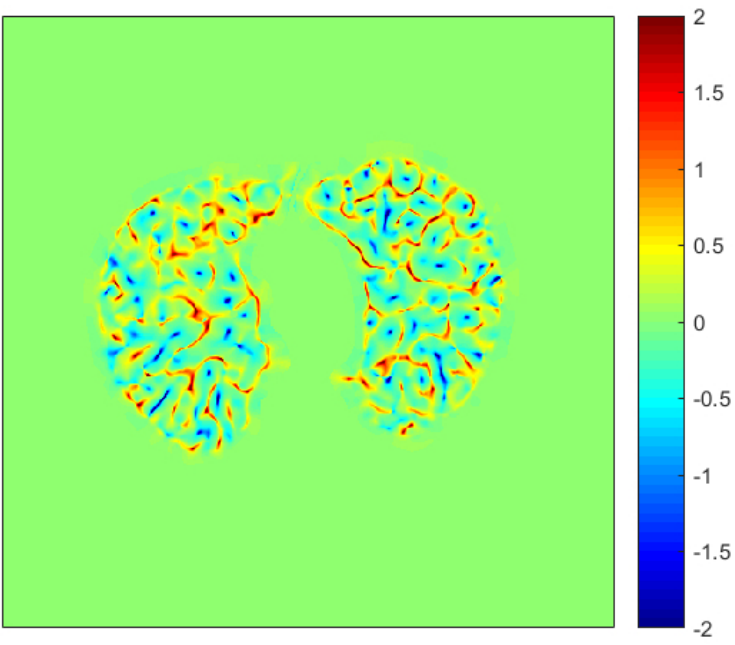
59x44mm (300 x 300 DPI)





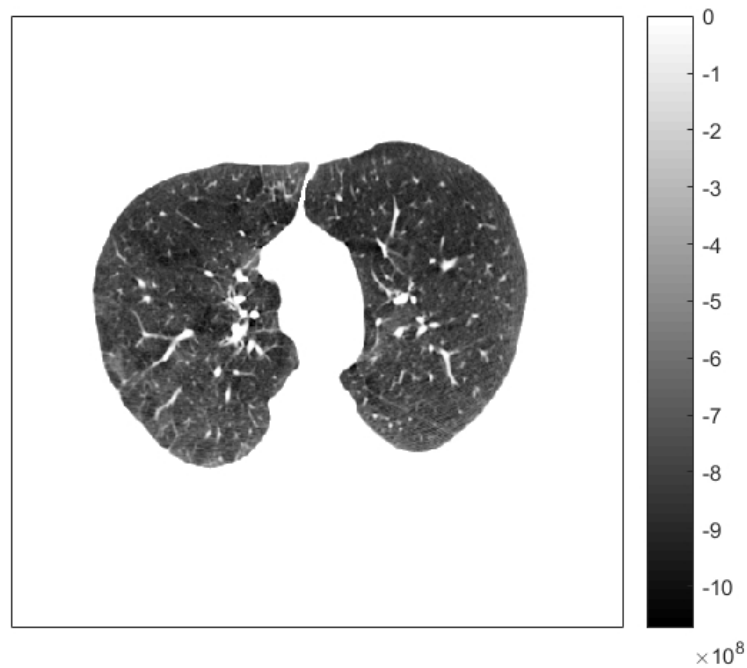
Supplementary Figure S2 Second row Left. Illustration of the invariance of MCI to invertible image transformations. Masked original axial CT slice.

59x44mm (300 x 300 DPI)



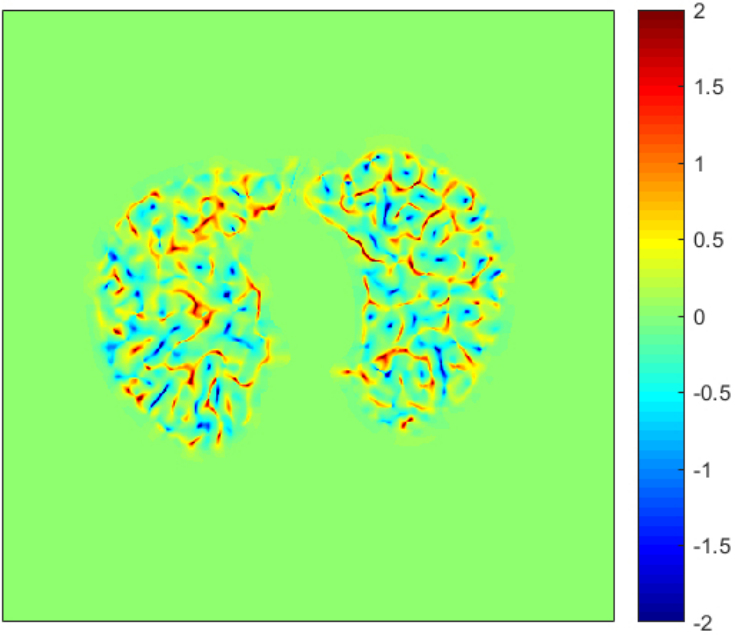
Supplementary Figure S2 Second row Right. Illustration of the invariance of MCI to invertible image transformations. Masked MCI image.

59x44mm (300 x 300 DPI)

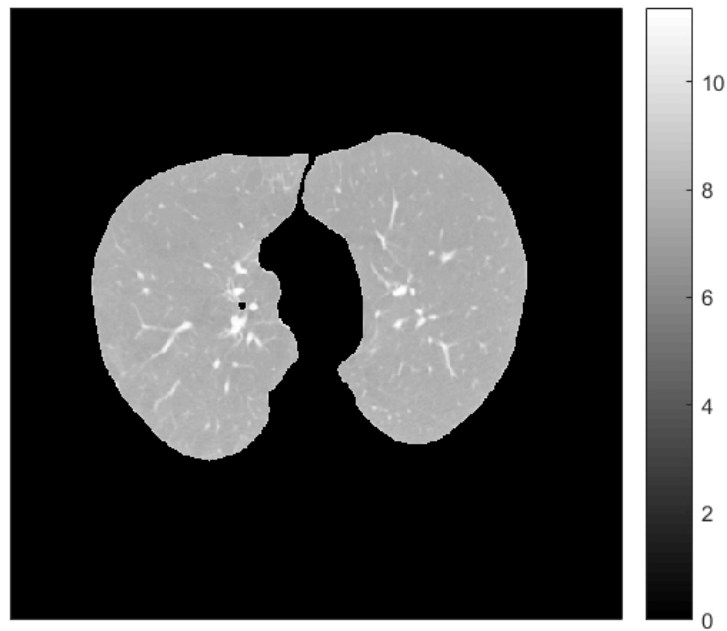


Supplementary Figure S2 Third row Left. Illustration of the invariance of MCI to invertible image transformations. Masked original axial CT slice to the power of 3.

59x44mm (300 x 300 DPI)

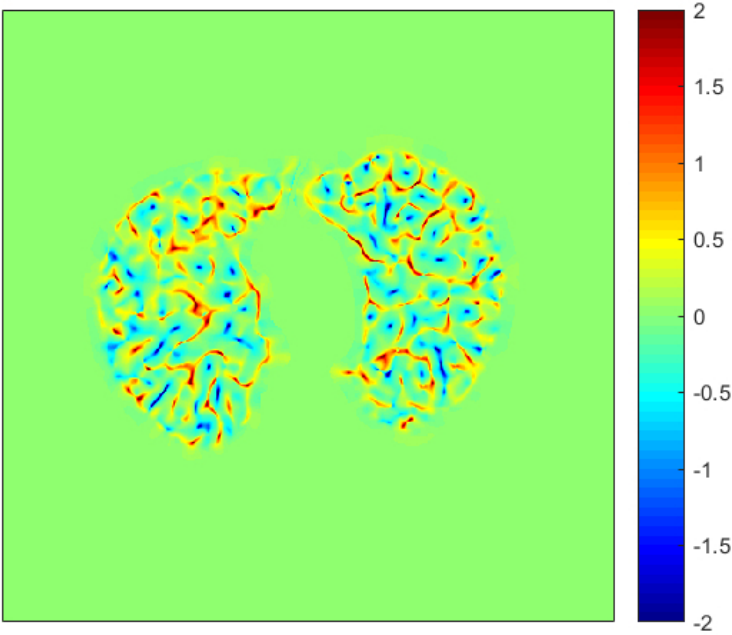


Supplementary Figure S2 Third row Right. Illustration of the invariance of MCI to invertible image transformations. Masked MCI image corresponding to the original CT slice to the power of 3.  
59x44mm (300 x 300 DPI)



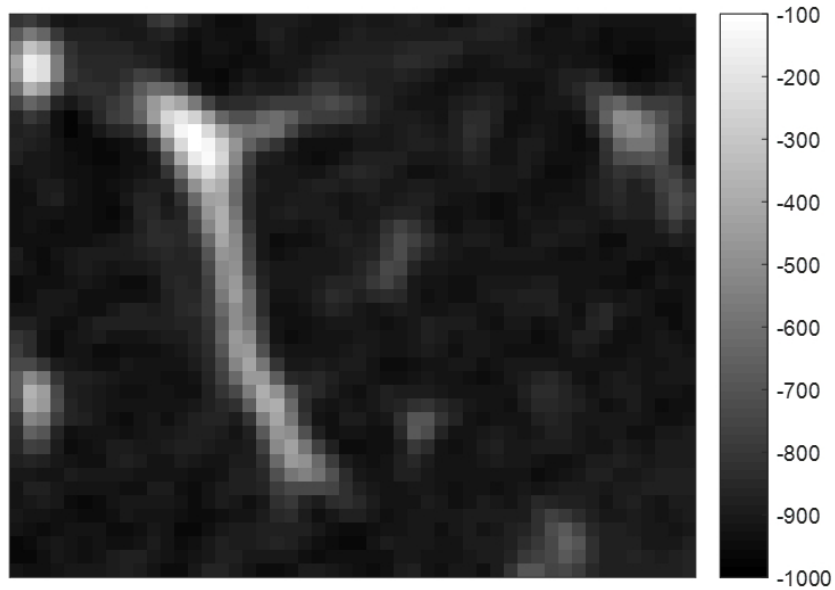
Supplementary Figure S2 Bottom row Left. Illustration of the invariance of MCI to invertible image transformations. Cube root of original masked CT slice.

59x44mm (300 x 300 DPI)



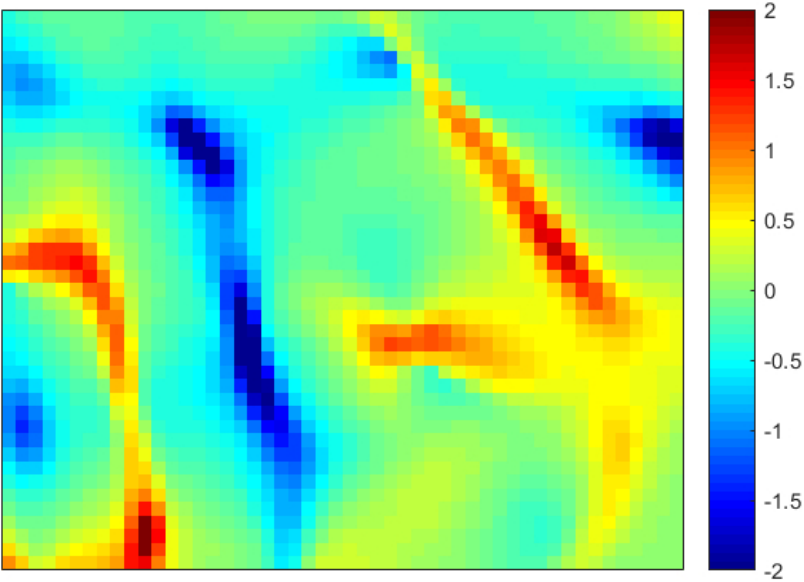
Supplementary Figure S2 Bottom row Right. Illustration of the invariance of MCI to invertible image transformations. Masked MCI image corresponding to the cube root of the original CT slice.

59x44mm (300 x 300 DPI)



Supplementary Figure S3. Left: A zoom-in on a region of the chest CT slice previously shown in the top row of Supplementary Figure S1.

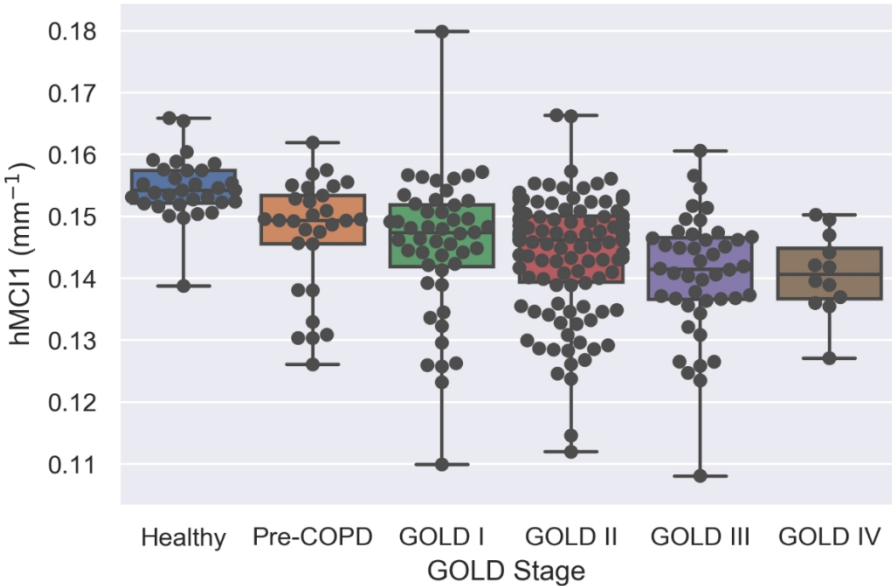
74x55mm (300 x 300 DPI)



Supplementary Figure S3. Right: A zoom-in on the corresponding MCI image. We remind the reader that bright structures in the CT image reflect blood vessels, and dark areas consist of air space and low density lung tissue such as alveoles. Note that the MCI computation is a 3D operation. While we show only a 2D slice, the MCI values reflect structures within a 3D neighborhood around each slice pixel, including slices above and below the one displayed.

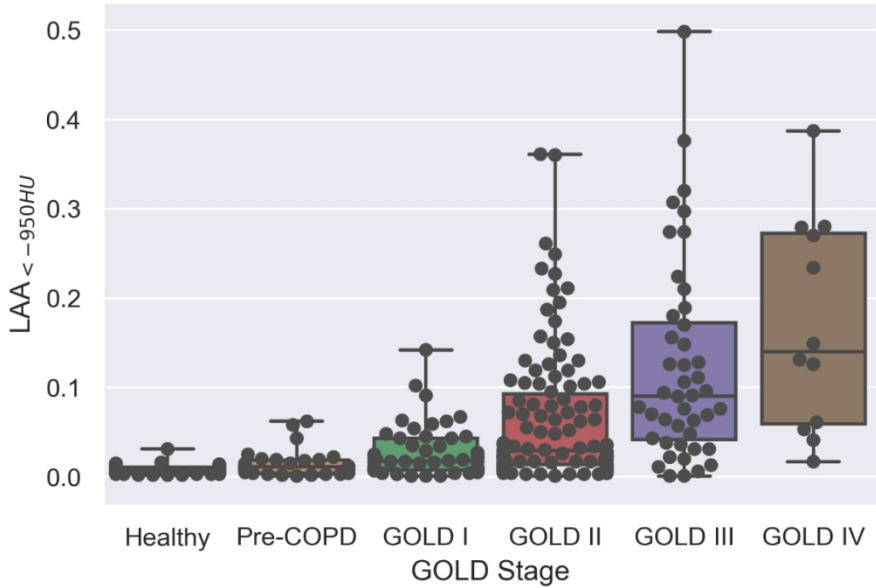
59x44mm (300 x 300 DPI)





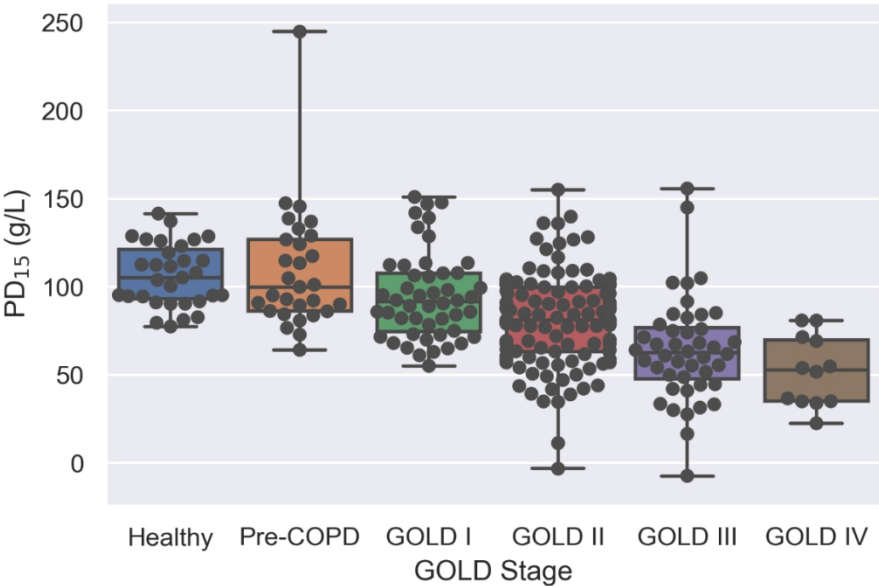
Supplementary Figure S4 Top:Swarmbox plots for different measures for our subjects separated by GOLD stage. hMCI1.

111x74mm (300 x 300 DPI)



Supplementary Figure S4 Middle:Swarmbox plots for different measures for our subjects separated by GOLD stage. LAA<-950HU.

111x74mm (300 x 300 DPI)



Supplementary Figure S4 Bottom:Swarmbox plots for different measures for our subjects separated by GOLD stage. PD15.

111x74mm (300 x 300 DPI)