

## Article

# A Comparative Evaluation of Computed Tomography Images for the Classification of Spirometric Severity of the Chronic Obstructive Pulmonary Disease with Deep Learning

Hiroyuki Sugimori <sup>1</sup>, Kaoruko Shimizu <sup>2,\*</sup>, Hironi Makita <sup>2,3</sup>, Masaru Suzuki <sup>2</sup> and Satoshi Konno <sup>2</sup><sup>1</sup> Faculty of Health Sciences, Hokkaido University, Sapporo 060-0812, Japan; sugimori@hs.hokudai.ac.jp<sup>2</sup> Department of Respiratory Medicine, Faculty of Medicine, Hokkaido University, Sapporo 060-8648, Japan; hiron@hokkaido.med.or.jp (H.M.); suzumasa@med.hokudai.ac.jp (M.S.); satkonno@med.hokudai.ac.jp (S.K.)<sup>3</sup> Hokkaido Medical Research Institute for Respiratory Diseases, Sapporo 064-0807, Japan

\* Correspondence: okaoru@med.hokudai.ac.jp; Tel.: +81-11-706-5911

**Abstract:** Recently, deep learning applications in medical imaging have been widely applied. However, whether it is sufficient to simply input the entire image or whether it is necessary to preprocess the setting of the supervised image has not been sufficiently studied. This study aimed to create a classifier trained with and without preprocessing for the Global Initiative for Chronic Obstructive Lung Disease (GOLD) classification using CT images and to evaluate the classification accuracy of the GOLD classification by confusion matrix. According to former GOLD 0, GOLD 1, GOLD 2, and GOLD 3 or 4, eighty patients were divided into four groups ( $n = 20$ ). The classification models were created by the transfer learning of the ResNet50 network architecture. The created models were evaluated by confusion matrix and AUC. Moreover, the rearranged confusion matrix for former stages 0 and  $\geq 1$  was evaluated by the same procedure. The AUCs of original and threshold images for the four-class analysis were  $0.61 \pm 0.13$  and  $0.64 \pm 0.10$ , respectively, and the AUCs for the two classifications of former GOLD 0 and  $\text{GOLD} \geq 1$  were  $0.64 \pm 0.06$  and  $0.68 \pm 0.12$ , respectively. In the two-class classification by threshold image, recall and precision were over 0.8 in  $\text{GOLD} \geq 1$ , and in the McNemar–Bowker test, there was some symmetry. The results suggest that the preprocessed threshold image can be possibly used as a screening tool for GOLD classification without pulmonary function tests, rather than inputting the normal image into the convolutional neural network (CNN) for CT image learning.



**Citation:** Sugimori, H.; Shimizu, K.; Makita, H.; Suzuki, M.; Konno, S. A Comparative Evaluation of Computed Tomography Images for the Classification of Spirometric Severity of the Chronic Obstructive Pulmonary Disease with Deep Learning. *Diagnostics* **2021**, *11*, 929. <https://doi.org/10.3390/diagnostics11060929>

Academic Editor: Cesar A. Moran

Received: 30 April 2021

Accepted: 19 May 2021

Published: 21 May 2021

**Publisher's Note:** MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Copyright:** © 2021 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

**Keywords:** image classification; chronic obstructive pulmonary disease; deep learning

## 1. Introduction

Chronic obstructive pulmonary disease (COPD) affects an estimated 384 million people worldwide [1] and is reported to be the third leading cause of death worldwide [2]. The exacerbation of COPD not only leads to a significant decline in lung function [3] but also may shorten life expectancy and increase the risk of death [4,5], as well as causing significant economic and social burden [6]. COPD is defined by the Global Initiative for Chronic Obstructive Lung Disease (GOLD) published by WHO and the National Heart, Lung, and Blood Institute (NHLBI) [7] and has been revised since then. It is important for the early diagnosis of COPD to perform spirometry to measure forced expiratory volume in 1 s ( $\text{FEV}_1$ )/forced vital capacity (FVC). In particular, there is an abundance of studies [8–14] using CT, which have reported important findings on the loss of lung function, progression, and prognosis over time. However, in patients with clinical findings or contact history suspected of recent COVID-19 infection, pulmonary function tests may be discontinued or postponed, and there is concern that performing procedures such as deep breathing or forced expiration with maximal effort may cause the spread of contaminated droplets and aerosols to the surrounding area, resulting in the spread of infection. Therefore, we believe

that it is important to provide diagnostic assistance without using pulmonary function tests. In the conventional diagnosis of emphysema in COPD with significant emphysema lesions, the lesion is seen on CT as a low attenuation area (LAA) [15,16], and the boundary is characterized by the absence of a clear wall. In early emphysema, small clusters of LAAs are formed, but as the lesion progresses, the fused LAAs occupy most of the lung field; the Goddard method [17] is representative for visual emphysema severity classification on CT. Additionally, in COPD with significant peripheral airway involvement, which is thought to be one of the main loci in the pathogenesis of COPD together with emphysema, it has been pointed out that central airways are also involved [18]. There are many reports on CT imaging in emphysema including COPD [10–14,19], and CT imaging is widely used as a simple imaging modality to obtain information on the whole lung.

Recently, the deep learning application in medical imaging has been actively promoted, and in the field of medicine, it has been widely used for image classification [20–23], object detection [24–26], semantic segmentation [27–30], and so on and has been widely applied from disease classification to detection. Deep learning has been reported for the automatic classification of emphysema patterns in CT images [31], but CT images were used as a simple input without any preprocessing. In image classification using convolutional neural networks (CNNs), features are extracted from the entire image and classified, but whether it is sufficient to simply input the entire image or whether preprocessing is required to set up a supervised image has not been sufficiently investigated. Additionally, there has been no attempt to classify the spirometric severity of COPD using CT images based on the diagnosis results of pulmonary function tests using image classification technology of artificial intelligence. This study aimed to create a classifier trained with and without preprocessing to detect the groups that require therapeutic intervention using CT images and to evaluate the recall, precision, overall accuracy, and AUC of different spirometric classifications using confusion matrix.

## 2. Materials and Methods

### 2.1. Subjects

A total of eighty patients (male: 77; female: 3, mean age  $\pm$  SD:  $69.8 \pm 7.9$  years) who underwent a chest CT scan on a 4-row CT (SOMATOM plus Volume Zoom; Siemens AG, Berlin, Germany) between 2003 and 2006 were classified into four spirometric classifications: without airflow limitation (FVC  $\geq$  70%, former GOLD 0) ( $n = 20$ ,  $66.4 \pm 7.2$  years), GOLD 1 ( $n = 20$ ,  $67.3 \pm 9.0$  years), GOLD 2 ( $n = 20$ ,  $72.5 \pm 6.8$  years), and GOLD 3 or 4 ( $n = 20$ ,  $72.4 \pm 7.2$  years) based on the values of post-bronchodilator pulmonary function tests by using a rolling seal Chestac-33 spirometer (Chest MI, Tokyo, Japan) on the same day as CT scans. The pulmonary function, based on the values of post-bronchodilator spirometry carbon monoxide diffusing capacity (DL<sub>CO</sub>) and transfer coefficient (K<sub>co</sub>), measured using the single breath method, was measured. The procedures met the requirements of the Japanese Respiratory Society Guidelines [32]. These subjects (Table 1) were participants in the Hokkaido COPD cohort study [33,34]. These subjects were included in this study because it is important to classify them using a group of subjects whose patient background has been diagnosed with accurate examinations in order to define accurate supervised images.

### 2.2. Preprocessing for the Lung CT Image

Two types of image data were created: the original image, which was converted from a DICOM image to a JPEG image with a pixel size of  $512 \times 512$ , and the threshold image, which was created by preprocessing a CT image to extract only CT values of  $-950$  or less in the lung field and filling it with red. The threshold setting of  $-950$  for the CT value was defined based on the report [35] that the threshold of  $-950$  HU for the CT value defining emphysema has the best correlation with pathological emphysematous lesions. For the next step, since CT lung field images include the entire chest, including areas where lung fields are not depicted, based on the positional relationship between the trachea and both

lung fields, we defined the lower lung field as the area where the upper border of the liver was not depicted and the upper lung field as the area where the trachea was not depicted anterior to the lung field. Based on the original image, slices were selected for the threshold image in the same procedure. These slices were selected based on the original image and the threshold image. With the above two types of preprocessing, we divided the data into training and test data sets and created four subsets to create a four-class classifier classified by stage so that fourfold cross validation could be performed (Table 2).

**Table 1.** Characteristics of subjects.

|                                                              | Former<br>GOLD 0<br>(N = 20) | GOLD 1<br>(N = 20)  | GOLD 2<br>(N = 20)  | GOLD 3 or 4<br>(N = 20) |
|--------------------------------------------------------------|------------------------------|---------------------|---------------------|-------------------------|
| Age, yr (Mean $\pm$ S.D.)                                    | 66.4 $\pm$ 7.2               | 67.3 $\pm$ 9.0      | 72.5 $\pm$ 6.8      | 72.4 $\pm$ 7.2          |
| Female sex, N (%)                                            | 1 (5)                        | 2 (10)              | 0                   | 0                       |
| Body mass index,<br>kg/m <sup>2</sup> (Mean $\pm$ S.D.)      | 22.9 $\pm$ 3.6               | 23.5 $\pm$ 2.6      | 22.8 $\pm$ 3.7      | 21.9 $\pm$ 2.4          |
| Current smoker, N (%)                                        | 8 (40)                       | 4 (20)              | 7 (35)              | 2 (10)                  |
| Smoking index,<br>pack-years [Median<br>(Q1-Q3)]             | 23.9 (45–86)                 | 20.3 (43.2–65.3)    | 29.2 (46.7–69.8)    | 19.3 (53.5–76)          |
| Lung function                                                |                              |                     |                     |                         |
| Post-bronchodilator<br>FVC, % predicted<br>(Mean $\pm$ S.D.) | 99.8 $\pm$ 10.1              | 119.9 $\pm$ 10.9    | 100.7 $\pm$ 12.2    | 88.2 $\pm$ 13.2         |
| FEV <sub>1</sub> , % predicted<br>(Mean $\pm$ S.D.)          | 91.8 $\pm$ 13.7              | 90.2 $\pm$ 9.8      | 61.3 $\pm$ 6        | 40.4 $\pm$ 7.5          |
| FEV <sub>1</sub> /FVC (Mean $\pm$<br>S.D.)                   | 0.75 $\pm$ 0.08              | 0.61 $\pm$ 0.07     | 0.49 $\pm$ 0.08     | 0.37 $\pm$ 0.07         |
| Reversibility of FEV <sub>1</sub> , %<br>[Median (Q1-Q3)]    | 4.1 (0.2–6.1)                | 8.8 (1.6–7.2)       | 12.2 (4.9–23.1)     | 12.9 (8.8–27.3)         |
| Reversibility of FEV <sub>1</sub> ,<br>ml [Median (Q1-Q3)]   | 96.6 (5–150)                 | 137.6<br>(40–162.5) | 135<br>(77.5–277.5) | 108.2<br>(67.5–225)     |
| DL <sub>CO</sub> , mmol/min/mm<br>Hg (Mean $\pm$ S.D.)       | 90.9 $\pm$ 16.5              | 86.7 $\pm$ 18.1     | 84.1 $\pm$ 17.5     | 81.3 $\pm$ 29.9         |
| Kco, mmol/min/mm<br>Hg/L (Mean $\pm$ S.D.)                   | 84.7 $\pm$ 20.2              | 75.6 $\pm$ 18.9     | 73.1 $\pm$ 19       | 74.6 $\pm$ 28.4         |

FVC: forced vital capacity, FEV<sub>1</sub>: forced expiratory volume in 1 sec, DL<sub>CO</sub>: carbon monoxide diffusing capacity, Kco: transfer coefficient, S.D.: standard deviation, Q1: first quartile, Q3: third quartile.

**Table 2.** The number of images in each subset.

|                  | Subset A | Subset B | Subset C | Subset D | Total  |
|------------------|----------|----------|----------|----------|--------|
| Former<br>GOLD 0 | 1545     | 1643     | 1578     | 1648     | 6414   |
| GOLD 1           | 1790     | 1659     | 1828     | 1507     | 6784   |
| GOLD 2           | 1716     | 1665     | 1651     | 1564     | 6596   |
| GOLD 3 or 4      | 1665     | 1678     | 1820     | 1837     | 7000   |
| Total            | 6716     | 6645     | 6877     | 6556     | 26,794 |

### 2.3. Training and Evaluation for Creating Models

The software for the deep learning technique was developed with in-house MATLAB software (The MathWorks, Inc., Natick, MA, USA) and the use of a desktop computer with an NVIDIA RTX 1080Ti or RTX 2080Ti graphics card (Nvidia Corporation, Santa Clara, CA, USA). ResNet50 was used as the CNN, and the hyperparameters were set to epoch 10 and base learning rate 0.0001. The training dataset consisting of each subset was used for data augmentation, and the images were rotated from  $-25$  degrees to  $25$  degrees. To create four classifiers for both the original image and the threshold image, we trained each

of these images. The classifiers were tested with test data for each dataset, and the results of classification into former GOLD 0, GOLD 1, GOLD 2, and GOLD 3 or 4 were combined into a confusion matrix to obtain the recall, precision, overall accuracy, and AUC (Figure 1). The confusion matrix was further divided into two classes, former GOLD 0 and  $GOLD \geq 1$ , and recall, precision, overall accuracy, and AUC were calculated for each class (Figure 2). The training was performed twice using RTX 1080Ti and RTX 2080Ti GPUs.

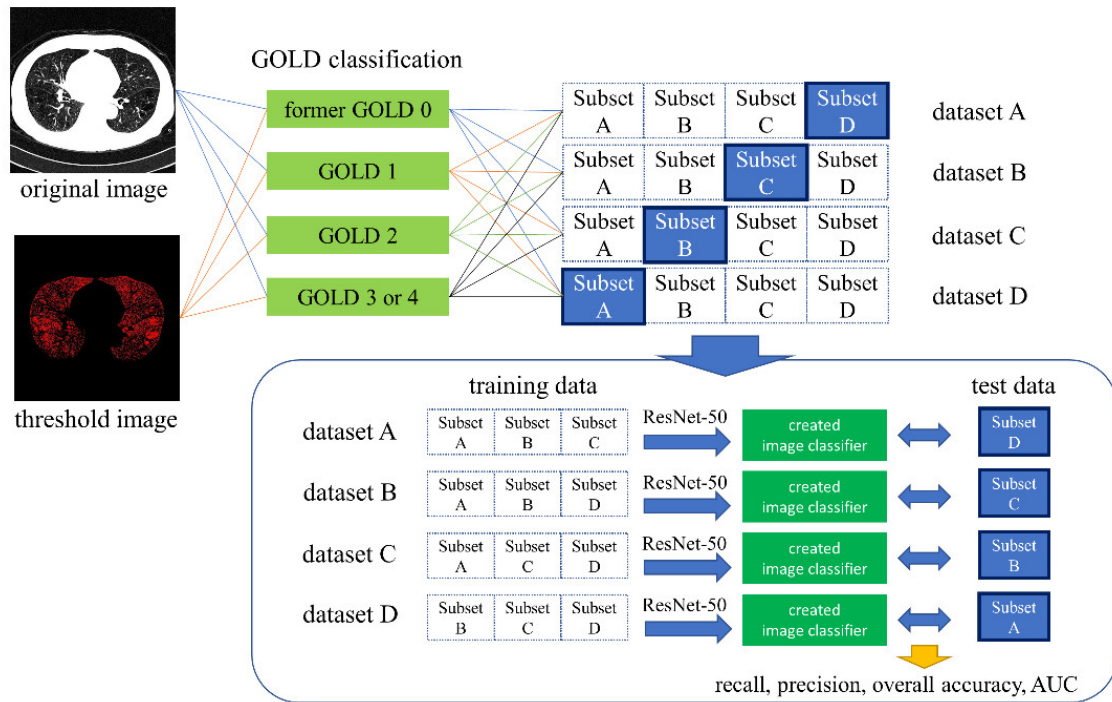


Figure 1. Schematic of the datasets and subsets for training and test.

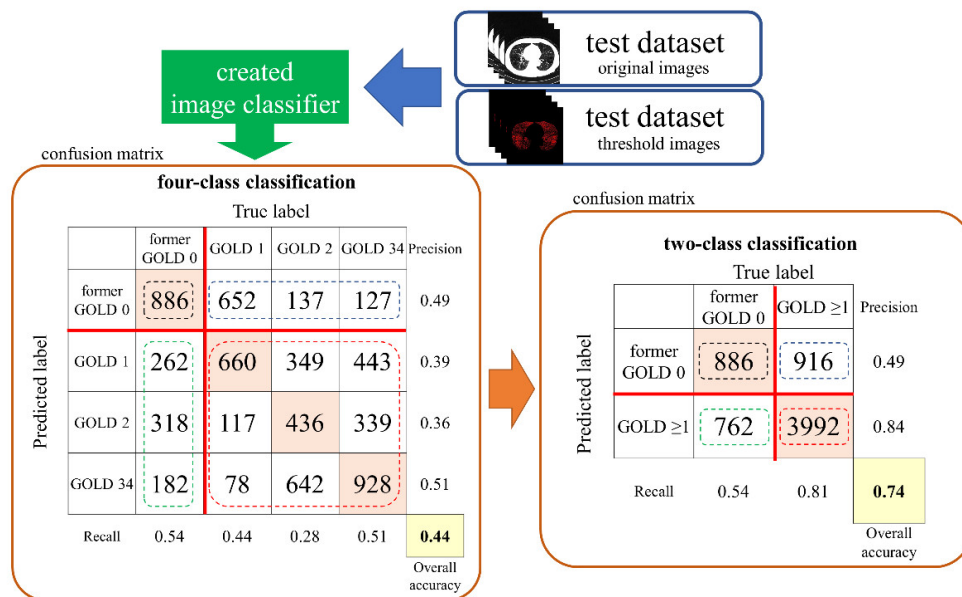


Figure 2. Confusion matrix for evaluating the recall, precision, and overall accuracy.

#### 2.4. Statistical Analysis

The obtained recall, precision, overall accuracy, and AUC were expressed as mean  $\pm$  standard deviation. The average of the calculation results on two GPUs per data set was used as the obtained results. The recall, precision, overall accuracy, and AUC were evaluated for each preprocessing. Firstly, we applied the Shapiro–Wilk test to the differences: if normality was accepted, the *t*-test was used; otherwise, the Wilcoxon signed ranked test was used. For comparing the methods of preprocessing, multiple comparisons between the spirometric classifications were performed to determine the optimal classification model. Levene’s homoscedasticity test was performed for the dose-dependence study. If homoscedasticity was confirmed, Dunnett’s multiple comparison test was used to assess between-group differences, while if homoscedasticity was not confirmed, the Steel–Dwass multiple comparison test was used. For comparing the methods of preprocessing, the McNemar–Bowker test was used to assess the symmetry of the confusion matrix. Differences in all statistical analyses were considered statistically significant if *p* was less than 0.05. All statistical analyses were performed using JMP software (version 14; SAS Institute Inc., Cary, NC, USA).

### 3. Results

#### 3.1. Four-Class Analysis

The detailed results are shown in Table 3. The recall of original and threshold images was  $0.35 \pm 0.15$  and  $0.39 \pm 0.17$ , respectively. The precision of original and threshold images was  $0.36 \pm 0.13$  and  $0.4 \pm 0.12$ , respectively. The AUC for original and threshold images was  $0.61 \pm 0.13$  and  $0.64 \pm 0.10$ , respectively. There were no significant differences in the mean values for recall, precision, overall accuracy, and AUC.

**Table 3.** Recall, precisions, AUC, and overall accuracy for four-class analysis.

|                  |               | Original Image  | Threshold Image |
|------------------|---------------|-----------------|-----------------|
| Recall           | Former GOLD 0 | $0.34 \pm 0.15$ | $0.4 \pm 0.19$  |
|                  | GOLD 1        | $0.28 \pm 0.12$ | $0.42 \pm 0.14$ |
|                  | GOLD 2        | $0.26 \pm 0.12$ | $0.25 \pm 0.16$ |
|                  | GOLD 3 or 4   | $0.51 \pm 0.09$ | $0.48 \pm 0.07$ |
| Precision        | Former GOLD 0 | $0.39 \pm 0.07$ | $0.42 \pm 0.14$ |
|                  | GOLD 1        | $0.26 \pm 0.1$  | $0.3 \pm 0.06$  |
|                  | GOLD 2        | $0.31 \pm 0.17$ | $0.39 \pm 0.1$  |
|                  | GOLD 3 or 4   | $0.46 \pm 0.09$ | $0.5 \pm 0.11$  |
| AUC              | Former GOLD 0 | $0.64 \pm 0.06$ | $0.68 \pm 0.12$ |
|                  | GOLD 1        | $0.56 \pm 0.13$ | $0.57 \pm 0.08$ |
|                  | GOLD 2        | $0.50 \pm 0.15$ | $0.60 \pm 0.07$ |
|                  | GOLD 3 or 4   | $0.72 \pm 0.08$ | $0.73 \pm 0.07$ |
| Overall accuracy |               | $0.35 \pm 0.11$ | $0.39 \pm 0.06$ |

A comparison between the original and thresholded images for each stage showed that, for the recall for the original image, there was a significant difference between GOLD 1 and GOLD 3 or 4 ( $p = 0.0202$ ) and between GOLD 2 and GOLD 3 or 4 ( $p = 0.0052$ ). For the precision for the original image, there was a significant difference between GOLD 1 and GOLD 3 or 4 ( $p = 0.0202$ ). For the precision for the threshold image, there was a significant difference between GOLD 1 and GOLD 3 or 4 ( $p = 0.0074$ ). There were no significant differences in the recall for the threshold image. For the AUC for the original image, there was a significant difference between GOLD 2 and GOLD 3 or 4 ( $p = 0.0052$ ). For the AUC for the threshold image, there was a significant difference between GOLD 1 and GOLD 3 or 4 ( $p = 0.0202$ ) and between GOLD 2 and GOLD 3 or 4 ( $p = 0.0372$ ). In all datasets, the McNemar–Bowker test showed that the *p* value was less than 0.05.



### 3.2. Two-Class Analysis

The detailed results are shown in Table 4. Both recall and precision were significantly higher for GOLD  $\geq 1$ , and symmetry was observed for dataset C in the McNemar–Bowker test ( $p = 0.5185$ ).

**Table 4.** Recall, precisions, AUC, and overall accuracy for two-class analysis.

|                  |               | Original Image  | Threshold Image |
|------------------|---------------|-----------------|-----------------|
| Recall           | Former GOLD 0 | $0.34 \pm 0.15$ | $0.40 \pm 0.19$ |
|                  | GOLD $\geq 1$ | $0.84 \pm 0.06$ | $0.84 \pm 0.02$ |
| Precision        | Former GOLD 0 | $0.39 \pm 0.07$ | $0.42 \pm 0.14$ |
|                  | GOLD $\geq 1$ | $0.80 \pm 0.02$ | $0.82 \pm 0.04$ |
| AUC              |               | $0.64 \pm 0.06$ | $0.68 \pm 0.12$ |
| Overall accuracy |               | $0.72 \pm 0.03$ | $0.74 \pm 0.04$ |

## 4. Discussion

### 4.1. Statement of Principal Findings

In this study, we developed and evaluated a classifier for spirometric classification using CT images with and without preprocessing prior to image learning, and the highest recall, precision, and AUC values were found in GOLD 3 or 4. As the percentage of emphysema in the lungs increases in severe COPD, the feature changes due to the enlarged alveolar spaces in the original image captured the feature values in CNN due to the destruction of alveolar septa and alveolar architecture. In the threshed image, due to the coarsening of the filled area of the airway caused by the threshold setting of CT value  $-950$ , the features were described. This is also true for the results of former GOLD 0 or GOLD  $\geq 1$ , and we comprehended that the results of GOLD 1 and above capture the structural changes caused by emphysema as a feature compared to former GOLD 0. Although there was no significant difference in this study, threshold images showed higher values than original images in all the result indices. The normal images are considered as positive data in CNN, which showed larger pixel values of lung structures including alveoli and arteriovenous veins than air. On the other hand, in the threshold image, features were learned from the positive data of the so-called air data, which were the data below the CT value of  $-950$  in the thorax. Although both images in this study have a negative-positive relationship with each other, in the image classification, the results captured the features of each spirometric classification. In the normal image, the body contour information was used as positive data, and the weights and biases were updated in the CNN, so it was possible that features other than lung structure were also calculated. In this respect, the results of the original image were lower than those of the threshold image. Additionally, the air in the lung field was used as positive data to obtain feature values; however, in the actual image data, the thickening of the airway wall was ignored in the preprocessing. Airway alterations from central to small airways, besides emphysema, contribute to airflow limitation in COPD [18]. In the current study, we focused on the preprocessing of the lung field on inspiratory CT for the initial attempt to perform deep learning in respiratory diseases. It is plausible that learning about airway structures using inspiratory CT or the combined use of inspiratory and expiratory CT may improve the accuracy of the classification of spirometric severity. Although there was no direct relationship between severe COPD exacerbation and visual structure on CT images [9], another report showed that quantitative assessment using fractal exponent D detected the changes associated with exacerbations [36]. Moreover, the lung condition depending on the severity of the disease is characterized on CT images [8], and the image classification by deep learning can be used for the spirometric severity classification. Additionally, in the two-class classification by threshold image, the recall and precision are over 0.8 in GOLD  $\geq 1$ , and symmetry is observed in dataset C in the McNemar–Bowker test. Additionally, the symmetry of dataset

C in the McNemar–Bowker test shows that the preprocessed threshold image can be used as a screening tool for spirometric classification without pulmonary function tests.

#### 4.2. Strengths and Weaknesses of the Study

There are some limitations to this study. The first limitation was that the images were evaluated on a per image basis rather than on a per patient basis. However, in this study, all the images from the upper to lower lung fields of the patients in each spirometric classification were given to the CNN as training data and were fed to the CNN as training data. Therefore, in mild cases such as GOLD 1, emphysema may not be present in the entire lung field, and when evaluated with a single image, an image equivalent to former GOLD 0 was also trained as GOLD 1. However, the creation of a dataset based on the diagnosed spirometric severity conducted in this study learns features even for regions that cannot be easily determined by humans, and it cannot necessarily be determined that the image is inappropriate. Secondly, the number of subjects was small. In this study, 20 subjects in each group were used. In addition, the subjects of this study were patients who entered the COPD cohort study and were classified according to the GOLD classification within cohort study. We simply compared the results with and without preprocessing using a fixed model of CT imaging to understand how the presence of preprocessing affects the accuracy of the classifier. Since medical images are said to have different image characteristics depending on the vendor and imaging conditions [37], it was necessary to minimize the effect of differences in conditions caused by using images from other facilities on classification accuracy. Since COPD is a systemic inflammatory disease with various complications, it is desirable to manage COPD to maintain the long-term activity of daily living using drug therapy and rehabilitation, especially in Japan, where the population is superaged. Inhalation drug clinical trials have shown that the decline in GOLD 2 is more effective than that in the advanced stage, suggesting that CT imaging can be used to detect patients who require therapeutic intervention.

## 5. Conclusions

In this study, we developed a classifier trained with and without preprocessing for spirometric classification using CT images and evaluated the recall, precision, overall accuracy, and AUC of different spirometric classifications by confusion matrix. In the two-class classification by threshold image, recall and precision exceeded 0.8 for  $\text{GOLD} \geq 1$ , and in the McNemar–Bowker test, there was some symmetry. The results suggest that preprocessed threshold images can be used as a possible screening tool for spirometric classification without pulmonary function tests.

**Author Contributions:** H.S. contributed to data analysis, algorithm construction, and the writing and editing of the original article; K.S. proposed the idea and contributed to data acquisition, performed project administration, and reviewed and edited the paper; H.M. contributed to data acquisition, supervision, and reviewing and edited the paper; M.S. and S.K. contributed to supervision, and reviewing and edited the paper. All authors have read and agreed to the published version of the manuscript.

**Funding:** The Hokkaido COPD cohort study is supported by a scientific research grant to the Hokkaido COPD Cohort Study from the Ministry of Education, Science, Culture and Sports of Japan (17390239 and 2139053 to M.N.), Nippon Boehringer Ingelheim, Pfizer, Inc., and a grant to the Respiratory Failure Research Group from the Ministry of Health, Labor, and Welfare, Japan.

**Institutional Review Board Statement:** The study was conducted according to the guidelines of the Declaration of Helsinki, and approved by the Ethics Committees of Hokkaido University School of Medicine (med 02-001: December 2002) and Hokkaido University Hospital (No.016-0495: May 2017).

**Informed Consent Statement:** Informed consent was obtained from all subjects involved in the study.

**Data Availability Statement:** The created models in this study are available on request from the corresponding author. However, the image datasets presented in this study are not publicly available

due to ethical reasons, e.g., containing information that could compromise the privacy of research participants.

**Acknowledgments:** We would like to express our appreciation for the support of Masaharu Nishimura, Department of Respiratory Medicine, Faculty of Medicine, Hokkaido University, for his advice and encouragement as we developed our vision for his help as we developed our proposal for the publication.

**Conflicts of Interest:** M.S. reports grants from GlaxoSmithKline, Novartis, and AstraZeneca outside the current work. None of these companies had a role in the design or analysis of the study or in the writing of the manuscript.

## References

- Adeloye, D.; Chua, S.; Lee, C.; Basquill, C.; Papan, A.; Theodoratou, E.; Nair, H.; Gasevic, D.; Sridhar, D.; Campbell, H.; et al. Global and regional estimates of COPD prevalence: Systematic review and meta-analysis. *J. Glob. Health* **2015**, *5*, 20415. [[CrossRef](#)]
- Quaderi, S.A.; Hurst, J.R. The unmet global burden of COPD. *Glob. Health Epidemiol. Genom.* **2018**, *3*, e4. [[CrossRef](#)] [[PubMed](#)]
- Halpin, D.M.G.; Decramer, M.; Celli, B.R.; Mueller, A.; Metzendorf, N.; Tashkin, D.P. Effect of a single exacerbation on decline in lung function in COPD. *Respir. Med.* **2017**, *128*, 85–91. [[CrossRef](#)]
- Ho, T.-W.; Tsai, Y.-J.; Ruan, S.-Y.; Huang, C.-T.; Lai, F.; Yu, C.-J. In-hospital and one-year mortality and their predictors in patients hospitalized for first-ever chronic obstructive pulmonary disease exacerbations: A nationwide population-based study. *PLoS ONE* **2014**, *9*, e114866. [[CrossRef](#)]
- Suissa, S.; Dell’Aniello, S.; Ernst, P. Long-term natural history of chronic obstructive pulmonary disease: Severe exacerbations and mortality. *Thorax* **2012**, *67*, 957–963. [[CrossRef](#)] [[PubMed](#)]
- Tachkov, K.; Kamusheva, M.; Pencheva, V.; Mitov, K. Social burden of copd. *Value Health* **2017**, *20*, A648–A649. [[CrossRef](#)]
- Pauwels, R.; Buist, S.; Calverley, P.; Jenkins, C.; Hurd, S. Global strategy for the diagnosis, management and prevention of chronic obstructive pulmonary disease. NHBLI/WHO global initiative for Chronic Obstructive Lung Disease (GOLD) workshop summary. *Rev. Port. Pneumol.* **2001**, *7*, 398–400. [[CrossRef](#)]
- Occhipinti, M.; Paoletti, M.; Bartholmai, B.J.; Rajagopalan, S.; Karwoski, R.A.; Nardi, C.; Inchingolo, R.; Larici, A.R.; Camiciottoli, G.; Lavorini, F.; et al. Spirometric assessment of emphysema presence and severity as measured by quantitative CT and CT-based radiomics in COPD. *Respir. Res.* **2019**, *20*, 101. [[CrossRef](#)]
- Hackx, M.; Ghaye, B.; Coche, E.; Van Muylem, A.; Gevenois, P.A. Severe COPD exacerbation: CT features. *COPD J. Chronic Obstr. Pulm. Dis.* **2015**, *12*, 38–45. [[CrossRef](#)]
- Lynch, D.A.; Austin, J.H.M.; Hogg, J.C.; Grenier, P.A.; Kauczor, H.U.; Bankier, A.A.; Barr, R.G.; Colby, T.V.; Galvin, J.R.; Gevenois, P.A.; et al. CT-definable subtypes of chronic obstructive pulmonary disease: A statement of the fleischner society<sup>1</sup>. *Radiology* **2015**, *277*, 192–205. [[CrossRef](#)] [[PubMed](#)]
- Jennergmann, T.; Humphries, S.M.; Austin, J.H.M.; Grenier, P.A.; Russell, P.; Beaty, T.H.; Curran-everett, D.; Hokanson, J.E.; Investigators, C. CT-based Visual Classification of Emphysema: Association with mortality in the COPD Gene Study. *Radiology* **2018**, *288*, 859–866.
- Feldhaus, F.W.; Theilig, D.C.; Hubner, R.H.; Kuhnigk, J.M.; Neumann, K.; Doellinger Feldhaus, F. Quantitative CT analysis in patients with pulmonary emphysema: Is lung function influenced by concomitant unspecific pulmonary fibrosis? *Int. J. Chronic Obstr. Pulm. Dis.* **2019**, *14*, 1583–1593. [[CrossRef](#)] [[PubMed](#)]
- Engel, M.; Endesfelder, D.; Schloter-Hai, B.; Kublik, S.; Granitsiotis, M.S.; Boschetto, P.; Stendardo, M.; Barta, I.; Dome, B.; Deleuze, J.F.; et al. Influence of lung CT changes in chronic obstructive pulmonary disease (COPD) on the human lung microbiome. *PLoS ONE* **2017**, *12*, e0180859.
- van Rikxoort, E.M.; de Jong, P.A.; Mets, O.M.; van Ginneken, B. Automatic classification of pulmonary function in COPD patients using trachea analysis in chest CT scans. *Med Imaging 2012 Comput. Diagn.* **2012**, *8315*, 83150P. [[CrossRef](#)]
- Mishima, M.; Hirai, T.; Itoh, H.; Nakano, Y.; Sakai, H.; Muro, S.; Nishimura, K.; Oku, Y.; Chin, K.; Ohi, M.; et al. Complexity of terminal airspace geometry assessed by lung computed tomography in normal subjects and patients with chronic obstructive pulmonary disease. *Proc. Natl. Acad. Sci. USA* **1999**, *96*, 8829–8834. [[CrossRef](#)]
- Shimizu, K.; Tanabe, N.; Van Tho, N.; Suzuki, M.; Makita, H.; Sato, S.; Muro, S.; Mishima, M.; Hirai, T.; Ogawa, E.; et al. Per cent low attenuation volume and fractal dimension of low attenuation clusters on CT predict different long-term outcomes in COPD. *Thorax* **2020**, *75*, 116–122. [[CrossRef](#)]
- Goddard, P.R.; Nicholson, E.M.; Laszlo, G.; Watt, I. Computed tomography in pulmonary emphysema. *Clin. Radiol.* **1982**, *33*, 379–387. [[CrossRef](#)]
- Tanabe, N.; Shimizu, K.; Terada, K.; Sato, S.; Suzuki, M.; Shima, H.; Oguma, A.; Oguma, T.; Konno, S.; Nishimura, M.; et al. Central airway and peripheral lung structures in airway disease dominant COPD. *ERJ Open Res.* **2021**, *7*, 00672–02020. [[CrossRef](#)] [[PubMed](#)]
- Cheplygina, V.; Pena, I.P.; Pedersen, J.H.; Lynch, D.A.; Sorensen, L.; De Bruijne, M. Transfer learning for multicenter classification of chronic obstructive pulmonary disease. *IEEE J. Biomed. Health Inform.* **2018**, *22*, 1486–1496. [[CrossRef](#)]



20. Ben-Eliezer, N.; Sodickson, D.K.; Block, K.T. Rapid and accurate T2 mapping from multi-spin-echo data using bloch-simulation-based reconstruction. *Magn. Reson. Med.* **2015**, *73*, 809–817. [[CrossRef](#)]
21. Barrett, T.; Mclean, M.; Priest, A.N.; Lawrence, E.M.; Patterson, A.J.; Koo, B.C.; Patterson, I.; Warren, A.Y.; Doble, A.; Gnanaprasam, V.J.; et al. Diagnostic evaluation of magnetization transfer and diffusion kurtosis imaging for prostate cancer detection in a re-biopsy population. *Eur. Radiol.* **2018**, *28*, 3141–3150. [[CrossRef](#)] [[PubMed](#)]
22. Bai, W.; Sinclair, M.; Tarroni, G.; Oktay, O.; Rajchl, M.; Vaillant, G.; Lee, A.M.; Aung, N.; Lukaschuk, E.; Sanghvi, M.M.; et al. Automated cardiovascular magnetic resonance image analysis with fully convolutional networks 08 Information and Computing Sciences 0801 Artificial Intelligence and Image Processing. *J. Cardiovasc. Magn. Reson.* **2018**, *20*, 65. [[CrossRef](#)] [[PubMed](#)]
23. Manabe, K.; Asami, Y.; Yamada, T.; Sugimori, H. Improvement in the convolutional neural network for computed tomography images. *Appl. Sci.* **2021**, *11*, 1505. [[CrossRef](#)]
24. Kawakami, M.; Hirata, K.; Furuya, S.; Kobayashi, K.; Sugimori, H.; Magota, K.; Katoh, C. Development of Combination methods for detecting malignant uptakes based on physiological uptake detection using object detection with PET-CT MIP images. *Front. Med.* **2020**, *7*, 616746. [[CrossRef](#)]
25. Sugimori, H.; Kawakami, M. Automatic detection of a standard line for brain magnetic resonance imaging using deep learning. *Appl. Sci.* **2019**, *9*, 3849. [[CrossRef](#)]
26. Sugimori, H.; Sugiyama, T.; Nakayama, N.; Yamashita, A.; Ogasawara, K. Development of a deep learning-based algorithm to detect the distal end of a surgical instrument. *Appl. Sci.* **2020**, *10*, 4245. [[CrossRef](#)]
27. Duong, M.T.; Rudie, J.D.; Wang, J.; Xie, L.; Mohan, S.; Gee, J.C.; Rauschecker, A.M. Convolutional neural network for automated FLAIR lesion segmentation on clinical brain MR imaging. *Am. J. Neuroradiol.* **2019**, *40*, 1282–1290. [[CrossRef](#)] [[PubMed](#)]
28. Khalili, N.; Lessmann, N.; Turk, E.; Claessens, N.; de Heus, R.; Kolk, T.; Viergever, M.A.; Benders, M.J.N.L.; Išgum, I. Automatic brain tissue segmentation in fetal MRI using convolutional neural networks. *Magn. Reson. Imaging* **2019**, *64*, 77–89. [[CrossRef](#)]
29. Liu, J.; Chen, F.; Pan, C.; Zhu, M.; Zhang, X.; Zhang, L.; Liao, H. A cascaded deep convolutional neural network for joint segmentation and genotype prediction of brainstem gliomas. *IEEE Trans. Biomed. Eng.* **2018**, *65*, 1943–1952. [[CrossRef](#)]
30. Dalmiş, M.U.; Litjens, G.; Holland, K.; Setio, A.; Mann, R.; Karssemeijer, N.; Gubern-Mérida, A. Using deep learning to segment breast and fibroglandular tissue in MRI volumes. *Med. Phys.* **2017**, *44*, 533–546. [[CrossRef](#)]
31. Humphries, S.M.; Notary, A.M.; Centeno, J.P.; Strand, M.J.; Crapo, J.D.; Silverman, E.K.; Lynch, D.A. Deep learning enables automatic classification of emphysema pattern at CT. *Radiology* **2020**, *294*, 434–444. [[CrossRef](#)] [[PubMed](#)]
32. Guideline of respiratory function tests-spirometry, flow-volume curve, diffusion capacity of the lung. *J. Jpn. Respir. Soc.* **2004**, 1–56. Available online: <https://pubmed.ncbi.nlm.nih.gov/15565748/> (accessed on 1 January 2021).
33. Nishimura, M.; Makita, H.; Nagai, K.; Konno, S.; Nasuhara, Y.; Hasegawa, M.; Shimizu, K.; Betsuyaku, T.; Ito, Y.M.; Fuke, S.; et al. Annual change in pulmonary function and clinical phenotype in chronic obstructive pulmonary disease. *Am. J. Respir. Crit. Care Med.* **2012**, *185*, 44–52. [[CrossRef](#)] [[PubMed](#)]
34. Suzuki, M.; Makita, H.; Konno, S.; Shimizu, K.; Kimura, H.; Kimura, H.; Nishimura, M. Asthma-like features and clinical course of chronic obstructive pulmonary disease: An analysis from the hokkaido COPD cohort study. *Am. J. Respir. Crit. Care Med.* **2016**, *194*, 1358–1365. [[CrossRef](#)]
35. Gevenois, P.A.; de Maertelaer, V.; De Vuyst, P.; Zanen, J.; Yernault, J.C. Comparison of computed density and macroscopic morphometry in pulmonary emphysema. *Am. J. Respir. Crit. Care Med.* **1995**, *152*, 653–657. [[CrossRef](#)] [[PubMed](#)]
36. Tanabe, N.; Muro, S.; Hirai, T.; Oguma, T.; Terada, K.; Marumo, S.; Kinose, D.; Ogawa, E.; Hoshino, Y.; Mishima, M. Impact of exacerbations on emphysema progression in chronic obstructive pulmonary disease. *Am. J. Respir. Crit. Care Med.* **2011**, *183*, 1653–1659. [[CrossRef](#)] [[PubMed](#)]
37. Yan, W.; Wang, Y.; Gu, S.; Huang, L.; Yan, F.; Xia, L.; Tao, Q. The domain shift problem of medical image segmentation and vendor-adaptation by Unet-GAN. *arXiv* **2019**, arXiv:1910.13681.