Computers in Biology and Medicine 125 (2020) 103962

Contents lists available at ScienceDirect

Computers in Biology and Medicine

journal homepage: www.elsevier.com/locate/compbiomed

A lung graph model for the radiological assessment of chronic thromboembolic pulmonary hypertension in CT

Oscar Jimenez-del-Toro [a,*,1], Yashin Dicente Cid [b,1], Alexandra Platon [c], Anne-Lise Hachulla [c,d], Frédéric Lador [d,e], Pierre-Alexandre Poletti [c], Henning Müller [a,c]

[a] MedGIFT group, Institute of Information Systems, University of Applied Sciences Western Switzerland (HES-SO), TechnoArk 3, Sierre 3960, Switzerland
[b] Data Science, WMG, University of Warwick, CV4 7AL Coventry, United Kingdom
[c] Division of Radiology, Geneva University Hospitals (HUG), Rue Gabrielle-Perret-Gentil 4, 1205 Geneva, Switzerland
[d] Pulmonary Hypertension Program, Geneva University Hospitals (HUG), Rue Gabrielle-Perret-Gentil 4, 1205 Geneva, Switzerland
[e] Division of Pneumology, Geneva University Hospitals (HUG), Rue Gabrielle-Perret-Gentil 4, 1205 Geneva, Switzerland

ARTICLE INFO

Keywords:
Pulmonary hypertension
Texture analysis
Pulmonary embolism
Lung graph model
Radiomics

ABSTRACT

Chronic thromboembolic pulmonary hypertension (CTEPH) is a possible complication of pulmonary embolism (PE), with poor prognosis if left untreated. Surgical curative treatment is available, particularly in the early stages of the disease. However, most cases are not diagnosed until specific symptoms become evident. A small number of computed tomography (CT) findings, such as a widened pulmonary artery and mosaicism in the lung parenchyma, have been correlated with pulmonary hypertension (PH). Quantitative texture analysis in the CT scans of these patients could provide complementary sub-visual information of the vascular changes taking place in the lungs. For this task, a lung graph model was developed with texture descriptors from 37 CT scans with confirmed CTEPH diagnosis and 48 CT scans from PE patients who did not develop PH. The probability of presenting CTEPH, computed with the graph model, outperformed a convolutional neural network approach using 10 different train/test splits of the data set. An accuracy of 0.76 was obtained with the proposed texture analysis, and was then compared to the visual assessment of CT findings, manually identified by a team of three expert radiologists, commonly associated with pulmonary hypertension. This graph-based score combined with the information attained from the radiological findings resulted in a Cohen's kappa coefficient of 0.47 when differentiating patients with confirmed CTEPH from those with PE who did not develop the disease. The proposed texture quantification could be an objective measurement, complementary to the current analysis of radiologists for the early detection of CTEPH and thus improve patient outcome.

1. Introduction

Chronic thromboembolic pulmonary hypertension, or CTEPH, is a subclass of pulmonary hypertension (PH), with a distinctive pathophysiology and treatment [1]. The disease is characterised by lung perfusion defects, believed to arise from one or multiple unresolved endothelialised pulmonary thrombi [2]. The reported incidence in the published literature ranges between 0.1% to 11.8% after a pulmonary embolism (PE), with a recent meta-analysis computing an incidence of 3.2% for all survivors of PE [3]. If left untreated, CTEPH may lead to right ventricular (RV) overload and ultimately to RV failure and death [4].

Pulmonary endarterectomy is a potential curative surgery for CTEPH once the diagnosis is confirmed [5]. Unfortunately, CTEPH is still misdiagnosed or undiagnosed as symptoms are non-specific or absent, particularly in the early stages of the disease [2,6]. Currently, a right heart catheterisation procedure is the study required to establish a definitive CTEPH diagnosis for patients with and without recorded history of PE [7]. Catheterisation is not only an invasive procedure but can occasionally also lead to moderate complications and rare fatal events [8]. On the other hand, when CTEPH is detected in its early stages, patient outcome has improved with a timely initiation of the treatment and better surgical possibilities [9].

Due to the lack of conclusive evidence, screening unselected survivors of PE is, at the present time, not recommended [10]. If a specific high-risk group of PE patients could be identified, valuable resources could be specifically targeted to this group, resulting in a better prognosis [11]. A recent strategy proposed for an early identification of CTEPH is to assess computed tomography (CT) findings of chronic PE or PH,

* Correspondence to: University of Applied Sciences Western Switzerland (HES-SO), TechnoArk 3, Sierre 3960, Switzerland.
E-mail address: oscar.jimenez@hevs.ch (O. Jimenez-del-Toro).
[1] O. Jimenez-del-Toro and Y. Dicente Cid contributed equally to this work.

https://doi.org/10.1016/j.compbiomed.2020.103962
Received 6 May 2020; Received in revised form 3 August 2020; Accepted 7 August 2020
Available online 14 August 2020

both at the time of an acute PE event and also in later symptomatic episodes of the disease [12]. In this respect, the measurement of an enlarged pulmonary artery (PA) diameter relative to the adjacent aorta, or PA:A ratio, is commonly used as a non-specific indicator of PH due to its simplicity and accessibility [13]. This parameter has shown a moderate sensitivity (0.79) and specificity (0.83) according to a recent meta-analysis in the detection of pulmonary hypertension [14]. Moreover, a widened PA has been associated with reduced survival in CTEPH, suggesting a potential role also as a monitoring and prognostic parameter [15]. A mosaic attenuation pattern in the lung parenchyma, or mosaicism, is another common CT finding in CTEPH patients, that has been reported with high sensitivity (*i.e.* 0.77 to 1.00) in some studies [16]. Mosaicism has additionally been regarded as a useful method for evaluating blood distribution in CTEPH patients, with a prognostic association to the pulmonary endarterectomy procedure as well [17]. Nevertheless, it has been suggested that these findings are routinely missed in the radiological assessment of acute PE patients [12]. On the other hand, an apparently normal CT cannot rule out CTEPH either, as there is insufficient evidence to support the use of CT for screening patients after acute PE [10].

Texture analysis can be regarded as an objective low-level quantitative biomarker, that encodes variations in the spatial relationships within image regions [18]. It has been shown to be a reproducible and promising approach for assessing lung ventilation and vascularisation, without relying on subjective interpretations of the images [19,20]. Previously, Foncubierta-Rodriguez et al. evaluated the benefits of multi-dimensional texture analysis using wavelet-based features for automatically detecting pulmonary embolisms [19]. More recently, Dicente Cid et al. proposed a graph model that quantified the tissue texture in the lung parenchyma, merging local and global radiomics of the lungs, to classify patients with vascular pathologies [20]. To the best of our knowledge, there is no study in literature that has evaluated a computerised lung texture approach in the assessment of CT scans from CTEPH patients. Moreover, the added value of quantifying the changes in the lung tissue from CTEPH patients together with their radiological CT findings has not been studied.

In this study, a lung graph model, built with state-of-the-art 3D texture descriptors, was developed to discriminate between CT scans with confirmed diagnosis of CTEPH and patients that did not develop PH after PE. The performance of this method was compared to the performance of a deep learning approach, *i.e.* a fine-tuned convolutional neural network, in a data set of 75 patients, with a total of 85 CT scans available. On the other hand, we measured the prevalence of the two most commonly used CT findings for detecting CTEPH, *i.e.* PA:A ratio and mosaicism, in the full data set. Finally, an optimal combination of both the CT findings and the computed graph-based score, resulting from the texture analysis, was calculated in the CT scans of patients at the time of diagnosis. The feasibility of performing an early detection of CTEPH, relying only on the initial CT scans taken in a real-clinical scenario is shown.

2. Methods

2.1. Data set

2.1.1. Study patients

The patient reports and imaging studies used in this work were acquired with the approval of the ethics committee from the University Hospitals of Geneva. Candidates for the PE group were retrospectively considered if they were referred for a computed tomography examination with suspicion of PE in the emergency radiology department from the Geneva University Hospitals between January 2012 and December 2016. Only the patients that showed an objectively proven PE were eligible to participate in the study and their electronic medical records were revised until May 2019. A selection of patients without any indication of PH after a follow up of at least 36 months was furthered evaluated for inclusion. In the end, 48 patients who had presented a symptomatic acute PE event, but who did not develop PH afterwards were included in this group. The 48 CT scans as well as the hospital records from these patients were screened by the authors to discard any major cardiopulmonary or malignant comorbidity.

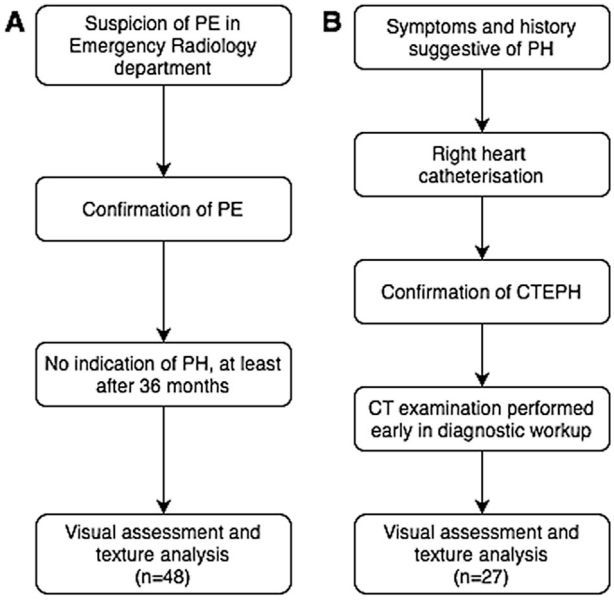

Fig. 1. Flowchart of the selection process for (A) PE cases and (B) CTEPH cases. PE = pulmonary embolism; DECT = dual energy computed tomography; PH = pulmonary hypertension; CTEPH = chronic thromboembolic pulmonary hypertension.

A second group of patients, that were diagnosed with pulmonary hypertension during the same period, were considered for the CTEPH group. Sixty-three patients with a confirmed diagnosis of PH and with a CT examination taken before the use of any targeted therapy were found. The diagnosis of PH was established for each patient at the time of their right heart catheterisation according to the available guidelines [2,21]. For this work, a group of three radiologists reanalysed the hospital record from each patient and jointly defined the corresponding PH group. Two examinations from this population could not be interpreted due to major metallic artefacts generated by spine fixation material. A total of 27 patients were then classified as CTEPH and differentiated from PH of other origins. Although most of these CTEPH patients were diagnosed relatively late, they are incident cases, meaning that the CT was performed early in the diagnostic workup. All CT scans taken at the time of the CTEPH diagnosis, *i.e.* one per patient, as well as 10 additional CT scans available after the diagnosis of some of these patients were included in the data set. This resulted in a final data set of 85 CT scans, with 48 CT scans from patients with PE and 37 CT scans from 27 patients with CTEPH (see Fig. 1).

2.1.2. CT data acquisition

The CT data were created with two main protocols: the first one from a GE Discovery 750 HD scanner, obtained following the protocol used in the emergency radiology unit for the screening of PE: automatic injection triggering of a bolus of 80 ml of contrast media (concentration 350 ml I/ml) 1.25 mm collimation, 1 mm interval, pitch 1.375, tube rotation 0.5 s, energy range between 80 to 140 kVp oscillating every 0.5 msec. The second one from the retrospective analysis of patients that were scanned on a Siemens Flash Definition scanner using the following parameters: automatic injection triggering of a bolus of 80 ml of contrast media (concentration 350 ml I/ml) 0.6 mm collimation, 0.7 mm interval, pitch 0.55, tube rotation 0.5 s, dual source, tube rotation 0.5 s, 100 and 140 kVp.

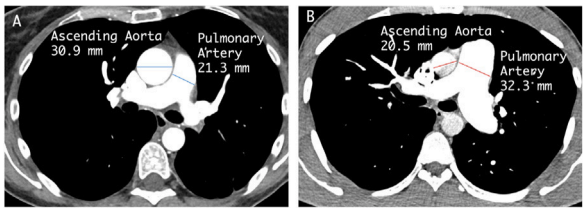

Fig. 2. Sample measurements of the pulmonary artery and ascending aorta diameters in axial slices from CT scans. The measurements are then used to compute the PA:A ratio. Patient (A) shows a normal PA:A ratio, unlike patient (B) who has a PA:A ratio larger than 1, suggestive of pulmonary hypertension.

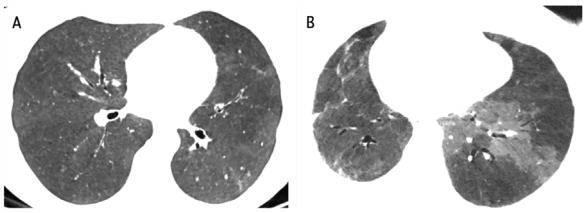

Fig. 3. Samples of mosaicism assessment in axial CT slices with minimum intensity projection reconstruction. No mosaic attenuation pattern is seen in patient (A). On the other hand, mosaicism is present in the lung parenchyma of patient (B).

The resolution of the CT slices in x- and *y*-directions varied from 0.5742 to 0.9766 mm, while the inter-slice distance was fixed at 1.00 mm. All images were anonymised and stored in the DICOM file format for further processing. CT assessment was carried out on the chest CT images obtained at 80 KeV, using mediastinal and lung window settings (window level -500 HU, window width 1500 HU).

2.2. CT assessment

The CT scans were retrospectively analysed and annotated in random order by a trained medical doctor and discussed in consensus with three expert radiologists. Morphological signs in CT classically associated with PH, *i.e.* main pulmonary artery to the aorta diameter ratio (PA:A) and lung parenchyma findings, *i.e.* mosaicism, were identified and measured (see Table 1) [22]. The main pulmonary artery diameter (MPAD) was defined as the largest diameter of the pulmonary trunk on a transverse image at the level of its bifurcation (see Fig. 2). The widest short-axis diameter was also obtained from the ascending aorta (AD). The PA:A ratio measurement was then calculated for each patient. A ratio greater than 1:1 has shown in previous studies a strong correlation with increased pulmonary pressures [2]. The right (RV) and left (LV) ventricular diameters were measured in the standard axial view perpendicular to the long axis of the heart. The right ventricle to left ventricle ratio (RV:LV) was calculated, searching for RV dysfunction. The right ventricle wall thickness (RVWT) was measured as well. The cardiothoracic ratio was calculated based on the maximum heart length divided by the thoracic cavity width. The left atrium area (LAA) traces the inner border, excluding the area under the mitral valve and the pulmonary veins. The presence of paradoxical inter-ventricular septal position, visualising the curvature of the septum was also reported. CT findings from the general aspect of the lung parenchyma were also assessed: PE detection, fibrosis, emphysema, pleural effusion, pulmonary edema and systemic arterial hypervascularisation. Using a minimum intensity projection, reconstructions were computed to detect mosaic perfusion patterns (see Fig. 3). A mosaic pattern was characterised by the presence of ground-glass attenuation with enlarged vascular segments, intermingled with areas of low attenuation and smaller vascular segments.

Table 1
Patient demographics, CT measurements and lung parenchyma assessment of the two target groups in this study: patients with pulmonary embolism (PE) who did not develop pulmonary hypertension, and patients with a CT taken at the time when the CTEPH diagnosis was made.

	PE n=48	CTEPH n=27
Gender		
Male	28(58.3%)	14(51.8%)
Female	20(41.6%)	13(48.1%)
Age	65 ± 16	62.3 ± 15
CT measurements		
MPAD, cm	28.39 ± 5.60	32.68 ± 5.71
AD, cm	33.00 ± 4.64	31.48 ± 6.22
PA:A, cm	0.87 ± 0.16	1.08 ± 0.31
RV:LV	1.27 ± 0.37	1.27 ± 0.43
RVWT, cm	3.77 ± 1.22	4.27 ± 1.29
CTR	0.50 ± 0.06	0.53 ± 0.07
LAA, cm^2	15.87 ± 4.95	17.90 ± 5.23
Lung parenchyma CT findings		
PIVSP	10(20.8%)	11(40.7%)
PE	48(100%)	27(100%)
Fibrosis	19(39.6%)	5(18.5%)
Emphysema	8(16.7%)	6(22.2%)
Pleural effusion	4(8.3%)	2(7.4%)
Pulmonary edema	4(8.3%)	2(7.4%)
SAV	3(6.3%)	7(25.9%)
Mosaicism	21(43.8%)	18(66.7%)

Abbreviations: *MPAD* main pulmonary artery diameter, *AD* aorta diameter, *PA:A* pulmonary artery to aorta ratio, *RV:LV* right ventricle to left ventricle ratio, *RVWT* right ventricle wall thickness, *CTR* cardiothoracic ratio, *LAA* left atrium area, *PIVSP* paradoxical inter-ventricular septal position, *PE* pulmonary embolism, *SAV* systemic arterial hypervascularisation

2.3. Texture-based graph model of the lungs

A mathematical model, hereafter termed a *lung graph model*, was developed to quantify the texture of the lung parenchyma of these patients. It characterises the entire lung parenchyma using information from local texture regions in the lung and their global correlations [20]. The method was built with state-of-the-art rotation-invariant 3D texture descriptors [23]. Due to the identical spatial resolution needed for the texture analysis of these data, the voxels of the CT scans were rescaled into isometric voxels, with a voxel size of 1 mm. After this initial step, we followed the pipeline introduced by Dicente Cid et al. in [20] to build a graph model of the lungs (see Fig. 4). As a first step, the lung fields were automatically extracted using an automatic lung segmentation algorithm detailed in [24]. Based on the 3D model of the human lung presented by Zrimec et al. the lung masks were geometrically divided into a 36-region atlas [25,26]. For each region r of this atlas, two texture descriptors were used to describe its regional texture: the Fourier histograms of oriented gradients (FHOG) [27] and the locally-oriented 3D Riesz-wavelet transform (3DRiesz) [28]. These descriptors are rotation-invariant and were successfully applied for multiple texture analysis applications [18,23,29]. We used the following previously validated parameters: for FHOG, we used 28 3D directions for the histogram, obtaining a 28-dimensional feature vector per image voxel v ($\mathbf{f}_H(v) \in \mathbb{R}^{28}$), and for 3DRiesz, we used the 3rd-order Riesz-wavelet transform, with 4 scales and 1st-order alignment (see [28]). The feature vector for a single voxel was defined as the absolute Riesz response along the 4 scales, obtaining a 40-dimensional feature vector ($\mathbf{f}_R(v) \in \mathbb{R}^{40}$). Finally, the average and standard deviation of these descriptors were used to describe each region r: $\boldsymbol{\mu}_H(r)$, $\boldsymbol{\sigma}_H(r)$, $\boldsymbol{\mu}_R(r)$, and $\boldsymbol{\sigma}_R(r)$ (see Fig. 4).

2.3.1. Graph model definition

In [20], the authors compared two graph definitions, *i.e.* directed and undirected, and tested them in four binary classification experiments including PH and PE patients. Their results show that the

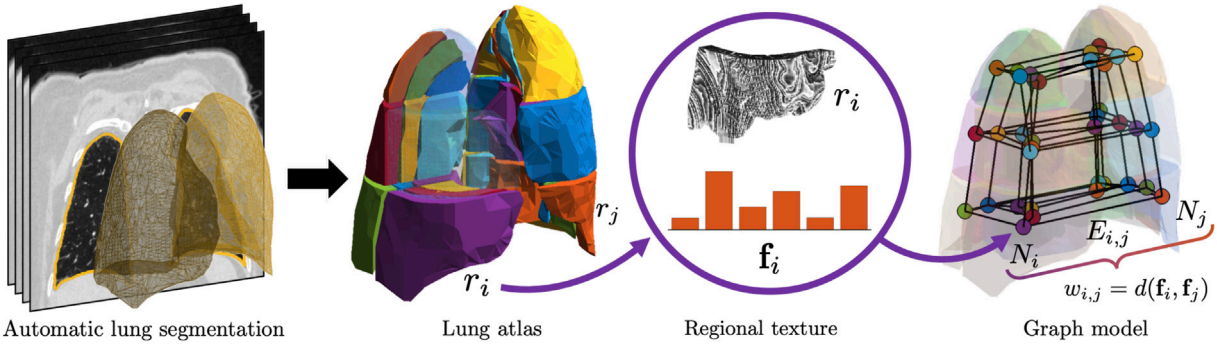

Fig. 4. Construction of the lung graph model: First, the lungs are automatically segmented from the CT image. Then, the lung fields are divided using a geometric atlas with 36 regions. Each atlas region r_i was described with a 3D texture descriptor $\mathbf{f}_i$. Finally, the graph is built based on the 3D adjacency of the regions (including left–right lung connections). The edges $E_{i,j}$ contain the similarities (or distances) between the regional descriptors ($w_{i,j} = d(\mathbf{f}_i, \mathbf{f}_j)$).

directed graph was overall better than the undirected graph. However, in the particular case of PH vs. PE, the undirected graph showed better performance than the directed one (an area under the receiver operating characteristic curve (AUC) of 0.78 vs. 0.71). Since CTEPH is a subclass of PH, we decided to use the undirected graph model in this work. The formal definition of the undirected graph model is: given a division (atlas) of the lungs with n regions $\{r_1, \ldots, r_n\}$, let $\mathcal{G}$ be a graph with a set of n nodes $\mathcal{N}$. Then, an undirected edge $E_{i,j}$ between nodes N_i and N_j exists if regions r_i and r_j are 3D adjacent in the atlas or symmetric with respect to the left–right division of the lungs (see Fig. 4). Considering $\mathbf{f}_i$ and $\mathbf{f}_j$ the texture-based feature vectors of regions r_i and r_j respectively, the weight $w_{i,j}$ of an edge $E_{i,j}$ is defined using the correlation distance, *i.e.* $w_{i,j} = 1 - corr(\mathbf{f}_i, \mathbf{f}_j) \in [0, 2]$. The use of rotation-invariant texture descriptors where each component corresponds to a texture-direction implies that regions with similar texture will be highly correlated, so with $w_{i,j} \approx 0$. The use of undirected edges implies that this graph has 84 weights since $w_{i,j} = w_{j,i}$.

2.3.2. Patient-specific graph-based lung descriptor

Four texture feature vectors were computed in each lung atlas region r (*i.e.* $\mu_H(r)$, $\sigma_H(r)$, $\mu_{\mathcal{R}}(r)$, and $\sigma_{\mathcal{R}}(r)$) providing complementary information about the local texture and its variability. Given a patient p, a graph $\mathcal{G}_{p,\mathbf{f}}$ was obtained from each of these feature vectors $\mathbf{f}$. Each of these graphs $\mathcal{G}_{p,\mathbf{f}}$ was described using the vector $\boldsymbol{w}_{p,\mathbf{f}}$ that contains the 84 weights in the graph $\mathcal{G}_{p,\mathbf{f}}$ ordered by their location in the graph. The final lung descriptor vector $\boldsymbol{W}_p$ used in our experiments is defined as the concatenation of the four vectors $\boldsymbol{w}_{p,\mathbf{f}}$, *i.e.*: $\boldsymbol{W}_p = (\boldsymbol{w}_{p,\mu_H} \| \boldsymbol{w}_{p,\sigma_H} \| \boldsymbol{w}_{p,\mu_{\mathcal{R}}} \| \boldsymbol{w}_{p,\sigma_{\mathcal{R}}})$. After this concatenation, $\boldsymbol{W}_p \in \mathbb{R}^{336}$. A graph-based score is henceforth defined as the output probability of a CT scan to be classified as CTEPH with the proposed lung graph model.

3. Experimental setup

Two major experiments were designed to evaluate and compare the performance of the lung graph model in the differentiation task PE vs. CTEPH, basing the classification only on CT examinations taken at the time of the diagnostic workup. In the first experiment, the two patient groups were classified both by the proposed approach and by two adaptations of a state-of-the-art deep learning approach for lung tissue characterisation. A second experiment was performed to estimate the contribution of the graph-based scores obtained with the proposed model in the identification of CT examinations with CTEPH. The performance of the model was compared to the presence of two radiological CT findings that have been previously associated with CTEPH: PA:A ratio > 1 and mosaicism.

3.1. Experiment 1: Lung graph model vs. convolutional neural network classification

In this experiment all the 85 CT scans from 75 patients (27 CTEPH and 48 PE) were considered for a binary classification task. Ten train/test splits (80%–20%) were generated, aiming at a balanced distribution of both classes and paying attention that CT scans belonging to the same patient were all grouped in either the train or the test set to avoid introducing a bias in the evaluation.

For the lung graph model, the patient-specific 336-dimensional descriptors $\boldsymbol{W}$ (see Section 2.3.2) were normalised based on the training set from each split. However, since each component of the vectors $\boldsymbol{w}_{\mathbf{f}}$ corresponds to an edge of the graph $\mathcal{G}_{\mathbf{f}}$, where $\mathbf{f}$ refers to the regional texture descriptor used to build the graph $\mathcal{G}_{\mathbf{f}}$ (see Section 2.3), these components cannot be seen independently. The normalisation was then performed simultaneously for all the components of a graph $\mathcal{G}_{\mathbf{f}}$, obtaining normalised descriptors $\hat{\boldsymbol{W}}$. To select the optimal decision threshold for the models (from 0.0 to 1.0 with a step of 0.1), a 10-fold cross-validation (CV) was performed on the training sets from the 10 grouping splits. To avoid overfitting, dimensionality reduction was applied to the normalised vectors by selecting the dimensions that correlated above average with the training labels. Both techniques (feature normalisation and dimensionality reduction) were computed for each the data split with the values obtained only from the training sets. The final feature space varied for each data split and had an average size of 149 dimensions. The final feature classification was done using a random forest (RF) classifier with 50 trees in MATLAB®.

A deep learning approach, that was previously tested on a lung classification task, was adapted for this study to have a strong baseline comparison [30]. The baseline approach was originally developed as a lung lesion-based descriptor and it was built using features extracted with a deep learning network. The approach achieved the top scores in an open challenge called 'ImageCLEF tuberculosis 2018', with a total of 39 approaches submitted [31]. The goal of this challenge was to the detect and classify tuberculosis lesions in a shared data set of 495 CT scans. The method proposed by Liauchuk et al. analysed individual 2D slice regions with the convolutional neural network (CNN) Alexnet [32]. In this work we used a deeper CNN, namely VGG19, which improves over AlexNet with a very uniform architecture and an overall better performance when extracting image features [33]. Two adaptations were applied for fine-tuning the VGG19 models, one with and one without data augmentation. Both were originally trained with ImageNet [34]. Artificial data augmentation is a common procedure performed in CNN model training that increases the amount of training samples to avoid overfitting, particularly when the number of training samples is relatively small. To obtain a fixed patch size of 224 × 224 × 3, the input sample patches were distorted by scaling, with a zoom range of 0.2, and randomly cropping regions within the rescaled samples each time they were fed into the network. Further data

augmentation was generated with random horizontal flipping of the cropped samples. Stochastic gradient descent was selected for weight optimisation with a learning rate of 0.00001, momentum of 0.9, a dropout regularisation ratio set to 0.2 and a fixed batch size of 64.

The evaluation metrics to describe the approaches in this experiment are presented as the average and standard error [35] obtained from the classification of all the test sets in the 10 data splits. Sensitivity, specificity, accuracy, F_1 score, AUC, Cohen's kappa coefficient, and the Matthews correlation coefficient (MCC) are reported. The statistical significance between the scores from the lung graph model and the best performing deep learning strategy was analysed using a student t-test, where a p-value <0.05 was considered significant.

3.2. Experiment 2: visual CT assessment vs. lung graph model

The commonly used CT findings to detect PH, *i.e.* PA:A ratio and mosaicism, were correlated to patient outcome. All were manually annotated for the available CT scans. The prevalence of these two CT findings is reported for both the CTEPH CT scans and the CT scans from acute PE cases. The classification performance of these indicators in the entire data set was both compared and combined with the score computed by our lung graph model. Unlike Experiment 1, for this experiment, the probability for presenting CTEPH in a CT scan was computed in a leave-one-patient-out (LOPO) setup, to obtain a single graph-based score for every patient in the data set. The following criterion was used to differentiate cases as PE or CTEPH in this experiment: (A) a PA:A ratio above 1, (B) the presence of mosaicism and (C) the graph-based score above the optimal threshold found in Experiment 1 of 0.4 (see Section 3.1). Additionally, the differentiation was also performed considering the presence of both CT findings (PA:A ratio above 1 and mosaicism), as well as the combination of both CT findings plus the graph-based scores.

The same performance metrics computed for Experiment 1 were also calculated in this experiment, only AUC was not used since the decision thresholds have been already fixed. Statistical p-values were calculated by paired Student t-tests. P-values lower than 0.05 were considered significant in the various binary classifications of the PH-related CT findings, the classification obtained from the lung graph model, as well as the intersection and combination of both the CT findings and the lung graph scores.

4. Results

4.1. Experiment 1: Lung graph model classification

Fig. 5 shows the mean classification metrics obtained with the lung graph model during the training phase, using a 10-fold CV setup and varying the decision threshold. The highest accuracy, *i.e.* 0.7636, was obtained both with the decision threshold of 0.4 and 0.5. However, the F_1 score, the kappa coefficient and the MCC had their top performance at 0.4. Since the kappa coefficient and MCC are more reliable than the accuracy when using unbalanced data sets, the optimal decision threshold for the graph model was fixed to 0.4. In the case of the two CNN approaches, the maximum performance in accuracy, kappa coefficient and MCC was obtained at 0.5. Therefore, the optimal decision threshold for the CNN approaches was set at 0.5.

Table 2 details the results obtained by the lung graph model and those from the two strategies using a convolutional neural network, in binary classification of PE vs. CTEPH, with an experimental setup using 10 train/test splits. The CTEPH class is considered the positive class, *i.e.* sensitivity refers to the proportion of CTEPH patients correctly detected. Using VGG19 features with data augmentation obtained the best scores out of both CNN methods. The specificity was significantly better (p<0.0001) with the CNN methods. However, for the rest of the evaluation metrics, the lung graph model clearly outperformed both deep learning approaches with a significant p<0.05, in all measures except in accuracy (p = 0.08). In particular, the sensitivity, F_1 score and kappa coefficient had a p<0.01.

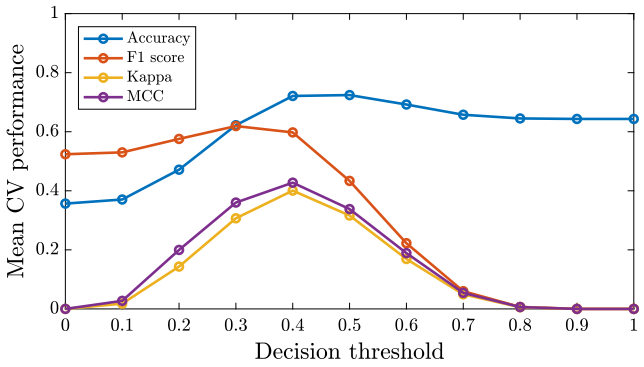

Fig. 5. Mean performance obtained during the training phase of the lung graph model when varying the decision threshold of the output score for the class CTEPH. The best overall performance was obtained with a 0.4 decision threshold.

Table 2
Detailed results obtained by the graph model and the deep learning approach in Experiment 1 (mean ± standard error in 10 train/test splits).

	VGG19	VGG19 wDataAugm	Lung graph
Sensitivity	0.42 ± 0.06	0.45 ± 0.06	**0.78 ± 0.07**
Specificity	0.91 ± 0.08	**0.91 ± 0.08**	0.75 ± 0.04
Accuracy	0.71 ± 0.05	0.72 ± 0.06	**0.76 ± 0.03**
F_1 score	0.53 ± 0.06	0.56 ± 0.07	**0.68 ± 0.05**
AUC	0.78 ± 0.06	0.79 ± 0.05	**0.84 ± 0.05**
Kappa	0.35 ± 0.10	0.39 ± 0.11	**0.49 ± 0.07**
MCC	0.42 ± 0.11	0.45 ± 0.11	**0.52 ± 0.07**

4.2. Experiment 2: visual assessment vs. lung graph model

Table 3 summarises the criteria tested in Experiment 2 along with the number of patients in each class satisfying the criteria for each group, PE and CTEPH. This table also shows a column with the p-values obtained with the criteria for the differentiation between PE vs. CTEPH. Considering all the CT examinations taken at the time of the diagnosis from this study, *i.e. 75*, a PA:A > 1 ratio is statistically significant to differentiate PE from CTEPH with a p of 0.04. On the other hand, the presence of mosaicism was not significant (p = 0.09) in this data set. Using the graph score resulted in a statistically significant trend for this task (p < 0.001). Taking into account all the cases identified using either the CT findings or the lung graph score, was also significant in this study.

Table 4 and Fig. 6 show the performance obtained when using each criterion. The graph model obtained similar results in this experiment with a LOPO setup (see Section 3.2) to the results from Experiment 1, that had a 80%–20% setup in 10 train/test splits (see Section 3.1). The accuracy obtained with the lung graph model in Experiment 2 was 0.74, F_1 score of 0.64, and a kappa coefficient of 0.44, respectively. The graph model alone showed better performance than any of the CT findings in terms of accuracy, F_1 score, kappa coefficient, and MCC. The CT findings together (PA:A ratio above 1 and presence of mosaicism) outperformed all other criteria in terms of specificity. However, for all the other measures, the addition of the lung graph criteria to these findings obtained the best performance. Finally, the combination of the CT findings (PA:A ratio > 1 and presence of mosaicism) obtained a correlation of 0.77 and 0.40 with the PA:A ratio > 1 and mosaicism respectively. On the other hand, the correlation of the CT findings with the lung graph criterion was only of 0.08.

5. Discussion

In this paper a texture-based lung graph model for the assessment of PE vs. CTEPH in CT examinations was developed and tested. This approach clearly outperformed deep learning approaches (using a CNN)

Table 3
Criteria selection in the two target groups of this study used in Experiment 2 and number of patients from each class satisfying the criteria. A LOPO approach was performed to obtain a lung graph score for each patient with an optimal threshold of 0.4 to be considered as CTEPH.

Criteria	PE (n = 48)	CTEPH (n = 27)	*p-value*
PA:A ratio > 1	7 (15%)	10 (37%)	0.04*
Mosaicism	21 (44%)	18 (67%)	0.09
PA:A > 1 and Mosaicism	3 (6%)	8 (30%)	<0.001**
Graph score > 0.4	10 (21%)	17 (63%)	<0.001**
(PA:A > 1 and Mosaicism) or Graph score > 0.4	13 (27%)	20 (74%)	<0.001**

PA:A ratio, pulmonary artery to aorta ratio.
*<0.05.
**<0.001.

Table 4
Performance metrics from Experiment 2.

	PA:A	Mosaic.	PA:A and Mosaic.	Graph	(PA:A and Mosaic.) or Graph
Sensitivity	0.38	0.69	0.31	0.65	**0.77**
Specificity	0.85	0.56	**0.94**	0.79	0.73
Accuracy	0.69	0.61	0.72	**0.74**	**0.74**
F_1 score	0.47	0.55	0.43	0.64	**0.68**
Kappa	0.26	0.23	0.28	0.44	**0.47**
MCC	0.27	0.24	0.33	0.44	**0.48**

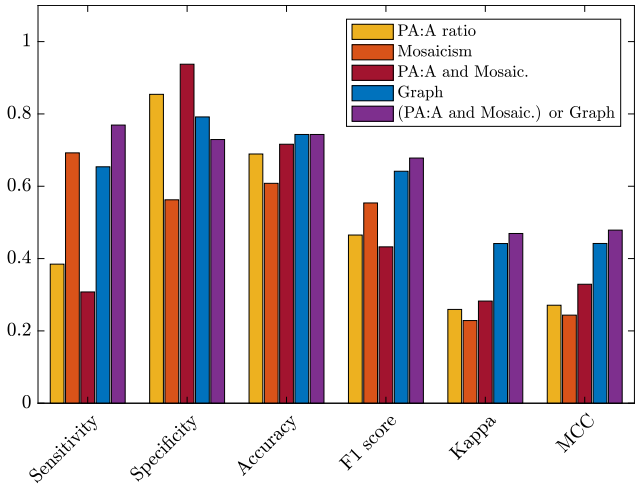

Fig. 6. Comparison of the criteria when detecting CTEPH. In order to obtain a graph-based score for each patient a leave-one-patient-out (LOPO) setup was used. The results of this setup are consistent with the ones reported in experiment 1 (see Table 2).

that analysed lung regions in an isolated manner. Moreover, the lung graph approach produced a higher sensitivity than known CT radiological findings suggestive of PH, *i.e.* PA:A ratio > 1 and mosaicism, in a data set of 75 cases. The combination of lung graph scores together with both CT findings resulted in the overall best performance for the differentiation between acute PE and CTEPH cases, using only the information obtained from the CT scans. All of the 27 CTEPH patients were incident cases, with the diagnosis confirmed by the results from a right heart catheterism and an early CT examination available.

In Experiment 1, the feasibility of using our novel lung graph model in the task of detecting PE vs. CTEPH was evaluated and compared to a state-of-the-art deep learning approach that had achieved the best scores in an open lung classification challenge with a large data set. The lung graph model obtained in this work an average AUC of 0.84 identifying CTEPH cases in 10 train/test splits, where the AUC of a random classifier is 0.50. Although the CNN strategy obtained a higher mean specificity (see Table 2), the lung graph model outperformed the deep learning method in the remaining evaluation metrics with a statistically significant result. It is important to highlight the mean sensitivity results, as our graph approach had a promising 0.78 score vs. 0.45 obtained using local CNN image features with data augmentation. The real contribution from using our novel graph framework is clear in this respect, as commonly used CNNs generally do not consider the global correlations of multiple local regions, in this case relying on independent lung tissue patches. Moreover, the lung graph model can be easily retrained for larger data sets in a robust and less time consuming process. This opens the door to perform a more thorough analysis of the changes in the lung parenchyma already present in the CT scans from these CTEPH patients, some of them at an earlier stage of the disease. The small standard error from the 10 train/test splits supports the robustness of our model to variations in the train and test sets. Moreover, this approach had a promising average sensitivity of 0.78, comparable to the previously reported sensitivities of CT findings associated with CTEPH: abnormal PA:A ratio (0.58–0.87) and presence of mosaic attenuation patterns (0.77–1.00) [14,17].

In Experiment 2, the performance of the lung graph model and each CT finding criterion was evaluated for the detection of CTEPH in scans taken at the time of the diagnosis. The CT findings were manually assessed for all patients by expert radiologists. A LOPO setup was used to compute an individual lung graph score for detecting CTEPH. Using this LOPO setup, the lung graph model obtained similar results as in Experiment 1 (see Tables 2 and 4), which further supports the results obtained with the initial experimental setup. Mosaicism was the most sensitive individual criterion, present in 69% of the CTEPH patients, however, it was also the least specific one, being present in 44% of the acute PE patients. On the other hand, the lung graph model offered the best trade-off between sensitivity and specificity with an accuracy of 0.74. The presence of both CT findings together had the highest specificity in the results. If both the CT findings and lung graph model are combined, then the majority of the evaluation metrics improve. The correlation between the two CT findings shows that most of the patients with a large PA:A ratio also presented mosaicism but not many patients with mosaicism had a large PA:A ratio. A closer analysis of the CTEPH patients detected with the lung graph model revealed that this approach is complementary to the radiological CT findings, with a relatively small correlation between the graph results and those from the coupling of both CT findings. Unlike the PA:A ratio, the lung graph score results from a direct analysis of the lung tissue, which might explain the differences in the cases detected from this CT finding and those with the proposed texture analysis. This is an interesting result from our study and a promising direction for the inclusion of lung texture analysis in future studies.

The inclusion of objective and quantitative image analysis techniques, such as the proposed graph model, could help by providing health experts with quantitative sub-visual information about the lung

parenchyma. More studies are needed to clarify this result for a better clinical interpretation, ideally relying on a larger number of cases and a more thorough evaluation of the resulting texture analysis for each individual patient. The proposed framework proved to be competitive against a state-of-the-art system using the same data set and obtained very good results in the comparison to some of the best CT findings that radiologists have to diagnose this disease. As future work, we intend to analyse prior CT scans of the same patients that were initially misclassified as acute PE. The hypothesis of discovering early changes associated to CTEPH, originally proposed by Guérin et al. in a follow-up of patients during 26 months after an acute PE event, who eventually developed CTEPH, could further be studied with the proposed method [12]. More recently, Klok et al. stressed the importance of distinguishing between 'true' acute PE and CTEPH at a (sub)acute clinical presentation [11], also motivating further texture analysis of the cases in our data set.

6. Conclusions

We proposed a texture-based lung graph model to enhance the characterisation and detection of CTEPH, based solely on CT scans taken during a symptomatic acute PE event. This objective texture biomarker was combined with CT findings suggestive of CTEPH, resulting in a considerably higher agreement with the reference diagnosis from the right heart catheterism. A full radiological assessment of the patient scans was performed and reported in the search for CT findings classically associated with PH. Overall, these patients could benefit from an early identification of CTEPH changes in the texture of the lung parenchyma, consequently distinguishing them from only acute PE cases.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgement

This work was partly supported by the Swiss National Science Foundation in the PH4D project (grant agreement 320030-146804).

References

[1] Heinrike Wilkens, Stavros Konstantinides, Irene M. Lang, Alexander C. Bunck, Mario Gerges, Felix Gerhardt, Aleksandar Grgic, Christian Grohé, Stefan Guth, Matthias Held, Jan B. Hinrichs, Marius M. Hoeper, Walter Klepetko, Thorsten Kramm, Ulrich Krüger, Mareike Lankeit, Bernhard C. Meyer, Karen M. Olsson, Hans-Joachim Schäfers, Matthias Schmidt, Hans-J Seyfarth, Silvia Ulrich, Christoph B. Wiedenroth, Eckhard Mayer, Chronic thromboembolic pulmonary hypertension (CTEPH): Updated recommendations from the cologne consensus conference 2018, Int. J. Cardiol. (2018).

[2] Nazzareno Galiè, Marc Humbert, Jean-Luc Vachiery, Simon Gibbs, Irene Lang, Adam Torbicki, Gérald Simonneau, Andrew Peacock, Anton Vonk Noordegraaf, Maurice Beghetti, Ardeschir Ghofrani, Miguel Angel Gomez Sanchez, Georg Hansmann, Walter Klepetko, Patrizio Lancellotti, Marco Matucci, Theresa McDonagh, Luc A. Pierard, Pedro T. Trindade, Maurizio Zompatori, Marius Hoeper, 2015 ESC/ERS guidelines for the diagnosis and treatment of pulmonary hypertension: the joint task force for the diagnosis and treatment of pulmonary hypertension of the European society of cardiology (ESC) and the European respiratory society (ERS): endorsed by: Association for European paediatric and congenital cardiology (AEPC), international society for heart and lung transplantation (ISHLT), Eur. Heart J. 37 (1) (2015) 67–119.

[3] Yvonne M. Ende-Verhaar, Suzanne C. Cannegieter, Anton Vonk Noordegraaf, Marion Delcroix, Piotr Pruszczyk, Albert T.A. Mairuhu, Menno V. Huisman, Frederikus A. Klok, Incidence of chronic thromboembolic pulmonary hypertension after acute pulmonary embolism: a contemporary view of the published literature, Eur. Respir. J. 49 (2) (2017) 1601792.

[4] Anton Vonk Noordegraaf, Berend E. Westerhof, Nico Westerhof, The relationship between the right ventricle and its load in pulmonary hypertension, J. Am. Coll. Cardiol. 69 (2) (2017) 236–243.

[5] Eckhard Mayer, David Jenkins, Jaroslav Lindner, Andrea D'armini, Jaap Kloek, Bart Meyns, Lars Bo Ilkjaer, Walter Klepetko, Marion Delcroix, Irene Lang, Joanna Pepke-Zaba, Gerald Simonneau, Philippe Dartevelle, Surgical management and outcome of patients with chronic thromboembolic pulmonary hypertension: results from an international prospective registry, J. Thorac. Cardiovasc. Surg. 141 (3) (2011) 702–710.

[6] Stavros V. Konstantinides, Adam Torbicki, Giancarlo Agnelli, Nicolas Danchin, David Fitzmaurice, Nazzareno Galiè, J. Simon R. Gibbs, Menno V. Huisman, Marc Humbert, et al., 2014 ESC guidelines on the diagnosis and management of acute pulmonary embolism: The task force for the diagnosis and management of acute pulmonary embolism of the European society of cardiology (ESC) endorsed by the European respiratory society (ERS), Eur. Heart J. 35 (43) (2014) 3033–3073.

[7] Frederic Lador, Philippe Herve, A practical approach of pulmonary hypertension in the elderly, Sem. Respir. Crit. Care Med. 34 (5) (2013) 654–664.

[8] Marius M. Hoeper, Stephen H. Lee, Robert Voswinckel, Massimiliano Palazzini, Xavier Jais, Alessandro Marinelli, Robyn J. Barst, Hossein A. Ghofrani, Zhi-Cheng Jing, Christian et al. Opitz, Complications of right heart catheterization procedures in patients with pulmonary hypertension in experienced centers, J. Am. Coll. Cardiol. 48 (12) (2006) 2546–2552.

[9] Joanna Pepke-Zaba, Marion Delcroix, Irene Lang, Eckhard Mayer, Pavel Jansa, David Ambroz, Carmen Treacy, Andrea M. D'armini, Marco Morsolini, Repke Snijder, et al., Chronic thromboembolic pulmonary hypertension (CTEPH) clinical perspective: Results from an international prospective registry, Circulation 124 (18) (2011) 1973–1981.

[10] Yvonne M. Ende-Verhaar, Menno V. Huisman, Frederikus A. Klok, To screen or not to screen for chronic thromboembolic pulmonary hypertension after acute pulmonary embolism, Thromb. Res. 151 (2017) 1–7.

[11] Frederikus A. Klok, Stefano Barco, Follow-up after acute pulmonary embolism, Hämostaseologie 38 (01) (2018) 22–32.

[12] Laurent Guérin, Francis Couturaud, Florence Parent, Marie-Pierre Revel, Florence Gillaizeau, Benjamin Planquette, Daniel Pontal, Marie Guégan, Gérald Simonneau, Guy Meyer, Olivier Sanchez, Prevalence of chronic thromboembolic pulmonary hypertension after acute pulmonary embolism, Thromb. Haemost. 112 (03) (2014) 598–605.

[13] Martin I. MacDonald, Broader implications for the widened pulmonary artery? Respirology 22 (8) (2017) 1495–1496.

[14] Yongchun Shen, Chun Wan, Panwen Tian, Yanqiu Wu, Xiaoou Li, Ting Yang, Jing An, Tao Wang, Lei Chen, Fuqiang Wen, CT–base pulmonary artery measurement in the detection of pulmonary hypertension: A meta–analysis and systematic review, Medicine 93 (27) (2014).

[15] Joanna Żyłkowska, Marcin Kurzyna, Michał Florczyk, Barbara Burakowska, Franciszek Grzegorczyk, Janusz Burakowski, Maria Wieteska, Karina Oniszh, Andrzej Biederman, Liliana Wawrzyńska, et al., Pulmonary artery dilatation correlates with the risk of unexpected death in chronic arterial or thromboembolic pulmonary hypertension, Chest 142 (6) (2012) 1406–1416.

[16] Marc Heinrich, Michael Uder, Dietmar Tscholl, Aleksandar Grgic, Bernhard Kramann, Hans-Joachim Schäfers, CT scan findings in chronic thromboembolic pulmonary hypertension: predictors of hemodynamic improvement after pulmonary thromboendarterectomy, Chest 127 (5) (2005) 1606–1613.

[17] Hajime Kasai, Nobuhiro Tanabe, Kazushi Fujimoto, Hiromasa Hoshi, Jun Naito, Ryuto Suzuki, Akane Matsumura, Toshihiko Sugiura, Seiichiro Sakao, Koichiro Tatsumi, Mosaic attenuation pattern in non–contrast computed tomography for the assessment of pulmonary perfusion in chronic thromboembolic pulmonary hypertension, Respir. Investig. 55 (5) (2017) 300–307.

[18] Adrien Depeursinge, Omar S. Al-Kadi, J. Ross Mitchell, Biomedical Texture Analysis: Fundamentals, Applications and Tools, in: Elsevier-MICCAI Society Book series, Elsevier, 2017, https://www.elsevier.com/books/title/author/9780128121337.

[19] Antonio Foncubierta-Rodríguez, Oscar Jimenez-del-Toro, Alexandra Platon, Pierre-Alexandre Poletti, Henning Müller, Adrien Depeursinge, Benefits of texture analysis of dual energy CT for computer–aided pulmonary embolism detection, in: The 35th Annual International Conference of the IEEE Engineering in Medicine and Biology Society, EMBC 2013, IEEE, 2013, pp. 3973–3976.

[20] Yashin Dicente Cid, Oscar Jimenez-del-Toro, Alexandra Platon, Henning Müller, Pierre-Alexandre Poletti, From local to global: A holistic lung graph model, in: Medical Image Computing and Computer–Assisted Intervention – MICCAI 2018, 2018.

[21] Nazzareno Galiè, Marius M. Hoeper, Marc Humbert, Adam Torbicki, Jean-Luc Vachiery, Joan Albert Barbera, Maurice Beghetti, Paul Corris, Sean Gaine, J Simon Gibbs, et al., Guidelines for the diagnosis and treatment of pulmonary hypertension: the task force for the diagnosis and treatment of pulmonary hypertension of the European society of cardiology (ESC) and the European respiratory society (ERS), endorsed by the international society of heart and lung transplantation (ISHLT), Eur. Heart J. 30 (20) (2009) 2493–2537.

[22] Claudia Grosse, Alexandra Grosse, CT findings in diseases associated with pulmonary hypertension: a current review, Radiographics 30 (7) (2010) 1753–1777.

[23] Anastasia Chalkidou, Michael J. O'Doherty, Paul K. Marsden, False discovery rates in PET and CT studies with texture features: A systematic review, PLoS One 10 (5) (2015) 1–18.

[24] Yashin Dicente Cid, Oscar Jimenez-del-Toro, Adrien Depeursinge, Henning Müller, Efficient and fully automatic segmentation of the lungs in CT volumes, in: Orcun Goksel, Jimenez-del-Toro, Oscar and Foncubierta–Rodriguez, Antonio and Müller, Henning (Eds.), Proceedings of the VISCERAL Challenge at ISBI, CEUR Workshop Proceedings, number 1390, 2015, pp. 31–35.

[25] Tatjana Zrimec, Sata Busayarat, Peter Wilson, A 3D model of the human lung with lung regions characterization, in: ICIP 2004 Proc. IEEE Int. Conf. on Image Processing, vol. 2, 2004, pp. 1149–1152.

[26] Adrien Depeursinge, Tatjana Zrimec, Sata Busayarat, Henning Müller, 3D lung image retrieval using localized features, in: Medical Imaging 2011: Computer-Aided Diagnosis, volume 7963, SPIE, 2011, p. 79632E.

[27] Kun Liu, Henrik Skibbe, Thorsten Schmidt, Thomas Blein, Klaus Palme, Thomas Brox, Olaf Ronneberger, Rotation-invariant HOG descriptors using Fourier analysis in polar and spherical coordinates, Int. J. Comput. Vis. 106 (3) (2014) 342–364.

[28] Yashin Dicente Cid, Henning Müller, Alexandra Platon, Pierre-Alexandre Poletti, Adrien Depeursinge, 3–D solid texture classification using locally–oriented wavelet transforms, IEEE Trans. Image Process. 26 (4) (2017) 1899–1910.

[29] Oscar Jimenez-del-Toro, Antonio Foncubierta-Rodríguez, Maria-Isabel Vargas Gomez, Henning Müller, Adrien Depeursinge, Epileptogenic lesion quantification in MRI using contralateral 3D texture comparisons, in: Medical Image Computing and Computer–Assisted Intervention – MICCAI 2013, in: Lecture Notes in Computer Science, vol. 16, Springer, Heidelberg, Germany, 2013, pp. 353–360.

[30] Vitali Liauchuk, Aleh Tarasau, Eduard Snezhko, Vassili Kovalev, Andrei Gabrielian, Alex Rosenthal, ImageCLEF 2018: Lesion-based TB-descriptor for CT image analysis, in: CLEF2018 Working Notes, in: CEUR Workshop Proceedings, CEUR-WS.org, Avignon, France, 2018, http://ceur-ws.org.

[31] Yashin Dicente Cid, Vitali Liauchuk, Vassili Kovalev, Henning Müller, Overview of imagecleftuberculosis 2018 - detecting multi-drug resistance, classifying tuberculosis type, and assessing severity score, in: CLEF2018 Working Notes, in: CEUR Workshop Proceedings, CEUR-WS.org, Avignon, France, 2018, http://ceur-ws.org.

[32] Alex Krizhevsky, Ilya Sutskever, Geoffrey E. Hinton, ImageNet classification with deep convolutional neural networks, in: F. Pereira, C.J.C. Burges, L. Bottou, K.Q. Weinberger (Eds.), Advances in Neural Information Processing Systems 25, Curran Associates, Inc., 2012, pp. 1097–1105.

[33] Karen Simonyan, Andrew Zisserman, Very deep convolutional networks for large-scale image recognition, 2014, arXiv preprint arXiv:1409.1556.

[34] Jia Deng, Wei Dong, Richard Socher, Li-Jia Li, Kai Li, Li Fei-Fei, Imagenet: A large-scale hierarchical image database, in: 2009 IEEE Conference on Computer Vision and Pattern Recognition, IEEE, 2009, pp. 248–255.

[35] Douglas G. Altman, J. Martin Bland, Standard deviations and standard errors, BMJ 331 (7521) (2005) 903, http://dx.doi.org/10.1136/bmj.331.7521.903.