

Fusing Learned Representations from Riesz and Deep CNNs for Lung Tissue Classification

Ranveer Joyseeree, Sebastian Otálora, Henning Müller, and Adrien Depeursinge

Abstract—We present a novel method to detect and classify several classes of diseased and healthy lung tissue in CT based on the fusion of Riesz and deep learning features. First, discriminative parametric lung tissue texture signatures are learned based on Riesz representations and a one-versus-one approach. The signatures are generated for four diseased tissue types and a healthy tissue class, all of which appear in Interstitial Lung Diseases (ILD). Because the Riesz wavelets are steerable, they can easily be made invariant to local image rotations, a property that is desirable when analyzing lung tissue micro-architectures in CT images. Second, features from deep Convolutional Neural Networks (CNN) are computed by fine-tuning the GoogLeNet architecture using an augmented version of the same dataset of ILDs. Because CNN features are both deep and non-parametric, they can accurately model virtually any pattern that is useful for tissue discrimination. However, invariance to local image rotations is not explicitly implemented can only be approximated with rotation-based data augmentation. This motivates the fusion of Riesz and deep CNN features. The two learned representations are combined in a joint softmax model for final classification, where early and late feature fusion schemes are compared. The experimental results show that a late fusion of the independent probabilities leads to significant improvements in classification performance when compared to each of the separate feature representations.

Index Terms—texture signatures, one-versus-one, classification, ILD, Deep Learning.

I. INTRODUCTION

THE White Book of the European Respiratory Society (ERS)¹ mentions respiratory diseases as one of the most common causes of premature mortality. In 2008, one out every six deaths worldwide was attributable to them. An associated annual cost of 380 billion Euros in the European Union was estimated in terms of loss of productive output, direct medical care and drugs. Battling these diseases is thus a top priority in the healthcare domain. To combat avoidable deaths and large costs, obtaining an early accurate diagnosis is essential. In such a scenario, clinicians may prescribe the correct treatment as early as possible.

Respiratory ailments affecting lung parenchyma are very common. One of the largest and most diverse group of such diseases is the set of Interstitial Lung Diseases (ILD). They

account for more than 200 pathologies affecting the alveoli, the small lung airways, and the pulmonary interstitium [1]. Information gathered from clinical, radiological, and pathological analyses are required to accurately diagnose ILDs. High Resolution Computed Tomography (HRCT) images are the radiological modality of choice for the characterization of ILDs [2]. Some of the ailments may be easily misdiagnosed due to their rarity and to the fact that radiologists are subjective when interpreting the content of the images [3, 4]. Therefore, computerized assistance yielding exhaustive and reproducible image analysis has been mentioned several times as beneficial for improving ILD management [5].

The task of classifying lung tissue pathologies benefits from recent advances made in the area of visual pattern recognition. In the particular context of texture and tissue characterization, the latter relies heavily on the local organization of image directions at different scales [6, 7], including local variations of pattern properties such as local anisotropy [8, 9]. Spatial domain representations of images provide insufficient information alone to properly examine the local organization of scales and directions. Therefore, to obtain a more complete overview of the relationships between them, intensity information needs to be complemented with information extracted in the frequency domain.

Information embedded in the local organization of scales and directions in images is exploited by several authors for pattern characterization and recognition. Grey-Level Co-occurrence Matrices (GLCM) [10], Histograms Of Gradients (HOG) [11] used in the Scale-Invariant Feature Transform (SIFT) [12], non-separable and separable wavelets [13], Run-Length Encoding (RLE) [14], and oriented filterbanks and wavelets (Gaussian, Gabor, Leung-Malik, Maximum Response [15–19] have been proposed for directional analysis. Unfortunately, separable wavelets suffer from bias along the vertical and horizontal axes [20], while the remainder require an arbitrary choice of image directions [8]. Using a sequence of pixels along perimeters of radius r , Local Binary Patterns (LBP) [21] are able to perform multi-directional analysis but they do not allow for multiresolution analysis. In addition, r is determined through costly optimization. Other methods exploit the local organization of directions and scales indirectly. Notable examples include Convolutional Neural Networks (CNN) [22, 23], Topographic Independent Component Analysis (TICA) [24] and the scattering transform [25, 26]. Despite their lack of interpretability, Deep Learning (DL) models, and specifically CNNs, are now *de facto* standard methods for solving challenging computer vision tasks due to the performance improvements they bring when com-

R. Joyseeree is with Swiss Federal Institute of Technology (ETH) Zürich, Switzerland and the MedGIFT group at the HES-SO Valais, Switzerland.

S. Otálora and H. Müller are with the University of Geneva and the MedGIFT group at HES-SO Valais, Switzerland.

A. Depeursinge is with the Biomedical Imaging Group at the Ecole polytechnique fédérale de Lausanne (EPFL) and the MedGIFT group at HES-SO Valais, Switzerland. E-mail: adrien.depeursinge@hevs.ch.

¹European lung white book, <http://www.erswhitebook.org/chapters/the-burden-of-lung-disease/>, as of February 2018.

pared with classical handcrafted features approaches. In recent years, these techniques have been successfully used in many medical-domain tasks showing promising results and opening different research avenues [27]. For the analysis of lung CT images, there are now several deep learning approaches that show the effectiveness of this technique when compared to the more classical design of handcrafted features approach. These results are now routinely encountered in the literature due to the capacity of the deep architectures to learn a wide range of filters that respond to complex patterns. Moreover, the increasing availability of medical datasets allows this method to have more robust and precise results [28–30]. For instance, CNNs have been used for lung disease classification in [28], where performance was not assessed explicitly, and in [31] with a reasonable degree of success. Due to the flexibility of the features learned with deep learning models having millions of parameters, they allow the representation of a large number of patterns present in the dataset, which usually exacerbates the risk of overfitting.

Another fundamental and general aspect that needs to be accounted for is that the same texture pattern can appear at several local orientations. Locally rotation-invariant features are desirable in such instances [32, 33]. LBP [21] and Rotation-covariant SIFT (RIFT) [34] possess such a property but they do not model discriminative patterns specifically (*i.e.*, they yield handcrafted representations) and require exhaustive calculations. Approaches based on steerable filters can achieve machine-precision multi-directional and multi-scale characterization with invariance to local rotations [35, 36].

Learned representations based on Riesz wavelets [37], as used in this work, can precisely model multi-scale and multi-directional information that is useful for tissue discrimination [38]. In addition, the obtained representation can easily be made invariant to local rotations using the steerability of the models (see Section II-B)). One drawback of Riesz representation is their reliance on parametric basis functions with potential lack of span. CNNs do not have invariance to local rotation explicitly implemented. This can be alleviated up to a certain degree with artificially augmented versions of the input after several rotations. Additionally, CNN kernels do not rely on parametric representations and the large number of learned filters can model virtually any pattern relevant to discrimination, under the condition that the dataset is large enough. The complementarity of the two approaches motivates the fusion of the two representations into a single model.

In summary, this paper describes a novel feature-fusion approach that exploits the complementarity of the learned representations from Riesz wavelets and fine-tuned deep CNNs to classify five tissue types associated with ILDs. We propose both early and late fusion strategies and estimate the performance with a four fold cross-validation setup. We compare all the methods using a softmax classifier with the same hyperparameters to focus on the discriminatory power of the extracted features.

The paper is organized as follows. In Section II, we first present the publicly available dataset used for the validation of our approach. We then review publications that used same data and comparable validation schemes. This is followed by

an in-depth description of the theoretic aspects of the proposed method. Section III presents the evaluation of classification performance, observed results, and statistical significance of the performance comparison. A thorough analysis and interpretation of all observed behavior is carried out in Section IV. Finally, conclusions are drawn based on the work done and we propose measures to improve performance in future work in Section V.

II. MATERIALS AND METHODS

A. Dataset and Validation Schemes

We use the publicly-available ILD dataset [5] to evaluate the performance of the proposed methods. A slice of an HRCT series belonging to that dataset is shown in Fig. 1. It depicts the lung parenchyma of a patient that has been annotated by an expert radiologist. That dataset has been used several times in past publications and these will be briefly covered in this section.

Five tissue types are often classified by the publications using the ILD dataset: healthy, emphysema, ground glass, fibrosis, and micronodules. Some of the earliest of such papers were published by our group [39, 40], where we combined image data with clinical parameters to carry out classification. This was followed by handcrafted steerable Riesz filterbanks [41], isotropic wavelet frames [42] and low-level localized features [43]. Thereafter, we used learned representations based on the Riesz transform [44]. Other authors include Song *et al.* [45] who first utilized feature-based image patch approximation. Li *et al.* [46] used then automatic feature learning followed by a customized CNN approach in [31], while in [47], a locality-constrained subcluster representation ensemble is used. Gao *et al.* use a deep CNN approach in [28].

The following publications use a different set of tissue types. Some classify six tissue types by including the consolidation type. Examples of such instances include Foncubierta *et al.* [48] who used multi-scale visual words for classification and retrieval. Shin *et al.* [29] used deep CNNs. Others used a radically different set of classes. For example, Anthimopoulos *et al.* [30] applied a deep CNN to the following classes: healthy, consolidation, honeycombing, micronodules, reticulation, ground glass, as well as a combination of reticulation and ground glass.

To facilitate the comparison of our work with the majority of other publications on the ILD dataset, we carry out supervised learning on the following five expert-annotated classes: healthy, emphysema, ground glass, fibrosis, and micronodules. Moreover, these classes are most common in the majority of ILD and therefore are relatively well represented in terms of number of annotated regions available. An illustration of their respective appearances is provided in Fig. 2. One may observe that the visual differences between them are quite subtle, especially when comparing the healthy, emphysema, and micronodules classes.

For the purposes of training and testing, four fold cross-validation is employed. This entails extracting as many patches as possible from the annotated ILD images. In the case of DL, a further augmentation step is taken whereby the patches

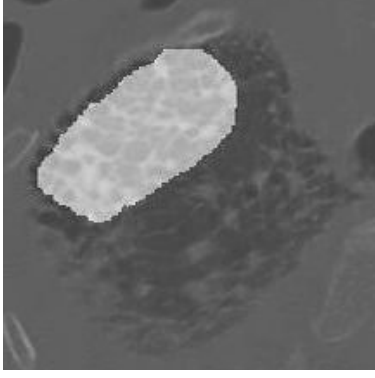


Fig. 1. Part of a slice taken from the ILD database that represents the left lung along with an expert annotation (overlay) corresponding to a Region Of Interest (ROI) are presented here.

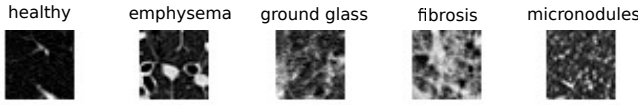


Fig. 2. The five tissue classes selected for our work representing healthy parenchyma as well as representatives of emphysema, ground glass, fibrosis, and micronodules.

previously obtained are rotated by 90, 180, and 270 degrees. We also reflected them along the vertical and horizontal axes. The new set of patches is then divided into four groups according to two strategies. In the first one, they are considered to be independent from each other, in line with what is usually encountered in the literature, and are divided into 4 equal sets. In the second strategy, we ensure that the patches originating from a certain patient only appear in one of the four groups to minimize the risk of latent bias. Two of the four groups are then concatenated and used for training the classification model. One of the remainder is used for validation where necessary and the last one is used for testing the trained model. This process is repeated four times to ensure that each group is once in the test set.

Finally, although the slice thickness and slice pixel dimensions of the HRCT protocol are all 1mm, the spacing between slices is 10mm. This implies that a considerable amount of information is missing between slices, which cannot be easily reconstructed. There is no justification, therefore, to consider 3D image analysis.

B. Tissue Characterization Using Riesz Filters

This section describes the approach to obtain learned discriminative and locally rotation-invariant texture representations from Riesz wavelets and is based on [37, 38]. We first introduce the Riesz transform and its combination with radial wavelets to derive steerable filterbanks in Section II-B1. Second, we describe in Section II-B2 how we learn one-versus-one class-specific discriminative texture signatures from the parametric Riesz representation using Support Vector Machines (SVM).

1) *Steerable Riesz filterbanks*: In a nutshell, Riesz filterbanks provide sets of image operators behaving like multi-

scale local partial image derivatives of any order. Let $f(\mathbf{x})$ represent the function that models the content of a patch where \mathbf{x} represents pixel coordinates x_1 and x_2 . In other words, $f: \mathbf{x} \rightarrow f(\mathbf{x}), \mathbf{x} \in \mathbb{R}^2$, where $\mathbf{x} = (x_1, x_2)$.

Since texture is encoded in the spatial transitions between the pixel values, the characterization of the imaging features may be achieved in the Fourier domain in terms of spatial frequencies. The Fourier domain representation of $f(\mathbf{x})$ is defined as

$$f(\mathbf{x}) \xleftrightarrow{\mathcal{F}} \hat{f}(\boldsymbol{\omega}) = \int_{\mathbb{R}^2} f(\mathbf{x}) e^{-j\langle \boldsymbol{\omega}, \mathbf{x} \rangle} d\mathbf{x}_1 d\mathbf{x}_2, \quad (1)$$

where $\boldsymbol{\omega} = (\omega_1, \omega_2)$ and $\langle \cdot, \cdot \rangle$ denotes the dot product.

The Riesz filterbanks needed for our work are based on the real Riesz transform [49]. The $N+1$ components of the N th-order Riesz transform are based on the collection of operators $\mathcal{R}^{(n, N-n)}\{\cdot\}$ as

$$\mathcal{R}^N \{f\}(\mathbf{x}) = \begin{pmatrix} \mathcal{R}^{(0, N)} \{f\}(\mathbf{x}) \\ \vdots \\ \mathcal{R}^{(n, N-n)} \{f\}(\mathbf{x}) \\ \vdots \\ \mathcal{R}^{(N, 0)} \{f\}(\mathbf{x}) \end{pmatrix}, \quad (2)$$

with $n = 0, 1, \dots, N$. A kernel $\mathcal{R}^{(n, N-n)} \{f\}(\mathbf{x})$ that represents the effect of a member of the filterbank on the input signal is defined in the spatial and Fourier domains as:

$$\mathcal{R}^{(n, N-n)} \{f\}(\mathbf{x}) \xleftrightarrow{\mathcal{F}} \widehat{\mathcal{R}^{(n, N-n)} \{f\}}(\boldsymbol{\omega}),$$

where

$$\widehat{\mathcal{R}^{(n, N-n)} \{f\}}(\boldsymbol{\omega}) = \sqrt{\frac{N}{n!(N-n)!}} \frac{(-j\omega_1)^n (-j\omega_2)^{N-n}}{\|\boldsymbol{\omega}\|^N} \hat{f}(\boldsymbol{\omega}). \quad (3)$$

According to Eq. (3), the product of $j\omega_1$ and/or $j\omega_2$ in the numerator followed by division with the norm of $\boldsymbol{\omega}$ produces allpass filters that only retain phase information that characterize directions [8, 50] and the order N controls the angular selectivity of the Riesz kernels. Therefore, the Riesz kernels behave like allpass N -th order partial image derivatives. Fig. 3 illustrates the Riesz filterbanks for $N = 1, \dots, 5$.

We also seek the steerability property of Riesz filterbanks [50, 51]. In essence, this implies that a linear combination of the filterbanks may model any local rotation. When looking at the maximum response over all possible orientations, steerability allows achieving local rotation invariance at a relatively cheap computational cost because it is not needed to re-convolve the image with rotated versions of the kernels.

For any rotation angle $\theta \in [0, 2\pi]$, a steering matrix A_θ determines the corresponding response of the kernels in the filterbank to $f(\mathbf{x})$ for a rotation around $\mathbf{0}$ as

$$\mathcal{R}^N \{f_\theta\}(\mathbf{0}) = A_\theta \mathcal{R}^N \{f\}(\mathbf{0}), \quad (4)$$

where f_θ denotes the rotation of f as $f_\theta(\mathbf{x}) = f(\mathbf{R}_\theta \mathbf{x})$ and \mathbf{R}_θ is a 2D rotation matrix.

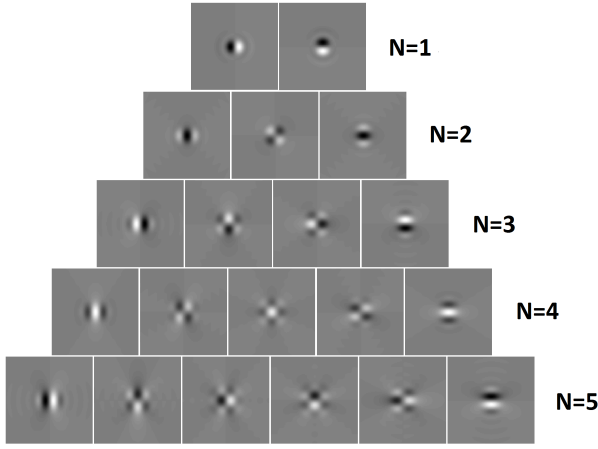


Fig. 3. Riesz filterbanks for orders up to 5 are shown here. To represent the filters on a finite spatial support, the Riesz transform was applied to an isotropic Gaussian function.

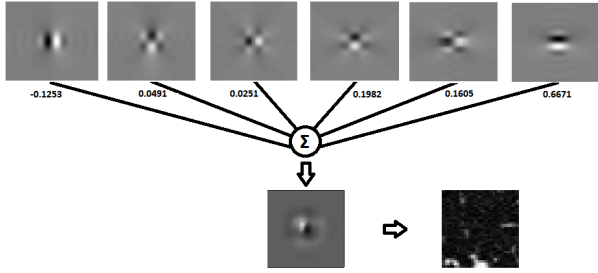


Fig. 4. A texture signature for the micronodule class is built after applying an appropriate weighing scheme for the components of a Riesz filterbank of order 5.

For multi-scale analysis, the Fourier domain is partitioned using wavelets into several consecutive dyadic bands of decreasing sizes based on Simoncelli's isotropic multiresolution framework [52]. The bands control the spatial support or scale of the (allpass) Riesz kernels.

2) *Parametric Discriminative Texture Signatures*: Learned representations based on class-specific texture steerable signatures are obtained by finding a weighting scheme for the Riesz filterbanks at each scale. Fig. 4 illustrates this for a Riesz filterbank of order 5, which is used to generate a signature for micronodules. We are looking for an optimal texture signature Γ_c^N of the class c from a linear combination of the Riesz kernels as

$$\begin{aligned} \Gamma_c^N &= \mathbf{w}^T \mathbf{R}^N \\ &= w_1 \mathcal{R}^{(0,N)} + w_2 \mathcal{R}^{(1,N-1)} + \dots + w_{N+1} \mathcal{R}^{(N,0)}, \end{aligned} \quad (5)$$

where \mathbf{w} contains the weights of the respective Riesz kernels. A multi-scale texture signature is obtained by extending Eq. (5) using multi-scale Riesz filterbanks [44] as

$$\begin{aligned} \Gamma_c^N &= w_1 \left(\mathcal{R}^{(0,N)} \right)_{s_1} + w_2 \left(\mathcal{R}^{(1,N-1)} \right)_{s_1} + \dots \\ &\quad + w_{J(N+1)} \left(\mathcal{R}^{(N,0)} \right)_{s_J}, \end{aligned} \quad (6)$$

where s_j , for $j = 1, \dots, J$ is the scale index.

We determine the weighing scheme using a one-versus-one SVM classification configuration. The filter energy responses

$E(\mathcal{R}^{(n,N-n)}\{f\}(x))$ are computed and regrouped for each class c versus each one from the remaining classes. SVMs then finds the optimal separation in terms of minimized structural risk [53, 54]. Each class benefits from a unique characterization with regard to each of the remaining classes. With five classes, the approach would be expected to lead to $5 \cdot 4 = 20$ separate optimal signatures. However, because the optimal separation between a class A and another class B is the same as that between class B and class A , the number of optimal signatures reduces to 10. The optimal weights $\mathbf{w} = (w_1, \dots, w_{N+1})$ are directly determined from the support vectors of the optimal separations [8].

C. Classification Using Riesz Filters

Through the approach defined previously, a class-wise texture signature Γ_c^N is obtained for each class c . Input images are then filtered using the steered texture signatures: they are steered at every position in the image to maximize their response, leading to a non-linear filtering operation. A feature space is spanned by the average energy of the steered filter responses and is complemented using a histogram of Hounsfield Units (HUs) of the patches in the spatial domain. This helps including the intensity information of the images which is not taken into account when only wavelet domain characteristics (*i.e.*, band-pass) are utilized. The feature space is completed by the number of pixels representing air in a patch as this also helps characterize lung tissue. In short, the feature space of a patch is made up of 10 filter responses, a histogram of HUs and the number of air pixels. After a series of initial investigations (not detailed in this paper) into an appropriate value for the Riesz order N , a value of 5 was chosen as it was providing a good trade-off between directional specificity and regularization. For $N=5$, 10 texture signatures of length $(N+1) \cdot J = 24$ each are obtained. To create the final feature vector, the $24 \cdot 10 = 240$ variables obtained in the previous step are concatenated with the 22 variables from the histogram and one variable for air content. As a result, a feature vector of length 263 is obtained for each patch.

A new supervised SVM model is trained using the features thus obtained. In this instance, a one-versus-all approach is adopted to find the optimal separation between a certain class and all remaining classes. Test patches are then classified using the obtained model. The strength of this approach is that it allows the analysis of the relative probabilities of a patch belonging to each of the 5 classes. This provides useful insights regarding the internal decision-making of the SVM and may be exploited for further classification improvements.

D. Tissue Characterization With Deep CNNs

DL has shown significant improvements for analyzing complex visual patterns, almost reaching human performance in various tasks. The CNN is the most prominent DL technique for computer vision which is a particular set of supervised multi-layer perceptron architectures. CNN is biologically inspired by the local activations of the visual cortex [55]. Similarly to Riesz filterbanks, these local activations can be thought as a bank of filters that act on certain areas of the input

(*i.e.*, receptive fields) and due to the overlap between them this allow to find local correlation, via convolutions [56].

Formally, given an input vector \mathbf{x} (which can be the output of an earlier layer), the computation of an unit \mathbf{a} in a layer of the neural network is a non-linear weighted sum as

$$a(\mathbf{x}) = \sigma(W\mathbf{x}) = \sigma\left(\sum_{j=1}^M w_j x_j + b\right), \quad (7)$$

where W is the weight matrix of the network for that layer with dimensionality M and b is the bias term. Several activation functions $\sigma(\cdot)$ are proposed in the DL literature. Nevertheless, Rectified Linear Units (ReLU), where $\sigma(x) = \max(0, x)$, are consistently used in many applications because of their efficient gradient propagation avoiding vanishing or exploding gradients and also for their efficient computation as they only require a comparison. In CNNs, one is interested in learning small filters g that captures the spatial correlation in the input, formally the output of a convolution unit h_j is computed as

$$h_j(\mathbf{x}) = \sum_{i=1}^C (f_i * g_{ij})(\mathbf{x}), \quad (8)$$

where the convolution is computed in a $P \times Q$ input window of the original image as

$$(f_i * g_{ij}) = \sum_{p=1}^P \sum_{q=1}^Q f_i(p, q) g_{ij}(x_1 - p, x_2 - q). \quad (9)$$

Usually, these operations are efficiently vectorized to leverage the parallel capabilities of the Graphical Processing Units (GPUs). In contrast with the Riesz filter analysis, where the local rotation invariance property is explicitly hardcoded in the model, DL will learn approximate rotation-invariance with directionally insentitive filters and multiple oriented versions of directional filters in a model with millions of free parameters [57]. This allows learning higher-level patterns thanks to the non-linear hierarchical composition of low-level features [47], with the drawback of being more prone to overfitting if not correctly trained.

Training such large networks for medical tasks could be unfeasible due to the lack of annotated data to train the model. An interesting alternative is to use knowledge gained in another tasks where this amount of data is available.³ This is known as transfer learning, domain adaptation, or even fine-tuning [58]. Particularly, the use of pre-trained models to recognize objects in natural image settings could be helpful in many medical tasks because of the two following aspects. First, the layers and units in the network that recognize primitive features (*e.g.*, edges and textures) are shared across different visual contents. Second, reusing a pre-trained deep network sets the state of the optimization problem near to a local optima which is beneficial for both the performance and earlier training time convergence. Transfer learning has also shown to be useful for faster convergence in medical scenarios where lack of annotated data is common [59].

For characterizing the high-level patterns in the 5 classes of our ILD dataset, we propose to extract a deep learning representation for all the ILD patches from the GoogLeNet deep

learning architecture. The latter computea representations in a multi-scale fashion by reusing the outputs of the first layers to feed later ones as well as intermediate convolution modules, thus keeping the computational burden under control [60].

The principal feature of the GoogLeNet architecture is the inception module that computes $a(\mathbf{x})$ in each layer. This module use filters g of 3×3 and 5×5 pixels, which are then arranged and concatenated with the help of 1×1 convolutions to shrink the number of channels of the input (previous layer output). They are then fed into the next unit by concatenating channel-wise all the output filters [60]. For augmenting the invariance detection capabilities of the network, we augment the number of labeled samples per class by producing 5 label preserving patches per sample. Three of them were generated rotating the original patch by 90, 180 and 270 degrees, the other two by reflecting the patch along the x and y axis respectively.

E. Combining Riesz Filters and Deep CNNs

To exploit the complementary of the two distinct learned representations, two combination approaches were tested: late probability fusion and early feature vector fusion. For a fair comparison of all four configurations (*i.e.*, Riesz, DL, early and late fusion), a softmax classifier with an intermediate hidden layer was trained using the same parameters. Softmax classifiers have been proven to be useful when combining features from different sources in medical imaging [61].

For the deep learning representation, a 1024-dimensional feature vector was extracted from the layer with the largest area covered in the input image (*i.e.*, with the largest receptive field). This happens to be the layer preceding the classification one: the *pool5* layer that carries all the different learned patterns from the previous layers.

For the early fusion of feature vectors Riesz and DL vectors are simply concatenated. The resulting vector is then fed to a standard softmax classifier with weight decay regularization. The output of the classifier represents the probability for a patch to be classified as one of each of the 5 ILD classes. To train the weights Θ of the softmax model mapping the fused features to class probabilities, the following cost function is minimized with a stochastic gradient descent procedure as

$$C(\Theta) = -\frac{1}{M} \left[\sum_{i=1}^M \sum_{j=1}^K \mathbb{I}\{y^{(i)} = j\} * \log\left(\frac{\exp \Theta_j x^{(i)}}{\sum_{l=1}^K \exp \Theta_l x^{(i)}}\right) \right] + \frac{\rho}{2} \sum_{i=1}^K \sum_{j=1}^N \Theta_{ij}^2, \quad (10)$$

where M stands for the number of samples, N for the number of units, and K is the number of classes. ρ is the weight decay parameter that penalizes large values for parameters. The fused representation of an unseen test patch $x^{(p)} \in \mathbb{R}^{1287}$ is classified as class c by calculating a probability:

$$p(y^c = 1 | x^{(p)}; \Theta) = \frac{\exp(\Theta_1 x^{(p)})}{\sum_{l=1}^K \exp(\Theta_l x^{(p)})}. \quad (11)$$

A patch belongs to the c class if $p(y^c = 1 | x^{(p)}; \Theta) > t$, where t is a threshold deciding firm class membership. As

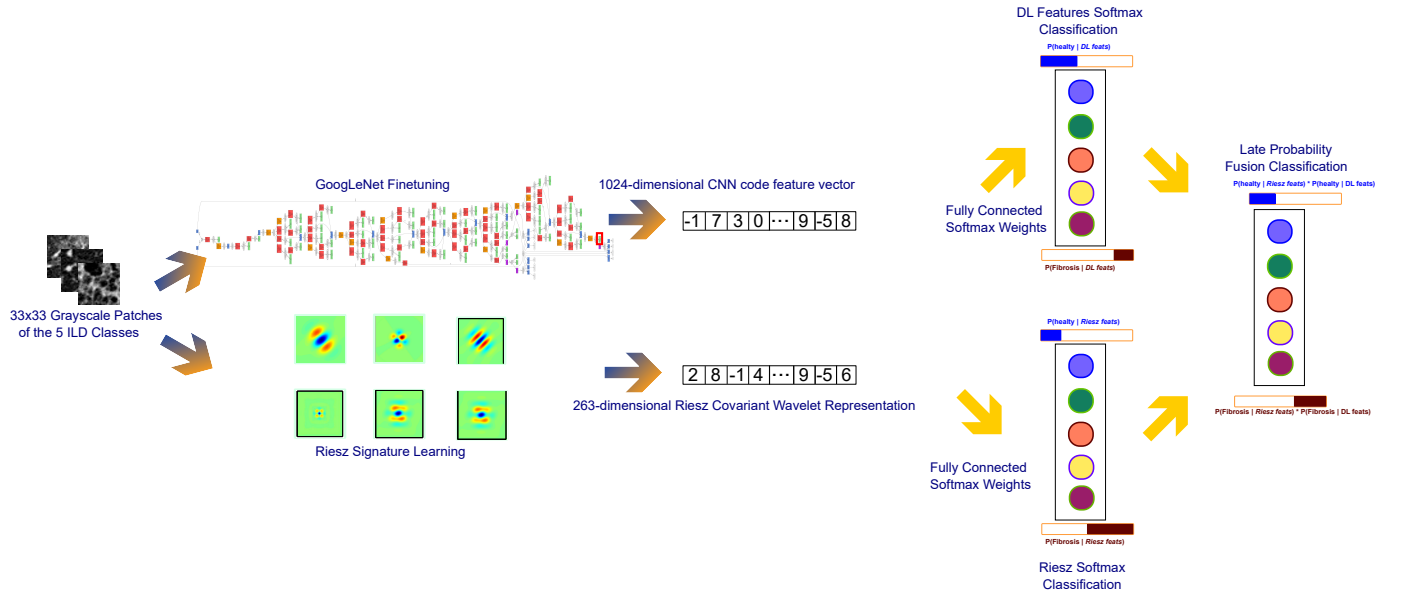


Fig. 5. Overall schema for the proposed feature combination approaches.

this varies across the folds, we report the average Receiver Operating Characteristic (ROC) curves and their respective Area Under the Curve (AUC) for each of the 5 classes and for each fold. The number of units N in the hidden layer was explored in the set $\{32, 64, 128\}$ with robust performance and we finally set it to 64 units for all experiments. The other parameters of the softmax classifier were the learning rate, decay and momentum that were set to 10^{-3} , 10^{-6} and 0.9, respectively. The accelerated gradient method of Nesterov was used as a parameter in the stochastic gradient descent optimizer.

Because the fused feature space is high-dimensional and the DL feature vector is approximately four times larger than the Riesz representation, the fused vector tends to reflect the performance of the DL classifier alone, leaving the complementary information out. To alleviate this, we performed a late fusion approach as follows. First, a single softmax classifier is trained for each representation. Then, the output probability vector of the two classifiers is multiplied element-wise to obtain a weighted probability vector to perform the final classification. The proposed combinations are depicted in Fig. 5.

F. Parallel Computing

Since the calculation of steered Riesz signatures implies iteration of all individual pixels in a patch, it is highly computationally expensive but embarrassingly parallel. On typical workstations, this step would take a prohibitively large amount of time. Using advances made in [62], we reduce the computation time by a factor of up to 30 times through a dedicated GPU-based implementation.

III. RESULTS

A. Experimental Results

The Caffe DL framework [63] was used to train and extract features from the GoogLeNet model. RGB replication of the

grayscale patches and a scaling from square patches of length 33 pixels to a length of 256 pixels was performed in order to be in line with the input layer of the architecture and use the pre-trained weights. The number of epochs was set to 30 but an early convergence up to the 5 epoch was achieved for all folds. The learning rate in all cases was initially set to 0.0001 and was decreased according to an exponential decay with $\sigma = 0.95$. We used the Keras² DL framework with the TensorFlow back-end for all the softmax models trained, using the hyper-parameters previously described in Section II-D.

MATLAB was used for the Riesz-related calculations. First, for each slice in the ILD database, square patches of length 33 pixels were extracted from the annotations present. A patch is defined as any 33×33 square region found to lie with at least 75% of it within the annotated region and the center of which is separated by at least half a patch length from the respective centers of other extracted patches. The patch size pixels was chosen after investigating patch sizes that yield good results and that generate a sufficient number of patches for training, validation, and testing. Fig. 7 illustrates the extraction of patches. In total, 14,594 patches are found in this manner and a breakdown of the set in terms of classes represented is provided in Table IV.

1) *Riesz Features Alone:* We first construct the feature vector representing each patch. The distribution of grayscale values between -1000 HU (value for air) and 650 HU (value for bone) in each patch is divided into 22 bins, which ensures a good balance of granularity in spatial-domain representation and low dimensionality. The number of air pixels in a patch is also noted. The energy of the filter response of each patch to each of the 10 weighted Riesz kernels at $J=4$ scales, which ensures sufficient coverage of spatial frequencies in the Fourier-domain, completes the feature space. In other words, for $N=5$, 10 texture signatures of length $(N+1) * J=24$ each

²<https://keras.io/>, as of February 2018.

TABLE I

THE CONFUSION MATRIX FOR THE HIGHEST CLASSIFICATION ACCURACY OBSERVED USING OUR METHOD IS SHOWN HERE.

True Label	Predicted Label				
	H	E	G	F	M
H	2186	18	21	19	767
E	48	233	7	59	60
G	124	2	1619	205	276
F	20	22	250	2441	229
M	441	9	152	144	5242

TABLE II

THE PER-CLASS SPECIFICITY, PRECISION, RECALL, F-SCORE, AND ACCURACY FOR THE EXPERIMENT YIELDING MAXIMUM OBSERVED CLASSIFICATION ACCURACY ARE SHOWN IN THE TABLE.

Class	Specificity	Precision	Recall	F-Score	Accuracy
H	0.9454	0.7755	0.7260	0.7499	0.9001
E	0.9964	0.8204	0.5725	0.6744	0.9846
G	0.9652	0.7901	0.7273	0.7574	0.9289
F	0.9633	0.8511	0.8241	0.8373	0.9350
M	0.8452	0.7974	0.8754	0.8346	0.8576

are extracted. Concatenating the $24 \cdot 10=240$ variables thus obtained with the 22 histogram bins and the air content value yields the final feature vector with 263 dimensions.

A Radial Basis Function (RBF) kernel SVM with 2 parameters (γ and cost C) is used to classify the feature vectors. γ controls the extent of influence of an individual training instance on the support vectors; high γ values correspond to a short-range influence. C determines the penalty for wrong classification; a high cost forces the algorithm to model the input more thoroughly, leading to an enhanced risk of overfit.

For $\gamma=1$ and $C=10^{-2}$, the highest accuracy value of 80.31% is observed. Table I details the confusion matrix of those results. Per-class specificity, precision, recall, f-score, and accuracy are shown in Table II.

2) *Deep CNN Features Alone*: The average accuracy for the four folds of the DL model was 77.1%. When using the weights of the model trained with the ImageNet dataset, the average accuracy slightly improved to 78.6%. Once the features were extracted the softmax model, we reached a fold-wise average AUC of 0.932.

3) *Combining Riesz and Deep CNN Representations*: The early fusion approach of concatenation of both DL and Riesz feature vectors yielded 78.1% average accuracy. An AUC performance of 0.931 was also noted, which is almost identical to the performance of the DL features alone. Because the fused feature space is high-dimensional and the DL feature vector is approximately four times larger than the Riesz representation, the fused vector tends to reflect the performance of the DL classifier alone, dismissing the complementary of the two representations. To alleviate this, we implemented a late fusion, which obtained the best AUC performance of 0.948 as depicted in Fig. 6 and this shows that it makes best use of both classifiers.

B. Statistical Significance of Performance Comparisons

To assess the statistical significance of the difference between the results of the classifiers, we computed the McNemar

TABLE III

p -VALUES FOR THE COMPARISON OF OUR FOUR APPROACHES

Comparison	p -value
DL vs Riesz	7.949e-14
(DL Riesz) early fusion vs (DL Riesz) late fusion	2.232e-85

TABLE IV

THE CLASSWISE DISTRIBUTION OF PATCHES EXTRACTED FROM THE ILD DATABASE IS SHOWN HERE.

Class	Number of patches
Healthy	3011
Emphysema	407
Ground Glass	2226
Fibrosis	2962
Micronodules	5988
Total	14594

test [64] under the null hypothesis of not having a significant difference between the classifier results, and the alternative hypothesis being the opposite, *i.e.*, the mean of their results are distinct enough and cannot be due to a random process.

We concatenated the class predictions for each classifier in each fold and computed the number of times that a certain classifier A has guessed the correct class and a certain classifier B did not. Then, we computed the same number after inverting the classifier predictions and these two sums were passed as parameters to the mid- p -test. If the p -value is less than 0.05, the results are considered to be statistically significant [65]. The results of the test are presented in Table III.

IV. DISCUSSIONS

The results obtained using our method are compared with the results obtained by other authors who used the same ILD dataset but with possibly slight variations in terms of the evaluation methodology. In fact, although the dataset used is the same, the exact validation scheme differs from one method to another according to selection of patches (percentage contained in the ROI), patch size, distribution of the classes, and cross-validation schemes. Notwithstanding, Table V details the accuracies for the different tissue types obtained by our method and reported by others.

A first observation is that when the dataset is carefully divided up in order to ensure that the same patient does not contribute patches to both the training and test sets (bias minimization), the performance of the same Riesz-based SVM classification method drops significantly on the whole (first and second rows of Table V). This proves beyond doubt that bias is present when the above separation step is not explicitly taken. Since many of the existing methods present no evidence of explicitly applying such a step, their reported performance values are at risk of being erroneously higher than they should be.

Table V shows that there is some room for improvement in the classification of the emphysema class. Indeed, only 407 patches with identified emphysema are encountered in the ILD database while the next least frequent disease class is ground glass with 2226 patches. That is a large disparity and we would

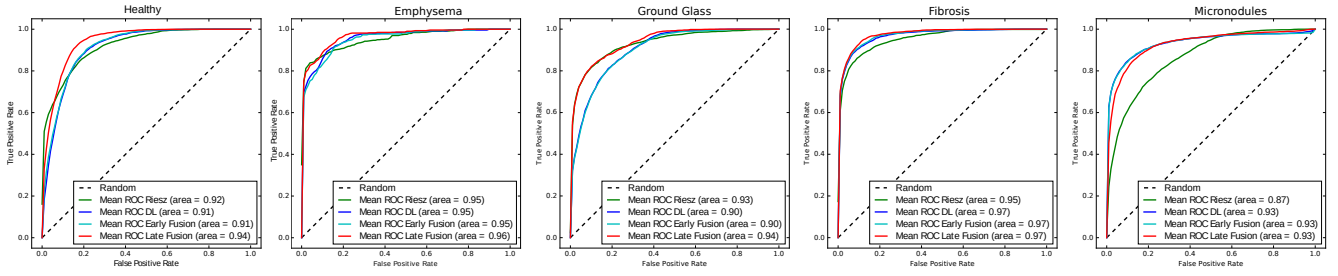


Fig. 6. ROC curves of the average performance in the 4 folds for each of the 5 classes using the late fusion approach depicted in Fig. 5. Healthy, Emphysema and Ground Glass classes are the ones that benefit the most from the fusion having a better AUC than the other approaches.

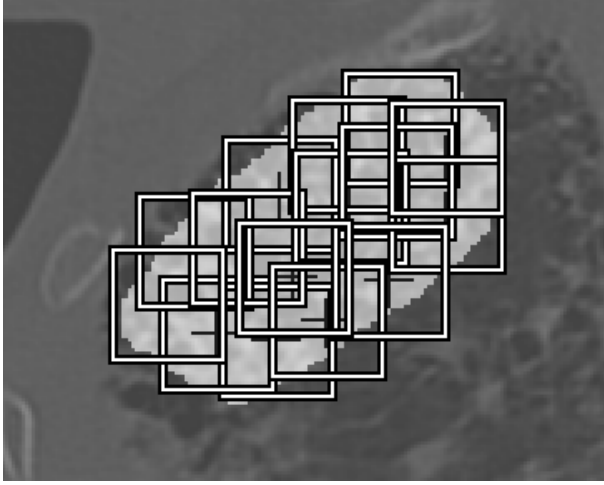


Fig. 7. Overlapping patches of size 33-by-33 pixels are extracted from the annotated slice that was shown in Figure 1.

TABLE V

THE DIAGONALS OF THE CONFUSION MATRICES OBTAINED FOR OUR METHODS VERSUS OTHERS IN THE LITERATURE ARE SHOWN IN PERCENT.

Method	H	E	Class G	F	M
Riesz (biased, SVM)	0.7260	0.5725	0.7273	0.8241	0.8754
Riesz (SVM)	0.6108	0.5553	0.7367	0.7606	0.7582
Riesz (softmax)	0.7556	0.3342	0.7071	0.8180	0.7258
DL only	0.4776	0.5455	0.7291	0.8467	0.8554
Early Fusion	0.4786	0.5160	0.7233	0.8549	0.8577
Late Fusion	0.6340	0.5430	0.7668	0.8805	0.8746
Song [45]	0.8760	0.8060	0.8270	0.8120	0.8110
Shin [29]	0.6800	0.9100	0.7000	0.8300	0.7900
Depeursinge [42]	0.6728	0.7872	0.7136	0.8273	0.8156
Depeursinge [44]	0.8270	0.7270	0.6840	0.8420	0.8350
Foncubierta [48]	0.0530	0.7450	0.4960	0.7460	0.519
Depeursinge [41]	0.7750	0.7330	0.7230	0.8450	0.8050
Li [46]	0.7600	0.6700	0.7000	0.7400	0.8400
Depeursinge [43]	0.7900	0.6920	0.5930	0.8050	0.7020
Gao [28]	0.9142	0.8270	0.8151	0.8910	0.8799
Song [47]	0.8850	0.7960	0.8000	0.8540	0.8720

argue that our learning approaches for emphysema are less well trained as compared to the other classes due to a much lower number of patches used for training. In addition, emphysema has very large intra-class variations and would require learning several steerable models or signatures per class. We contend that the use of more patches belonging to that class for training and the use of more than one signature for emphysema in subsequent work would significantly improve the overall classification accuracy.

We finally compared the classification accuracy of four different approaches: deep CNN features alone, Riesz features alone, early fusion of the feature vectors and late fusion of the class probability for each classifier. The comparison was made on the same basis using a softmax layer with the same hyperparameters. Furthermore, we assessed the statistical significance of the results of the classifiers by applying the McNemar test. The observed p -values are much lower than 0.05, demonstrating high statistical significance of the presented comparisons. Nevertheless, there is a bigger gap, thus a smaller p -value, between the early and late fusion classifiers. This could be due to the fact that when the early fusion of the feature vectors is performed, the representations are merged together in the intermediate layer of the softmax classifier, leading to more aligned representations than in the case of the late fusion, where we multiply both independent probabilities. On the other hand, for the separate feature classifiers, some of the learned features in the early layers of DL classifier likely resemble the filterbanks learnt using Riesz aligned texture signatures. This would explain the similar predictions in that particular case.

V. CONCLUSIONS

In this paper, we show that fusing learned tissue representations based on Riesz and Deep CNNs for texture characterization yields performance gains over each approach separately after bias minimization. We believe that further performance gains can be achieved by investigating new methods of fusing Riesz-based and DL-based texture learning approaches [66].

ACKNOWLEDGMENT

This work was supported by the Swiss National Science Foundation (under grant PZ00P2_154891) and partially

funded by the European Commission under the CP– Collaborative Project funding scheme through the FP7–ICT MD–PAEDIGREE project (ID: 600932).

REFERENCES

- [1] M. Kreuter, F. J. Herth, M. Wacker, R. Leidl, A. Hellmann, M. Pfeifer, J. Behr, S. Witt, D. Kauschka, M. Mall *et al.*, “Exploring clinical and epidemiological characteristics of interstitial lung diseases: Rationale, aims, and design of a nationwide prospective registrythe exciting-ild registry,” *BioMed research international*, vol. 2015, 2015.
- [2] R. Barr, D. Parr, and J. Vogel-Claussen, *Imaging*. European Respiratory Society, 2016.
- [3] T. Watadani, F. Sakai, T. Johkoh, S. Noma, M. Akira, K. Fujimoto, A. A. Bankier, K. S. Lee, N. L. Müller, J.-W. Song *et al.*, “Interobserver variability in the ct assessment of honeycombing in the lungs,” *Radiology*, vol. 266, no. 3, pp. 936–944, 2013.
- [4] Z. A. Aziz, A. U. Wells, D. M. Hansell, G. A. Bain, S. J. Copley, S. R. Desai, S. M. Ellis, F. V. Gleeson, S. Grubnic, A. G. Nicholson, S. P. Padley, K. S. Pointon, J. H. Reynolds, R. J. Robertson, and M. B. Rubens, “HRCT diagnosis of diffuse parenchymal lung disease: inter-observer variation,” *Thorax*, vol. 59, no. 6, pp. 506–511, Jun. 2004.
- [5] A. Depeursinge, A. Vargas, A. Platon, A. Geissbuhler, P.-A. Poletti, and H. Müller, “Building a reference multimedia database for interstitial lung diseases,” *Computerized Medical Imaging and Graphics*, vol. 36, no. 3, pp. 227–238, Apr. 2012.
- [6] C. Blakemore and F. W. Campbell, “On the existence of neurones in the human visual system selectively sensitive to the orientation and size of retinal images,” *The Journal of Physiology*, vol. 203, no. 1, pp. 237–260, 1969.
- [7] B. M. ter Haar Romeny, “Multi-scale and multi-orientation medical image analysis,” in *Biomedical Image Processing*. Springer, 2010, pp. 177–196.
- [8] A. Depeursinge, A. Foncubierta-Rodríguez, D. Van De Ville, and H. Müller, “Three-dimensional solid texture analysis and retrieval in biomedical imaging: review and opportunities,” *Medical Image Analysis*, vol. 18, no. 1, pp. 176–196, 2014.
- [9] A. Depeursinge, “Multi-Scale and Multi-Directional Biomedical Texture Analysis: Finding the Needle in the Haystack,” in *Biomedical Texture Analysis: Fundamentals, Applications and Tools*, ser. Elsevier-MICCAI Society Book series. Elsevier, 2017, pp. 29–53.
- [10] R. M. Haralick, K. Shanmugam, and I. Dinstein, “Textural features for image classification,” *IEEE Transactions on Systems, Man and Cybernetics*, vol. 3, no. 6, pp. 610–621, Nov. 1973.
- [11] N. Dalal and B. Triggs, “Histograms of oriented gradients for human detection,” in *Proceedings of the 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR’05)*, ser. CVPR ’05, vol. 1. Washington, DC, USA: IEEE Computer Society, 2005, pp. 886–893.
- [12] D. G. Lowe, “Distinctive image features from scale-invariant keypoints,” *International Journal of Computer Vision*, vol. 60, no. 2, pp. 91–110, 2004.
- [13] P. Jeng-Shyang and W. Jing-Wein, “Texture segmentation using separable and non-separable wavelet frames,” *IEICE transactions on fundamentals of electronics, communications and computer sciences*, vol. 82, no. 8, pp. 1463–1474, 1999.
- [14] D.-H. Xu, A. S. Kurani, J. Furst, and D. S. Raicu, “Run-length encoding for volumetric texture,” in *The 4th IASTED International Conference on Visualization, Imaging, and Image Processing – VIIP 2004*, Marbella, Spain, Sep. 2004.
- [15] O. G. Cula and K. J. Dana, “3D texture recognition using bidirectional feature histograms,” *International Journal of Computer Vision*, vol. 59, no. 1, pp. 33–60, 2004.
- [16] T. Leung and J. Malik, “Representing and recognizing the visual appearance of materials using three-dimensional textons,” *International Journal of Computer Vision*, vol. 43, no. 1, pp. 29–44, 2001.
- [17] R. Porter and N. Canagarajah, “Robust rotation-invariant texture classification: wavelet, Gabor filter and GMRF based schemes,” *IEE Proceedings on Vision, Image and Signal Processing*, vol. 144, no. 3, pp. 180–188, jun 1997.
- [18] T. Randen and J. H. Husoy, “Filtering for texture classification: a comparative study,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 21, no. 4, pp. 291–310, Apr. 1999.
- [19] Y. Xu, X. Yang, H. Ling, and H. Ji, “A new texture descriptor using multifractal analysis in multi-orientation wavelet pyramid,” in *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 2010, pp. 161–168.
- [20] S. G. Mallat, “A theory for multiresolution signal decomposition: the wavelet representation,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 11, no. 7, pp. 674–693, July 1989.
- [21] T. Ojala, M. Pietikäinen, and T. Mäenpää, “Multiresolution gray-scale and rotation invariant texture classification with local binary patterns,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 24, no. 7, pp. 971–987, Jul. 2002.
- [22] Y. LeCun, F. J. Huang, and L. Bottou, “Learning methods for generic object recognition with invariance to pose and lighting,” in *Computer Vision and Pattern Recognition, 2004. CVPR 2004. Proceedings of the 2004 IEEE Computer Society Conference on*, vol. 2. IEEE, 2004, pp. II–97.
- [23] Y. LeCun, K. Kavukcuoglu, and C. Farabet, “Convolutional networks and applications in vision,” in *IEEE International Symposium on Circuits and Systems (ISCAS)*, 2010, pp. 253–256.
- [24] A. Hyvärinen, P. Hoyer, and M. Inki, “Topographic independent component analysis,” *Neural computation*, vol. 13, no. 7, pp. 1527–1558, 2001.
- [25] M. J. Ablowitz, D. J. Kaup, and A. C. Newell, “The inverse scattering transform-fourier analysis for nonlinear problems,” *Studies in Applied Mathematics*, vol. 53, no. 4, pp. 249–315, 1974.
- [26] M. J. Ablowitz and H. Segur, *Solitons and the inverse scattering transform*. SIAM, 1981, vol. 4.
- [27] H. Greenspan, B. van Ginneken, and R. M. Summers, “Guest editorial deep learning in medical imaging: Overview and future promise of an exciting new technique,” *IEEE Transactions on Medical Imaging*, vol. 35, no. 5, pp. 1153–1159, 2016.
- [28] M. Gao, U. Bagci, L. Lu, A. Wu, M. Buty, H.-C. Shin, H. Roth, G. Z. Papadakis, A. Depeursinge, R. M. Summers *et al.*, “Holistic classification of ct attenuation patterns for interstitial lung diseases via deep convolutional neural networks,” *Computer Methods in Biomechanics and Biomedical Engineering: Imaging & Visualization*, pp. 1–6, 2016.
- [29] H.-C. Shin, H. R. Roth, M. Gao, L. Lu, Z. Xu, I. Nogues, J. Yao, D. Mollura, and R. M. Summers, “Deep convolutional neural networks for computer-aided detection: Cnn architectures, dataset characteristics and transfer learning,” *IEEE transactions on medical imaging*, vol. 35, no. 5, pp. 1285–1298, 2016.
- [30] M. Anthimopoulos, S. Christodoulidis, L. Ebner, A. Christe, and S. Mougiakakou, “Lung pattern classification for interstitial lung diseases using a deep convolutional neural network,” *IEEE transactions on medical imaging*, vol. 35, no. 5, pp. 1207–1216, 2016.
- [31] Q. Li, W. Cai, X. Wang, Y. Zhou, D. D. Feng, and M. Chen, “Medical image classification with convolutional neural network,” in *Control Automation Robotics & Vision (ICARCV), 2014 13th International Conference on*. IEEE, 2014, pp. 844–848.
- [32] C. Schmid, “Constructing models for content-based image retrieval,” in *IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, ser. CVPR, vol. 2, 2001, pp. 39–45.
- [33] A. Depeursinge, J. Fageot, and O. S. Al-Kadi, “Fundamentals of Texture Processing for Biomedical Image Analysis: A General Definition and Problem Formulation,” in *Biomedical Texture Analysis: Fundamentals, Applications and Tools*, ser. Elsevier-MICCAI Society Book series. Elsevier, 2017, pp. 1–27.
- [34] S. Lazebnik, C. Schmid, and J. Ponce, “A sparse texture representation using local affine regions,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 27, no. 8, pp. 1265–1278, August 2005.
- [35] A. Depeursinge, Z. Püspöki, J.-P. Ward, and M. Unser, “Steerable Wavelet Machines (SWM): Learning Moving Frames for Texture Classification,” *IEEE Transactions on Image Processing*, vol. 26, no. 4, pp. 1626–1636, 2017.
- [36] M. N. Do and M. Vetterli, “Rotation invariant texture characterization and retrieval using steerable wavelet-domain hidden markov models,” *IEEE Transactions on Multimedia*, vol. 4, no. 4, pp. 517–527, Dec. 2002.
- [37] A. Depeursinge, A. Foncubierta-Rodríguez, D. Van De Ville, and H. Müller, “Rotation-covariant texture learning using steerable Riesz wavelets,” *IEEE Transactions on Image Processing*, vol. 23, no. 2, pp. 898–908, Feb. 2014.
- [38] R. Joyseere, H. Müller, and A. Depeursinge, “Rotation-covariant tissue analysis for interstitial lung diseases using learned steerable filters: Performance evaluation and relevance for diagnostic aid,” *Computerized Medical Imaging and Graphics*, vol. 64, pp. 1–11, 2018.
- [39] A. Depeursinge, D. Sage, A. Hidki, A. Platon, P.-A. Poletti, M. Unser, and H. Müller, “Lung tissue classification using Wavelet frames,” in *29th*

- Annual International Conference of the IEEE Engineering in Medicine and Biology Society, 2007. EMBS 2007.* Lyon, France: IEEE Computer Society, 2007, pp. 6259–6262.
- [40] A. Depeursinge, J. Iavindrasana, G. Cohen, A. Platon, P.-A. Poletti, and H. Müller, “Lung tissue classification in HRCT data integrating the clinical context,” in *21th IEEE Symposium on Computer-Based Medical Systems (CBMS)*, Jyväskylä, Finland, June 2008, pp. 542–547.
- [41] A. Depeursinge, A. Foncubierta-Rodríguez, D. Van De Ville, and H. Müller, “Lung texture classification using locally-oriented Riesz components,” in *Medical Image Computing and Computer Assisted Intervention – MICCAI 2011*, ser. Lecture Notes in Computer Science, G. Fichtinger, A. Martel, and T. Peters, Eds., vol. 6893. Springer Berlin / Heidelberg, Sep. 2011, pp. 231–238.
- [42] A. Depeursinge, D. Van De Ville, A. Platon, A. Geissbuhler, P.-A. Poletti, and H. Müller, “Near-affine-invariant texture learning for lung tissue analysis using isotropic wavelet frames,” *IEEE Transactions on Information Technology in BioMedicine*, vol. 16, no. 4, pp. 665–675, Jul. 2012.
- [43] A. Depeursinge, T. Zrimec, S. Busayarat, and H. Müller, “3D lung image retrieval using localized features,” in *Medical Imaging 2011: Computer-Aided Diagnosis*, vol. 7963. SPIE, Feb. 2011, p. 79632E.
- [44] A. Depeursinge, A. Foncubierta-Rodríguez, D. Van De Ville, and H. Müller, “Multiscale lung texture signature learning using the Riesz transform,” in *Medical Image Computing and Computer-Assisted Intervention MICCAI 2012*, ser. Lecture Notes in Computer Science, vol. 7512. Springer Berlin / Heidelberg, Oct. 2012, pp. 517–524.
- [45] Y. Song, W. Cai, Y. Zhou, and D. D. Feng, “Feature-based image patch approximation for lung tissue classification,” *IEEE transactions on medical imaging*, vol. 32, no. 4, pp. 797–808, 2013.
- [46] Q. Li, W. Cai, and D. D. Feng, “Lung image patch classification with automatic feature learning,” in *Engineering in Medicine and Biology Society (EMBC), 2013 35th Annual International Conference of the IEEE.* IEEE, 2013, pp. 6079–6082.
- [47] Y. Song, W. Cai, H. Huang, Y. Zhou, Y. Wang, and D. D. Feng, “Locality-constrained subcluster representation ensemble for lung image classification,” *Medical image analysis*, vol. 22, no. 1, pp. 102–113, 2015.
- [48] A. Foncubierta-Rodríguez, A. Depeursinge, and H. Müller, “Using multiscale visual words for lung texture classification and retrieval,” in *Medical Content-based Retrieval for Clinical Decision Support*, ser. MCBR-CDS 2011, H. Greenspan, H. Müller, and T. Syeda Mahmood, Eds., vol. 7075. Lecture Notes in Computer Sciences (LNCS), Sep. 2012, pp. 69–79.
- [49] M. Unser, N. Chenouard, and D. Van De Ville, “Steerable pyramids and tight wavelet frames in $L_2(\mathbb{R}^d)$,” *IEEE Transactions on Image Processing*, vol. 20, no. 10, pp. 2705–2721, Oct. 2011.
- [50] M. Unser and D. Van De Ville, “Wavelet steerability and the higher-order Riesz transform,” *IEEE Transactions on Image Processing*, vol. 19, no. 3, pp. 636–652, March 2010.
- [51] W. T. Freeman and E. H. Adelson, “The design and use of steerable filters,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 13, no. 9, pp. 891–906, September 1991.
- [52] E. P. Simoncelli and W. T. Freeman, “The steerable pyramid: a flexible architecture for multi-scale derivative computation,” in *Proceedings of International Conference on Image Processing, 1995.*, vol. 3, October 1995, pp. 444–447.
- [53] V. N. Vapnik, *The Nature of Statistical Learning Theory*. New York: Springer, November 1995.
- [54] I. Guyon, J. Weston, S. Barnhill, and V. Vapnik, “Gene selection for cancer classification using support vector machines,” *Machine Learning*, vol. 46, no. 1–3, pp. 389–422, 2002.
- [55] Y. LeCun, Y. Bengio, and G. Hinton, “Deep learning,” *Nature*, vol. 521, no. 7553, pp. 436–444, 2015.
- [56] A. Depeursinge, O. S. Al-Kadi, and J. R. Mitchell, *Biomedical Texture Analysis: Fundamentals, Applications and Tools*, ser. Elsevier-MICCAI Society Book series. Elsevier, Oct. 2017. [Online]. Available: <https://www.elsevier.com/books/title/author/9780128121337>
- [57] D. M. Gonzalez, M. Volpi, and D. Tuia, “Learning rotation invariant convolutional filters for texture classification,” *CoRR*, vol. abs/1604.0, 2016.
- [58] S. M. Rozantsev, A. and P. Fua, “Beyond Sharing Weights for Deep Domain Adaptation,” *ArXiv e-prints*, Mar. 2016.
- [59] A. Janowczyk and A. Madabhushi, “Deep learning for digital pathology image analysis: A comprehensive tutorial with selected use cases,” *Journal of Pathology Informatics*, vol. 7, 2016.
- [60] C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov, D. Erhan, V. Vanhoucke, and A. Rabinovich, “Going deeper with convolutions,” in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2015, pp. 1–9.
- [61] S. Otálor, A. Cruz Roa, J. Arevalo, M. Atzori, A. Madabhushi, A. Judkins, F. González, H. Müller, and A. Depeursinge, “Combining unsupervised feature learning and riesz wavelets for histopathology image representation: Application to identifying anaplastic medulloblastoma,” in *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015*, ser. Lecture Notes in Computer Science, N. Navab, J. Hornegger, W. M. Wells, and A. Frangi, Eds. Springer International Publishing, Oct. 2015, vol. 9349, pp. 581–588.
- [62] A. Vizitiu, L. Itu, R. Joyseere, A. Depeursinge, H. Müller, and C. Suciu, “Gpu-accelerated texture analysis using steerable riesz wavelets,” in *24th IEEE Euromicro International Conference on Parallel, Distributed, and Network-Based Processing*, ser. PDP2016, 2016.
- [63] Y. Jia, E. Shelhamer, J. Donahue, S. Karayev, J. Long, R. B. Girshick, S. Guadarrama, and T. Darrell, “Caffe: Convolutional architecture for fast feature embedding,” in *Proceedings of the ACM International Conference on Multimedia, MM’14*, 2014, pp. 675–678.
- [64] T. G. Dietterich, “Approximate statistical test for comparing supervised classification learning algorithms,” *Neural Computation*, vol. 10, no. 7, pp. 1895–1923, Oct. 1998.
- [65] M. W. Fagerland, S. Lydersen, and P. Laake, “The mcnemar test for binary matched-pairs data: mid-p and asymptotic are better than exact conditional,” *BMC medical research methodology*, vol. 13, no. 1, p. 91, 2013.
- [66] A. Depeursinge and H. Müller, “Fusion techniques for combining textual and visual information retrieval,” in *ImageCLEF*, ser. The Springer International Series On Information Retrieval, H. Müller, P. Clough, T. Deselaers, and B. Caputo, Eds. Springer Berlin Heidelberg, 2010, vol. 32, pp. 95–114.