Fusing Learned Representations from Riesz Filters and Deep CNN for Lung Tissue Classification

Ranveer Joyseeree[a,*], Sebastian Otálora[b,c], Henning Müller[b], Adrien Depeursinge[b]

[a] *ETH Zürich, Zürich, Rämistrasse 101, 8092 Zurich, Switzerland*
[b] *HES-SO Valais, Technopôle 3, 3960 Sierre, Switzerland*
[c] *University of Geneva, Geneva, Switzerland*

Abstract

A novel method to detect and classify several classes of diseased and healthy lung tissue in CT (Computed Tomography), based on the fusion of Riesz and deep learning features, is presented. First, discriminative parametric lung tissue texture signatures are learned from Riesz representations using a one–versus–one approach. The signatures are generated for four diseased tissue types and a healthy tissue class, all of which frequently appear in the publicly available Interstitial Lung Diseases (ILD) dataset used in this article. 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 Inception V3 architecture using an augmented version of the same ILD dataset. Because CNN features are both deep and non–parametric, they can accurately model virtually any pattern that is useful for tissue discrimination, and they are the

*Corresponding author
Email address: `ranveer@icloud.com` (Ranveer Joyseeree)

de facto standard for many medical imaging tasks. However, invariance to local image rotations is not explicitly implemented and can only be approximated with rotation–based data augmentation. This motivates the fusion of Riesz and deep CNN features, as the two techniques are very complementary. 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 and also compared to an ensemble of deep learning approaches.

Keywords: texture signatures, classification, ILD, Deep Learning

1. Introduction

The White Book of the European Respiratory Society (ERS) 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 annual cost of 380 billion Euros was associated with them in the European Union alone, and this figure was estimated by taking into account the loss of productive output, and the costs of direct medical care and drugs[1]. Battling these diseases is thus a priority in the healthcare domain. To combat avoidable deaths and significant costs, obtaining an early accurate diagnosis is essential. In such a scenario, clinicians may prescribe the correct treatment as early as possible and thus limit disease progression.

[1] European lung white book, http://www.erswhitebook.org/chapters/the-burden-of-lung-disease/, as of April 2018.

Respiratory ailments affecting the lung parenchyma are prevalent. One of the largest and most diverse groups of such diseases is the set of Interstitial Lung Diseases (ILDs). They account for more than 200 pathologies affecting the alveoli, the small lung airways, and the pulmonary interstitium (Kreuter et al., 2015). Information gathered from clinical, radiological, and pathological analyses are required to diagnose them accurately. In particular, High-Resolution Computed Tomography (HRCT) images are the radiological modality of choice for their characterization (Barr et al., 2016). Some of the ailments may easily be misdiagnosed due to their rarity and to the fact that radiologists are subjective when interpreting the content of the images (Aziz et al., 2004; Watadani et al., 2013). Therefore, computerized assistance yielding exhaustive and reproducible image analysis has been mentioned several times as beneficial for improving ILD management (Depeursinge et al., 2012c).

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 (Blakemore and Campbell, 1969; ter Haar Romeny, 2010), including local variations of pattern properties such as local anisotropy (Depeursinge et al., 2014b; Depeursinge, 2017). Spatial domain representations of images alone provide insufficient information to examine the local organization of scales and directions properly. Therefore, to obtain a complete overview of the relationships between them, intensity information needs to be complemented with information extracted in the frequency domain.

Several authors exploit information embedded in the local organization of scales and directions in images for pattern characterization and recognition. Grey–Level Co–occurrence Matrices (GLCM) (Haralick et al., 1973), Histograms Of Gradients (HOG) (Dalal and Triggs, 2005) used in the Scale–Invariant Feature Transform (SIFT) (Lowe, 2004), non–separable and separable wavelets (Jeng-Shyang and Jing-Wein, 1999), Run–Length Encoding (RLE) (Xu et al., 2004), and oriented filterbanks and wavelets (Gaussian, Gabor, Leung–Malik, Maximum Response (Cula and Dana, 2004; Leung and Malik, 2001; Porter and Canagarajah, 1997; Randen and Husoy, 1999; Xu et al., 2010) have been proposed for directional analysis. Unfortunately, separable wavelets suffer from bias along the vertical and horizontal axes (Mallat, 1989), while the remainder requires an arbitrary choice of image directions (Depeursinge et al., 2014b). Using a sequence of pixels along perimeters of radius r, Local Binary Patterns (LBP) (Ojala et al., 2002) can perform multi–directional analysis but they do not allow for multiresolution analysis, easily. 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) (LeCun et al., 2004, 2010), Topographic Independent Component Analysis (TICA) (Hyvärinen et al., 2001) and the scattering transform (Ablowitz et al., 1974; Ablowitz and Segur, 1981). 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 compared with most classic handcrafted feature approaches. In recent years, these techniques have been successfully in many medical–

domain tasks showing promising results and opening different research avenues (Greenspan et al., 2016; Litjens et al., 2017), particularly applied to ILD as discussed later in the text below.

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 (Anthimopoulos et al., 2016; Gao et al., 2016; Shin et al., 2016). For instance, CNNs have been used for lung disease classification in (Gao et al., 2016), where performance was not assessed explicitly and in (Li et al., 2014) 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. Features that are locally rotation–invariant are desirable in such instances (Depeursinge et al., 2017b; Schmid, 2001). LBPs (Ojala et al., 2002) and Rotation–covariant SIFT (RIFT) (Lazebnik et al., 2005) 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 (Depeursinge et al., 2017c; Do and Vetterli, 2002).

Learned representations based on Riesz wavelets (Depeursinge et al., 2014a), as used in this work, can precisely model multi–scale and multi–directional information that is important for tissue discrimination (Joyseeree et al., 2018). Also, the obtained representation can easily be made invariant to local rotations using the steerability of the models (see Section 2.2). One drawback of Riesz representations is its reliance on parametric basis functions with a 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 with several rotations, but usually, this is hard to do for small degrees of rotation. Additionally, CNN kernels do not rely on parametric representations, and a large number of learned filters can model virtually any pattern relevant to discrimination, under the condition that the training dataset is large enough. The complementarity of the two approaches motivates the fusion of the two representations into a single model, which is also the main contribution of this article.

Five tissue types are often classified by the automatic methods found in the literature using a publicly-available ILD dataset (Depeursinge et al., 2012c), as they have a more significant number of annotated regions than other patterns: healthy, emphysema, ground glass, fibrosis, and micronodules. Some of the earliest of these papers (Depeursinge et al., 2007, 2008) combined image data with clinical parameters to carry out classification. This was followed by handcrafted steerable Riesz filterbanks (Depeursinge et al., 2011a), low–level localized features (Depeursinge et al., 2011b) and isotropic wavelet frames (Depeursinge et al., 2012b). After that, learned representations based on the Riesz transform (Depeursinge et al., 2012a) were utilized. Recently, in (Joyseeree et al., 2018) a rotation-covariant approach learning a class–wise texture signature using Riesz wavelets were proposed. Here we build upon this work by complementing the features extracted using a deep–learning network to extract high–level patterns not captured by texture signatures. Other authors working on the same data include Song *et al.* (Song et al., 2013) who first employed feature–based image patch approximation. Li *et al.* (Li et al., 2013) then used automatic feature learning followed by a customized CNN approach in (Li et al., 2014), while in (Song et al., 2015), a locality–constrained subcluster representation ensemble is used. Gao *et al.* use a deep CNN approach in (Gao et al., 2016).

The following publications use a different set of tissue types. A few classify six tissue types by including the consolidation type. Examples of such instances include Foncubierta *et al.* (Foncubierta-Rodríguez et al., 2012) who used multi–scale visual words for classification and retrieval. Shin *et al.* (Shin et al., 2016) used deep CNNs. Others used a significantly different set of classes. For example, Anthimopoulous *et al.* (Anthimopoulos et al., 2016) 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 the best of our knowledge, there are not work in the literature exploiting the joint discriminative power of rotation invariant and deep learning representations for ILD classification. 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 2, we first present the publicly available dataset used for the validation of our approach. This is followed by an in–depth description of the theoretical aspects of the proposed method. Section 3 presents the evaluation of classification performance, observed results, and statistical significance of the performance comparison. A thorough analysis and interpretation of the observed behavior is carried out in Section 4. Finally, conclusions are drawn based on the work done, and we propose measures to improve performance in future work in Section 5.

2. Materials and Methods

2.1. Dataset and Validation Schemes

We use the most frequently used publicly–available ILD dataset (Depeursinge et al., 2012c) 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 was annotated by an expert radiologist. The data set was used several times in past publications and these past approaches on the same data set will be briefly covered in this section.

To facilitate the comparison of our work with the majority of other techniques used 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 the ILDs and are therefore relatively well represented regarding the 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 subtle, especially when comparing the healthy, emphysema, and micronodules classes.

For 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 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 of each other, in line with what is often encountered in the literature, and are divided into four equal sets. In the second strategy, we ensure that the patches originating from an individual patient only appear in one of the four groups to minimize the risk of inherent bias. Two of the four groups are then concatenated and used for training the classification model. One of the remaining sets 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 possibility, therefore, to consider full 3D image analysis, which might lead to better results if the data are available.

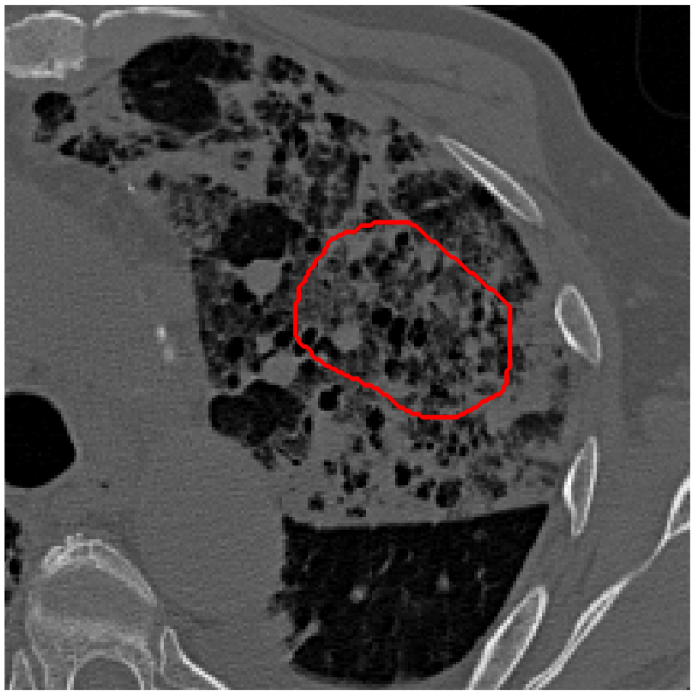


Figure 1: Part of a slice taken from the ILD database that represents the right lung along with an expert annotation (red delineation) corresponding to a Region Of Interest (ROI) are presented here.

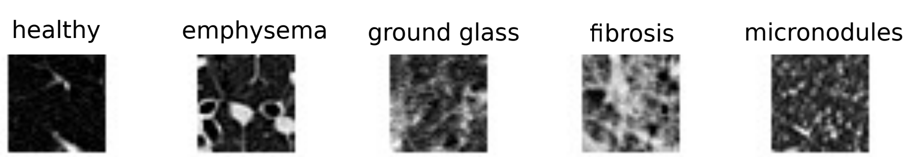


Figure 2: The five tissue classes selected for our work representing healthy parenchyma as well as emphysema, ground glass, fibrosis, and micronodules.

2.2. 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 (Depeursinge et al., 2014a; Joyseeree et al., 2018). We first introduce the Riesz transform and its combination with radial wavelets to derive steerable filterbanks in Section 2.2.1. Second, we describe in Section 2.2.2 how we learn one–versus–one class–specific discriminative texture signatures from the parametric Riesz representation using Support Vector Machines (SVM).

2.2.1. 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 $f(\boldsymbol{x})$ represent the function that models the content of a patch where $\boldsymbol{x}$ represents pixel coordinates x_1 and x_2. In other words, $f\colon \boldsymbol{x} \rightarrow f(\boldsymbol{x}), \boldsymbol{x} \in \mathbb{R}^2$, where $\boldsymbol{x} = (x_1, x_2)$.

Since the 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 $f(\boldsymbol{x})$ is defined as

$$f(\boldsymbol{x}) \xleftrightarrow{\mathcal{F}} \hat{f}(\boldsymbol{\omega}) = \int_{\mathbb{R}^2} f(\boldsymbol{x}) e^{-j\langle \boldsymbol{\omega}, \boldsymbol{x} \rangle} \, \mathrm{d}x_1 \mathrm{d}x_2, \tag{1}$$

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 (Unser et al., 2011). The $N + 1$ components of the Nth–order

Riesz transform are based on the collection of operators $\mathcal{R}^{(n,N-n)}\{\cdot\}$ as

$$\boldsymbol{\mathcal{R}}^{N}\left\{f\right\}(\boldsymbol{x})=\begin{pmatrix}\mathcal{R}^{(0,N)}\left\{f\right\}(\boldsymbol{x})\\ \vdots\\ \mathcal{R}^{(n,N-n)}\left\{f\right\}(\boldsymbol{x})\\ \vdots\\ \mathcal{R}^{(N,0)}\left\{f\right\}(\boldsymbol{x})\end{pmatrix}, \tag{2}$$

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

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

where

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

According to Eq. (3), the product of $\mathrm{j}\omega_1$ and/or $\mathrm{j}\omega_2$ in the numerator followed by division with the norm of $\boldsymbol{\omega}$ produces allpass filters that only retain phase information that characterizes directions (Depeursinge et al., 2014b; Unser and Van De Ville, 2010) 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 $N=1,\ldots,5$.

We also seek the steerability property of Riesz filterbanks (Freeman and Adelson, 1991; Unser and Van De Ville, 2010). In essence, this implies that a linear combination of the filterbanks may model any local rotation. When

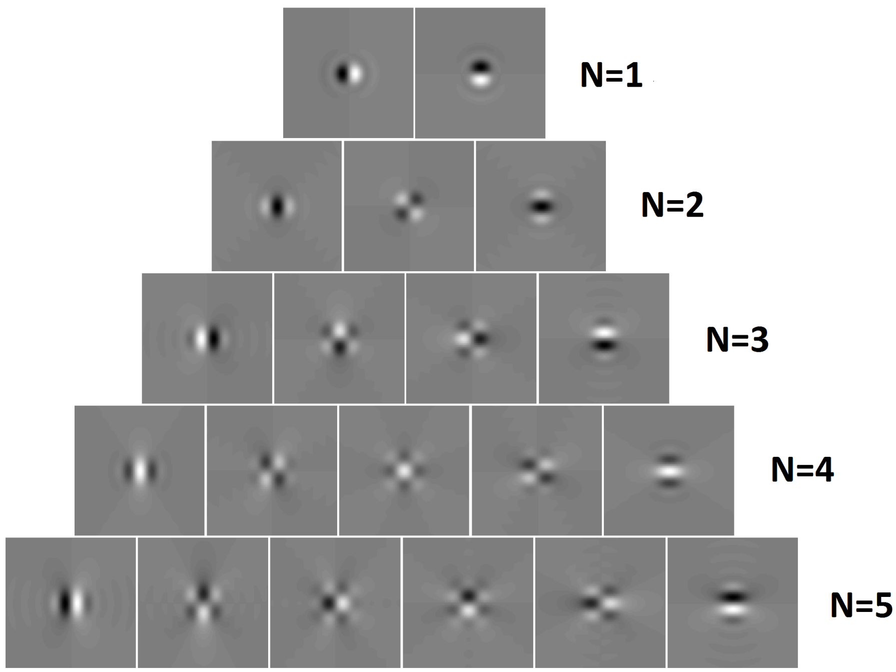


Figure 3: Riesz filterbanks for orders up to 5 are shown here. To represent the filters on a finite spatial support, the Riesz transform was applied to an isotropic Gaussian function.

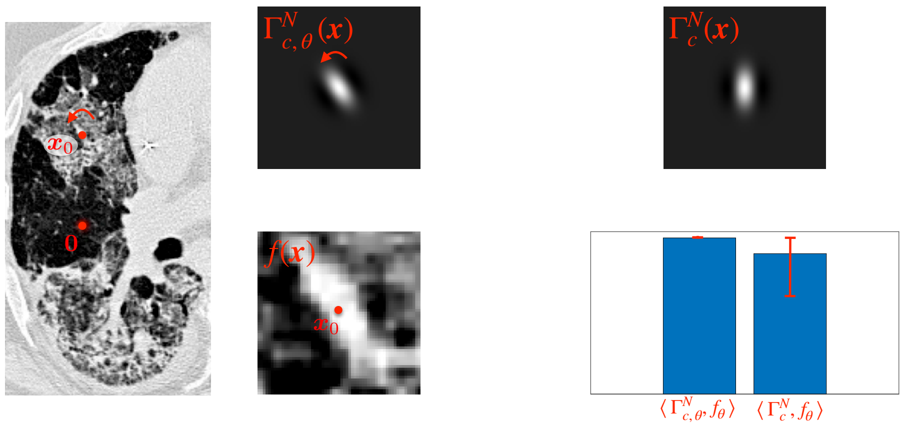


Figure 4: Steering the texture signatures Γ^N_c with an angle θ allows reducing the variability in feature values when compared to linear filtering used by e.g. CNNs. This example demonstrates that the responses of the unsteered signature (i.e., $\langle \Gamma^N_c, f_\theta \rangle$) on a patch rotated with θ varies strongly, which creates noise in the feature representation. However, the response of the steered signature (i.e. $\langle \Gamma^N_{c,\theta}, f_\theta \rangle$) is invariant to rotations of f.

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 $f(\boldsymbol{x})$ for a rotation around $\mathbf{0}$ as

$$\boldsymbol{\mathcal{R}}^N \{f_\theta\}(\mathbf{0}) = \mathrm{A}_\theta \boldsymbol{\mathcal{R}}^N \{f\}(\mathbf{0}), \tag{4}$$

where f_θ denotes the rotation of f as $f_\theta(\boldsymbol{x}) = f(\mathrm{R}_\theta \boldsymbol{x})$ and R_θ is a 2D rotation matrix. The use of steerability to reduce variability in feature values caused by rotations of the input patches is illustrated in Fig. 4.

For multi–scale analysis, the Fourier domain is partitioned using wavelets

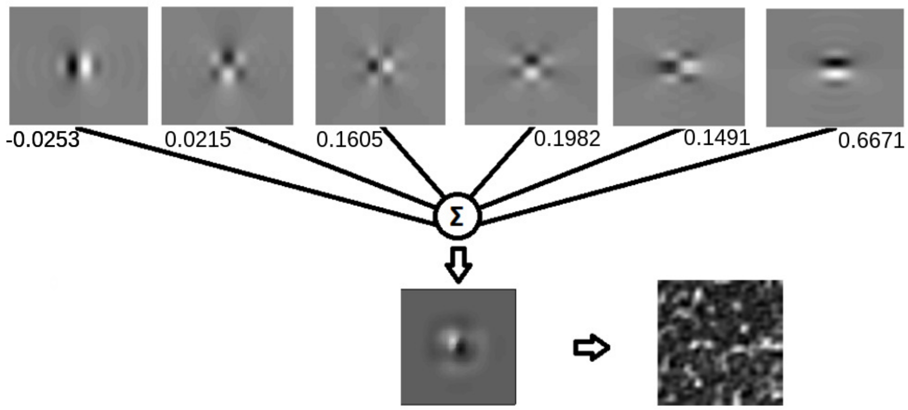


Figure 5: A texture signature for the micronodule class is built after applying an appropriate weighing scheme for the components of a Riesz filterbank of order 5.

into several progressive dyadic bands of decreasing sizes based on Simoncelli's isotropic multiresolution framework (Simoncelli and Freeman, 1995). The bands control the spatial support or scale of the (allpass) Riesz kernels.

2.2.2. Parametric Discriminative Texture Signatures

Learned representations based on class–specific steerable texture signatures are obtained by finding a weighting scheme for the Riesz filterbanks at each scale. Fig. 5 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

$$\begin{aligned}\Gamma_c^N &= \boldsymbol{w}^{\mathrm{T}}\boldsymbol{\mathcal{R}}^N \\ &= w_1\mathcal{R}^{(0,N)} + w_2\mathcal{R}^{(1,N-1)} + \cdots + w_{N+1}\mathcal{R}^{(N,0)},\end{aligned} \tag{5}$$

where $\boldsymbol{w}$ contains the weights of the respective Riesz kernels. A multi–scale texture signature is obtained by extending Eq. (5) using multi–scale Riesz filterbanks (Depeursinge et al., 2012a) as

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

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 $E\left(\mathcal{R}^{(n,N-n)}\{f\}(\boldsymbol{x})\right)$ are computed and regrouped for each class c versus each one from the remaining classes. SVMs then find the optimal separation in terms of minimized structural risk (Guyon et al., 2002; Vapnik, 1995). Each class benefits from a unique characterization with regard to each of the remaining classes. With five classes, the approach is expected to lead to $5 \cdot 4{=}20$ separate optimal signatures. However, because of 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 reduce to 10. The optimal weights $\boldsymbol{w} = (w_1, \dots, w_{N+1})$ are directly determined from the support vectors of the optimal separations (Depeursinge et al., 2014b).

2.2.3. 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 (for example in the case of emphysema). 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 $(N+1)*J$=24 each are obtained. To create the final input feature vector, the $24 \cdot 10$=240 variables obtained in the previous step are concatenated with the 22 variables from the histogram and also with one last variable for air content. As a result, a feature vector of length 263 is obtained for each patch. Once the final Riesz feature vector is built, we train a Softmax classifier that maps from the 263 Riesz feature vector to the 5 ILD classes. The softmax classifier uses the same hyperparameters for the training of the DL feature classification alone, in order to evaluate the discriminatory power of the features itself. The details of the softmax classifier are discussed in section 2.4.

2.3. Tissue Characterization With Deep CNNs

DL has shown significant improvements for analyzing complex visual patterns, reaching human performance in various tasks. The CNN is the most prominent DL technique for computer vision. A CNN is a particular set of supervised multi–layer perceptron architectures. CNNs are biologically inspired by the local activations of the visual cortex (LeCun et al., 2015). Similarly to Riesz filterbanks, these local activations can be thought of like a bank of filters that act on certain areas of the input (*i.e.*, receptive fields). Due to overlap, one may find local correlation via convolutions (Depeursinge et al., 2017a).

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

$$a(\boldsymbol{x}) = \sigma(\mathrm{W}\boldsymbol{x}) = \sigma(\sum_{j=1}^{M} w_j x_j + b), \tag{7}$$

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. Rectified Linear Units (ReLUs), where $\sigma(x) = \max(0, x)$, are consistently used in many applications because of their efficient gradient propagation that avoids 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 capture the spatial correlation in the input. Formally, the output of a convolution unit h_j is computed as

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

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

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

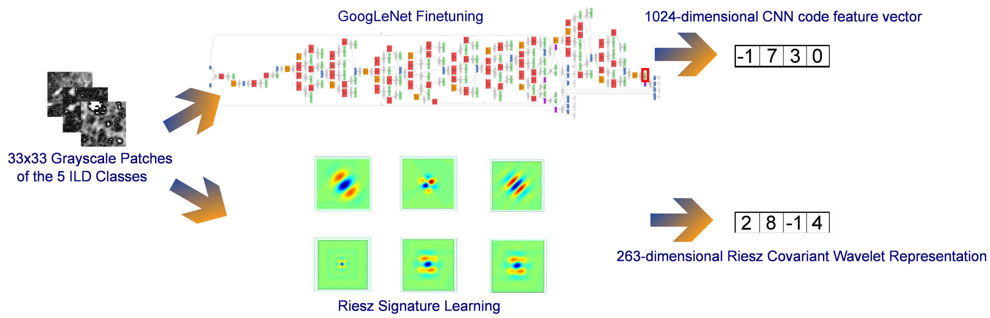
These matrix operations are efficiently vectorized to leverage the parallel capabilities of the Graphical Processing Units (GPUs). In contrast with the Riesz filter analysis, where the local rotation invariance is explicitly hard-coded in the model, DL learns relative rotation–invariance with directionally insensitive filters and multiple oriented versions of directional filters in a model with millions of free parameters (Gonzalez et al., 2016). This permits learning higher–level patterns thanks to the non–linear hierarchical composition of low–level features (Song et al., 2015) at the cost of being more prone to overfitting if the model is not regularized accordingly.

Training such large networks for medical tasks can be unfeasible due to the lack of annotated data to train the model. An exciting alternative is to use knowledge gained in other tasks where a large amount of data is available. This is known as transfer learning, where a model that was initially trained using a large amount of labeled data is then fine–tuned (Rozantsev and Fua, 2016) to a new dataset where a fewer annotated samples are available, thus leveraging the filters learned in the first dataset to serve as a starting point to learn the optimal filters in the new dataset. Notably, the use of pre–trained models to recognize objects in natural image settings could be helpful in many medical tasks because of the following two 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 a local optimum which is beneficial for both the performance and earlier training convergence. Transfer learning has also shown to be useful for faster convergence in medical scenarios where a lack of annotated data is common (Janowczyk and Madabhushi, 2016).

For characterizing the high–level patterns in the five classes of our ILD dataset, we propose the extraction of a deep learning representation of all ILD patches from the Inception V3 deep learning architecture. This network computes 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 (Szegedy et al., 2015).

The principal feature of the inception architecture is the module that computes $a(\boldsymbol{x})$ in each layer. This module uses filters g of sizes 3×3 and 5×5 pixels that are then arranged and concatenated with the help of 1×1 convolutions to shrink the number of channels of the input (or previous layer output). They are then fed into the next unit by channel–wise concatenating all the output filters (Szegedy et al., 2015). For augmenting the invariance of the network, we augment the number of labeled samples per class by producing five labels–preserving patches per sample. Three of them were generated by rotating the original patch by 90, 180 and 270 degrees; the other two were obtained by reflecting the patch along the x and y axes respectively.

In the experimental setup, training all the weights from the network from random initializations yields slightly worse results than fine-tuning the network with pre-trained weights on ImageNet, confirming the previous results of Yosinski et al. (2014) where the authors study that transferring features even from seemingly distant tasks can be better than using a random feature initialization. Thus, in the following subsections when we write DL features, we make a reference to fine-tuned ImageNet pre-trained weights from the inception V3 architecture. The only change necessary in our setup in order to extract the features with the fine-tuned weights is to up-sample each original patch to 256×256 pixels and repeating the gray-scale value matrix in three different channels, to match the input size of the pre-trained architecture.



(a) Feature vectors are separately extracted from image patches using a Deep Learning and a Riesz-wavelet-based approach.

(b) The separate feature vectors are combined into a single vector after applying softmax classification.

Figure 6: The overall schema for the proposed late fusion approach is presented in two inter-connected parts: (a) and (b).

2.4. Combining Riesz Filters and Deep CNNs

Our approach to combine the Riesz and DL features is related to the one of learning a mixture of experts Masoudnia and Ebrahimpour (2014); Jacobs et al. (1991), particularly the mixture of MLP experts that learn a linear combination of the output vectors of multi-layered perceptron *experts* that specializes in a local region of the space of possible input vectors. Furthermore, we also performed experiments to assess at which fusion level the classifier better exploits the complementary information:

Early feature vector fusion: Given the two feature vectors, a simple approach is to concatenate them into a single image representation and to train a supervised classifier C on top of this joint representation, i.e., $x_f^i = [x_R^i, x_D^i]$, where $x_R^i \in \mathbb{R}^{263}$ is the Riesz feature vector and $x_D^i \in \mathbb{R}^{1024}$ is an extracted embedding vector of the DL architecture as described below. Formally, this corresponds to the direct sum of the Riesz and the DL feature vector spaces: $X_f = X_R \oplus X_D$ thus, dim (X_f) =dim(X_R)+ dim (X_D) . In this fusion scheme, the interaction between the features is expected to help the individual classifiers.

Late probability fusion: This approach consists of simply multiplying the output probability of each of the classifiers.

For a fair comparison of all five configurations (*i.e.*, Riesz, DL, early and late fusion, a softmax classifier with an intermediate hidden layer was trained using the same hyper-parameters. Softmax classifiers have proven to be useful when combining features from several sources in medical imaging (Otálora et al., 2015).

For the deep learning representation, a 1024–dimensional feature vector is 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 layer: the *pool5* layer that carries all the different learned patterns from the previous layers.

The output of the classifier represents the probability for a patch to be classified as one of each of the 5 ILD classes. To assess that the performance gains come from the interaction of either the representations or the output probabilities, and not from the random initializations of the weight matrices that might lead to different local minima, an ensemble of three individual classifiers (trained with Riesz or DL features) are trained to have more robust predictions.

2.5. Softmax Classifier

To train the weights Θ of the softmax model mapping the feature vectors to class probabilities, the following cost function is minimized with a stochastic gradient descent procedure:

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

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 representation of an unseen test patch $x \in \mathbb{R}^{DIM}$, where DIM is the dimensionality of the feature space, which can be either 1287, 1024 or 263, for the early fusion, DL and Riesz feature vectors respectively, is then classified as class c by calculating a probability:

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

A patch belongs to the c class if $p(y^c = 1|x;\Theta) > t$, where t is a threshold deciding firm class membership. As 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 five classes and 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; they 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. 6.

2.6. Parallel Computing

Since the calculation of steered Riesz signatures implies an iteration of all individual pixels in a patch, it is highly computationally expensive but also highly parallelizable. On typical workstations, this step can take a prohibitively large amount of time. Using advances made in (Vizitiu et al., 2016), we reduce the computation time by a factor of up to 30 times through a dedicated GPU–based implementation.

3. Results

3.1. Experimental Results

The Caffe DL framework (Jia et al., 2014) was used to train and extract features from the Inception V3 model. RGB replication of the grayscale patches and 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 five 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 Keras[2] DL framework with the TensorFlow back–end for all the softmax models trained, using the hyperparameters previously described in subsection 2.5.

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 in pixels was chosen after

[2] `https://keras.io/`, as of February 2018.

investigating patch sizes that yield good results and that generate a sufficient number of patches for training, validation, and testing. Fig. 9 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 2.

3.1.1. Riesz features only

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 right 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 ten 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 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. After all the feature vectors are computed, the softmax classifier is trained, and its performance on the test patches is evaluated. The recall for the Riesz features alone is displayed in the second row in Table 3, the model reached a fold–wise average AUC of 0.924 and an average accuracy of 74.4%.

3.1.2. Deep CNN features only

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 features generalized more achieving a slightly improved average accuracy

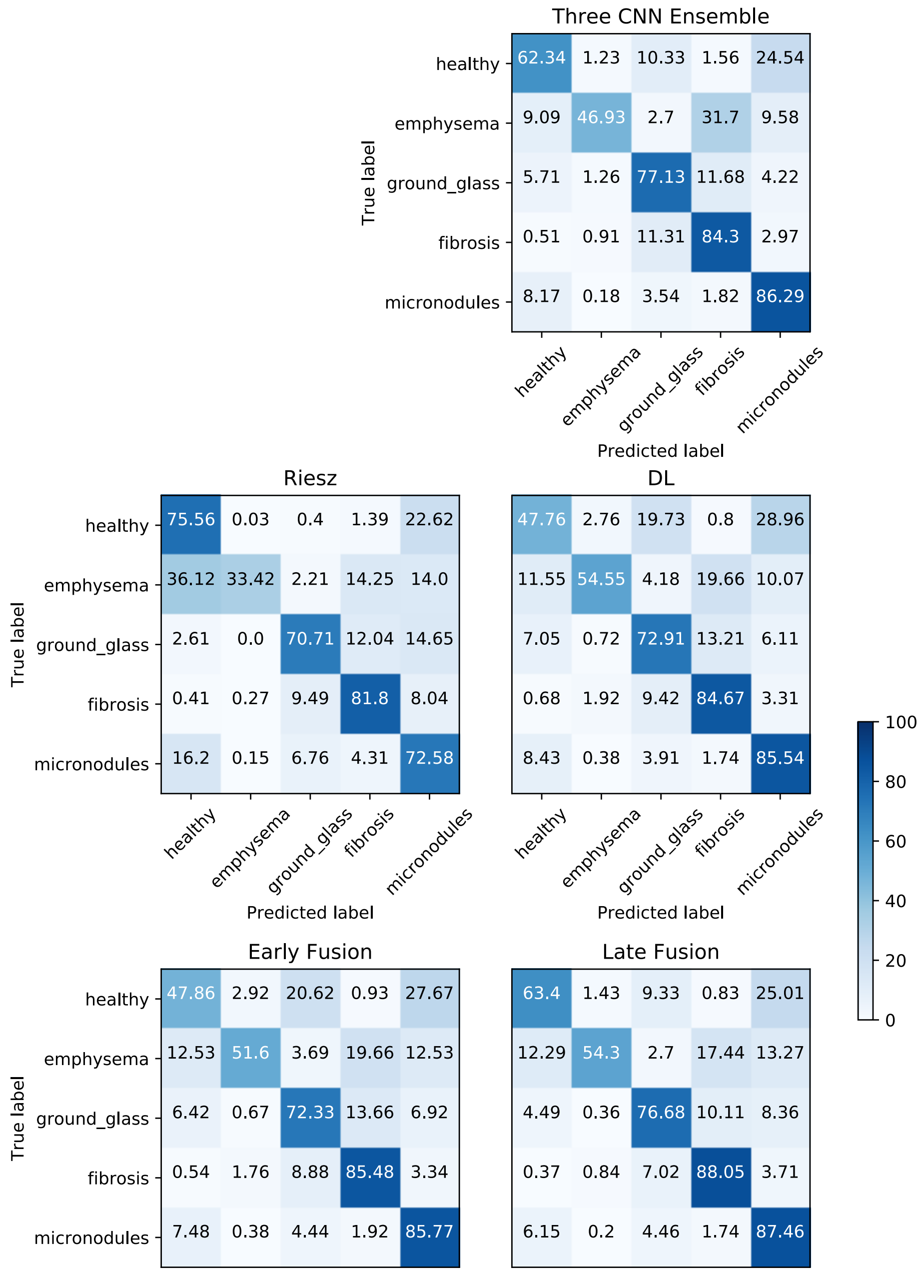


Figure 7: The aggregated confusion matrices for the four folds, of the five compared methods, is displayed. Percentages of the total number of patches are inside each cell.

of 78.6%. Once the features were extracted, we trained the softmax model using the hyperparameters described in subsection 2.5, the model reached a fold–wise average AUC of 0.932.

3.1.3. Ensemble of three CNN models

An ensemble model consisting of the fusion of three weight initialization of the InceptionV3 architecture was trained to have a better estimate of the performance of the CNN features, not relying only on the optima found in one single CNN model training. This model takes as input the concatenated vectors $x \in \mathbb{R}^{3072}$ of the three trained CNN's representation. The same softmax gating model architecture of the single DL model is trained with the fused vectors. The average accuracy was 77.9%, and the fold–wise average AUC in the ensemble model was 0.937. The confusion matrix as displayed in Figure 7 shows a better performance for the healthy, ground glass, and micronodules classes while slightly worsening the results of the single DL model in the fibrosis and emphysema classes. This result suggests that the results of the single model are relatively robust to the weight initialization of the network, and it stands to reason to not use more than one DL feature vector in the fusion with the Riesz features since the dimension of the combination will increase unnecessarily.

3.1.4. Combining Riesz and Deep CNN Representations

The early fusion approach of the concatenation of both DL and Riesz feature vectors yielded 78.1% average fold 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. 8 and this shows that it makes the best use of both classifiers.

3.1.5. Combining Riesz and Deep CNN Representations

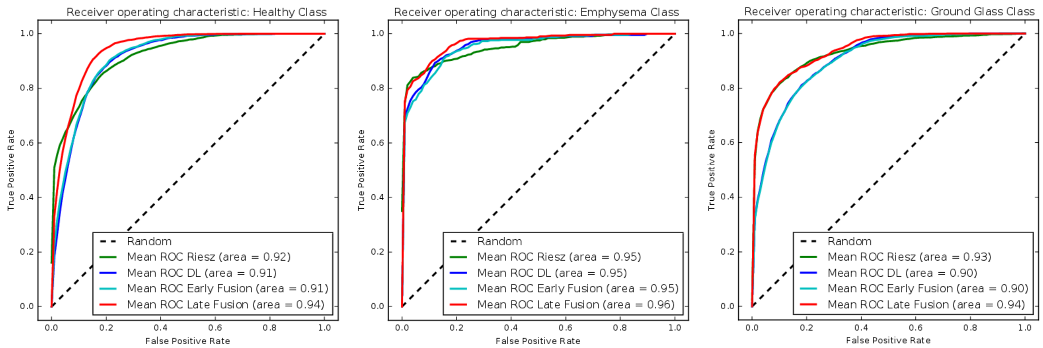
3.2. Statistical Significance of the Performance Comparisons

To assess the statistical significance of the difference between the results of the classifiers for all the classes together, we computed the McNemar test (Dietterich, 1998). For the test, the null hypothesis is 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.

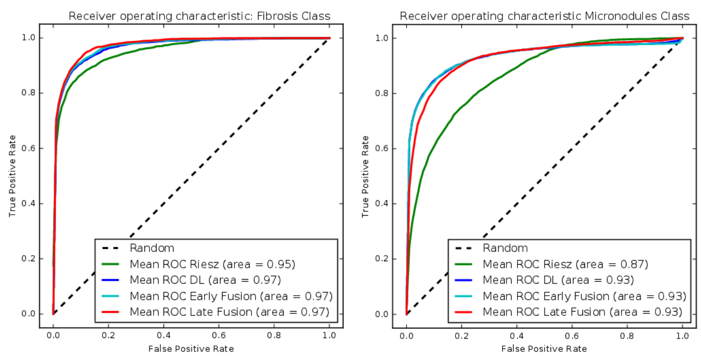
We concatenated the class predictions for each classifier in each fold and computed the number of times that a specific classifier A has guessed the correct class and a certain classifier B did not. Then, we computed the same number after inverting the classifier predictions, and these two sums were 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 (Fagerland et al., 2013). The results of the test are presented in Table 1.

4. Discussions

The results obtained using our method are compared with the results obtained by other authors who used the same ILD dataset but with possible



(a) ROC curves for the healthy, emphysema, and ground glass classes.



(b) ROC curves for fibrosis and micronodules.

Figure 8: ROC curves of the average performance in the four folds for each class using the late fusion approach depicted here. Healthy, Emphysema and Ground Glass classes, having a better AUC than the other approaches, benefit the most from the fusion.

Table 1: 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

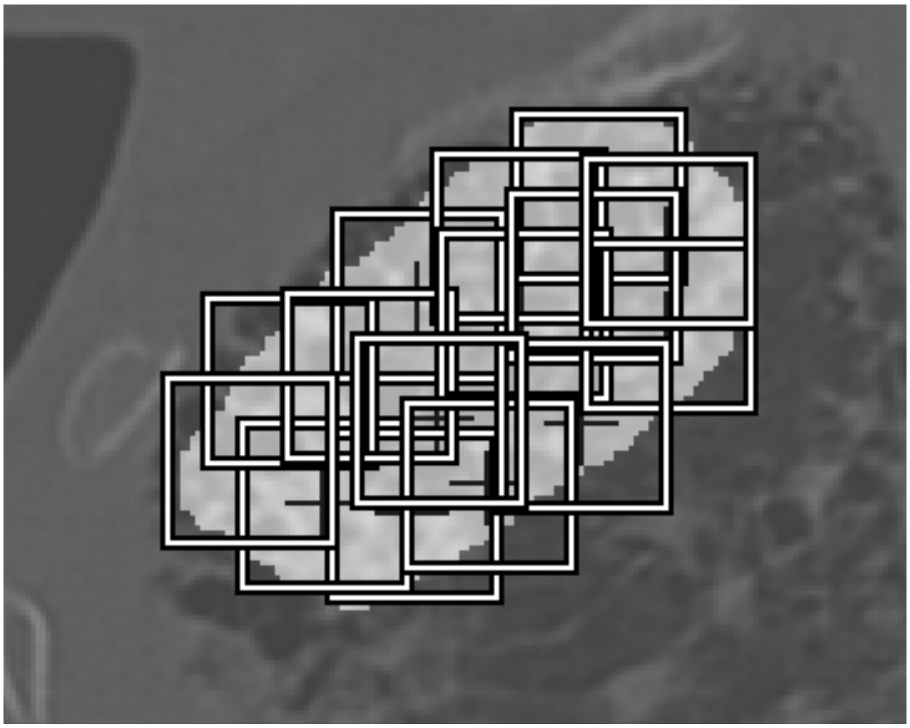


Figure 9: Overlapping patches of size 33–by–33 pixels are extracted from an annotated slice.

Table 2: 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

Table 3: Recalls obtained for our methods versus others in the literature are shown here.

	Class				
Method	H	E	G	F	M
Riesz (biased)	0.726	0.573	0.727	0.824	0.875
Riesz	0.756	0.334	0.707	0.818	0.726
DL only	0.478	0.546	0.729	0.847	0.855
Early Fusion	0.479	0.516	0.723	0.855	0.858
Late Fusion	0.634	0.543	0.767	0.881	0.875
(Song et al., 2013)	0.876	0.806	0.827	0.812	0.811
(Shin et al., 2016)	0.680	0.910	0.700	0.830	0.790
(Depeursinge et al., 2012b)	0.673	0.787	0.714	0.827	0.816
(Depeursinge et al., 2012a)	0.827	0.727	0.684	0.842	0.835
(Foncubierta-Rodríguez et al., 2012)	0.053	0.745	0.496	0.746	0.519
(Depeursinge et al., 2011a)	0.775	0.733	0.723	0.845	0.805
(Li et al., 2013)	0.760	0.670	0.700	0.740	0.840
(Depeursinge et al., 2011b)	0.790	0.692	0.593	0.805	0.702
(Gao et al., 2016)	0.914	0.827	0.815	0.891	0.880
(Song et al., 2015)	0.885	0.796	0.800	0.854	0.872

variations in terms of the evaluation methodology. Although the dataset used is the same, the exact validation scheme differs from one method to another according to the selection of patches (percentage lying within the ROI), patch size, distribution of the classes, and cross–validation schemes. Notwithstanding, Table 3 details the accuracies for the different tissue types obtained by our method and reported by others but they need to be read with care due to the differences in exact evaluation.

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, the performance of the same Riesz–based SVM classification method drops significantly on the whole (first versus second rows of Table 3). 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 3 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. This is a notable disparity, and we would 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. Besides, 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.

Finally, the classification accuracy of four different approaches are compared: deep CNN features alone, Riesz features alone, early fusion of the feature vectors and late fusion of the class probability for each classifier. Moreover, an ensemble of three deep CNN architectures was found to only add a negligible improvement to the results of a single network. An approach that uses multiple architectures at training time, or using the dropout technique Srivastava et al. (2014) in some of the layers could lead to more significant improvements. 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 important statistical significance of the presented differences. Nevertheless, there is a more significant gap, thus a smaller p–value, between the early and late fusion classifiers. This could be because when the early fusion of the feature vectors is performed, the representations are merged 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 filter-banks learned using Riesz aligned texture signatures. This would explain the similar predictions in that particular case.

5. Conclusions

In this paper, we show that late–fusing learned tissue representations based on Riesz and Deep CNN's for texture characterization yields performance gains over each approach separately or even early fusion. We showed that this is because is not dependent on the feature vector dimensionality but only on the independent probability of the classifiers.We believe that further performance gains can be achieved by investigating new methods of fusing Riesz–based and DL–based features, taking advantage of the complementarity of both sources of visual content from the ILD patches (Depeursinge and Müller, 2010).

Acknowledgment

This work was supported by the Swiss National Science Foundation (under grants PZ00P2_154891 and 205320_179069) and partially funded by the European Commission under the CP– Collaborative Project funding scheme through the FP7–ICT MD–PAEDIGREE project (ID: 600932). Sebastian Otálora thanks Colciencias for its partial support through call 756 for Ph.D. programs.

References

Ablowitz, M.J., Kaup, D.J., Newell, A.C., 1974. The inverse scattering transform-fourier analysis for nonlinear problems. Studies in Applied Mathematics 53, 249–315.

Ablowitz, M.J., Segur, H., 1981. Solitons and the inverse scattering transform. volume 4. SIAM.

Anthimopoulos, M., Christodoulidis, S., Ebner, L., Christe, A., Mougiakakou, S., 2016. Lung pattern classification for interstitial lung diseases using a deep convolutional neural network. IEEE transactions on medical imaging 35, 1207–1216.

Aziz, Z.A., Wells, A.U., Hansell, D.M., Bain, G.A., Copley, S.J., Desai, S.R., Ellis, S.M., Gleeson, F.V., Grubnic, S., Nicholson, A.G., Padley, S.P., Pointon, K.S., Reynolds, J.H., Robertson, R.J., Rubens, M.B., 2004.

HRCT diagnosis of diffuse parenchymal lung disease: inter–observer variation. Thorax 59, 506–511.

Barr, R., Parr, D., Vogel-Claussen, J., 2016. Imaging. European Respiratory Society.

Blakemore, C., Campbell, F.W., 1969. On the existence of neurones in the human visual system selectively sensitive to the orientation and size of retinal images. The Journal of Physiology 203, 237–260.

Cula, O.G., Dana, K.J., 2004. 3D texture recognition using bidirectional feature histograms. International Journal of Computer Vision 59, 33–60.

Dalal, N., Triggs, B., 2005. Histograms of oriented gradients for human detection, in: Proceedings of the 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05), IEEE Computer Society, Washington, DC, USA. pp. 886–893.

Depeursinge, A., 2017. Multi-Scale and Multi-Directional Biomedical Texture Analysis: Finding the Needle in the Haystack, in: Biomedical Texture Analysis: Fundamentals, Applications and Tools. Elsevier. Elsevier-MICCAI Society Book series, pp. 29–53.

Depeursinge, A., Al-Kadi, O.S., Mitchell, J.R., 2017a. Biomedical Texture Analysis: Fundamentals, Applications and Tools. Elsevier-MICCAI Society Book series, Elsevier. URL: `https://www.elsevier.com/books/title/author/9780128121337`.

Depeursinge, A., Fageot, J., Al-Kadi, O.S., 2017b. Fundamentals of Texture Processing for Biomedical Image Analysis: A General Definition and

Problem Formulation, in: Biomedical Texture Analysis: Fundamentals, Applications and Tools. Elsevier. Elsevier-MICCAI Society Book series, pp. 1–27.

Depeursinge, A., Foncubierta-Rodríguez, A., Van De Ville, D., Müller, H., 2011a. Lung texture classification using locally–oriented Riesz components, in: Fichtinger, G., Martel, A., Peters, T. (Eds.), Medical Image Computing and Computer Assisted Intervention – MICCAI 2011, Springer Berlin / Heidelberg. pp. 231–238.

Depeursinge, A., Foncubierta-Rodríguez, A., Van De Ville, D., Müller, H., 2012a. Multiscale lung texture signature learning using the Riesz transform, in: Medical Image Computing and Computer–Assisted Intervention MICCAI 2012, Springer Berlin / Heidelberg. pp. 517–524.

Depeursinge, A., Foncubierta-Rodríguez, A., Van De Ville, D., Müller, H., 2014a. Rotation–covariant texture learning using steerable Riesz wavelets. IEEE Transactions on Image Processing 23, 898–908.

Depeursinge, A., Foncubierta-Rodríguez, A., Van De Ville, D., Müller, H., 2014b. Three–dimensional solid texture analysis and retrieval in biomedical imaging: review and opportunities. Medical Image Analysis 18, 176–196.

Depeursinge, A., Iavindrasana, J., Cohen, G., Platon, A., Poletti, P.A., Müller, H., 2008. Lung tissue classification in HRCT data integrating the clinical context, in: 21th IEEE Symposium on Computer-Based Medical Systems (CBMS), Jyväskylä, Finland. pp. 542–547.

Depeursinge, A., Müller, H., 2010. Fusion techniques for combining textual and visual information retrieval, in: Müller, H., Clough, P., Deselaers, T., Caputo, B. (Eds.), ImageCLEF. Springer Berlin Heidelberg. volume 32 of *The Springer International Series On Information Retrieval*, pp. 95–114.

Depeursinge, A., Püspöki, Z., Ward, J.P., Unser, M., 2017c. Steerable Wavelet Machines (SWM): Learning Moving Frames for Texture Classification. IEEE Transactions on Image Processing 26, 1626–1636.

Depeursinge, A., Sage, D., Hidki, A., Platon, A., Poletti, P.A., Unser, M., Müller, H., 2007. Lung tissue classification using Wavelet frames, in: 29th Annual International Conference of the IEEE Engineering in Medicine and Biology Society, 2007. EMBS 2007., IEEE Computer Society, Lyon, France. pp. 6259–6262.

Depeursinge, A., Van De Ville, D., Platon, A., Geissbuhler, A., Poletti, P.A., Müller, H., 2012b. Near–affine–invariant texture learning for lung tissue analysis using isotropic wavelet frames. IEEE Transactions on Information Technology in BioMedicine 16, 665–675.

Depeursinge, A., Vargas, A., Platon, A., Geissbuhler, A., Poletti, P.A., Müller, H., 2012c. Building a reference multimedia database for interstitial lung diseases. Computerized Medical Imaging and Graphics 36, 227–238.

Depeursinge, A., Zrimec, T., Busayarat, S., Müller, H., 2011b. 3D lung image retrieval using localized features, in: Medical Imaging 2011: Computer-Aided Diagnosis, SPIE. p. 79632E.

Dietterich, T.G., 1998. Approximate statistical test for comparing supervised classification learning algorithms. Neural Computation 10, 1895–1923.

Do, M.N., Vetterli, M., 2002. Rotation invariant texture characterization and retrieval using steerable wavelet–domain hidden markov models. IEEE Transactions on Multimedia 4, 517–527.

Fagerland, M.W., Lydersen, S., Laake, P., 2013. The mcnemar test for binary matched-pairs data: mid-p and asymptotic are better than exact conditional. BMC medical research methodology 13, 91.

Foncubierta-Rodríguez, A., Depeursinge, A., Müller, H., 2012. Using multiscale visual words for lung texture classification and retrieval, in: Greenspan, H., Müller, H., Syeda Mahmood, T. (Eds.), Medical Content–based Retrieval for Clinical Decision Support, Lecture Notes in Computer Sciences (LNCS). pp. 69–79.

Freeman, W.T., Adelson, E.H., 1991. The design and use of steerable filters. IEEE Transactions on Pattern Analysis and Machine Intelligence 13, 891–906.

Gao, M., Bagci, U., Lu, L., Wu, A., Buty, M., Shin, H.C., Roth, H., Papadakis, G.Z., Depeursinge, A., Summers, R.M., et al., 2016. Holistic classification of ct attenuation patterns for interstitial lung diseases via deep convolutional neural networks. Computer Methods in Biomechanics and Biomedical Engineering: Imaging & Visualization , 1–6.

Gonzalez, D.M., Volpi, M., Tuia, D., 2016. Learning rotation invariant convolutional filters for texture classification. CoRR abs/1604.0.

Greenspan, H., van Ginneken, B., Summers, R.M., 2016. Guest editorial deep learning in medical imaging: Overview and future promise of an exciting new technique. IEEE Transactions on Medical Imaging 35, 1153–1159.

Guyon, I., Weston, J., Barnhill, S., Vapnik, V., 2002. Gene selection for cancer classification using support vector machines. Machine Learning 46, 389–422.

ter Haar Romeny, B.M., 2010. Multi-scale and multi-orientation medical image analysis, in: Biomedical Image Processing. Springer, pp. 177–196.

Haralick, R.M., Shanmugam, K., Dinstein, I., 1973. Textural features for image classification. IEEE Transactions on Systems, Man and Cybernetics 3, 610–621.

Hyvärinen, A., Hoyer, P., Inki, M., 2001. Topographic independent component analysis. Neural computation 13, 1527–1558.

Jacobs, R.A., Jordan, M.I., Nowlan, S.J., Hinton, G.E., 1991. Adaptive mixtures of local experts. Neural computation 3, 79–87.

Janowczyk, A., Madabhushi, A., 2016. Deep learning for digital pathology image analysis: A comprehensive tutorial with selected use cases. Journal of Pathology Informatics 7.

Jeng-Shyang, P., Jing-Wein, W., 1999. Texture segmentation using separable and non-separable wavelet frames. IEICE transactions on fundamentals of electronics, communications and computer sciences 82, 1463–1474.

Jia, Y., Shelhamer, E., Donahue, J., Karayev, S., Long, J., Girshick, R.B., Guadarrama, S., Darrell, T., 2014. Caffe: Convolutional architecture for fast feature embedding, in: Proceedings of the ACM International Conference on Multimedia, MM'14, pp. 675–678.

Joyseeree, R., Müller, H., Depeursinge, A., 2018. Rotation-covariant tissue analysis for interstitial lung diseases using learned steerable filters: Performance evaluation and relevance for diagnostic aid. Computerized Medical Imaging and Graphics 64, 1–11.

Kreuter, M., Herth, F.J., Wacker, M., Leidl, R., Hellmann, A., Pfeifer, M., Behr, J., Witt, S., Kauschka, D., Mall, M., et al., 2015. Exploring clinical and epidemiological characteristics of interstitial lung diseases: Rationale, aims, and design of a nationwide prospective registrythe exciting-ild registry. BioMed research international 2015.

Lazebnik, S., Schmid, C., Ponce, J., 2005. A sparse texture representation using local affine regions. IEEE Transactions on Pattern Analysis and Machine Intelligence 27, 1265–1278.

LeCun, Y., Bengio, Y., Hinton, G., 2015. Deep learning. Nature 521, 436–444.

LeCun, Y., Huang, F.J., Bottou, L., 2004. 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, IEEE. pp. II–97.

LeCun, Y., Kavukcuoglu, K., Farabet, C., 2010. Convolutional networks and applications in vision, in: IEEE International Symposium on Circuits and Systems (ISCAS), pp. 253–256.

Leung, T., Malik, J., 2001. Representing and recognizing the visual appearance of materials using three–dimensional textons. International Journal of Computer Vision 43, 29–44.

Li, Q., Cai, W., Feng, D.D., 2013. 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. pp. 6079–6082.

Li, Q., Cai, W., Wang, X., Zhou, Y., Feng, D.D., Chen, M., 2014. Medical image classification with convolutional neural network, in: Control Automation Robotics & Vision (ICARCV), 2014 13th International Conference on, IEEE. pp. 844–848.

Litjens, G., Kooi, T., Bejnordi, B.E., Setio, A.A.A., Ciompi, F., Ghafoorian, M., van der Laak, J.A., Van Ginneken, B., Sánchez, C.I., 2017. A survey on deep learning in medical image analysis. Medical image analysis 42, 60–88.

Lowe, D.G., 2004. Distinctive image features from scale–invariant keypoints. International Journal of Computer Vision 60, 91–110.

Mallat, S.G., 1989. A theory for multiresolution signal decomposition: the wavelet representation. IEEE Transactions on Pattern Analysis and Machine Intelligence 11, 674–693.

Masoudnia, S., Ebrahimpour, R., 2014. Mixture of experts: a literature survey. Artificial Intelligence Review 42, 275–293.

Ojala, T., Pietikäinen, M., Mäenpää, T., 2002. Multiresolution gray–scale and rotation invariant texture classification with local binary patterns. IEEE Transactions on Pattern Analysis and Machine Intelligence 24, 971–987.

Otálora, S., Cruz Roa, A., Arevalo, J., Atzori, M., Madabhushi, A., Judkins, A., González, F., Müller, H., Depeursinge, A., 2015. Combining unsupervised feature learning and riesz wavelets for histopathology image representation: Application to identifying anaplastic medulloblastoma, in: Navab, N., Hornegger, J., Wells, W.M., Frangi, A. (Eds.), Medical Image Computing and Computer-Assisted Intervention — MICCAI 2015. Springer International Publishing. volume 9349 of *Lecture Notes in Computer Science*, pp. 581–588.

Porter, R., Canagarajah, N., 1997. Robust rotation–invariant texture classification: wavelet, Gabor filter and GMRF based schemes. IEE Proceedings on Vision, Image and Signal Processing 144, 180–188.

Randen, T., Husoy, J.H., 1999. Filtering for texture classification: a comparative study. IEEE Transactions on Pattern Analysis and Machine Intelligence 21, 291–310.

Rozantsev, A., S.M., Fua, P., 2016. Beyond Sharing Weights for Deep Domain Adaptation. ArXiv e-prints `arXiv:1603.06432`.

Schmid, C., 2001. Constructing models for content–based image retrieval, in: IEEE Computer Society Conference on Computer Vision and Pattern Recognition, pp. 39–45.

Shin, H.C., Roth, H.R., Gao, M., Lu, L., Xu, Z., Nogues, I., Yao, J., Mollura, D., Summers, R.M., 2016. Deep convolutional neural networks for computer-aided detection: Cnn architectures, dataset characteristics and transfer learning. IEEE transactions on medical imaging 35, 1285–1298.

Simoncelli, E.P., Freeman, W.T., 1995. The steerable pyramid: a flexible architecture for multi–scale derivative computation, in: Proceedings of International Conference on Image Processing, 1995., pp. 444–447.

Song, Y., Cai, W., Huang, H., Zhou, Y., Wang, Y., Feng, D.D., 2015. Locality-constrained subcluster representation ensemble for lung image classification. Medical image analysis 22, 102–113.

Song, Y., Cai, W., Zhou, Y., Feng, D.D., 2013. Feature-based image patch approximation for lung tissue classification. IEEE transactions on medical imaging 32, 797–808.

Srivastava, N., Hinton, G., Krizhevsky, A., Sutskever, I., Salakhutdinov, R., 2014. Dropout: a simple way to prevent neural networks from overfitting. The Journal of Machine Learning Research 15, 1929–1958.

Szegedy, C., Liu, W., Jia, Y., Sermanet, P., Reed, S., Anguelov, D., Erhan, D., Vanhoucke, V., Rabinovich, A., 2015. Going deeper with convolutions, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 1–9.

Unser, M., Chenouard, N., Van De Ville, D., 2011. Steerable pyramids and tight wavelet frames in $L_2(\mathbb{R}^d)$. IEEE Transactions on Image Processing 20, 2705–2721.

Unser, M., Van De Ville, D., 2010. Wavelet steerability and the higher–order Riesz transform. IEEE Transactions on Image Processing 19, 636–652.

Vapnik, V.N., 1995. The Nature of Statistical Learning Theory. Springer, New York.

Vizitiu, A., Itu, L., Joyseeree, R., Depeursinge, A., Müller, H., Suciu, C., 2016. Gpu-accelerated texture analysis using steerable riesz wavelets, in: 24th IEEE Euromicro International Conference on Parallel, Distributed, and Network-Based Processing, pp. 431–434.

Watadani, T., Sakai, F., Johkoh, T., Noma, S., Akira, M., Fujimoto, K., Bankier, A.A., Lee, K.S., Müller, N.L., Song, J.W., et al., 2013. Interobserver variability in the ct assessment of honeycombing in the lungs. Radiology 266, 936–944.

Xu, D.H., Kurani, A.S., Furst, J., Raicu, D.S., 2004. Run–length encoding for volumetric texture, in: The 4th IASTED International Conference on Visualization, Imaging, and Image Processing – VIIP 2004, Marbella, Spain. pp. 105–110.

Xu, Y., Yang, X., Ling, H., Ji, H., 2010. 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, pp. 161–168.

Yosinski, J., Clune, J., Bengio, Y., Lipson, H., 2014. How transferable are features in deep neural networks?, in: Advances in neural information processing systems, pp. 3320–3328.