

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 drugs¹. 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.

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

12 Respiratory ailments affecting the lung parenchyma are prevalent. One
13 of the largest and most diverse groups of such diseases is the set of Inter-
14 stitial Lung Diseases (ILDs). They account for more than 200 pathologies
15 affecting the alveoli, the small lung airways, and the pulmonary intersti-
16 tium (Kreuter et al., 2015). Information gathered from clinical, radiologi-
17 cal, and pathological analyses are required to diagnose them accurately. In
18 particular, High-Resolution Computed Tomography (HRCT) images are the
19 radiological modality of choice for their characterization (Barr et al., 2016).
20 Some of the ailments may easily be misdiagnosed due to their rarity and
21 to the fact that radiologists are subjective when interpreting the content of
22 the images (Aziz et al., 2004; Watadani et al., 2013). Therefore, computer-
23 ized assistance yielding exhaustive and reproducible image analysis has been
24 mentioned several times as beneficial for improving ILD management (De-
25 peursinge et al., 2012c).

26 The task of classifying lung tissue pathologies benefits from recent ad-
27 vances made in the area of visual pattern recognition. In the particular
28 context of texture and tissue characterization, the latter relies heavily on
29 the local organization of image directions at different scales (Blakemore and
30 Campbell, 1969; ter Haar Romeny, 2010), including local variations of pattern
31 properties such as local anisotropy (Depeursinge et al., 2014b; Depeursinge,
32 2017). Spatial domain representations of images alone provide insufficient in-
33 formation to examine the local organization of scales and directions properly.
34 Therefore, to obtain a complete overview of the relationships between them,
35 intensity information needs to be complemented with information extracted
36 in the frequency domain.

37 Several authors exploit information embedded in the local organization
38 of scales and directions in images for pattern characterization and recogni-
39 tion. Grey-Level Co-occurrence Matrices (GLCM) (Haralick et al., 1973),
40 Histograms Of Gradients (HOG) (Dalal and Triggs, 2005) used in the Scale-
41 Invariant Feature Transform (SIFT) (Lowe, 2004), non-separable and sep-
42 arable wavelets (Jeng-Shyang and Jing-Wein, 1999), Run-Length Encoding
43 (RLE) (Xu et al., 2004), and oriented filterbanks and wavelets (Gaussian, Ga-
44 bor, Leung-Malik, Maximum Response (Cula and Dana, 2004; Leung and
45 Malik, 2001; Porter and Canagarajah, 1997; Randen and Husoy, 1999; Xu
46 et al., 2010) have been proposed for directional analysis. Unfortunately, sep-
47 arable wavelets suffer from bias along the vertical and horizontal axes (Mal-
48 lat, 1989), while the remainder requires an arbitrary choice of image direc-
49 tions (Depeursinge et al., 2014b). Using a sequence of pixels along perimeters
50 of radius r , Local Binary Patterns (LBP) (Ojala et al., 2002) can perform
51 multi-directional analysis but they do not allow for multiresolution analysis,
52 easily. In addition, r is determined through costly optimization. Other meth-
53 ods exploit the local organization of directions and scales indirectly. Notable
54 examples include Convolutional Neural Networks (CNN) (LeCun et al., 2004,
55 2010), Topographic Independent Component Analysis (TICA) (Hyvärinen
56 et al., 2001) and the scattering transform (Ablowitz et al., 1974; Ablowitz
57 and Segur, 1981). Despite their lack of interpretability, Deep Learning (DL)
58 models, and specifically CNNs, are now *de facto* standard methods for solv-
59 ing challenging computer vision tasks due to the performance improvements
60 they bring when compared with most classic handcrafted feature approaches.
61 In recent years, these techniques have been successfully in many medical-

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

65 These results are now routinely encountered in the literature due to the
66 capacity of the deep architectures to learn a wide range of filters that respond
67 to complex patterns. Moreover, the increasing availability of medical datasets
68 allows this method to have more robust and precise results (Anthimopoulos
69 et al., 2016; Gao et al., 2016; Shin et al., 2016). For instance, CNNs have been
70 used for lung disease classification in (Gao et al., 2016), where performance
71 was not assessed explicitly and in (Li et al., 2014) with a reasonable degree
72 of success. Due to the flexibility of the features learned with deep learning
73 models having millions of parameters, they allow the representation of a large
74 number of patterns present in the dataset, which usually exacerbates the risk
75 of overfitting.

76 Another fundamental and general aspect that needs to be accounted
77 for is that the same texture pattern can appear at several local orienta-
78 tions. Features that are locally rotation-invariant are desirable in such in-
79 stances (Depeursinge et al., 2017b; Schmid, 2001). LBPs (Ojala et al., 2002)
80 and Rotation-covariant SIFT (RIFT) (Lazebnik et al., 2005) possess such
81 a property, but they do not model discriminative patterns specifically (*i.e.*,
82 they yield handcrafted representations) and require exhaustive calculations.
83 Approaches based on steerable filters can achieve machine-precision, multi-
84 directional and multi-scale characterization with invariance to local rota-
85 tions (Depeursinge et al., 2017c; Do and Vetterli, 2002).

86 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

112 learning a class-wise texture signature using Riesz wavelets were proposed.
113 Here we build upon this work by complementing the features extracted us-
114 ing a deep-learning network to extract high-level patterns not captured by
115 texture signatures. Other authors working on the same data include Song
116 *et al.* (Song et al., 2013) who first employed feature-based image patch ap-
117 proximation. Li *et al.* (Li et al., 2013) then used automatic feature learning
118 followed by a customized CNN approach in (Li et al., 2014), while in (Song
119 et al., 2015), a locality-constrained subcluster representation ensemble is
120 used. Gao *et al.* use a deep CNN approach in (Gao et al., 2016).

121 The following publications use a different set of tissue types. A few clas-
122 sify six tissue types by including the consolidation type. Examples of such
123 instances include Foncubierta *et al.* (Foncubierta-Rodríguez et al., 2012) who
124 used multi-scale visual words for classification and retrieval. Shin *et al.* (Shin
125 et al., 2016) used deep CNNs. Others used a significantly different set of
126 classes. For example, Anthimopoulous *et al.* (Anthimopoulos et al., 2016)
127 applied a deep CNN to the following classes: healthy, consolidation, honey-
128 combing, micronodules, reticulation, ground glass, as well as a combination
129 of reticulation and ground glass.

130 To the best of our knowledge, there are not work in the literature ex-
131 ploiting the joint discriminative power of rotation invariant and deep learn-
132 ing representations for ILD classification. In summary, this paper describes
133 a novel feature-fusion approach that exploits the complementarity of the
134 learned representations from Riesz wavelets and fine-tuned deep CNNs to
135 classify five tissue types associated with ILDs. We propose both early and
136 late fusion strategies and estimate the performance with a four-fold cross-

137 validation setup. We compare all the methods using a softmax classifier
138 with the same hyperparameters to focus on the discriminatory power of the
139 extracted features.

140 The paper is organized as follows. In Section 2, we first present the
141 publicly available dataset used for the validation of our approach. This is
142 followed by an in–depth description of the theoretical aspects of the proposed
143 method. Section 3 presents the evaluation of classification performance, ob-
144 served results, and statistical significance of the performance comparison. A
145 thorough analysis and interpretation of the observed behavior is carried out
146 in Section 4. Finally, conclusions are drawn based on the work done, and we
147 propose measures to improve performance in future work in Section 5.

148 **2. Materials and Methods**

149 *2.1. Dataset and Validation Schemes*

150 We use the most frequently used publicly–available ILD dataset (De-
151 peursinge et al., 2012c) to evaluate the performance of the proposed meth-
152 ods. A slice of an HRCT series belonging to that dataset is shown in Fig. 1.
153 It depicts the lung parenchyma of a patient that was annotated by an expert
154 radiologist. The data set was used several times in past publications and
155 these past approaches on the same data set will be briefly covered in this
156 section.

157 To facilitate the comparison of our work with the majority of other tech-
158 niques used on the ILD dataset, we carry out supervised learning on the
159 following five expert–annotated classes: healthy, emphysema, ground glass,
160 fibrosis, and micronodules. Moreover, these classes are most common in the

161 majority of the ILDs and are therefore relatively well represented regarding
162 the number of annotated regions available. An illustration of their respec-
163 tive appearances is provided in Fig. 2. One may observe that the visual
164 differences between them are subtle, especially when comparing the healthy,
165 emphysema, and micronodules classes.

166 For training and testing, four-fold cross-validation is employed. This en-
167 tails extracting as many patches as possible from the annotated ILD images.
168 In the case of DL, a further augmentation step is taken whereby the patches
169 previously obtained are rotated by 90, 180, and 270 degrees. We also reflected
170 them along the vertical and horizontal axes. The new set of patches is then
171 divided into four groups according to two strategies. In the first one, they
172 are considered to be independent of each other, in line with what is often
173 encountered in the literature, and are divided into four equal sets. In the
174 second strategy, we ensure that the patches originating from an individual
175 patient only appear in one of the four groups to minimize the risk of inherent
176 bias. Two of the four groups are then concatenated and used for training the
177 classification model. One of the remaining sets is used for validation where
178 necessary, and the last one is used for testing the trained model. This process
179 is repeated four times to ensure that each group is once in the test set.

180 Finally, although the slice thickness and slice pixel dimensions of the
181 HRCT protocol are all 1mm, the spacing between slices is 10mm. This
182 implies that a considerable amount of information is missing between slices,
183 which cannot be easily reconstructed. There is no possibility, therefore, to
184 consider full 3D image analysis, which might lead to better results if the data
185 are available.

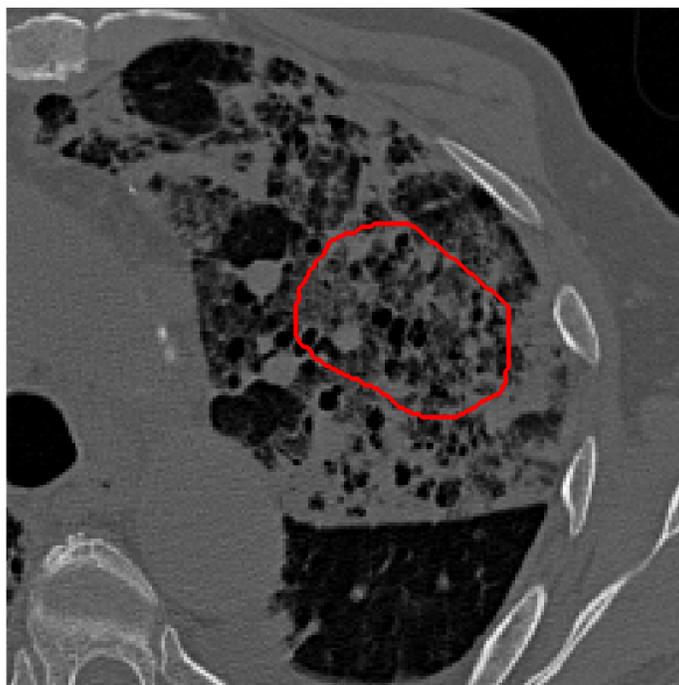


Figure 1: Part of a slice taken from the ILD database that represents the right lung along with an expert annotation (red delineation) corresponding to a Region Of Interest (ROI) are presented here.



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

186 *2.2. Tissue Characterization Using Riesz Filters*

187 This section describes the approach to obtain learned discriminative and
188 locally rotation-invariant texture representations from Riesz wavelets and is
189 based on (Depeursinge et al., 2014a; Joyseeree et al., 2018). We first intro-
190 duce the Riesz transform and its combination with radial wavelets to derive
191 steerable filterbanks in Section 2.2.1. Second, we describe in Section 2.2.2
192 how we learn one-versus-one class-specific discriminative texture signatures
193 from the parametric Riesz representation using Support Vector Machines
194 (SVM).

195 *2.2.1. Steerable Riesz filterbanks*

196 In a nutshell, Riesz filterbanks provide sets of image operators behaving
197 like multi-scale local partial image derivatives of any order. Let $f(\mathbf{x})$ rep-
198 resent the function that models the content of a patch where \mathbf{x} represents
199 pixel coordinates x_1 and x_2 . In other words, $f: \mathbf{x} \rightarrow f(\mathbf{x}), \mathbf{x} \in \mathbb{R}^2$, where
200 $\mathbf{x} = (x_1, x_2)$.

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

$$205 \quad f(\mathbf{x}) \xleftrightarrow{\mathcal{F}} \hat{f}(\boldsymbol{\omega}) = \int_{\mathbb{R}^2} f(\mathbf{x}) e^{-j\langle \boldsymbol{\omega}, \mathbf{x} \rangle} dx_1 dx_2, \quad (1)$$

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

207 The Riesz filterbanks needed for our work are based on the real Riesz
208 transform (Unser et al., 2011). The $N + 1$ components of the N th-order

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

$$210 \quad \mathcal{R}^N \{f\}(\mathbf{x}) = \begin{pmatrix} \mathcal{R}^{(0,N)} \{f\}(\mathbf{x}) \\ \vdots \\ \mathcal{R}^{(n,N-n)} \{f\}(\mathbf{x}) \\ \vdots \\ \mathcal{R}^{(N,0)} \{f\}(\mathbf{x}) \end{pmatrix}, \quad (2)$$

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:

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

211 where

$$212 \quad \widehat{\mathcal{R}^{(n,N-n)} \{f\}}(\boldsymbol{\omega}) = \sqrt{\frac{N}{n!(N-n)!}} \frac{(-j\omega_1)^n (-j\omega_2)^{N-n}}{\|\boldsymbol{\omega}\|^N} \hat{f}(\boldsymbol{\omega}). \quad (3)$$

213 According to Eq. (3), the product of $j\omega_1$ and/or $j\omega_2$ in the numerator
 214 followed by division with the norm of $\boldsymbol{\omega}$ produces allpass filters that only
 215 retain phase information that characterizes directions (Depeursinge et al.,
 216 2014b; Unser and Van De Ville, 2010) and the order N controls the angular
 217 selectivity of the Riesz kernels. Therefore, the Riesz kernels behave like
 218 allpass N -th order partial image derivatives. Fig. 3 illustrates the Riesz
 219 filterbanks for $N = 1, \dots, 5$.

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

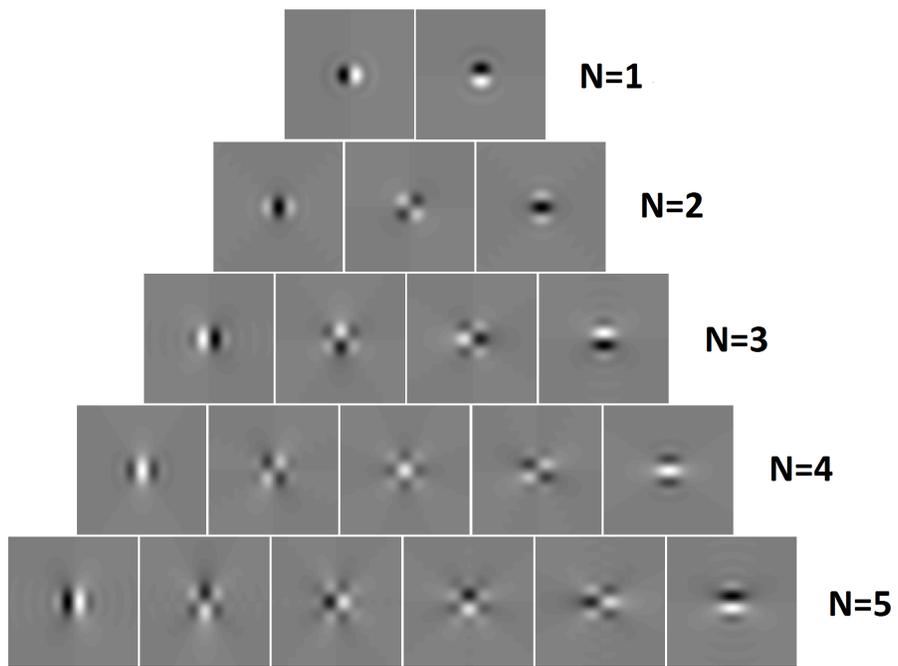


Figure 3: Riesz filterbanks for orders up to 5 are shown here. To represent the filters on a finite spatial support, the Riesz transform was applied to an isotropic Gaussian function.

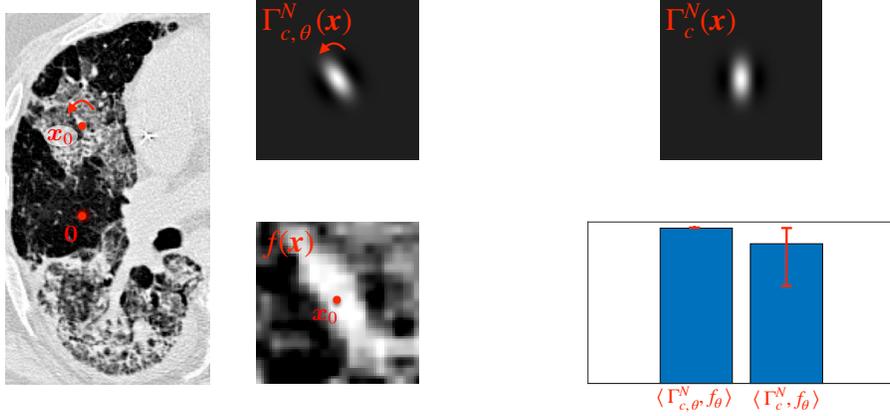


Figure 4: Steering the texture signatures Γ_c^N with an angle θ allows reducing the variability in feature values when compared to linear filtering used by e.g. CNNs. This example demonstrates that the responses of the unsteered signature (i.e., $\langle \Gamma_c^N, f_\theta \rangle$) on a patch rotated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

223 looking at the maximum response over all possible orientations, steerability
 224 allows achieving local rotation invariance at a relatively cheap computational
 225 cost because it is not needed to re-convolve the image with rotated versions
 226 of the kernels.

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

$$230 \quad \mathcal{R}^N \{f_\theta\}(\mathbf{0}) = A_\theta \mathcal{R}^N \{f\}(\mathbf{0}), \quad (4)$$

231 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
 232 matrix. The use of steerability to reduce variability in feature values caused
 233 by rotations of the input patches is illustrated in Fig. 4.

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

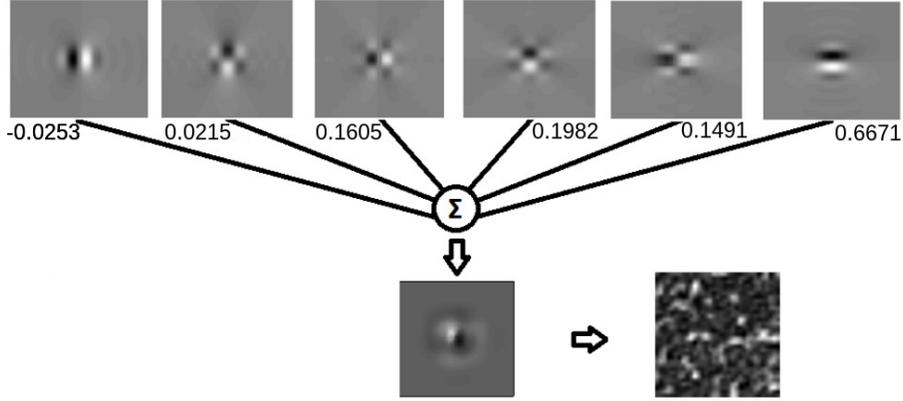


Figure 5: A texture signature for the micronodule class is built after applying an appropriate weighing scheme for the components of a Riesz filterbank of order 5.

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

238 2.2.2. Parametric Discriminative Texture Signatures

239 Learned representations based on class-specific steerable texture signa-
 240 tures are obtained by finding a weighting scheme for the Riesz filterbanks at
 241 each scale. Fig. 5 illustrates this for a Riesz filterbank of order 5, which is
 242 used to generate a signature for micronodules. We are looking for an optimal
 243 texture signature Γ_c^N of the class c from a linear combination of the Riesz
 244 kernels as

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

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

248 filterbanks (Depeursinge et al., 2012a) as

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

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

251 We determine the weighing scheme using a one-versus-one SVM classi-
 252 fication configuration. The filter energy responses $E(\mathcal{R}^{(n,N-n)}\{f\}(\mathbf{x}))$ are
 253 computed and regrouped for each class c versus each one from the remain-
 254 ing classes. SVMs then find the optimal separation in terms of minimized
 255 structural risk (Guyon et al., 2002; Vapnik, 1995). Each class benefits from
 256 a unique characterization with regard to each of the remaining classes. With
 257 five classes, the approach is expected to lead to $5 \cdot 4 = 20$ separate optimal sig-
 258 natures. However, because of the optimal separation between a class A and
 259 another class B is the same as that between class B and class A , the number
 260 of optimal signatures reduce to 10. The optimal weights $\mathbf{w} = (w_1, \dots, w_{N+1})$
 261 are directly determined from the support vectors of the optimal separa-
 262 tions (Depeursinge et al., 2014b).

263 2.2.3. Classification Using Riesz Filters

264 Through the approach defined previously, a class-wise texture signature
 265 Γ_c^N is obtained for each class c . Input images are then filtered using the
 266 steered texture signatures: they are steered at every position in the image to
 267 maximize their response, leading to a non-linear filtering operation. A fea-
 268 ture space is spanned by the average energy of the steered filter responses and
 269 is complemented using a histogram of Hounsfield Units (HUs) of the patches
 270 in the spatial domain. This helps including the intensity information of the

271 images which is not taken into account when only wavelet domain charac-
272 teristics (*i.e.*, band-pass) are utilized. The feature space is completed by
273 the number of pixels representing air in a patch as this also helps character-
274 ize lung tissue (for example in the case of emphysema). In short, the feature
275 space of a patch is made up of 10 filter responses, a histogram of HUs and the
276 number of air pixels. After a series of initial investigations (not detailed in
277 this paper) into an appropriate value for the Riesz order N , a value of 5 was
278 chosen, as it was providing a good trade-off between directional specificity
279 and regularization. For $N=5$, 10 texture signatures of length $(N + 1) * J=24$
280 each are obtained. To create the final input feature vector, the $24 \cdot 10=240$
281 variables obtained in the previous step are concatenated with the 22 variables
282 from the histogram and also with one last variable for air content. As a re-
283 sult, a feature vector of length 263 is obtained for each patch. Once the final
284 Riesz feature vector is built, we train a Softmax classifier that maps from
285 the 263 Riesz feature vector to the 5 ILD classes. The softmax classifier uses
286 the same hyperparameters for the training of the DL feature classification
287 alone, in order to evaluate the discriminatory power of the features itself.
288 The details of the softmax classifier are discussed in section 2.4.

289 *2.3. Tissue Characterization With Deep CNNs*

290 DL has shown significant improvements for analyzing complex visual pat-
291 terns, reaching human performance in various tasks. The CNN is the most
292 prominent DL technique for computer vision. A CNN is a particular set
293 of supervised multi-layer perceptron architectures. CNNs are biologically
294 inspired by the local activations of the visual cortex (LeCun et al., 2015).
295 Similarly to Riesz filterbanks, these local activations can be thought of like

296 a bank of filters that act on certain areas of the input (*i.e.*, receptive fields).
 297 Due to overlap, one may find local correlation via convolutions (Depeursinge
 298 et al., 2017a).

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

$$302 \quad a(\mathbf{x}) = \sigma(\mathbf{W}\mathbf{x}) = \sigma\left(\sum_{j=1}^M w_j x_j + b\right), \quad (7)$$

303 where \mathbf{W} is the weight matrix of the network for that layer with dimensional-
 304 ity M , and b is the bias term. Several activation functions $\sigma(\cdot)$ are proposed in
 305 the DL literature. Rectified Linear Units (ReLUs), where $\sigma(x) = \max(0, x)$,
 306 are consistently used in many applications because of their efficient gradient
 307 propagation that avoids vanishing or exploding gradients and also for their
 308 efficient computation as they only require a comparison. In CNNs, one is
 309 interested in learning small filters g that capture the spatial correlation in
 310 the input. Formally, the output of a convolution unit h_j is computed as

$$311 \quad h_j(\mathbf{x}) = \sum_{i=1}^C (f_i * g_{ij})(\mathbf{x}), \quad (8)$$

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

$$314 \quad (f_i * g_{ij}) = \sum_{p=1}^P \sum_{q=1}^Q f_i(p, q) g_{ij}(x_1 - p, x_2 - q). \quad (9)$$

315 These matrix operations are efficiently vectorized to leverage the parallel ca-
 316 pabilities of the Graphical Processing Units (GPUs). In contrast with the
 317 Riesz filter analysis, where the local rotation invariance is explicitly hard-
 318 coded in the model, DL learns relative rotation-invariance with directionally

319 insensitive filters and multiple oriented versions of directional filters in a
320 model with millions of free parameters (Gonzalez et al., 2016). This permits
321 learning higher-level patterns thanks to the non-linear hierarchical composi-
322 tion of low-level features (Song et al., 2015) at the cost of being more prone
323 to overfitting if the model is not regularized accordingly.

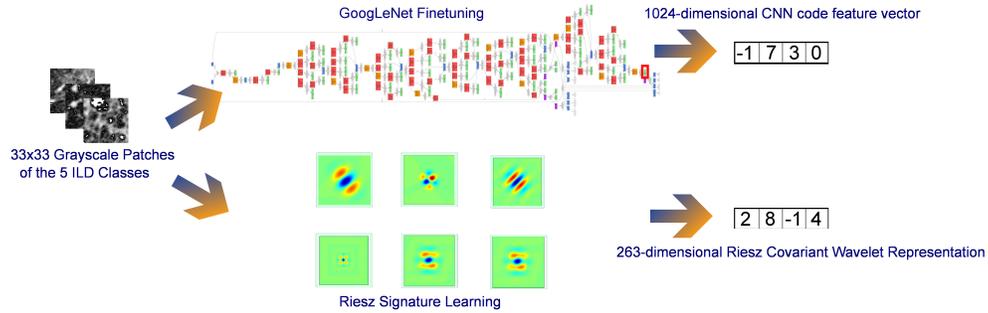
324 Training such large networks for medical tasks can be unfeasible due to
325 the lack of annotated data to train the model. An exciting alternative is to
326 use knowledge gained in other tasks where a large amount of data is avail-
327 able. This is known as transfer learning, where a model that was initially
328 trained using a large amount of labeled data is then fine-tuned (Rozant-
329 sev and Fua, 2016) to a new dataset where a fewer annotated samples are
330 available, thus leveraging the filters learned in the first dataset to serve as
331 a starting point to learn the optimal filters in the new dataset. Notably,
332 the use of pre-trained models to recognize objects in natural image settings
333 could be helpful in many medical tasks because of the following two aspects.
334 First, the layers and units in the network that recognize primitive features
335 (*e.g.*, edges and textures) are shared across different visual contents. Second,
336 reusing a pre-trained deep network sets the state of the optimization prob-
337 lem near a local optimum which is beneficial for both the performance and
338 earlier training convergence. Transfer learning has also shown to be useful
339 for faster convergence in medical scenarios where a lack of annotated data is
340 common (Janowczyk and Madabhushi, 2016).

341 For characterizing the high-level patterns in the five classes of our ILD
342 dataset, we propose the extraction of a deep learning representation of all
343 ILD patches from the Inception V3 deep learning architecture. This network

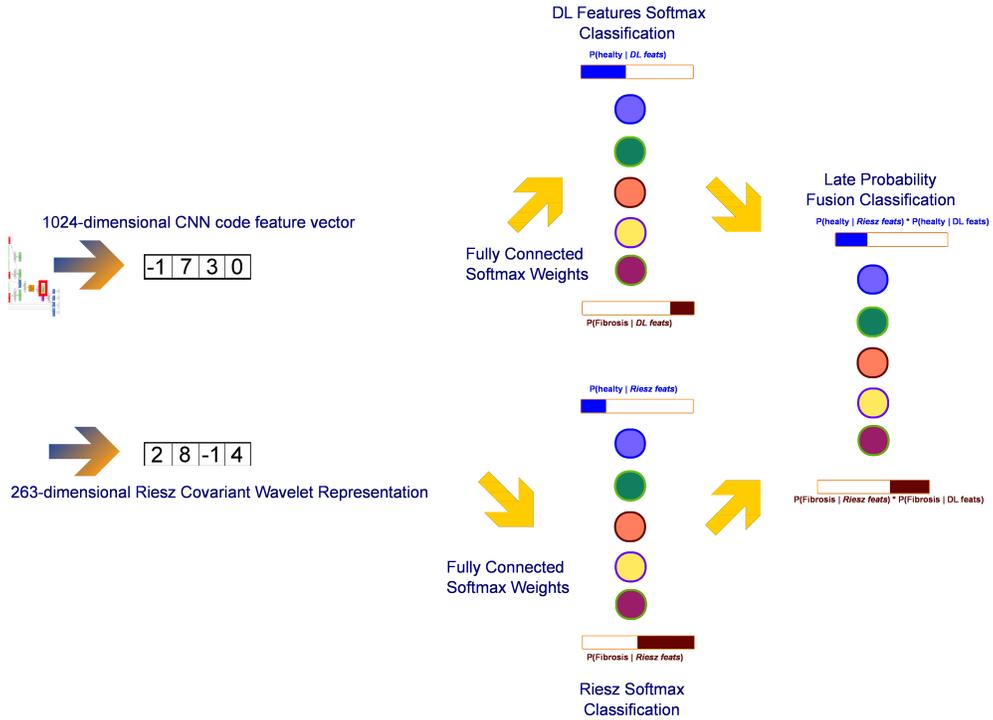
344 computes representations in a multi-scale fashion by reusing the outputs of
345 the first layers to feed later ones as well as intermediate convolution modules,
346 thus keeping the computational burden under control (Szegedy et al., 2015).

347 The principal feature of the inception architecture is the module that
348 computes $a(\mathbf{x})$ in each layer. This module uses filters g of sizes 3×3 and
349 5×5 pixels that are then arranged and concatenated with the help of 1×1
350 convolutions to shrink the number of channels of the input (or previous layer
351 output). They are then fed into the next unit by channel-wise concatenating
352 all the output filters (Szegedy et al., 2015). For augmenting the invariance
353 of the network, we augment the number of labeled samples per class by
354 producing five labels-preserving patches per sample. Three of them were
355 generated by rotating the original patch by 90, 180 and 270 degrees; the
356 other two were obtained by reflecting the patch along the x and y axes
357 respectively.

358 In the experimental setup, training all the weights from the network from
359 random initializations yields slightly worse results than fine-tuning the net-
360 work with pre-trained weights on ImageNet, confirming the previous results
361 of Yosinski et al. (2014) where the authors study that transferring features
362 even from seemingly distant tasks can be better than using a random feature
363 initialization. Thus, in the following subsections when we write DL features,
364 we make a reference to fine-tuned ImageNet pre-trained weights from the in-
365 ception V3 architecture. The only change necessary in our setup in order to
366 extract the features with the fine-tuned weights is to up-sample each original
367 patch to 256×256 pixels and repeating the gray-scale value matrix in three
368 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).

369 *2.4. Combining Riesz Filters and Deep CNNs*

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

377 **Early feature vector fusion:** Given the two feature vectors, a simple
378 approach is to concatenate them into a single image representation and to
379 train a supervised classifier C on top of this joint representation, i.e., $x_f^i =$
380 $[x_R^i, x_D^i]$, where $x_R^i \in \mathbb{R}^{263}$ is the Riesz feature vector and $x_D^i \in \mathbb{R}^{1024}$ is
381 an extracted embedding vector of the DL architecture as described below.
382 Formally, this corresponds to the direct sum of the Riesz and the DL feature
383 vector spaces: $X_f = X_R \oplus X_D$ thus, $\dim(X_f) = \dim(X_R) + \dim(X_D)$. In
384 this fusion scheme, the interaction between the features is expected to help
385 the individual classifiers.

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

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

393 For the deep learning representation, a 1024-dimensional feature vector

394 is extracted from the layer with the largest area covered in the input image
 395 (*i.e.*, with the largest receptive field). This happens to be the layer preceding
 396 the classification layer: the *pool5* layer that carries all the different learned
 397 patterns from the previous layers.

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

405 2.5. Softmax Classifier

406 To train the weights Θ of the softmax model mapping the feature vec-
 407 tors to class probabilities, the following cost function is minimized with a
 408 stochastic gradient descent procedure:

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

410 where M stands for the number of samples, N for the number of units, and
 411 K is the number of classes. ρ is the weight decay parameter that penalizes
 412 large values for parameters. The representation of an unseen test patch
 413 $x \in \mathbb{R}^{DIM}$, where DIM is the dimensionality of the feature space, which can
 414 be either 1287, 1024 or 263, for the early fusion, DL and Riesz feature vectors

415 respectively, is then classified as class c by calculating a probability:

$$416 \quad p(y^c = 1|x; \Theta) = \frac{\exp(\Theta_1 x)}{\sum_{l=1}^K \exp(\Theta_l x)}. \quad (11)$$

417 A patch belongs to the c class if $p(y^c = 1|x; \Theta) > t$, where t is a threshold
418 deciding firm class membership. As this varies across the folds, we report the
419 average Receiver Operating Characteristic (ROC) curves and their respective
420 Area Under the Curve (AUC) for each of the five classes and each fold. The
421 number of units N in the hidden layer was explored in the set $\{32, 64, 128\}$
422 with robust performance and we finally set it to 64 units for all experiments.
423 The other parameters of the softmax classifier were the learning rate, decay,
424 and momentum; they were set to 10^{-3} , 10^{-6} and 0.9, respectively. The
425 accelerated gradient method of Nesterov was used as a parameter in the
426 stochastic gradient descent optimizer.

427 Because the fused feature space is high-dimensional and the DL feature
428 vector is approximately four times larger than the Riesz representation, the
429 fused vector tends to reflect the performance of the DL classifier alone, leaving
430 the complementary information out. To alleviate this, we performed a late
431 fusion approach as follows. First, a single softmax classifier is trained for each
432 representation. Then, the output probability vector of the two classifiers is
433 multiplied element-wise to obtain a weighted probability vector to perform
434 the final classification. The proposed combinations are depicted in Fig. 6.

435 *2.6. Parallel Computing*

436 Since the calculation of steered Riesz signatures implies an iteration of
437 all individual pixels in a patch, it is highly computationally expensive but

438 also highly parallelizable. On typical workstations, this step can take a pro-
439 hibitively large amount of time. Using advances made in (Vizitiu et al.,
440 2016), we reduce the computation time by a factor of up to 30 times through
441 a dedicated GPU-based implementation.

442 **3. Results**

443 *3.1. Experimental Results*

444 The Caffe DL framework (Jia et al., 2014) was used to train and extract
445 features from the Inception V3 model. RGB replication of the grayscale
446 patches and scaling from square patches of length 33 pixels to a length of
447 256 pixels was performed in order to be in line with the input layer of the
448 architecture and use the pre-trained weights. The number of epochs was set
449 to 30, but an early convergence up to the five epoch was achieved for all folds.
450 The learning rate in all cases was initially set to 0.0001 and was decreased
451 according to an exponential decay with $\sigma = 0.95$. We used the Keras² DL
452 framework with the TensorFlow back-end for all the softmax models trained,
453 using the hyperparameters previously described in subsection 2.5.

454 MATLAB was used for the Riesz-related calculations. First, for each
455 slice in the ILD database, square patches of length 33 pixels were extracted
456 from the annotations present. A patch is defined as any 33×33 square region
457 found to lie with at least 75% of it within the annotated region and the center
458 of which is separated by at least half a patch length from the respective
459 centers of other extracted patches. The patch size in pixels was chosen after

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

460 investigating patch sizes that yield good results and that generate a sufficient
461 number of patches for training, validation, and testing. Fig. 9 illustrates the
462 extraction of patches. In total, 14,594 patches are found in this manner, and
463 a breakdown of the set in terms of classes represented is provided in Table 2.

464 3.1.1. Riesz features only

465 We first construct the feature vector representing each patch. The dis-
466 tribution of grayscale values between -1000 HU (value for air) and 650 HU
467 (value for bone) in each patch is divided into 22 bins, which ensures a right
468 balance of granularity in spatial-domain representation and low dimension-
469 ality. The number of air pixels in a patch is also noted. The energy of the
470 filter response of each patch to each of the ten weighted Riesz kernels at $J=4$
471 scales, which ensures sufficient coverage of spatial frequencies in the Fourier-
472 domain, completes the feature space. In other words, for $N=5$, 10 texture
473 signatures of length $(N + 1) * J=24$ each are extracted. Concatenating the
474 $24 \cdot 10=240$ variables thus obtained with the 22 histogram bins and the air
475 content value yields the final feature vector with 263 dimensions. After all
476 the feature vectors are computed, the softmax classifier is trained, and its
477 performance on the test patches is evaluated. The recall for the Riesz fea-
478 tures alone is displayed in the second row in Table 3, the model reached a
479 fold-wise average AUC of 0.924 and an average accuracy of 74.4%.

480 3.1.2. Deep CNN features only

481 The average accuracy for the four folds of the DL model was 77.1%.
482 When using the weights of the model trained with the ImageNet dataset,
483 the features generalized more achieving a slightly improved average accuracy

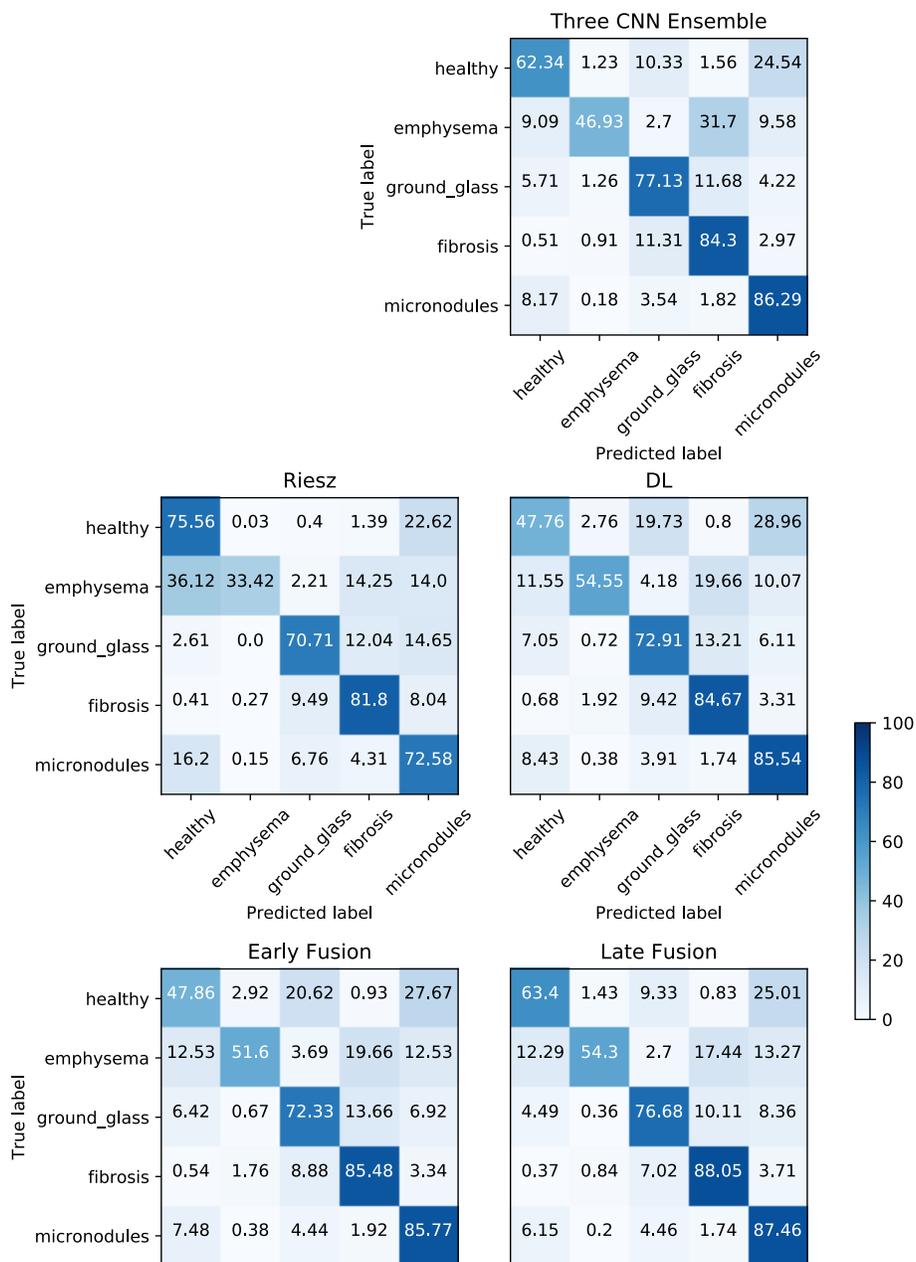


Figure 7: The aggregated confusion matrices for the four folds, of the five compared methods, is displayed. Percentages of the total number of patches are inside each cell.

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

487 3.1.3. Ensemble of three CNN models

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

503 3.1.4. Combining Riesz and Deep CNN Representations

504 The early fusion approach of the concatenation of both DL and Riesz
505 feature vectors yielded 78.1% average fold accuracy. An AUC performance
506 of 0.931 was also noted, which is almost identical to the performance of
507 the DL features alone. Because the fused feature space is high-dimensional

508 and the DL feature vector is approximately four times larger than the Riesz
509 representation, the fused vector tends to reflect the performance of the DL
510 classifier alone, dismissing the complementary of the two representations. To
511 alleviate this, we implemented a late fusion, which obtained the best AUC
512 performance of 0.948 as depicted in Fig. 8 and this shows that it makes the
513 best use of both classifiers.

514 *3.1.5. Combining Riesz and Deep CNN Representations*

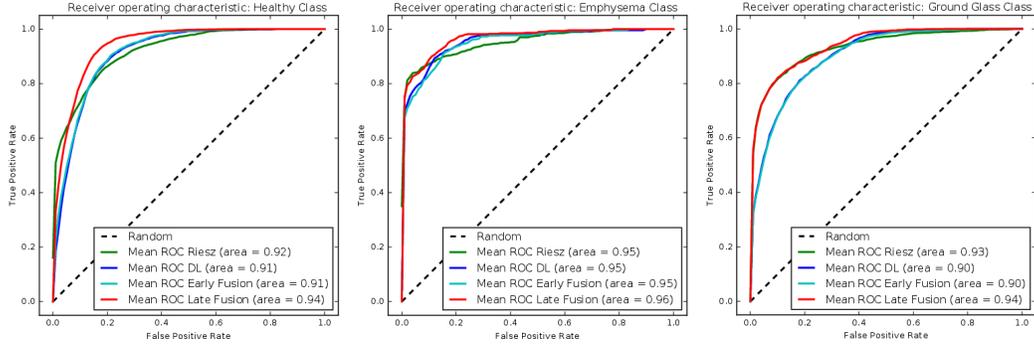
515 *3.2. Statistical Significance of the Performance Comparisons*

516 To assess the statistical significance of the difference between the results of
517 the classifiers for all the classes together, we computed the McNemar test (Di-
518 etterich, 1998). For the test, the null hypothesis is not having a significant
519 difference between the classifier results, and the alternative hypothesis being
520 the opposite, *i.e.*, the mean of their results are distinct enough and cannot
521 be due to a random process.

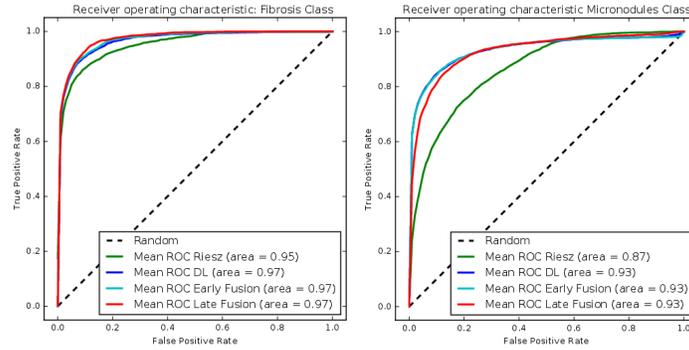
522 We concatenated the class predictions for each classifier in each fold and
523 computed the number of times that a specific classifier A has guessed the
524 correct class and a certain classifier B did not. Then, we computed the same
525 number after inverting the classifier predictions, and these two sums were
526 passed as parameters to the mid- p -test. If the p -value is less than 0.05, the
527 results are considered to be statistically significant (Fagerland et al., 2013).
528 The results of the test are presented in Table 1.

529 **4. Discussions**

530 The results obtained using our method are compared with the results
531 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

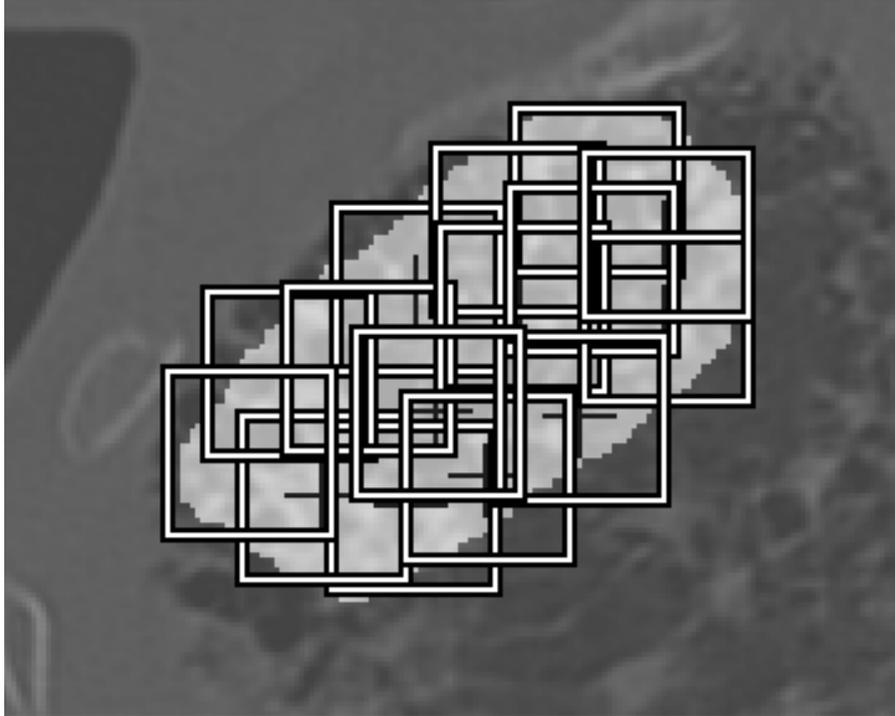


Figure 9: Overlapping patches of size 33-by-33 pixels are extracted from an annotated slice.

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.

Method	Class				
	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

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

539 A first observation is that when the dataset is carefully divided up in
540 order to ensure that the same patient does not contribute patches to both
541 the training and test sets, the performance of the same Riesz-based SVM
542 classification method drops significantly on the whole (first versus second
543 rows of Table 3). This proves beyond doubt that bias is present when the
544 above separation step is not explicitly taken. Since many of the existing
545 methods present no evidence of explicitly applying such a step; their reported
546 performance values are at risk of being erroneously higher than they should
547 be.

548 Table 3 shows that there is some room for improvement in the classifica-
549 tion of the emphysema class. Indeed, only 407 patches with identified em-
550 physema are encountered in the ILD database while the next least frequent
551 disease class is ground glass with 2226 patches. This is a notable disparity,
552 and we would argue that our learning approaches for emphysema are less
553 well trained as compared to the other classes due to a much lower number
554 of patches used for training. Besides, emphysema has very large intra-class
555 variations and would require learning several steerable models or signatures
556 per class. We contend that the use of more patches belonging to that class for
557 training and the use of more than one signature for emphysema in subsequent
558 work would significantly improve the overall classification accuracy.

559 Finally, the classification accuracy of four different approaches are com-
560 pared: deep CNN features alone, Riesz features alone, early fusion of the
561 feature vectors and late fusion of the class probability for each classifier.
562 Moreover, an ensemble of three deep CNN architectures was found to only
563 add a negligible improvement to the results of a single network. An approach

564 that uses multiple architectures at training time, or using the dropout tech-
565 nique Srivastava et al. (2014) in some of the layers could lead to more sig-
566 nificant improvements. The comparison was made on the same basis using a
567 softmax layer with the same hyperparameters. Furthermore, we assessed the
568 statistical significance of the results of the classifiers by applying the McNe-
569 mar test. The observed p -values are much lower than 0.05, demonstrating
570 important statistical significance of the presented differences. Nevertheless,
571 there is a more significant gap, thus a smaller p -value, between the early and
572 late fusion classifiers. This could be because when the early fusion of the fea-
573 ture vectors is performed, the representations are merged in the intermediate
574 layer of the softmax classifier, leading to more aligned representations than
575 in the case of the late fusion, where we multiply both independent proba-
576 bilities. On the other hand, for the separate feature classifiers, some of the
577 learned features in the early layers of DL classifier likely resemble the filter-
578 banks learned using Riesz aligned texture signatures. This would explain the
579 similar predictions in that particular case.

580 **5. Conclusions**

581 In this paper, we show that late-fusing learned tissue representations
582 based on Riesz and Deep CNN's for texture characterization yields perfor-
583 mance gains over each approach separately or even early fusion. We showed
584 that this is because is not dependent on the feature vector dimensionality but
585 only on the independent probability of the classifiers. We believe that further
586 performance gains can be achieved by investigating new methods of fusing
587 Riesz-based and DL-based features, taking advantage of the complementar-

588 ity of both sources of visual content from the ILD patches (Depeursinge and
589 Müller, 2010).

590 **Acknowledgment**

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

597 **References**

- 598 Ablowitz, M.J., Kaup, D.J., Newell, A.C., 1974. The inverse scattering
599 transform-fourier analysis for nonlinear problems. *Studies in Applied*
600 *Mathematics* 53, 249–315.
- 601 Ablowitz, M.J., Segur, H., 1981. *Solitons and the inverse scattering trans-*
602 *form.* volume 4. SIAM.
- 603 Anthimopoulos, M., Christodoulidis, S., Ebner, L., Christe, A.,
604 Mouggiakakou, S., 2016. Lung pattern classification for interstitial lung
605 diseases using a deep convolutional neural network. *IEEE transactions on*
606 *medical imaging* 35, 1207–1216.
- 607 Aziz, Z.A., Wells, A.U., Hansell, D.M., Bain, G.A., Copley, S.J., Desai,
608 S.R., Ellis, S.M., Gleeson, F.V., Grubnic, S., Nicholson, A.G., Padley,
609 S.P., Pointon, K.S., Reynolds, J.H., Robertson, R.J., Rubens, M.B., 2004.

610 HRCT diagnosis of diffuse parenchymal lung disease: inter-observer vari-
611 ation. *Thorax* 59, 506–511.

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

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

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

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

623 Depeursinge, A., 2017. Multi-Scale and Multi-Directional Biomedical Tex-
624 ture Analysis: Finding the Needle in the Haystack, in: *Biomedical Tex-*
625 *ture Analysis: Fundamentals, Applications and Tools*. Elsevier. Elsevier-
626 MICCAI Society Book series, pp. 29–53.

627 Depeursinge, A., Al-Kadi, O.S., Mitchell, J.R., 2017a. *Biomedical Texture*
628 *Analysis: Fundamentals, Applications and Tools*. Elsevier-MICCAI So-
629 ciety Book series, Elsevier. URL: [https://www.elsevier.com/books/
630 title/author/9780128121337](https://www.elsevier.com/books/title/author/9780128121337).

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

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

636 Depeursinge, A., Foncubierta-Rodríguez, A., Van De Ville, D., Müller, H.,
637 2011a. Lung texture classification using locally-oriented Riesz components,
638 in: Fichtinger, G., Martel, A., Peters, T. (Eds.), Medical Image Computing
639 and Computer Assisted Intervention – MICCAI 2011, Springer Berlin /
640 Heidelberg. pp. 231–238.

641 Depeursinge, A., Foncubierta-Rodríguez, A., Van De Ville, D., Müller, H.,
642 2012a. Multiscale lung texture signature learning using the Riesz trans-
643 form, in: Medical Image Computing and Computer-Assisted Intervention
644 MICCAI 2012, Springer Berlin / Heidelberg. pp. 517–524.

645 Depeursinge, A., Foncubierta-Rodríguez, A., Van De Ville, D., Müller, H.,
646 2014a. Rotation-covariant texture learning using steerable Riesz wavelets.
647 IEEE Transactions on Image Processing 23, 898–908.

648 Depeursinge, A., Foncubierta-Rodríguez, A., Van De Ville, D., Müller, H.,
649 2014b. Three-dimensional solid texture analysis and retrieval in biomedical
650 imaging: review and opportunities. Medical Image Analysis 18, 176–196.

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

- 655 Depeursinge, A., Müller, H., 2010. Fusion techniques for combining textual
656 and visual information retrieval, in: Müller, H., Clough, P., Deselaers, T.,
657 Caputo, B. (Eds.), ImageCLEF. Springer Berlin Heidelberg. volume 32 of
658 *The Springer International Series On Information Retrieval*, pp. 95–114.
- 659 Depeursinge, A., Püspöki, Z., Ward, J.P., Unser, M., 2017c. Steerable
660 Wavelet Machines (SWM): Learning Moving Frames for Texture Classi-
661 fication. *IEEE Transactions on Image Processing* 26, 1626–1636.
- 662 Depeursinge, A., Sage, D., Hidki, A., Platon, A., Poletti, P.A., Unser, M.,
663 Müller, H., 2007. Lung tissue classification using Wavelet frames, in: 29th
664 Annual International Conference of the IEEE Engineering in Medicine and
665 Biology Society, 2007. EMBS 2007., IEEE Computer Society, Lyon, France.
666 pp. 6259–6262.
- 667 Depeursinge, A., Van De Ville, D., Platon, A., Geissbuhler, A., Poletti, P.A.,
668 Müller, H., 2012b. Near-affine-invariant texture learning for lung tissue
669 analysis using isotropic wavelet frames. *IEEE Transactions on Information*
670 *Technology in BioMedicine* 16, 665–675.
- 671 Depeursinge, A., Vargas, A., Platon, A., Geissbuhler, A., Poletti, P.A.,
672 Müller, H., 2012c. Building a reference multimedia database for interstitial
673 lung diseases. *Computerized Medical Imaging and Graphics* 36, 227–238.
- 674 Depeursinge, A., Zrimec, T., Busayarat, S., Müller, H., 2011b. 3D lung image
675 retrieval using localized features, in: *Medical Imaging 2011: Computer-*
676 *Aided Diagnosis*, SPIE. p. 79632E.

- 677 Dietterich, T.G., 1998. Approximate statistical test for comparing supervised
678 classification learning algorithms. *Neural Computation* 10, 1895–1923.
- 679 Do, M.N., Vetterli, M., 2002. Rotation invariant texture characterization
680 and retrieval using steerable wavelet–domain hidden markov models. *IEEE*
681 *Transactions on Multimedia* 4, 517–527.
- 682 Fagerland, M.W., Lydersen, S., Laake, P., 2013. The mcnemar test for bi-
683 nary matched-pairs data: mid-p and asymptotic are better than exact
684 conditional. *BMC medical research methodology* 13, 91.
- 685 Foncubierta-Rodríguez, A., Depeursinge, A., Müller, H., 2012. Using
686 multiscale visual words for lung texture classification and retrieval, in:
687 Greenspan, H., Müller, H., Syeda Mahmood, T. (Eds.), *Medical Content–*
688 *based Retrieval for Clinical Decision Support, Lecture Notes in Computer*
689 *Sciences (LNCS)*. pp. 69–79.
- 690 Freeman, W.T., Adelson, E.H., 1991. The design and use of steerable filters.
691 *IEEE Transactions on Pattern Analysis and Machine Intelligence* 13, 891–
692 906.
- 693 Gao, M., Bagci, U., Lu, L., Wu, A., Buty, M., Shin, H.C., Roth, H., Pa-
694 padakis, G.Z., Depeursinge, A., Summers, R.M., et al., 2016. Holistic
695 classification of ct attenuation patterns for interstitial lung diseases via
696 deep convolutional neural networks. *Computer Methods in Biomechanics*
697 *and Biomedical Engineering: Imaging & Visualization* , 1–6.
- 698 Gonzalez, D.M., Volpi, M., Tuia, D., 2016. Learning rotation invariant con-
699 volutional filters for texture classification. *CoRR* abs/1604.0.

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

703 Guyon, I., Weston, J., Barnhill, S., Vapnik, V., 2002. Gene selection for
704 cancer classification using support vector machines. *Machine Learning* 46,
705 389–422.

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

708 Haralick, R.M., Shanmugam, K., Dinstein, I., 1973. Textural features for
709 image classification. *IEEE Transactions on Systems, Man and Cybernetics*
710 3, 610–621.

711 Hyvärinen, A., Hoyer, P., Inki, M., 2001. Topographic independent compo-
712 nent analysis. *Neural computation* 13, 1527–1558.

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

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

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

- 721 Jia, Y., Shelhamer, E., Donahue, J., Karayev, S., Long, J., Girshick, R.B.,
722 Guadarrama, S., Darrell, T., 2014. Caffe: Convolutional architecture for
723 fast feature embedding, in: Proceedings of the ACM International Confer-
724 ence on Multimedia, MM'14, pp. 675–678.
- 725 Joyseeree, R., Müller, H., Depeursinge, A., 2018. Rotation-covariant tissue
726 analysis for interstitial lung diseases using learned steerable filters: Perfor-
727 mance evaluation and relevance for diagnostic aid. *Computerized Medical*
728 *Imaging and Graphics* 64, 1–11.
- 729 Kreuter, M., Herth, F.J., Wacker, M., Leidl, R., Hellmann, A., Pfeifer, M.,
730 Behr, J., Witt, S., Kauschka, D., Mall, M., et al., 2015. Exploring clinical
731 and epidemiological characteristics of interstitial lung diseases: Rationale,
732 aims, and design of a nationwide prospective registrythe exciting-ild reg-
733 istry. *BioMed research international* 2015.
- 734 Lazebnik, S., Schmid, C., Ponce, J., 2005. A sparse texture representation
735 using local affine regions. *IEEE Transactions on Pattern Analysis and*
736 *Machine Intelligence* 27, 1265–1278.
- 737 LeCun, Y., Bengio, Y., Hinton, G., 2015. Deep learning. *Nature* 521, 436–
738 444.
- 739 LeCun, Y., Huang, F.J., Bottou, L., 2004. Learning methods for generic
740 object recognition with invariance to pose and lighting, in: *Computer*
741 *Vision and Pattern Recognition, 2004. CVPR 2004. Proceedings of the*
742 *2004 IEEE Computer Society Conference on, IEEE*. pp. II–97.

- 743 LeCun, Y., Kavukcuoglu, K., Farabet, C., 2010. Convolutional networks and
744 applications in vision, in: IEEE International Symposium on Circuits and
745 Systems (ISCAS), pp. 253–256.
- 746 Leung, T., Malik, J., 2001. Representing and recognizing the visual appear-
747 ance of materials using three-dimensional textons. *International Journal*
748 *of Computer Vision* 43, 29–44.
- 749 Li, Q., Cai, W., Feng, D.D., 2013. Lung image patch classification with au-
750 tomatic feature learning, in: Engineering in Medicine and Biology Society
751 (EMBC), 2013 35th Annual International Conference of the IEEE, IEEE.
752 pp. 6079–6082.
- 753 Li, Q., Cai, W., Wang, X., Zhou, Y., Feng, D.D., Chen, M., 2014. Med-
754 ical image classification with convolutional neural network, in: Control
755 Automation Robotics & Vision (ICARCV), 2014 13th International Con-
756 ference on, IEEE. pp. 844–848.
- 757 Litjens, G., Kooi, T., Bejnordi, B.E., Setio, A.A.A., Ciompi, F., Ghafoorian,
758 M., van der Laak, J.A., Van Ginneken, B., Sánchez, C.I., 2017. A survey
759 on deep learning in medical image analysis. *Medical image analysis* 42,
760 60–88.
- 761 Lowe, D.G., 2004. Distinctive image features from scale-invariant keypoints.
762 *International Journal of Computer Vision* 60, 91–110.
- 763 Mallat, S.G., 1989. A theory for multiresolution signal decomposition: the
764 wavelet representation. *IEEE Transactions on Pattern Analysis and Ma-*
765 *chine Intelligence* 11, 674–693.

- 766 Masoudnia, S., Ebrahimpour, R., 2014. Mixture of experts: a literature
767 survey. *Artificial Intelligence Review* 42, 275–293.
- 768 Ojala, T., Pietikäinen, M., Mäenpää, T., 2002. Multiresolution gray-scale
769 and rotation invariant texture classification with local binary patterns.
770 *IEEE Transactions on Pattern Analysis and Machine Intelligence* 24, 971–
771 987.
- 772 Otálora, S., Cruz Roa, A., Arevalo, J., Atzori, M., Madabhushi, A., Judkins,
773 A., González, F., Müller, H., Depeursinge, A., 2015. Combining unsuper-
774 vised feature learning and riesz wavelets for histopathology image represen-
775 tation: Application to identifying anaplastic medulloblastoma, in: Navab,
776 N., Hornegger, J., Wells, W.M., Frangi, A. (Eds.), *Medical Image Com-
777 puting and Computer-Assisted Intervention — MICCAI 2015*. Springer
778 International Publishing. volume 9349 of *Lecture Notes in Computer Sci-
779 ence*, pp. 581–588.
- 780 Porter, R., Canagarajah, N., 1997. Robust rotation-invariant texture classi-
781 fication: wavelet, Gabor filter and GMRF based schemes. *IEE Proceedings
782 on Vision, Image and Signal Processing* 144, 180–188.
- 783 Randen, T., Husoy, J.H., 1999. Filtering for texture classification: a com-
784 parative study. *IEEE Transactions on Pattern Analysis and Machine In-
785 telligence* 21, 291–310.
- 786 Rozantsev, A., S.M., Fua, P., 2016. Beyond Sharing Weights for Deep Domain
787 Adaptation. ArXiv e-prints [arXiv:1603.06432](https://arxiv.org/abs/1603.06432).

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

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

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

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

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

804 Srivastava, N., Hinton, G., Krizhevsky, A., Sutskever, I., Salakhutdinov, R.,
805 2014. Dropout: a simple way to prevent neural networks from overfitting.
806 *The Journal of Machine Learning Research* 15, 1929–1958.

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

- 811 Unser, M., Chenouard, N., Van De Ville, D., 2011. Steerable pyramids and
812 tight wavelet frames in $L_2(\mathbb{R}^d)$. IEEE Transactions on Image Processing
813 20, 2705–2721.
- 814 Unser, M., Van De Ville, D., 2010. Wavelet steerability and the higher-order
815 Riesz transform. IEEE Transactions on Image Processing 19, 636–652.
- 816 Vapnik, V.N., 1995. The Nature of Statistical Learning Theory. Springer,
817 New York.
- 818 Vizitiu, A., Itu, L., Joyseeree, R., Depeursinge, A., Müller, H., Suciu, C.,
819 2016. Gpu-accelerated texture analysis using steerable riesz wavelets, in:
820 24th IEEE Euromicro International Conference on Parallel, Distributed,
821 and Network-Based Processing, pp. 431–434.
- 822 Watadani, T., Sakai, F., Johkoh, T., Noma, S., Akira, M., Fujimoto, K.,
823 Bankier, A.A., Lee, K.S., Müller, N.L., Song, J.W., et al., 2013. Inter-
824 observer variability in the ct assessment of honeycombing in the lungs.
825 Radiology 266, 936–944.
- 826 Xu, D.H., Kurani, A.S., Furst, J., Raicu, D.S., 2004. Run-length encod-
827 ing for volumetric texture, in: The 4th IASTED International Conference
828 on Visualization, Imaging, and Image Processing – VIIP 2004, Marbella,
829 Spain. pp. 105–110.
- 830 Xu, Y., Yang, X., Ling, H., Ji, H., 2010. A new texture descriptor using mul-
831 tifractal analysis in multi-orientation wavelet pyramid, in: Proceedings of
832 the IEEE Computer Society Conference on Computer Vision and Pattern
833 Recognition, pp. 161–168.

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