



Fusing visual and clinical information for lung tissue classification in high-resolution computed tomography

Adrien Depeursinge^{a,*}, Daniel Racoceanu^b, Jimison Iavindrasana^a, Gilles Cohen^a, Alexandra Platon^c, Pierre-Alexandre Poletti^c, Henning Müller^{a,d}

^a Medical Informatics Service, Geneva University Hospitals and University of Geneva (HUG), Geneva, Switzerland

^b Image & Pervasive Access Lab (IPAL), Institute for Infocomm Research (I2R), Singapore

^c Emergency Radiology Service, Geneva University Hospitals and University of Geneva (HUG), Geneva, Switzerland

^d Business Information Systems, University of Applied Sciences Western Switzerland (HES-SO), Sierre, Switzerland

ARTICLE INFO

Article history:

Received 19 September 2008

Received in revised form 3 September 2009

Accepted 29 March 2010

Keywords:

Multimodal information fusion
Contextual image analysis
Feature ranking
Support vector machines
Wavelet-based texture analysis
Lung tissue classification
Computer-aided diagnosis
High-resolution computed tomography
Interstitial lung diseases

ABSTRACT

Objective: We investigate the influence of the clinical context of high-resolution computed tomography (HRCT) images of the chest on tissue classification.

Methods and materials: 2D regions of interest in HRCT axial slices from patients affected with an interstitial lung disease are automatically classified into five classes of lung tissue. Relevance of the clinical parameters is studied before fusing them with visual attributes. Two multimedia fusion techniques are compared: early versus late fusion. Early fusion concatenates features in one single vector, yielding a true multimedia feature space. Late fusion consisting of the combination of the probability outputs of two support vector machines.

Results and conclusion: The late fusion scheme allowed a maximum of 84% correct predictions of testing instances among the five classes of lung tissue. This represents a significant improvement of 10% compared to a pure visual-based classification. Moreover, the late fusion scheme showed high robustness to the number of clinical parameters used, which suggests that it is appropriate for mining clinical attributes with missing values in clinical routine.

© 2010 Elsevier B.V. All rights reserved.

1. Introduction

The interpretation of high-resolution computed tomography (HRCT) images of the chest from patients affected with interstitial lung diseases (ILDs) is challenging and time-consuming even for experienced radiologists. The term interstitial lung disease accounts for around 150 illnesses of which many forms are rare. Images play an important role for confirming the diagnosis and patients may not require surgical lung biopsy when the clinical and radiographic impression is consistent with a safe diagnosis [1]. The most common imaging procedure used is the chest X-ray because of its low cost and weak radiation exposure. However, chest X-rays are negative in a large proportion of diseases and often unspecific where HRCT of the chest contains essential visual data for the characterization of lung tissue patterns associated with ILDs [2]. The three-dimensional form of HRCT data requires significant reading time, effort, and experience for a correct interpretation. Owing to this intrinsic complexity of the interpretation of HRCTs, an image-based computerized diagnostic aid tool (CAD) can bring

quick and precious information, particularly to less experienced radiologists and non-chest experts [3–5]. Moreover, a radiologist's ability to interpret HRCT images is likely to change based on the domain-specific experience, human factors and time of the day where computerized image analysis is 100% reproducible.

1.1. The clinical context of HRCT images

When analyzing an image, one interprets its content according to a given context. This is particularly true when analyzing medical images. Radiologists do never interpret HRCT images without taking into account the clinical context. For example, discovering some fibrotic findings in a lung belonging to an 80-year-old patient is not as surprising as finding some in a lung of a 25-years-old young person. Several clinical parameters – in particular the age of the patient (see Fig. 1) – have a major influence on the visual aspect (density) of HRCT images of the chest [6,7]. In Fig. 1, one can see that healthy tissue from the 88-year-old man has lower mean density with more pre-fibrotic lesions compared to the homogeneous healthy tissue of the 25-year-old man. To accurately analyze HRCT images an image-based computerized diagnostic aid system for ILDs must integrate the clinical context of the images.

* Corresponding author. Tel.: +41 22 372 8875; fax: +41 22 372 8879.
E-mail address: Adrien.Depeursinge@sim.hcuge.ch (A. Depeursinge).

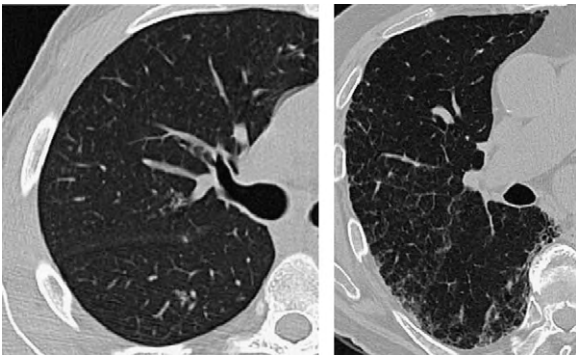


Fig. 1. Healthy tissue from a 25-year-old man on the left, and from an 88-year-old man on the right. Both images have identical window level settings.

1.2. Contextual image analysis

Although fundamental in almost every medical field, the context is rarely used in computer vision applications. On the one hand, collecting contextual information beside images is usually time-consuming, requiring the help of a specialist. On the other hand, a high-level of knowledge of the application domain is required to find relevant contextual parameters [8]. The selection of parameters for contextual medical image analysis has to be carried out based on domain-specific literature along with knowledge bases of computer-based diagnostic decision support systems.

Contextual image analysis implies the fusion of multimodal information sources. When carrying out contextual image analysis of lung tissue in HRCT data, the visual information extracted from HRCT image series is combined with the clinical parameters of the corresponding patient. Integrating information from multiple modalities consists of two major steps [9]. Firstly, the best modalities have to be identified. The best modalities have to be informative according to the considered classes along with being complementary among each other. Each modality k is represented by a set of features ν_k . Secondly, the information from the best modalities must be combined with an optimal scheme in order to allow for synergy. The so-called “fusion” can be carried out according to two main strategies [10,11]:

- early fusion, where features are concatenated into one vector $\nu = (\nu_1, \dots, \nu_k)$ to create one unique feature space,
- late fusion, where multiple classifiers h_k are built on each modality $\{\nu_1, \dots, \nu_k\}$ and the fusion is carried out at the decision level.

Early fusion allows for a true multimedia (images and clinical data) representation. One single classifier can learn from all information sources. However this method is confronted with the *curse of dimensionality* because the dimension of the resulting feature space ν is equal to the sum of the dimensions of the subspaces ν_k . Even associated with feature weighting, high-dimensional spaces tend to scatter the homogeneous clusters of instances belonging to the same class. This is particularly true when negative synergies occur among features [12,13].

Combining classifiers has been a very active domain over the past ten years [14,15]. Wide interest for the latter mainly relies on the assumption that the heterogeneity of classifiers h_k leads to better results [16]. Late fusion is a special case of classifier combination where heterogeneity occurs in the input spaces. It allows for a reduction of the data dimensionality by dividing the feature space without neglecting information contained in features that would be discarded by classical feature selection methods [17].

1.3. Related work

Context has been used in content-based image retrieval (CBIR) where information from textual annotations of images was fused with image features [18] (i.e. grey-level histograms and texture features). In [19], a CBIR system combined visual statistics with textual statistics directly in the feature vector space representation. Inter-media medical image retrieval was carried out in [20] using textual features semantically parsed and described with the Unified Medical Language System (UMLS) along with color and texture features. The visual and textual information was combined in the calculation of the similarity measure. Investigation of the effectiveness of combining text and image for retrieval including medical image retrieval is one of the main goals of the CBIR benchmarking campaign ImageCLEF¹ [21,22].

Combined decision of classifiers constructed on sequentially selected sets of features were tested on four datasets including medical data in [17]. Best results were obtained when the combination approach was applied on top of feature selection. Unfortunately, experiments were carried out with homogeneous datasets which did not contain heterogeneous features such as visual, textual, audio, etc.

Early fusion of clinical parameters and genetic factors was used to predict the risk of coronary artery disease in [23]. Interaction among features were studied using Bayesian network representations. Although visually identified feature groups were in accordance with medical knowledge, no quantitative analysis of the interactions was performed.

A combination of radiologic findings on chest radiographs and clinical parameters to provide probability output of 11 possible ILDs using an artificial neural network is carried out in [24]. By using these probabilities, radiologists were able to significantly improve their diagnosing accuracies. However, automatic detection of relevant patterns in the chest radiographs was not investigated.

Utilization of knowledge of disease location to improve detection of *fibrosis* patterns in HRCT data was carried out in [25]. The locations of the patterns showed to significantly improve detection performance but require an accurate segmentation of the anatomy of the lung.

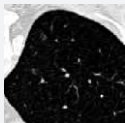
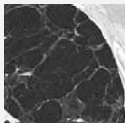

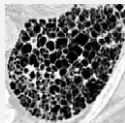
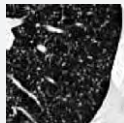
Many image-based diagnostic aid systems for ILDs achieved high classification accuracy of lung tissue patterns in HRCT data [4,26,27] and showed to be effective in clinical routine [5]. Yet, most of these systems are based on visual data only. To our knowledge no system attempted to integrate clinical parameters for automatic detection of lung tissue patterns associated with ILDs in HRCT data. Texture analysis of lung images using Wavelet frames was investigated in [28,29] and support vector machines (SVM) showed to be optimal for the categorization of lung tissue using quincunx Wavelet frames in [30]. In this paper, we study the influence of the integration of the clinical context of HRCT images on classification performance of 2D regions of interest (ROI) drawn by two radiologists in axial slices from patients affected with an ILD.

The paper is structured as follows. In Section 2, the dataset and software used for evaluating the influence of the clinical context on lung tissue classification in HRCT images is described. Section 3 is divided into two parts. Section 3.1 describes the composition of the multimodal feature space whereas the comparison of the fusing techniques is carried out Section 3.2. Results are interpreted and discussed in Section 4, future work is proposed in Section 5 and final conclusions are drawn in Section 6.

¹ <http://www.imageclef.org/> (Accessed: 28 August 2009).

Table 1

Distribution of the ROIs and patients per class of lung tissue pattern. Each patient may contain several kinds of lung tissue patterns.

Visual aspect					
Class	Healthy	Emphysema	Ground glass	Fibrosis	Micronodules
# of ROIs	63	58	148	312	155
# of patients	5	4	14	28	5

2. Methods

The dataset used to investigate the influence of clinical parameters on classification accuracy of lung tissue is part of an internal multimedia database of ILD cases containing HRCT images created in Talisman.² 99 Relevant clinical parameters were chosen according to the 15 most frequent ILDs [31] based on the literature [32,33], along with knowledge bases of computer-based diagnostic decision support systems. Discussions and remarks from lung specialists, radiologists and the medical informatics service (SIM³) at the University Hospitals of Geneva (HUG) allowed an iterative review of the selected parameters as well as standardized units and data format to be used. The parameters that were not available from the electronic health record (EHR) were removed. An HTML form and PHP scripts were used to collect the clinical parameters and to store them into a MySQL database. When multiples instances of clinical parameters (e.g. laboratory data) were available in the EHR, the instance as close as possible to HRCT examinations was retained. 96 Patients with confirmed diagnosis were retrospectively collected at the HUG between 2003 and 2006. For each patient, a physician filled as many clinical parameters as possible and a total of 1104 ROIs of lung tissue patterns were drawn in full-resolution DICOM images by two experienced radiologists. The slice thickness of the images is limited to 1 mm as this is the current clinical protocol. A graphical user interface implemented in Java was developed in order to meet the needs of the radiologists for the various annotation tasks.

736 ROIs from healthy and four pathologic lung tissue patterns belonging to 48 patients with filled clinical parameters were selected for this study (see Table 1). Patterns that are represented by less than 4 patients are left aside. The selected patterns are *healthy*, *emphysema*, *ground glass*, *fibrosis* and *micronodules*. Distributions of the classes are highly imbalanced as the largest class *fibrosis* contains 312 ROIs and the smallest class only 58 ROIs. There is a mean of 147.2 ROIs per class.

Implementation of the SVMs' C-support vector classification is taken from the open source Java library *Weka*⁴ using a wrapper for *LIBSVM*.⁵ The image feature extraction and the optimization of SVMs is implemented in Java. Quincunx Wavelet frames are implemented in Java [34].

3. Results

The first part of this section describes the composition of the features in each modality used for the classification of the lung tissue patterns. The second part describes and details the results of the fusion of the two modalities.

3.1. Modalities

The two modalities used for classifying the lung tissue patterns are made up of the visual information from HRCT images of the chest as well as the corresponding textual information describing the clinical state of the patient at the time of the disease episode.

3.1.1. Clinical features c_n

The clinical parameters entered in the MySQL database are not directly usable for data-mining. Pre-processing steps are required to build a workable feature space. Nominal variables are divided into binary features. Textual variables and binary variables that contained one single modality are left aside. Since leaving aside cases with missing values is not conceivable, variables with less than 50% of the values filled were removed and average values were substituted. *unknown* was used when the clinical parameter was not detailed in the EHR. After having gathered binary and continuous variables, the created clinical feature space contains 72 attributes (63 binary and 9 continuous). For example, over the 48 selected patients, the parameter *host_HIV* has 3 *yes*, 43 *no* and 2 *unknown* values. *yes* values are coded with 1, *no* with 0 and thus the missing values are substituted by the mean: 0.065. The continuous features were not discretized as it is preferable not to group their values together into categories with a further purpose of separating clusters of instances in the feature space. The mean filling rate of the retained attributes is 88.7%.

3.1.2. Visual features t_m

Visual features consist of grey level features with 22 bins of grey-level histograms of Hounsfield Units (H.U.) within the ROIs, along with texture features using quincunx Wavelet frame (QWF) coefficients extracted at 8 scales. The distributions of the wavelet coefficients in each subband i are characterized through the parameters of mixtures of two Gaussians. With fixed means $\mu_{1,2}^i = \mu^i$, the standard-deviations $\sigma_{1,2}^i$ ($\sigma_1^i > \sigma_2^i$) are estimated using the expectation-maximization (EM) algorithm. An additional feature *airpix* measuring the number of pixels of the ROI with value inferior to -1000 H.U. (which corresponds to the density of air) is used. In total, the visual feature set is composed of 47 attributes. A complete description and evaluation of the visual feature space can be found in [29].

3.2. Combining features: early versus late fusion

In order to study the effect of the integration of the clinical context of HRCT images on the classification accuracy of the lung tissue patterns, optimized SVMs with a Gaussian kernel are used to categorize ROIs from the multimodal feature space. SVMs with a Gaussian kernel have shown to be effective to categorize lung tissue patterns from visual features in [30] and are adapted to mine clinical parameters as shown in [35]. Two methods for combining visual and clinical attributes are compared: early versus late fusion (Sections 3.2.2 and 3.2.3). The relevance of clinical attributes is studied in Section 3.2.1.

² TALISMAN: Texture Analysis of Lung ImageS for Medical diagnosis AssistaNce, http://www.sim.hcuge.ch/medgift/01_Talisman_EN.htm (Accessed: 28 August 2009).

³ <http://www.sim.hcuge.ch/> (Accessed: 28 August 2009).

⁴ <http://www.cs.waikato.ac.nz/ml/weka/> (Accessed: 28 August 2009).

⁵ <http://www.csie.ntu.edu.tw/~cjlin/libsvm/> (Accessed: 28 August 2009).

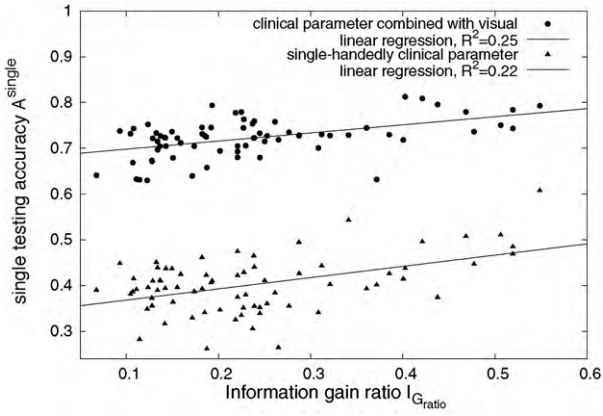


Fig. 2. Correlation of $I_{G_{ratio}}$ and A^{single} .

3.2.1. Ranking the clinical attributes

Integrating the clinical context in lung tissue classification implicitly assumes that clinical parameters contain relevant information to predict the types of lung tissue contained in HRCT image series of a patient affected with an ILD. Although parameters such as *age* are clearly related to the visual aspect of the lung tissue (see Section 1.1), dependencies between clinical attributes and classes of lung tissue must be investigated before any fusion with the visual features. Indeed, due to missing values, binarization or irrelevance according to the studied diseases, some features might introduce noise by scattering homogeneous clusters of instances in the feature space. High presence of binary attributes increases the risk of obtaining XOR configurations of instances, which leads to highly non-linear decision boundaries. Moreover, it is preferable to keep as few as possible features to limit the curse of dimensionality, especially for early fusion. A feature ranking is thus required to build an effective set of attributes.

Two measures are compared in their ability to rank the clinical attributes for lung tissue classification: the information gain ratio $I_{G_{ratio}}$ and the single testing accuracy A^{single} . $I_{G_{ratio}}$ is derived from the information gain measure I_G originally used by Quinlan in decision trees in [36]. The information gain $I_G(Y|X)$ of a given attribute X

with respect to the class attribute Y quantifies the change in information entropy when the value of X is revealed:

$$I_G(Y|X) = H(Y) - H(Y|X) \quad (1)$$

The information entropy $H(Y)$ measures the uncertainty about the value of Y and the conditional information entropy $H(Y|X)$ measures the uncertainty about the value of Y when the value of X is known:

$$H(Y) = - \sum_{y \in \mathcal{Y}} p(y) \log p(y) \quad (2)$$

$$H(Y|X) = - \sum_{x \in \mathcal{X}, y \in \mathcal{Y}} p(x, y) \log p(y|x) \quad (3)$$

The information gain ratio $I_{G_{ratio}}$ is derived from I_G using

$$I_{G_{ratio}}(Y|X) = \frac{I_G(Y|X)}{- \sum_{i=1}^l |T_i|/|T| \log(|T_i|/|T|)} \quad (4)$$

with T the training set and l the number of possible values of X . Compared to I_G the gain ratio will not give advantage to attributes with a high range of possible values [37,38]. As the clinical feature space is populated with binary as well as continuous attributes, it is highly preferable to use the $I_{G_{ratio}}$ for ranking. Another measure proposed for ranking the clinical attributes is the single testing accuracy A^{single} . A^{single} is defined as the classification accuracy with SVMs (see Section 3.2.4) based on a feature vector concatenating all visual features t_m with the studied clinical feature c_n :

$$v = (t_1 \dots t_M c_n) \quad (5)$$

The correlation of $I_{G_{ratio}}$ and A^{single} obtained using the experimental setup described in Section 3.2.4 is studied in Fig. 2. Table 2 lists the first 20 clinical attributes with highest A^{single} value. The correlation matrix of the feature space containing the visual features along with the first 20 clinical attributes is shown in Fig. 3.

3.2.2. Early fusion: feature concatenation

In order to create a multimodal feature space, clinical attributes and visual features are normalized and concatenated into one single feature vector v as follows:

$$v = (t_1 \dots t_M \quad c_1 \dots c_N) \quad (6)$$

with t_m , $m \in [1; M]$ the visual features and c_n , $n \in [1; N]$ the clinical attributes. Using all 72 clinical attributes, the maximum

Table 2

List of the first 20 clinical attributes with highest A^{single} when combined with visual features. Abbreviations: HTA: arterial hypertension, subOAP: acute pulmonary edema, LDH: serum lactate dehydrogenase.

Rank	A^{single}	$I_{G_{ratio}}$	Name	Type
1	0.813	0.403	laboratory_hematocrit	Continuous
2	0.809	0.421	age	Continuous
3	0.796	0.438	laboratory_hemoglobin	Continuous
4	0.794	0.193	past_medical_allergy	Binary
5	0.793	0.549	findings_physical_generals_lymph	binary
6	0.784	0.519	past_medical_lymphom	Binary
7	0.779	0.469	findings_physical_generals_fever	Binary
8	0.779	0.225	medication_cordarone	Binary
9	0.778	0.218	host_diabetes	Binary
10	0.763	0.227	biopsy_bronchoscopy_transbronchial_eosinophil	Binary
11	0.759	0.239	past_medical_HTA	Binary
12	0.758	0.261	past_medical_dyspnea_attack	Binary
13	0.754	0.237	past_medical_subOAP	Binary
14	0.752	0.124	findings_physical_respiratory_tachypnea	Binary
15	0.751	0.506	host_chemotherapy	Binary
16	0.746	0.182	past_medical_wheezing	Binary
17	0.745	0.192	findings_physical_abdominals_liver	Binary
18	0.745	0.227	biopsy_bronchoscopy_transbronchial_interstitial_fibrosis	Binary
19	0.745	0.361	laboratory_LDH	Continuous
20	0.743	0.52	host_hemopathy	Binary

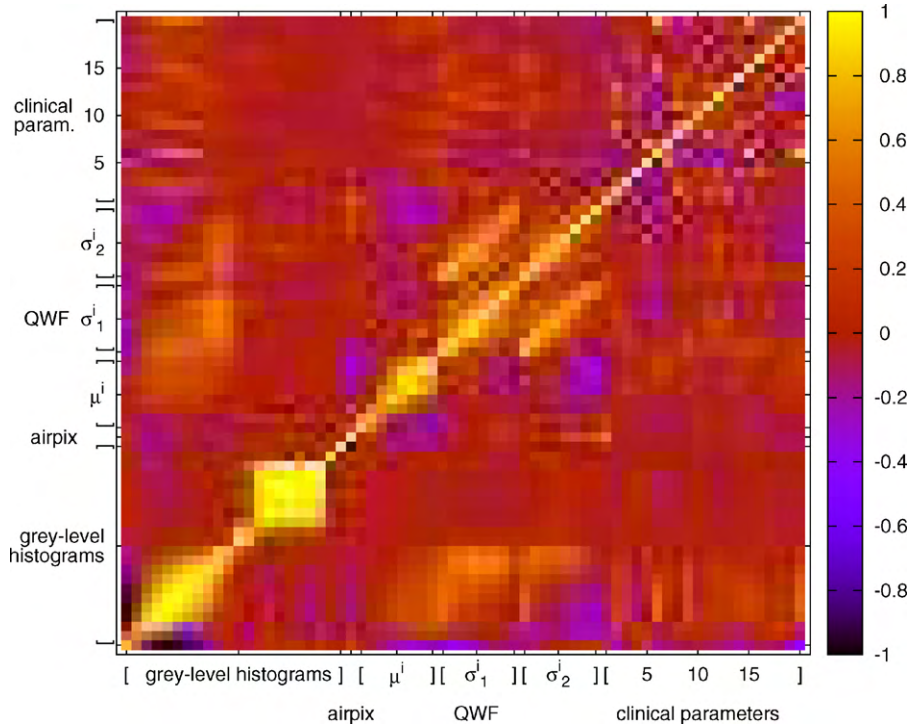


Fig. 3. Correlation matrix of the combined feature space. Indexes of the clinical parameters corresponds to their rank described in Table 2.

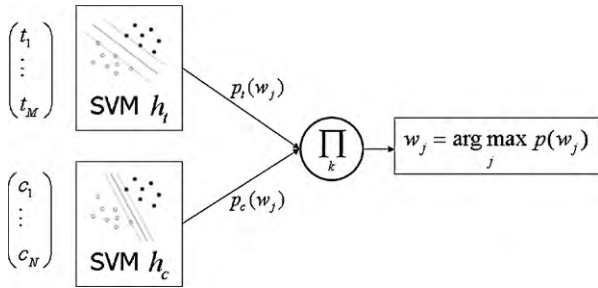


Fig. 4. Classification scheme for late fusion. Two expert SVM classifiers h_t and h_c output probabilities $p_k(w_j)$ which are multiplied to obtain the final probability of each class w_j .

dimensionality of the multimodal feature space reaches 119 with 56 continuous and 63 binary features. v is used as input of SVMs which directly output the predicted class using one versus one multiclass approach. Testing accuracies are obtained using the experimental setup described in Section 3.2.4.

3.2.3. Late fusion: combination of SVM classifiers

Two SVM classifiers h_t and h_c are trained using visual features t and clinical attributes c respectively. Attributes in t and c are normalized (within each group) in order to give equal importance to each of them. Both SVM output probabilities p_t and p_c using pairwise coupling [39]. For each class w_j ($j = 1, \dots, 5$), probabilities are multiplied to compute final probability $p(w_j) = p_t(w_j) \cdot p_c(w_j)$. The final predicted class w_j is given by $\text{argmax}_j p(w_j)$. Using the product of probabilities for predicting the final class assumes that the modalities t and c are conditionally statistically independent [14,40], which is admissible as mean correlation value ρ_{mean} of each feature pair (t_m, c_n) is equal to 0.0143 (see Fig. 3). The late fusion scheme is summarized in Fig. 4.

Smaller feature subspaces have the advantage to reduce the computational complexity for solving the quadratic problem of finding the maximum margin hyperplane of SVMs. Moreover the

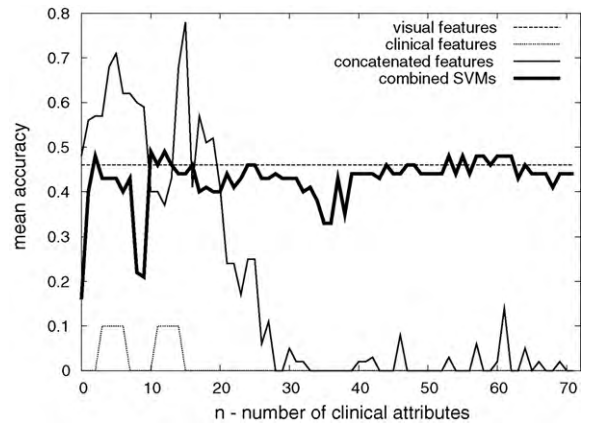
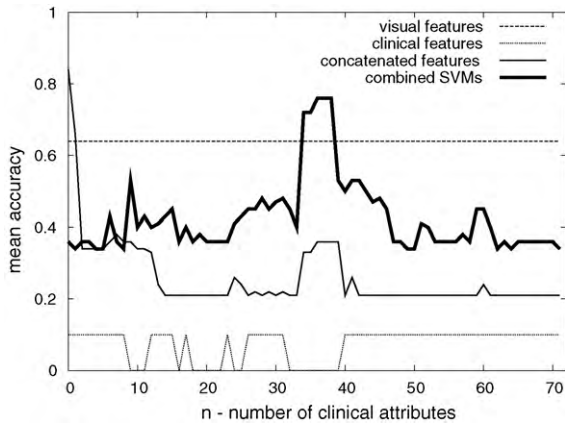
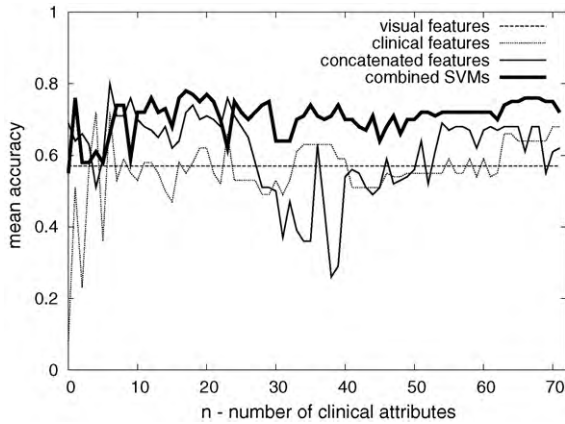


Fig. 5. Classification accuracies for class healthy.

subspaces can be processed in parallel to allow for faster training of the SVMs.

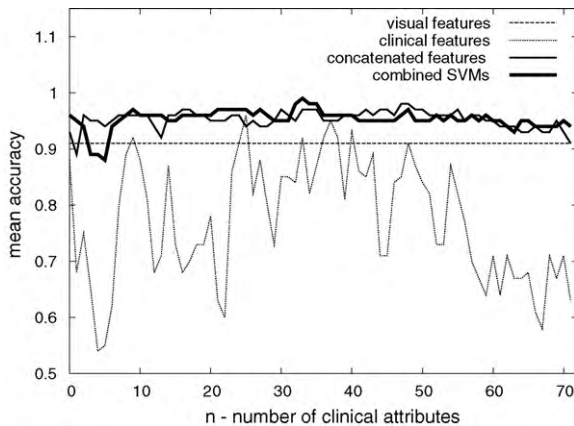
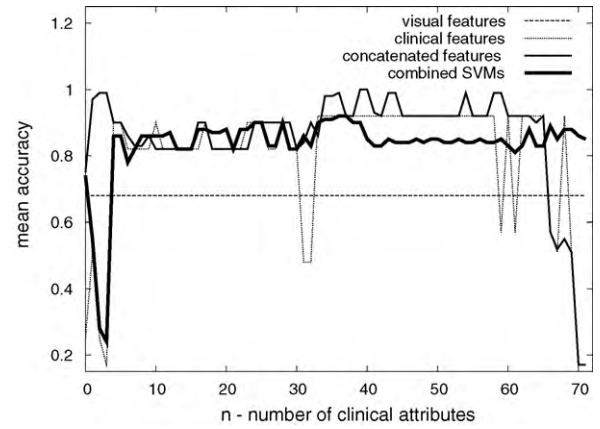
3.2.4. Experimental setup

In order to test the influence of clinical parameters on classification accuracy of the 2D ROIs, a leave-one-patient-out cross-validation was used. The latter is in accordance with a clinical usage of the CAD. Each of the ROIs belonging to one same patient are left aside for testing and all remaining ROIs are used to train and optimize the SVMs. Indeed, the training set is used both for grid search for optimal parameters and adjustment of the maximum-margin hyperplane of the SVMs. Optimized parameters of the SVM are the cost of the errors C and the width σ of the Gaussian kernel. A grid search is carried out within the intervals $C \in [1; 100]$ and $\sigma \in [10^{-2}; 10^2]$. For every coordinate of the grid, a 10-fold cross-validation (CV) is carried out on the training set. Optimal parameters (C_{opt}, σ_{opt}) that allowed best mean CV accuracy A^{CV} are used to train the final model on the entire

Fig. 6. Accuracies for class *emphysema*.Fig. 7. Accuracies for class *ground glass*.

training set. A preliminary coarse grid search was performed to locate regions of the space with high A^{CV} values.

As the toughness of the classification task can strongly vary depending on the draw of the testing patient, the global experimentation is repeated for all 48 patients to obtain reliable accuracy values. In order to study the optimal number n of clinical attributes to be used, mean classification accuracies over the 48 patients are computed for each $n \in [1; 72]$, with clinical attributes ordered by A^{single} values (see Section 4.1). Mean classification accuracies according to n obtained with the test set using visual features only, clinical features only and combined features with

Fig. 8. Accuracies for class *fibrosis*.Fig. 9. Accuracies for class *micronodules*.

early and late fusion are shown in Figs. 5–9 for each class. Global accuracies of each method are summarized in Fig. 10. Mean accuracies over n values as well as classification based on principal component analysis (PCA) are contained in Table 3. First, a PCA transform is applied only on the clinical feature set c_n to be used with combined SVMs. Second, a PCA transform is applied on the whole concatenated feature set. The number of principal components P kept is chosen according to [41]:

$$P > 1 + 2\sqrt{\frac{N-1}{K-1}} \quad (7)$$

with N the number of features and K the number of instances. In both cases, $P = 2$ is chosen based on (7).

4. Interpretation

The first part of this section discusses the limitations of the ranking measures $I_{G_{ratio}}$ and A^{single} . The second part verifies the relevance of the 20 clinical attributes with highest A^{single} value to the medical domain. The third part studies the consistency and the complementarity of the multimodal feature space through the correlation matrix presented in Fig. 3. The last section interprets the performances of classification of the two fusion designs presented in Table 3 and Figs. 5–10.

4.1. Measures for ranking

Fig. 2 shows that $I_{G_{ratio}}$ is little correlated to both A^{single} and testing accuracy obtained with each single-handed clinical

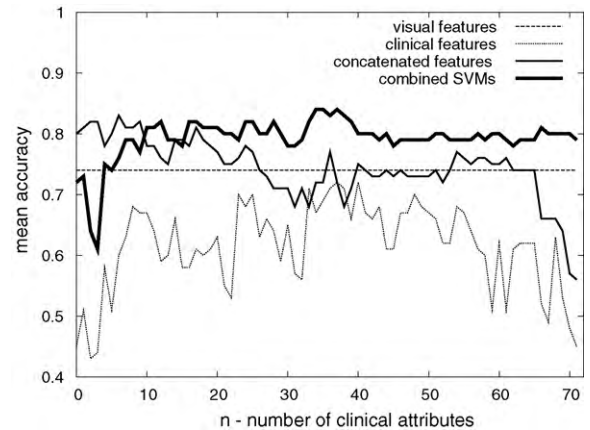


Fig. 10. Global classification accuracies.

Table 3

Averaged accuracies obtained with the various techniques. Best performances are highlighted in bold.

	Visual features	Clinical features	Concatenated features	Combined SVMs	Combined SVMs, PCA on clinical features	PCA on concatenated features
Healthy	0.46	0.01	0.19	0.43	0.48	0.22
Emphysema	0.64	0.08	0.26	0.43	0.47	0.78
Ground glass	0.57	0.56	0.6	0.71	0.62	0.5
Fibrosis	0.91	0.77	0.95	0.95	0.95	0.87
Micronodules	0.68	0.8	0.86	0.83	0.45	0.21
Global	0.74	0.61	0.74	0.79	0.72	0.58

parameter. Pearson's coefficient of regression R^2 is below 0.25 for both comparisons. Even if A^{single} is averaged over 30 experiments, the values obtained still have high variance according to the drawdowns of the training and testing sets. This is a first explication for having low values of R^2 . A second explication comes with the definition of $I_{G_{ratio}}$ which measures the relevance of each separated single attribute. One feature generally distinguishes classes in combination with other features [42], which suggests that $I_{G_{ratio}}$ is not convenient to rank the clinical attributes with a purpose of fusing them with visual features. $I_{G_{ratio}}$ is also known to be unstable as it is very sensitive to small changes in the training set [16] which is not desirable for ranking attributes from a high-dimensional set of heterogeneous features. Due to the several drawbacks of $I_{G_{ratio}}$, A^{single} is used for ranking the clinical attributes.

4.2. Relevance of clinical attributes

The relevance of the clinical attributes for classifying lung tissue patterns in HRCT data is subject to many external factors such as the availability of the parameters in the EHR, its binarization required to be added to \mathbf{v} and relevance according to the studied diseases. Indeed a parameter such as the result of a lung biopsy is obviously highly informative for characterizing the lung tissue but is rarely carried out and available in the EHR. The categorization and binarization has also major influence on the quality of clinical data. At last, the relevance of the parameter according to the studied diseases is of course primordial.

As observed in Section 1.1, the age has an important influence on the visual aspect of lung tissue (see Fig. 1) and this is confirmed by finding it at the 2nd rank in Table 2. The presence of the parameter *laboratory_hematocrit* at the top of the list is a bit more subtle. An explication for this is that large homogeneous regions of air, characterizing *emphysema* patterns, will cause hypoxia and may elicit an increased production of red blood cells by the kidney, and thus increase the level of hematocrit. This phenomenon is indeed commonly observed in cases affected with chronic obstructive pulmonary disease (COPD) [43], characterized by HRCT images showing *emphysema* patterns. The latter observation is firmly confirmed by looking at the correlation matrix in Fig. 3, where the first clinical parameter (*laboratory_hematocrit*) is strongly anticorrelated with the means μ_i of the QWF and highly correlated to *airpix* the number of pixels of air within the ROIs. Indeed the means μ_i of the QWF have high values for inhomogeneous patterns, where *emphysema* patterns are very homogeneous due to absence of lung tissue. Coherently, the 3rd rank is occupied by the parameter *laboratory_hemoglobin* which is also involved in the transport of oxygen. Indeed, hemoglobin is the protein contained in red blood cells that is responsible for delivery of oxygen to the tissues. To ensure adequate tissue oxygenation, a sufficient hemoglobin level must be maintained.

The presences of parameters *findings_physical_general_lymph* (enlargement of lymph node(s)) and *findings_physical_general_fever* at the 5th and 7th ranks are not surprising as they usually

highlights the presence of a host illness. Finding the parameter *medication_cordarone* at the 8th rank is in accordance with the well-known side effect of the cordarone drug creating pulmonary fibrosis on the long range.

4.3. Consistency of the multimodal feature space

The study of the correlation of the multimodal feature space is carried out in Fig. 3. A first look at the correlation matrix shows that clinical features have little correlation with the visual features. This is confirmed as mean correlation ρ_{mean} is equal to 0.0143. Several homogeneous groups can be identified within the visual features. The first histogram bin representing pixels values within $[-1050; -975]$ H.U. is of course highly correlated with *airpix* and is anticorrelated with bins in range $[-900; -450]$. This partly due to patterns with low-density tissues (mostly *emphysema*) are mainly composed by air and thus do contain few pulmonary tissue in range $[-900; -450]$. Globally, histograms are logically correlated in contiguous pairs. Bins 14–20 form a strongly correlated group ($\rho_{mean} = 0.95$) which shows that high-density tissues with H.U. values in $[-75; 375]$ only occur together, most probably in *fibrosis* and *ground glass* patterns. Bins 21 and 22 are very sparse and thus not correlated to any other attribute. Within the QWF features, two groups can be identified: the means μ^i and the standard-deviations $\mu_{1,2}^i \cdot \rho_{mean}$ is equal to -0.04 between the two groups. Within the groups, it is not surprising to observe that means and standard-deviations of two consecutive Wavelet subbands are correlated.

Within the clinical parameters, 3 are highly correlated: *past_medical_lymphom*, *host_chemotherapy* and *host_hemopathy* ($\rho_{mean} = 0.84$). This is not surprising as *past_medical_lymphom*, which stands for having had a lymphoma or leukemia, which is a type of hemopathy (blood cancer), is treated with chemotherapy. Those three parameters are all involved in ILDs as chemotherapy can induce diverse injuries of the lung tissue [44].

4.4. Influence of the clinical context on lung tissue classification: early versus late fusion

Influence of the clinical context of HRCT images on lung tissue classification accuracy is studied in Figs. 5–10. As baseline performance, let us consider the accuracy achieved by using visual features only which has a mean global value of 74%. The last line of Table 3 shows that integrating the clinical context of the images allows for significant global improvements of the classification accuracies. In the mean, 5% is gained in global accuracy using combined SVMs compared to using visual features only. Classification accuracies of combined modalities are on average always superior to single modalities as observed in Table 3. However, the clinical features can harm the classification accuracy if they are not integrated using an appropriate fusing technique. With early fusion, positive interactions between clinical and visual features are allowed as shown in Fig. 9 and for low n values in Fig. 5. However, concatenating all features in a single vector has the

drawback that less informative attributes scatter homogeneous clusters of instances in the feature space. Figs. 5, 6 and 10 confirm this phenomenon, where the curve of the concatenated features drops when adding more noisy clinical attributes with low discriminatory power. Some clinical features have negative interactions among themselves as well as with the visual ones. Separating visual and clinical features for mining using late fusion avoid interactions between the feature groups which show more stable performances compared to early fusion. This is particularly true when clinical attributes carry little information as it is the case for classes *healthy*. Moreover, the combined SVMs show high robustness towards the number n of clinical parameters used and allowed the best accuracy of 84% correct predictions of testing instances (ROIs) among the five classes of lung tissue with an optimal number of clinical attributes $n = 35$. Although the PCA transform allows best results for classes where clinical attributes have low discriminatory power (see Table 3), it does not improve the global accuracy.

5. Future work

We believe that negative synergies still occur among features using a late fusion scheme. Groups of features with positive synergy [13], which allow for homogeneous clusters of instances belonging to the same class have to be identified and mined into separated subspaces. Indeed, part of the fluctuations among the performances according to n are the results of interactions among the various groups of features. An approach for identifying feature groups with positive synergy based on mutual information is described in [12]. Identifying the groups should also include medical knowledge. A visual approach based on Bayesian networks is proposed in [23].

6. Conclusions

In this paper, the influence of the clinical context on lung tissue classification from HRCT data is investigated. Correlation analysis of the multimedia feature space shows that the dataset is in accordance with medical knowledge. Two fusion schemes of the modalities were studied: early versus late fusion. The combination of two SVM classifiers achieved highest classification accuracies and allowed a mean of 79% and a maximum of 84% correct predictions of testing instances among the five classes of lung tissue. This represents a significant benefit of 10% compared to a pure visual classification. Late fusion shows robustness towards the number of clinical parameters used, which suggests that it is appropriate for mining clinical attributes with missing values. Accuracy values are trustworthy for further usage in clinical routine as we never train and test with ROIs that belongs to the same patient. In addition, the leave-one-patient-out cross-validation testing approach is in accordance with a clinical usage of the CAD where the radiologist analyzes one patient at a time and the system was trained with all the previously analyzed patients. We believe that the late fusion scheme can still be improved by identifying groups of attributes with high synergy and mining them separately in order to preserve homogeneous clusters of instances in the feature space. Moreover, the combination rules among the modalities have to be investigated in order to allow optimal complementarity of the modalities.

Acknowledgments

This work was supported by the Swiss National Science Foundation (FNS) with grant 200020-118638/1, the equalization fund of the University and Hospitals of Geneva (grant 05-9-II), the

EU 6th Framework Program in the context of the KnowARC project (IST 032691) and the ONCO-MEDIA ICT Asia project.⁶

References

- [1] Flaherty Kevin R, King Jr Talmadge Ela, Raghu Ganesh, Lynch III Joseph Psp, Colby Thomas V, Travis William D, et al. Idiopathic interstitial pneumonia: what is the effect of a multidisciplinary approach to diagnosis? *Am J Respir Crit Care Med* 2004;170(July):904–10.
- [2] Stark Paul. High resolution computed tomography of the lungs. UpToDate. Denise S. Basow edition: Waltham, MA; August 2008.
- [3] Doi K. Current status and future potential of computer-aided diagnosis in medical imaging. *Br J Radiol* 2005;78:3–19.
- [4] Shyu Chi-Ren, Brodley Carla E, Kak Avinash C, Kosaka Akio, Aisen Alex M, Broderick Lynn S. ASSERT: a physician-in-the-loop content-based retrieval system for HRCT image databases. *Comput Vision Image Understand* 1999;750(July/August (1/2)):111–32 (Special issue on content-based access for image and video libraries).
- [5] Aisen Alex M, Broderick Lynn S, Winer-Muram Helen, Brodley Carla E, Kak Avinash C, Pavlopoulou Christina, et al. Automated storage and retrieval of thin-section CT images to assist diagnosis: system description and preliminary assessment. *Radiology* 2003;2280(July (1)):265–70.
- [6] Ryu Jay H, Olson Eric J, Midthun David E, Swensen Stephen J. Diagnostic approach to the patient with diffuse lung disease. *Mayo Clin Proc* 2002;770(November (11)):1221–7.
- [7] Mitsunobu F, Mifune T, Ashida K, Hosaki Y, Tsugeno H, Okamoto M, et al. Influence of age and disease severity on high resolution CT lung densitometry in asthma. *Thorax* 2001;560(November (11)):851–6.
- [8] Toussaint Godfried T. The use of context in pattern recognition. *Pattern Recognit* 1978;100(January (3)):189–204.
- [9] Wu Yi, Chang Edward Y, Chen-Chuan Chang Kevin, Smith John R. Optimal multimodal fusion for multimedia data analysis. In: MULTIMEDIA '04: proceedings of the 12th annual ACM international conference on Multimedia; October 2004.p. 572–9.
- [10] Snoek Cees GM, Worring Marcel, Smeulders Arnold WM. Early versus late fusion in semantic video analysis. In: MULTIMEDIA '05: Proceedings of the 13th annual ACM international conference on Multimedia; November 2005. p. 399–402.
- [11] Gunes Hatice, Piccardi Massimo. Affect recognition from face and body: early fusion vs. late fusion. In: 2005 IEEE international conference on systems, man and cybernetics, vol. 4; October 2005.p. 3437–43.
- [12] Kludas Jana, Bruno Eric, Marchand-Maillet Stephane. Information fusion in multimedia information retrieval. In: Proceedings of 5th international workshop on adaptive multimedia retrieval (AMR), vol. 4918; July 2007. p. 147–59.
- [13] Bell Anthony J. The co-information lattice. In: Proceedings of the 4th international symposium on independent component analysis and blind signal separation (ICA2003); April 2003. p. 921–6.
- [14] Kittler Josef, Hafez Mohamad, Duin Robert PW, Matas Jiri. On combining classifiers. *IEEE Trans Pattern Anal Mach Intell* 1998;20(March (3)):226–39.
- [15] Lam Louisa. Classifier combinations: implementations and theoretical issues. In: MCS'00: proceedings of the first international workshop on multiple classifier systems; 2000.p. 77–86.
- [16] Dietterich Thomas G. Ensemble methods in machine learning. In: MCS'00: proceedings of the first international workshop on multiple classifier systems; 2000.p. 1–15.
- [17] Skurichina Marina, Duin Robert PW. Combining feature subsets in feature selection. In: Multiple classifier systems, vol. 3541, Lecture Notes in Computer Science. Springer; June 2005. p. 165–75.
- [18] Westerveld Thijs. Image retrieval: content versus context. In: Recherche d'Informations Assistée par Ordinateur (RIA0'2000) computer-assisted information retrieval, vol. 1. Paris, France: CID; April 2000. p. 276–84.
- [19] La Cascia Marco, Sethi Saratendu, Sclaroff Stan. Combining textual and visual cues for content-based image retrieval on the world wide web. In: Content-based access of image and video libraries. Proceedings. IEEE workshop on, Washington, DC, USA: IEEE Computer Society; 1998. p. 24–8.
- [20] Lacoste Caroline, Chevallet Jean-Pierre, Lim Joo-Hwee, Thi Hoang Le Diem, Xiong Wei, Racoceanu Daniel, et al. Inter-media concept-based medical image indexing and retrieval with UMLS at IPAL. In: Peters Carol, Clough Paul, Gey Fredric C., Karlgren Jussi, Magnini Bernardo, Oard Douglas W., de Rijke Maarten, Stempfhuber Maximilian, editors, CLEF, Lecture Notes in Computer Science, vol. 4730. Springer; 2007. p. 694–701.
- [21] Müller Henning, Michoux Nicolas, Bandon David, Geissbuhler Antoine. A review of content-based image retrieval systems in medicine—clinical benefits and future directions. *Int J Med Inform* 2004;73(1):1–23.
- [22] Hersh William, Kalpathy-Cramer Jayashree, Jensen Jeffery. Medical image retrieval and automated annotation: OHSU at ImageCLEF 2006. In: Peters Carol, Clough Paul, Gey Fredric C., Karlgren Jussi, Magnini Bernardo, Oard Douglas W., de Rijke Maarten, Stempfhuber Maximilian, editors, CLEF, Lecture Notes in Computer Science, vol. 4730. Springer; 2007. p. 660–9.
- [23] Chen Qiongyu, Li Guoliang, Leong Tze-Yun, Heng Chew-Kiat. Predicting coronary artery disease with medical profile and gene polymorphisms data. In:

⁶ ONCO-MEDIA: Ontology and context related medical image distributed intelligent access, <http://www.onco-media.com/> (Accessed: 28 August 2009).

- Kuhn Klaus A, Warren James R, Leong Tze-Yun, editors. *Medinfo 2007: Proceedings of the 12th world congress on health (medical) informatics*. Brisbane, Australia: IOS Press; August 2007. p. 1219–1224.
- [24] Abe Hiroyuki, Ashizawa Kazuto, Li Feng, Matsuyama Naohiro, Fukushima Aya, Shiraishi Junji, et al. Artificial neural networks (ANNs) for differential diagnosis of interstitial lung disease: results of a simulation test with actual clinical cases. *Acad Radiol* 2004;11(June (1)):29–37.
- [25] Zrimec Tatjana, Wong James S. Improving computer aided disease detection using knowledge of disease appearance. *Stud Health Technol Inf* 2007; 129:1324–8.
- [26] Uppaluri Renuka, Hoffman Eric A, Sonka Milan, Hartley Patrick G, Hunnighake Gary W, McLennan Geoffrey. Computer recognition of regional lung disease patterns. *Am J Respir Crit Care Med* 1999;160(August(2)):648–54.
- [27] Yoshikazu Uchiyama, Shigehiko Katsuragawa, Hiroyuki Abe, Junji Shiraishi, Feng Li, Qiang Li, et al. Quantitative computerized analysis of diffuse lung disease in high-resolution computed tomography. *Med Phys* 2003;30(September(9)):2440–54.
- [28] Depeursinge Adrien, Sage Daniel, Hidki Asmâa, Platon Alexandra, Poletti Pierre-Alexandre, et al. Lung tissue classification using Wavelet frames. In: *Engineering in Medicine and Biology Society. EMBS 2007. 29th Annual international conference of the IEEE*. Lyon, France, August 2007. IEEE Computer Society; 2007. p. 6259–6262.
- [29] Depeursinge Adrien, Van De Ville Dimitri, Unser Michael, Müller Henning. Lung tissue analysis using isotropic polyharmonic B-spline wavelets. In: *MICCAI 2008 workshop on pulmonary image analysis*. New York, USA: Lulu; September 2008. p. 125–134.
- [30] Depeursinge Adrien, Iavindrasana Jimison, Hidki Asmâa, Cohen Gilles, Geissbuhler Antoine, Platon Alexandra, et al. A classification framework for lung tissue categorization. In: *Andriole Katherine P, Siddiqui Khan M, editors, Medical imaging 2008: PACS and imaging informatics*, vol. 6919. San Diego, CA, USA: SPIE; 2008. p. 69190C.
- [31] Hidki Asmâa, Müller Henning, Depeursinge Adrien, Poletti Pierre-Alexandre, Geissbuhler Antoine. Putting the image into perspective: the need for domain knowledge when performing image-based diagnostic aid. In: *Swiss conference on medical informatics (SSIM 2006)*, Basel, Switzerland; April 2006.
- [32] King Talmagde E. *Approach to the adult with interstitial lung disease*. UpToDate. Denise S. Basow edition: Waltham, MA; 2008.
- [33] Webb W. Richard, Müller Nestor L, Naidich David P., editors. *High-resolution CT of the Lung*. Philadelphia, PA, USA: Lippincott Williams & Wilkins; 2001.
- [34] Van De Ville Dimitri, Blu Thierry, Unser Michael. Isotropic polyharmonic B-splines: scaling functions and wavelets. *IEEE Trans. Image Process* 2005; 14(November (11)):1798–813.
- [35] Cohen Gilles, Hilario Melanie, Sax Hugo, Hugonnet Stephane, Geissbuhler Antoine. Learning from imbalanced data in surveillance of nosocomial infection. *Artif Intell Med* 2006;37(May (1)):7–18.
- [36] Quinlan Ross J. Induction of decision trees. *Mach Learn* 1986;1(March (1)):81–106.
- [37] Quinlan Ross J. *C4. 5: programs for machine learning*. San Francisco, CA, USA: Morgan Kaufmann Publishers Inc.; 1993.
- [38] Perner Petra, Belikova Tatjana P, Yashunskaya Nadeszda I. Knowledge acquisition by symbolic decision tree induction for interpretation of digital images in radiology. In *SSPR '96: Proceedings of the 6th international workshop on advances in structural and syntactical pattern recognition*, vol. 1121. London, UK: Springer-Verlag; 1996. p. 208–19.
- [39] Wu Ting-Fan, Lin Chih-Jen, Weng Ruby C. Probability estimates for multi-class classification by pairwise coupling. *J Mach Learn Res* 2004;5(August):975–1005.
- [40] Benmokhtar Rachid, Huet Benoit. Classifier fusion: combination methods for semantic indexing in video content. In: *ICANN 2006: international conference on artificial neural networks*. Springer; 2006. p. 65–74.
- [41] Pechenizkiy Mykola, Tsybal Alexey, Puuronen Seppo. PCA-based feature transformation for classification: issues in medical diagnostics. In: *CBMS '04: proceedings of the 17th IEEE symposium on computer-based medical systems*. IEEE Computer Society; 2004. p. 535–40.
- [42] Hong Se June. Use of contextual information for feature ranking and discretization. *IEEE Trans Knowledge Data Eng* 1997;9(5):718–30.
- [43] Chambellan Arnaud, Chailleux Edmond, Similowski Thomas. Prognostic value of the hematocrit in patients with severe COPD receiving long-term oxygen therapy. *Chest* 2005;128(September (3)):1201–8.
- [44] Limper Andrew H. Chemotherapy-induced lung disease. *Clin Chest Med* 2004;25(March (1)):53–64.