

UNIVERSITÉ DE GENÈVE  
Département d'Informatique  
UNIVERSITÉ DE GENÈVE  
Service d'Informatique Médicale  
UNIVERSITÉ DE GENÈVE  
Service d'Informatique Médicale

---

FACULTÉ DES SCIENCES  
Professeur Christian Pellegrini  
FACULTÉ DE MÉDECINE  
Professeur Antoine Geissbuhler  
FACULTÉ DE MÉDECINE  
Privat Docent Dr. Henning Müller

---

# Affine-invariant texture analysis and retrieval of 3D medical images with clinical context integration

## THÈSE

présentée à la Faculté des sciences de l'Université de Genève  
pour obtenir le grade de Docteur ès sciences, mention informatique

par

Adrien Depeursinge

de  
Bièvre (VD)

Thèse N° 4186

GENÈVE  
2010

La Faculté des sciences, sur le préavis de Messieurs C. PELLEGRINI, professeur ordinaire et co-directeur de thèse (Département d'informatique), A. GEISSBUHLER, professeur ordinaire et co-directeur de thèse (Faculté de médecine — Service d'informatique médicale), H. MÜLLER, docteur privat docent et co-directeur de thèse (Faculté de médecine — Service d'informatique médicale), M. UNSER, professeur (École Polytechnique Fédérale de Lausanne — Laboratoire d'imagerie biomédicale — Lausanne, Suisse), P.-A. POLETTI, docteur privat docent (Hôpitaux Universitaires de Genève — Service de radiologie), et Madame T. ZRIMEC, professeure (University of New South Wales — Centre for Health Informatics — Sydney, Australia), autorise l'impression de la présente thèse, sans exprimer d'opinion sur les propositions qui y sont énoncées.

Genève, le 22 février, 2010

**Thèse - 4186 -**



Le Doyen, Jean-Marc TRISCONE

# Contents

<b>Résumé</b>	<b>v</b>
<b>Abstract</b>	<b>vii</b>
<b>Acknowledgements</b>	<b>ix</b>
<b>1 Introduction to computer vision of medical images</b>	<b>1</b>
1.1 Computer tools for managing visual medical information . . . . .	2
1.1.1 Event detection and quantification in medical images . . . . .	2
1.1.2 Automated indexation of medical visual information . . . . .	3
1.2 Computer vision: inspirations from human vision and cognition . . . . .	5
1.2.1 Eyes and visual cortex: feature extraction . . . . .	7
1.2.2 Trained brain: supervised machine learning . . . . .	12
1.3 CAD in HRCT imaging of the chest . . . . .	16
1.3.1 Interstitial lung diseases . . . . .	16
1.3.2 Texture-based CADs for lung tissue analysis in thin-section CT . . . . .	20
1.4 Organization and scientific contributions of this thesis . . . . .	27
1.4.1 Thesis overview . . . . .	27
1.4.2 Scientific contributions . . . . .	28
<b>2 Multimedia database of interstitial lung diseases</b>	<b>29</b>
2.1 Existing databases of CT imaging of the lung . . . . .	31
2.2 Scope of the database . . . . .	32
2.2.1 Selection of the histological diagnoses . . . . .	32
2.2.2 Selection of the clinical parameters . . . . .	32
2.3 Data collection . . . . .	32
2.3.1 Selection of the cases at the University Hospitals of Geneva . . . . .	33
2.3.2 Annotations . . . . .	33
2.3.3 Data entry . . . . .	36
2.4 Contents . . . . .	36
2.4.1 Numbers and statistics . . . . .	36
2.4.2 Control of Quality . . . . .	36
2.5 Discussions . . . . .	39
<b>3 Features</b>	<b>41</b>
3.1 Visual features . . . . .	41
3.1.1 Gray-level histograms . . . . .	42
3.1.2 Wavelet transforms for lung tissue analysis . . . . .	42
3.1.3 Evaluation of the visual features . . . . .	48
3.1.4 Discussions on the visual features . . . . .	56
3.2 Clinical and multimodal features . . . . .	60
3.2.1 Extraction: from MySQL variables to a consistent feature space . . . . .	60
3.2.2 Evaluation . . . . .	61

3.2.3 Interpretation . . . . .	62
<b>4 Classification and retrieval</b>	<b>67</b>
4.1 Choice of the model based on visual features . . . . .	67
4.1.1 Methodology for comparing classifier families . . . . .	67
4.1.2 Discussion on classifier performances . . . . .	72
4.2 Clinical context integration for medical image analysis . . . . .	73
4.2.1 Related work in contextual image analysis . . . . .	74
4.2.2 The clinical context of HRCT images from patients with ILDs . . . . .	75
4.2.3 Modality fusion . . . . .	76
4.2.4 Evaluation: early versus late fusion . . . . .	76
4.2.5 Discussion on the fusion schemes . . . . .	78
4.3 Case-based retrieval . . . . .	80
4.3.1 3D multimodal retrieval . . . . .	81
4.3.2 Evaluation . . . . .	82
4.3.3 Discussions on the retrieval performance . . . . .	82
<b>5 Applications and evaluation</b>	<b>83</b>
5.1 Use cases, visualization and software . . . . .	83
5.1.1 Multimedia database browsing . . . . .	84
5.1.2 3D lung tissue categorization and case-based retrieval . . . . .	85
5.2 Evaluation . . . . .	91
5.2.1 Validation strategies . . . . .	92
5.2.2 Metrics . . . . .	94
5.2.3 Causes of error and bias . . . . .	97
5.2.4 Qualitative evaluation . . . . .	98
<b>6 Conclusions and Perspectives</b>	<b>99</b>
6.1 Achievements and limitations . . . . .	99
6.1.1 Literature analysis . . . . .	99
6.1.2 Multimedia library . . . . .	99
6.1.3 Features . . . . .	100
6.1.4 Machine learning . . . . .	100
6.1.5 Information retrieval . . . . .	100
6.1.6 Application . . . . .	101
6.1.7 Evaluation of image-based CAD systems . . . . .	101
6.2 Perspectives . . . . .	101
6.2.1 Potential improvements and extensions of the CAD system . . . . .	101
6.2.2 Classification and retrieval in $N$ dimensions . . . . .	103
6.2.3 Towards indexing the whole PACS . . . . .	103
6.3 Better than clinicians? . . . . .	104
<b>A CAD in HRCT of the lungs</b>	<b>105</b>
<b>B HTML form for clinical data</b>	<b>115</b>
<b>Notation</b>	<b>119</b>
<b>Glossary</b>	<b>123</b>
<b>List of Figures</b>	<b>126</b>
<b>List of Tables</b>	<b>128</b>
<b>Bibliography</b>	<b>130</b>

<b>Index</b>	<b>150</b>
--------------	------------



# Résumé

Au cours de la dernière décennie, le nombre d'images produites par jour dans les hôpitaux a augmenté de manière exponentielle. Les développements considérables de la physique de l'imagerie ont apporté aux cliniciens toute une variété d'outils de diagnostic indispensables. L'explosion de la quantité et de la diversité de l'information visuelle médicale a eu pour conséquence fâcheuse de submerger les radiologues avec les tâches d'interprétation de ces images. Récemment, la forme digitale et le stockage standardisé des images médicales ont donné naissance à une nouvelle discipline, au point de rencontre entre la vision par ordinateur et la médecine: le diagnostic assisté par ordinateur (DAO) basé sur l'image. En réponse à l'avancée de la physique de l'imagerie, les DAOs basés sur l'image représentent un important potentiel d'amélioration, tant de la quantité que de la qualité de la productivité des radiologues.

Il y a deux types de DAOs basés sur l'image: les DAOs basés sur la détection et les DAOs basés sur la recherche d'images par le contenu (RIC). Le premier analyse les séries d'images de manière exhaustive afin de détecter et caractériser les anomalies alors que le second a pour but de comparer et récupérer des images à partir de similarités visuelles. Les DAOs basés sur la détection permettent d'obtenir un rapide aperçu des régions pathologiques, réduisant ainsi le risque d'omission et assurant la reproductibilité du diagnostic en attirant l'attention du radiologue sur les lésions importantes dans les séries d'images médicales. Les DAOs basés sur la RIC fournissent quant à eux de précieuses informations pour le diagnostic et la planification du traitement grâce à des exemples de cas dont le diagnostic a été préalablement confirmé. Conceptuellement, les configurations de DAO reposent toutes deux sur des caractéristiques visuelles décrivant le contenu visuel des images. À partir de l'espace des caractéristiques visuelles, le DAO basé sur la détection décide à quelle classe (par exemple "normal" ou "malade") l'image d'entrée (ou une région de l'image) appartient, alors que les DAOs basés sur la RIC calculent une mesure de similarité entre deux images à partir de laquelle une liste d'instances ordonnée peut être créée.

La présente thèse envisage différentes méthodes de création d'un système DAO hybride basé tant sur la détection que la RIC, avec un accent sur l'analyse des textures dans des images de tomodensitométrie à haute résolution (TDHR) chez des patients atteints de maladies interstitielles pulmonaires (MIP). Plusieurs analogies entre la vision humaine et la vision par ordinateur sont proposées et l'importance des représentations invariantes aux transformations affines pour une description adéquate et robuste des tissus et organes est mise en évidence. En conséquence, un ensemble de caractéristiques visuelles invariantes aux transformations affines est développé à partir d'une transformée en ondelettes spécifique. Afin de choisir parmi une grande variété d'algorithmes de classification existants responsables de la prise de décision pour les DAOs basés sur la détection, une méthodologie de comparaison de modèles de classification basée sur le test statistique de McNemar est exposée. À ce stade, le DAO basé sur la détection repose sur de solides caractéristiques visuelles et algorithmes de classification. Le contexte clinique des images médicales est ensuite intégré dans le système de classification et permet d'améliorer significativement la performance de détection, ce qui est en adéquation avec l'approche naturelle des radiologues pour l'interprétation des images. La stratégie optimale de combinaison de l'information visuelle avec les paramètres cliniques est étudiée. À partir des résultats de l'interprétation des images du DAO basé sur la détection, une mesure de similarité multimodale est introduite, permettant la récupération de cas similaires, ce qui donne lieu à un système DAO hybride basé sur la détection ainsi que la RIC. En parallèle, une base de données multimédia de cas de MIP a été créée aux Hôpitaux Universitaires

de Genève (HUG). Elle contient une sélection de paramètres cliniques associés aux MIP et des séries d'images TDHR avec des régions annotées qui servent de référence pour l'évaluation des algorithmes du DAO et constitue une collection de cas de MIP de référence pour l'enseignement. Des hypothèses d'utilisation pour la navigation dans la base de données, la catégorisation automatique tridimensionnelle des tissus pulmonaires dans les images TDHR et la récupération de cas similaires sont définies et implémentées dans une application Internet riche contentant des interfaces utilisateur graphiques. Afin de mesurer la performance du système DAO proposé, une méthodologie d'évaluation visant à reproduire les conditions cliniques est introduite et discutée.

Les expériences et résultats des recherches présentées dans cette thèse suggèrent que, bien qu'elle soit encore loin des formidables capacités de la vision humaine, la vision par ordinateur a dépassé le stade expérimental; des contributions de recherche multidisciplinaire sont maintenant nécessaires afin d'exploiter le plein potentiel des nouvelles technologies de vision par ordinateur et de les rendre disponibles aux cliniciens. Dans cette perspective, les cliniciens et les ordinateurs seront amenés à former équipe dans un avenir proche.

# Abstract

The number of images produced per day in modern hospitals followed an exponential growth during the last decade. The mature field of imaging physics brought a large variety of incontrovertible diagnosis tools to the clinicians. As a consequence, the explosion of the quantity and variety of medical visual information has the undesirable effect to overwhelm the radiologists with image interpretation tasks. Recently, the digital form of medical images along with their standardized storage gave birth to a new domain at the crossing of computer vision and medicine: image-based computer-aided diagnosis (CAD). As a response to the imaging physics breakaway, image-based CAD systems have the potential to improve both the quantity and the quality of the productivity of the radiologists.

Image-based CAD comes into two declensions: detection-based CAD and content-based image retrieval (CBIR)-based CAD. The former exhaustively analyzes image series to detect and characterize anomalies whereas the latter aims at retrieving images based on the visual similarity. Detection-based CAD can quickly provide an overview of the diseased regions, which reduces the risk of omission and ensures the reproducibility of the diagnosis by drawing the radiologists' attention to diagnostically interesting events in medical image series. CBIR-based CAD delivers quick and precious information for diagnosis aid and treatment planning through examples of cases with beforehand confirmed diagnosis. Conceptually, both CAD configurations relies on visual features describing the visual contents of the images. Based on a visual feature space, the detection-based CAD decides to which predefined class (e.g. “normal” or “diseased”) the input instance (i.e. image or image region) belongs, whereas CBIR-based CADs output a similarity score between two instances from which a ranked list of images can be built.

In this thesis, methods for building a hybrid detection-CBIR-based CAD system are proposed, evaluated and discussed with a focus on texture analysis of high-resolution computed tomography (HRCT) images of patients affected with interstitial lung diseases (ILDs). Several analogies between human and computer vision are proposed and highlight the importance of using affine-invariant visual descriptors to obtain robust representations of tissue or organs. Consequently, an affine-invariant set of texture features is developed based on tailored wavelet transforms. In order to choose among the large variety of existing classification algorithms that are responsible for the decision-making of the detection-based CAD, a methodology for comparing classification models based on McNemar's statistical test is proposed. At this step, the detection-based CAD system relies on robust visual features and classification algorithms. Then, the clinical context of medical images is introduced in the classification framework and allowed to significantly improve the detection performance, which is in accordance with the natural approach of the radiologists to image interpretation. The optimal strategy for combining the visual information with the clinical parameters is studied. Based on the output of the detection-based CAD, a multimodal similarity measure is introduced to enable case-based retrieval, yielding a hybrid detection-CBIR-based CAD system. In parallel, a multimedia library of ILD cases was built at the University Hospitals of Geneva. It contains a selection of clinical parameters associated with ILDs and HRCT image series with annotated regions and aims at providing ground truth for the evaluation of the CAD algorithms and constitutes a reference library for teaching. Use cases for database browsing, three-dimensional automated categorization of the lung tissue in HRCT and case-based retrieval are defined and implemented in a rich Internet application (RIA) with graphical user interfaces (GUIs). To assess the performance and robustness of the proposed CAD system, an evaluation

methodology aiming at reproducing clinical conditions is proposed and discussed.

The experiences and outcomes of the research presented in this thesis suggest that although computer vision is still far behind the tremendous skills of human vision, the end of its early days has been reached and multidisciplinary research efforts are required to bring the full potential of emerging computer vision technologies to the clinicians. In that perspective, clinicians and computers are expected to team up in a very close future.

# Acknowledgements

My deepest gratitude goes to my advisor Henning Müller. Since my arrival at the University Hospitals of Geneva, he has always been an example to follow to me. He has been present to provide me advice at every moment of this work and taught me a solid research methodology that gave me the means to cope with the various challenges of research demands. I learned enormously from his near-perfect management of the technical and human aspects. He has been a constant source of inspiration from the start of this adventure with many ideas that always made sense to me. For that, I would like to address him my warmest thanks.

I also would like to thank my thesis director professor Christian Pellegrini for trusting me and for allowing me the required time to carry out this thesis. I am also deeply grateful to professor Antoine Geissbühler for welcoming me into the Medical Informatics Service of the University Hospitals of Geneva and allowed me to discover the fascinating world of medical informatics during the four years of the preparation of this thesis.

I address my lively thanks to professor Tatjana Zrimec from the University of New South Wales in Australia for accepting to be part of the board of examiners of this thesis.

I would like to address particular thanks to professor Dimitri Van de Ville for his very valuable insights and comments concerning wavelets. I also thank professor Michael Unser for welcoming me into his group for half a year before the beginning of this work and for introducing me to Henning Müller, and the members of the Biomedical Imaging Group with whom I had several collaborations concerning image processing aspects.

Particular thanks are addressed to our clinical partners Dr. Pierre-Alexandre Poletti and Dr. Alexandra Platon for their fantastic help with the creation of the library of cases of interstitial lung diseases as well as many precious insights concerning the approach of radiologists to image analysis. I also would like warmly thank the three medical doctors that carried out a enormous amount of work for the collection of ILD cases and helped me for all the medical aspects: Alejandro Vargas, Asmâa Hidki and Herizo Andriambololoniaina.

I also would like to thank professor Daniel Racoceanu for welcoming me at the National University of Singapore for three months of fruitful collaboration on contextual medical image analysis.

I thank our two secretaries Dominique Guérin and Martine Burford who helped me whenever I needed it as well as the people with whom I shared the office in a nice atmosphere and provided me technical help on programming aspects. In particular, I warmly thank Dr. Gilles Cohen and Jimison Iavindrasana which helped me with several aspects of machine learning as well as Xin Zhou and Marko Niinimäki that gave me valuable assistance for the parallelization of several parts of my code. I thank Frédéric Gaillard for the development of several parts of the rich Internet application. I hope that every person whose name is missing here know that (s)he equally have her/his place here.

I am also very grateful to my brothers Xavier and Vincent as well as my parents Florence and Christian who supported me during the entirety of this thesis. In particular, my father Christian awoke my scientific curiosity from the earliest moments that I can remember through to his passion to research. Similarly, the passion to medicine of my mother Florence always fascinated me. At last, I would like to thank my lovely Camille who turns to happiness every minute spent at her side and who actively supported me during the writing process of this thesis.

# Chapter 1

## Introduction to computer vision of medical images

The role of medical images increases in importance for establishing diagnoses, decision making and treatment planning. The standardized imaging protocols allow to faithfully confirm the presence of abnormalities and quantitatively evaluate their extent. They are thereby perfectly aligned with evidence-based medicine aiming to draw decisions based on proven facts [221]. The imaging techniques evolved in order to be able to assess the visual appearance of almost every organ with both high spatial and temporal resolution. Functional imaging was also more recently proposed to study the behavior of certain organs. All the specific requirements for imaging each organ bred a wide range of imaging techniques producing multidimensional data in the forms of signals, images (2D), images series (3D volumes, videos or 4D volumetric videos) and combination of modalities ( $N$ -D). Examples of the most common signals and imaging techniques in medicine are listed in Table 1.1.

As a consequence, the amount of visual information created in modern hospitals exploded in the last decade. For instance, the mean number of images stored in the PACS (Picture Archiving and Communication System) at the University Hospitals of Geneva (HUG) per day follows an exponential increase with 114'000 images produced per day in 2009 (see Figure 1.1). Following this trend, the predicted number of images produced per day in the year 2020 would be around 3 million. Such a tremendous amount of visual information calls upon computerized aid both at a large-scale management level of the whole content of the PACS [170] and at the diagnosis level for the interpretation of single image series [67]. Furthermore, recent advances in medical informatics enabled access to most of the radiological exams to all clinicians through the electronic health record (EHR) and the PACS. This change of the medical workflow calls upon computer expert systems able to bring the right information to the right people at the right time.

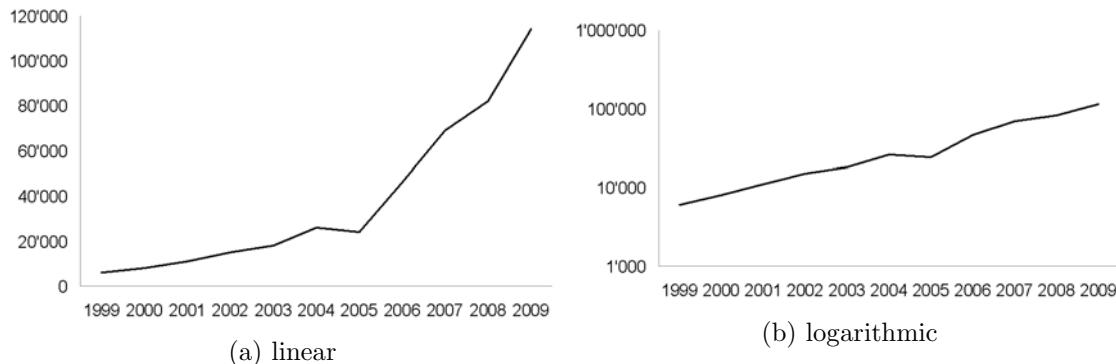


Figure 1.1: Average number of images produced per day at the HUG during the last decade.

Table 1.1: Most common signals and imaging techniques in medicine.

signals	electrocardiogram (ECG), electromyogram (EMG), electroencephalogram (EEG), magnetoencephalogram (MEG), ...
2D images	X-ray, microscopic imaging, ...
3D volumes, videos	X-ray computed tomography (CT), magnetic resonance imaging (MRI), positron emission tomography (PET), single photon emission CT (SPECT), ultrasound imaging (US), ...
4D imaging	temporal CT, temporal MRI, 3D US, ...
$N$ -D combination of modalities	functional MRI (fMRI), PET-CT, PET-MRI, SPECT-CT, ...

## 1.1 Computer tools for managing visual medical information

The increasingly important digital form of the medical images along with their storage on image servers such as PACS in the standardized DICOM<sup>1</sup> format enables automatic analysis and content-based indexing as diagnostic aid. Consequently, computer-aided diagnosis (CAD) has become one of the major research subject in medical imaging and diagnostic radiology [67, 68]. Goal of CAD is to use computer vision to assist radiologists in focusing their attention on diagnostically interesting events. The CAD system can be used as first reader in order to improve the radiologists' productivity and reduce reading fatigue [170, 176]. Whereas the radiologists' ability to interpret visual information is likely to change based on the domain-specific experience, human factors and time of the day, computerized classification of lung tissue patterns is 100% reproducible. CAD was also proposed in several fields of medicine with non-visual systems such as QMR (quick medical reference) or ILIAD [81, 159].

Early studies on computerized analysis of medical images (mostly in X-ray imaging) were reported in the 1960s [126, 158]. These systems aimed at replacing the radiologists as the originators relied on the assumption that computers are better at performing certain tasks than are human beings. However, it quickly became clear that physicians and radiologists have to take the final decision and outputs of CAD systems must be used as “second opinions” and information providers [73, 155].

Recently, CAD systems have been used in clinical routine in the rather mature field of cancer screening in mammograms and allowed to improve the detection of nonpalpable cancerous masses [17, 80]. Image-based CAD systems are expected to be introduced in clinical routine for several other organs such as the chest, colon, brain, liver, skeletal and vascular systems [2, 67, 176]. Table 1.2 lists the number of papers presented at the Annual Meeting of the Radiological Society of North America (RSNA) related to CAD from year 2000 to 2005 (source [67]).

Mainly two approaches are proposed in the literature for assessing image-based computerized diagnostic aid [32, 222]:

- event detection and quantification in medical images,
- content-based image retrieval (CBIR).

The two approaches are presented and discussed in Sections 1.1.1 and 1.1.2.

### 1.1.1 Event detection and quantification in medical images

According to the number of imaging modalities existing, the large variety of organs studied and the high amount of visual information that modern imaging devices deliver, the radiologists require expert CAD systems that provide an exhaustive scan and categorization of medical image

---

<sup>1</sup>Digital imaging and communications in medicine (DICOM)

Table 1.2: Number of papers presented at the Annual Meeting of RSNA related to CAD from year 2000 to 2005 (source [67]).

	2000	2001	2002	2003	2004	2005	Total
Chest	22	37	53	94	70	48	324
Breast	23	28	32	37	48	49	217
Colon	4	10	21	17	15	30	164
Brain	–	4	2	10	9	15	70
Liver	3	–	5	9	9	9	29
Skeletal	2	7	7	9	8	5	38
Vascular	5	–	12	15	2	7	41
Total	59	86	134	191	161	163	883

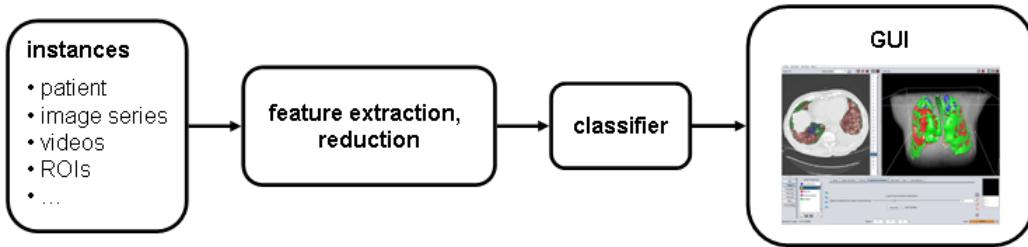


Figure 1.2: General scheme of CAD for event detection and quantification in medical image series.

series. This is particularly important in stressful environments such as emergency radiology where the radiologists must provide diagnostics rapidly, usually alone and at any time of the day or night. The computerized interpretation of the images is complementary to the human's one which is not devoid of limitations [202]. Such detection-based CADs can help to reduce the risk of omission and to ensure the reproducibility of the diagnostic by drawing the radiologists' attention to diagnostically interesting events in medical image series.

The classical scheme for event detection and quantification in medical image series is depicted in Figure 1.2. The instances constitutes the input of the CAD system and can be patients, images series, videos, regions of interest (ROI),.... From the instances, a set of features aiming to compactly describe the diagnostically relevant visual content is extracted. When required, a feature reduction is applied to cut down the number of features. The resulting set of attributes constitutes the feature space in which the classifier draws decision boundaries to predict the class(es) of the instance(s) that are displayed in the appropriate format to the user via a graphical user interface (GUI).

An introduction to feature extraction is given in Section 1.2.1 and the features used in the CAD proposed in this work are detailed in Chapter 3. The classifier's underlying mechanisms of machine learning are introduced in Section 1.2.2 and the classification algorithms used in this work are detailed in Chapter 4.

### 1.1.2 Automated indexation of medical visual information

In the context of medical image analysis, providing quick and precious information to the clinician is not limited to automatic detection and quantification of abnormal tissue and/or structure. In clinical routine, the approach of the clinician to a diagnosis when he has little experience in a particular domain is to compare the image under investigation with typical cases with confirmed diagnosis listed in textbooks or contained in personal collections. This permits to rule out diagnoses and, in association with clinical parameters, prevents the reader from mixing diagnoses with similar radiological findings. This process allows the clinician to partly replace a lack of experience but has two major drawbacks: searching for similar images is time-consuming and the notion of

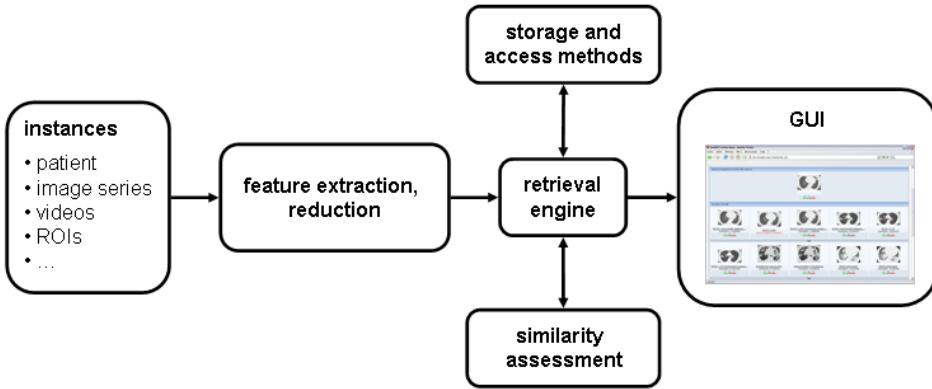


Figure 1.3: General scheme of CBIR-based CAD (inspired from [170]).

similarity may be context- or time-dependent [139, 164].

Content-based medical image retrieval (CBIR) aims at finding objectively visually similar images in large standardized image collections (e.g. PACS) [170]. The main components of CBIR-based CADs are detailed in Figure 1.3. In CBIR, the notion of similarity is usually established from a set of visual features describing the content of the instances (i.e. image series, videos, ROIs,...). Features can be identical to the one used in detection-based CADs (see Section 1.1.1) and can vary from low-level measures such as the histogram quantification of the colors to high-level semantically-related features describing the anatomical content of images. A retrieval engine manages the storage and access methods to the features and computes the distance measures for similarity assessment between the instances described in terms of feature vectors. The most similar instances to the query are shown in an appropriate format to the clinician in a GUI.

The automatic indexation of medical visual content with CBIR enables systematic storage of the ever-increasing production of medical images in todays modern hospitals. Medical data repositories represent potentially rich knowledge bases. However, the access to medical image series in the PACS is most often based on the identification of the patient, which is not suitable to search for similar cases based on a disease or on the visual similarity of the images. Thereby, CBIR allows to quickly find similar images according to objective criteria in large image collections to assist the radiologists in their diagnosis workup [172, 213].

Concepts for integration of CBIR in medical practice were often proposed in the literature [170, 179, 269]. However, few CBIR systems have been integrated and evaluated in clinical practice. A major challenge is to find features and distance measures that are matching the clinicians requirements according to a particular application. The discrepancy between the user's intentions when looking for a particular image and the visual information that the features are able to model is called the "semantic gap" in the literature [226] and is usually a bottleneck in medical CBIR. Nevertheless, the few CBIR-based CADs evaluated in clinical practice showed very promising results as that they can be accepted by the clinicians as a helpful and easy-to-use tool and allows significant improvement in the diagnosis accuracy, especially for little experienced radiologists [4, 28, 111, 220].

A possible extension to CBIR is to carry out case-based retrieval. Most often, the clinician actually looks for similar cases as he considers the image within the context of a patient with a personal history, findings on the physical examination, laboratory tests, etc. At last, CBIR can complementary to event detection where the features can be used both for the categorization of the image series and retrieval of similar instances.

In this work, a case-based retrieval system is proposed in Section 4.3. The latter is based on the output of detection-based CAD.

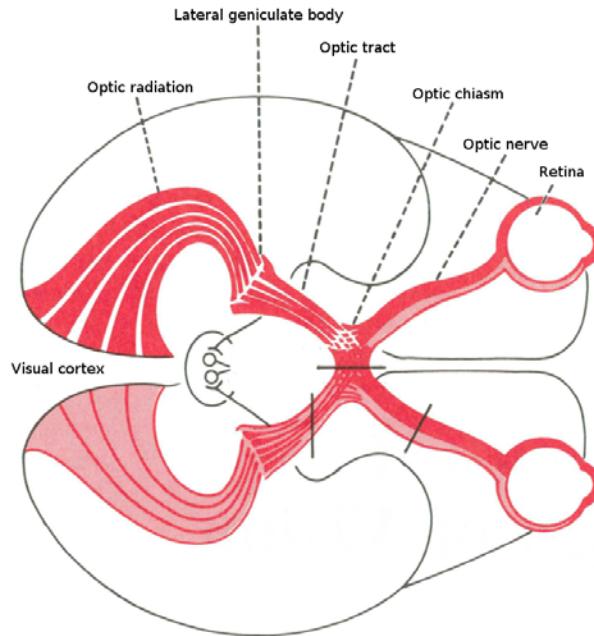


Figure 1.4: Visual pathways in the human visual system (source [86]).

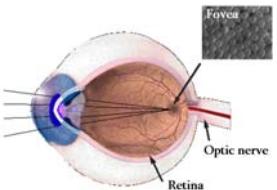
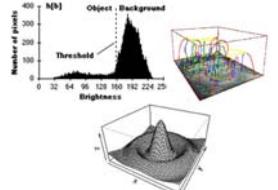
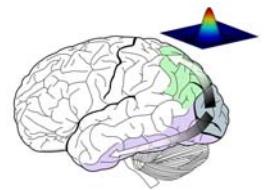
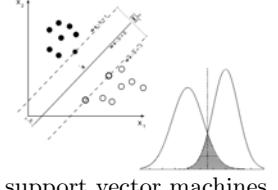
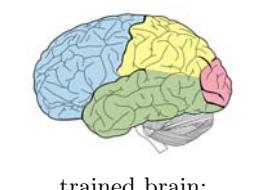
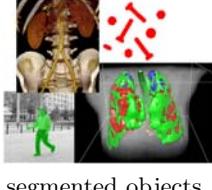
## 1.2 Computer vision: inspirations from human vision and cognition

Computers have near-unlimited computational power compared to human beings and are incomparably more rapid and precise for a wide range of applications from simple arithmetics to non-linear multivariate simulations. Nonetheless, vision is an exception to this where human beings are by far more accurate and rapid to recognize objects and textures in scenes. This suggests that computer vision has not yet reached its full potential and further research contributions are required.

The human vision is a complicated process that results from the interactions of numerous components of the human eye and brain [86]. The visual pathways start from the eyes, in particular from the retinas, and ends at the visual cortex. In terms of signal propagation, the nerve impulses are generated in the two retinas and passed backwards through the optic nerves, cross themselves at the optic chiasm, and synapse in the corresponding dorsal lateral geniculate nucleus (LGN) through the optic tracts. Then, both LGN generate nerve impulses that converge to the visual cortex through the optic radiations (see Figure 1.4). The latter is responsible for the perception of all aspects of visual forms and colors in humans. It extracts features about the form and orientation of objects. These features yield a compact representation of the observed scene that is further processed by complex cognitive schemes involving several lobes of the brain. As an output of the visual information pathway, objects, textures, motion, and other hi-level visual concepts are identified to define, along with the context of the observed image, a whole scene from which decisions can be drawn [16, 62, 153, 267]. In radiology for instance, the identified textures and objects contained in the images are interpreted according to the clinical context of the patient to establish the radiological diagnostic.

Computer vision has been widely inspired by the human visual system at many levels [12, 16, 45, 62, 119, 153, 154, 192, 240, 267]. Analogies between natural and artificial worlds are proposed in Sections 1.2.1 and 1.2.2 where the several components of computer vision are defined. The proposed analogies are summarized in Table 1.3.

Table 1.3: Analogies between computer vision and human vision.

	Computer vision	Human vision analogy	Output
Image capture	 digital camera, scanner, ...	 human eye and retina	 sampled images (2D, 3D, ...)
Feature extraction	 thresholding, filtering, zero-crossing, multiscale, ...	 visual cortex: colors, gradient of contrast, ...	 affine-invariant visual features
Supervised machine learning	 support vector machines, naive Bayes, ...	 trained brain: cognitive schemes	 segmented objects, decisions, ...

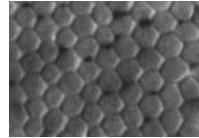


Figure 1.5: Organization of cone cells in the foveal region (source [46, 267]).

### 1.2.1 Eyes and visual cortex: feature extraction

Several analogies between the way our visual system extracts salient features from the observed scene and image processing techniques can be found. In image processing, feature extraction is a procedure to uncover salient features from an image. Comparing images pixel by pixel is extremely impractical, like comparing two objects atom by atom. In contrast, features are more compact and efficient. For example, to compare two objects, our visual system uses simple features such as the shape (square, round, etc...) and color. Similarly, a computer can rely on salient features in the image, such as corners, edges, textures, and colors to derive a compact representation that allows for efficient comparison with other images.

A first notable similarity between human and computer vision is already in the capture of the image: the spatial sampling. When cameras and scanners capture and store visual information on regular 2D or 3D sampling grids through the conception of the sensor or reconstruction algorithms, the retina sample values of the observed scene at the cone positions which are organized in mosaics (see Figure 1.5). According to Nyquist's sampling theorem, an observed continuous signal of bandwidth  $B$  can be reconstructed without loss when the distance between two samples is equal to  $B/2$ , corresponding to a sampling frequency  $f_s = 2B$ . Yellott showed that when the observed signal contains frequencies higher than  $f_s/2$  aliasing effects are perceived by humans [282]. As a consequence, we can consider that both human and computers are processing visual information starting from a sampled image [45].

#### Affine invariance

From a fixed point of view, the projections of three-dimensional objects and textures can be obtained by affine transformations of the original version as illustrated in Figure 1.6 [12]. An affine transformation between two vectorial spaces consists of a linear transformation followed by a translation:

$$\mathbf{x} \mapsto A\mathbf{x} + \mathbf{b}. \quad (1.1)$$

The vector  $\mathbf{x}$  is linearly transformed by the matrix  $A$  and translated by  $\mathbf{b}$ . The elementary linear transformations are: rotation by an angle  $\theta$ , scaling by factor  $a$ , reflexion and shearing (vertical and by a factor  $m$  in the example). Their corresponding transformation matrices in  $\mathbb{R}^2$  are:

$$A_{rotation} = \begin{pmatrix} \cos(\theta) & -\sin(\theta) \\ \sin(\theta) & \cos(\theta) \end{pmatrix}, \quad A_{scaling} = \begin{pmatrix} a & 0 \\ 0 & a \end{pmatrix},$$

$$A_{xReflexion} = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}, \quad A_{shear} = \begin{pmatrix} 1 & m \\ 0 & 1 \end{pmatrix}.$$

A projection of any object composed of a set of vectors  $\mathbf{x}$  can be obtained by a combination of several affine transformations, which remains an affine transformation itself. Our assumption is that the description of the observed object (or texture) has to be affine-invariant in order to optimize the recognition accuracy of any projection of the original object [184]. In other words, affine-invariant visual features will allow best discrimination among projected versions of distinct objects or textures.

The human visual system itself is affine-invariant. Indeed, as we can experience ourselves by glancing at Figure 1.6, our visual system is able to recognize most of the affine transformations of the leopard. Insights of the mechanisms allowing affine invariance in human vision are proposed

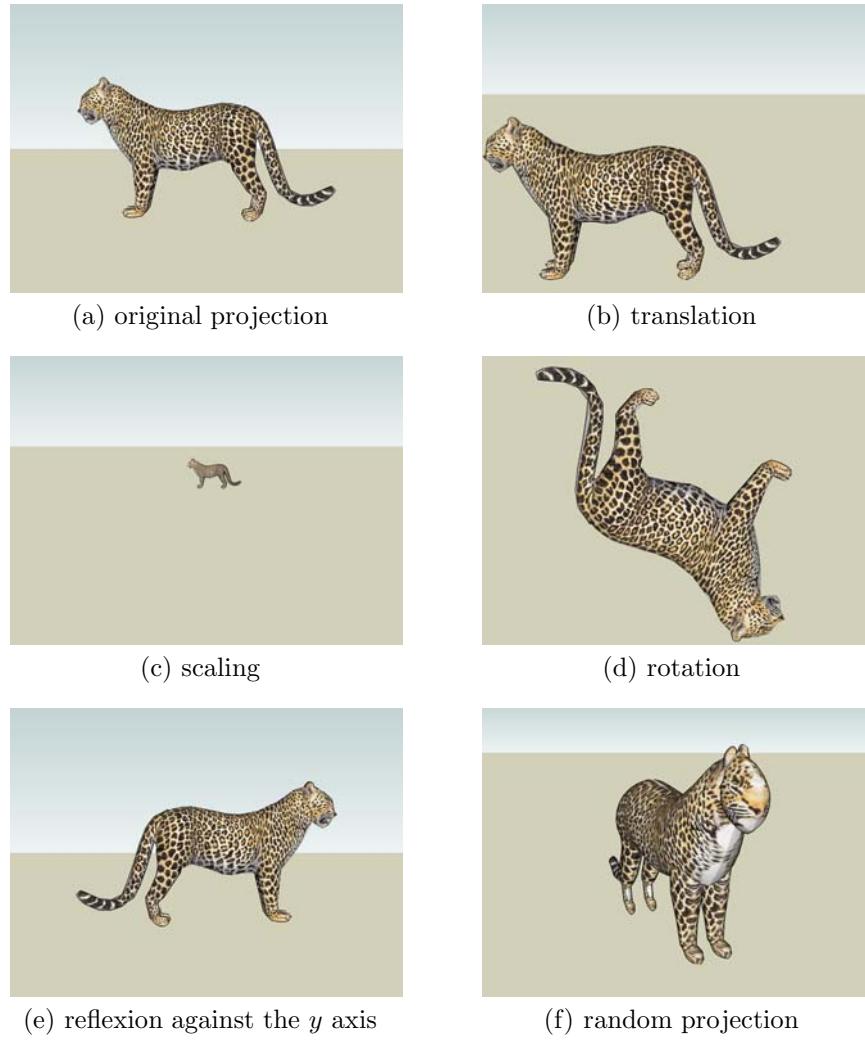


Figure 1.6: Affine transformations of 3D objects and textures. Our visual system is able to recognize most of the affine transformations of the leopard. (f) shows a combination of linear transformations of (a).

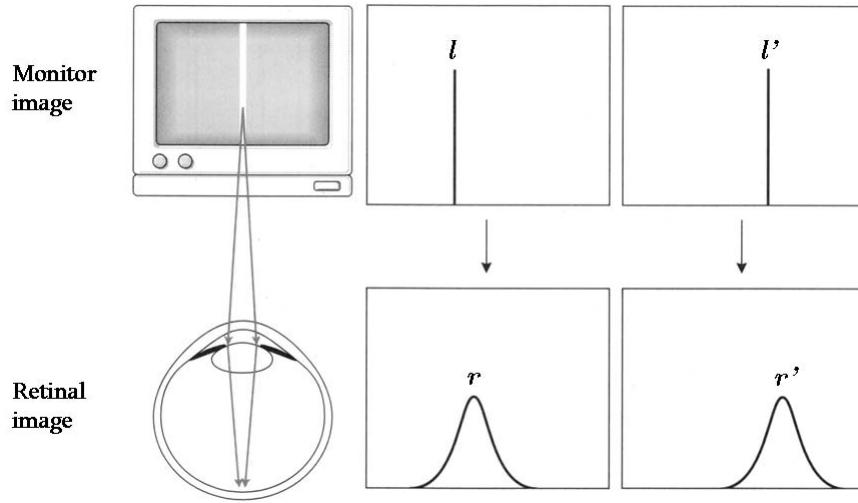


Figure 1.7: Shift invariance of the human visual system: retinal images of two parallel lines displayed on a monitor are shifted versions of the same image (source [267]).

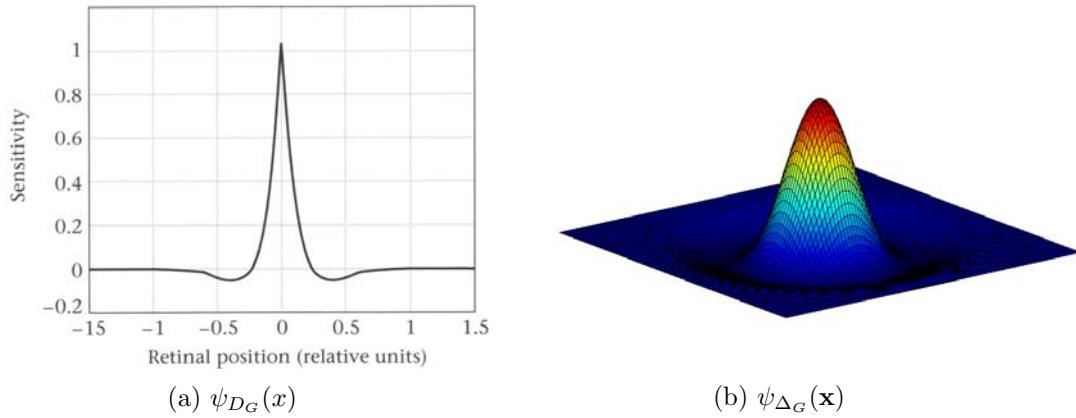


Figure 1.8:  $\psi_{fovea}$  is best modeled by the 2D Laplacian of Gaussian  $\psi_{\Delta G}(\mathbf{x})$ .

in the next paragraph with a focus on the foveal region, which is the main source of pattern information.

Shift invariance is a special property of the optics of the eye. By looking at the image projected on the retina (retinal image) near the fovea, it was shown that two parallel lines  $l$  and  $l'$  observed on a monitor will create identical (but shifted) retinal images  $r$  and  $r'$  (see Figure 1.7) [27, 267].

Let then consider the array of neurons in the foveal region as a shift-invariant linear receptive field which is equivalent to the convolution kernel of the shift-invariant mapping to a so-called neural image. This convolution kernel is also called the psychophysical linespread function  $\psi_{fovea}$  [267]. That is, the motion of our eyes screens the observed scene with fovea which is similar to a random convolution<sup>2</sup> of  $\psi_{fovea}$  with the observed scene or image. In 1956, Schade suggested that the general shape of  $\psi_{fovea}$  can be described using the difference of two Gaussian functions  $\psi_{DG}$  of the same mean  $\mu$  [214] (see Figure 1.8 (a)):

$$\psi_{DG}(x) = \frac{1}{\sigma_1 \sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma_1^2}} - \frac{1}{\sigma_2 \sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma_2^2}}. \quad (1.2)$$

This suggestion was supported by Rodieck in 1965 [203] and then Enroth-Cugell *et al.* in 1966 [74]. Later, it was found that  $\psi_{fovea}$  was best modeled by the two-dimensional Laplacian of the Gaussian

<sup>2</sup>Under certain conditions, random convolutions allow the perfect reconstruction of a sparse signal [205].

kernel  $\psi_{\Delta_G}$  (Eq. 1.3, Figure 1.8 (b)), from which the difference of two Gaussians is actually an approximation when  $\sigma_2/\sigma_1 \cong 1.6$  [119, 154].

$$\psi_{\Delta_G}(\mathbf{x}) = -\frac{1}{\pi\sigma^4} \left(1 - \frac{\mathbf{x}^2}{2\sigma^2}\right) e^{-\frac{\mathbf{x}^2}{2\sigma^2}} \quad (1.3)$$

In the context of their theory on edge detection describing the early human visual perception (so-called *raw primal sketch*), Marr *et al.* (1980) showed that  $\psi_{\Delta_G}$  has indeed several desirable properties for image analysis. Considered as a band-pass filter, it offers an appropriate trade-off between spatial localization and bandwidth which allows to locally examine a portion of the spectrum of the image. They suggest that the *raw primal sketch* is constituted by a collection of images filtered with  $\psi_{\Delta_G}$  with different values of  $\sigma$  in which the zero-crossings are detected (i.e. the locations where the filtered images turn from a negative value to a positive one, and inversely). This calculation might be implemented in the nervous system by using neurons with a variety of receptive-field properties in the visual cortex. Because the individual component images obtained from convolutions with  $\psi_{\Delta_G}$  using several values of  $\sigma$  are assumed to represent different spatial frequency resolutions, the neural image is a multiresolution representation which is scale-invariant [240]. Indeed, we know from the wavelet theory that a continuous signal can be perfectly reconstructed from the coefficients of several convolutions with an appropriate filter each covering successive octaves of the spectrum [150]. The analogy to the wavelet theory goes even further: it can be proved easily that  $\psi_{\Delta_G}$  fulfills the conditions for being an appropriate filter: a wavelet.

The two Nobel prizes Hubel and Wiesel discovered in 1959 that, unlike receptive fields in the LGN, most of the neurons of the visual cortex are orientation-selective [101, 267]. By combining several neurons that are selective to distinct orientations, our visual system is rotation-invariant and is able to recognize any rotated version of an object. However, we will discover in Section 3.1.2 that when the analyzed patterns do not have an a priori prevailing orientation, an interesting solution to obtain rotation-invariant descriptions of the patterns is to use isotropic filters [154, 255].

In this section, it was stated that affine-invariant features allow an accurate and compact representation for further classification of the objects and textures and correspond to our perception of visual information. In this thesis, this assumption will be followed to develop affine-invariant texture features that will be evaluated on high-resolution computed tomography (HRCT) pulmonary images of patients affected with interstitial lung diseases (ILD) in Chapter 3.

### Texture descriptors in computer vision

Texture analysis in digital image processing has been an active research domain over more than thirty years [87, 199, 288]. In [87], texture in digital images is defined as nonconfigurative and cellularly organized areas of pixels. Such patterns can be described by a given spatial organization of gray levels (e.g., random, periodic). However, despite many attempts, there are no appropriate mathematical models for describing homogeneous image textures [150]. Indeed, a texture region is considered as homogeneous if it is perceived as being homogeneous by a human observer.

Texture analysis is recognized as being particularly relevant to describe the visual content of medical images that is often not characterized by well-defined objects but has well-defined visual aspects [30, 244].

**Early texture descriptors** Early examples of texture features are the autocorrelation function, textural edginess, measurements derived from mathematical morphology (e.g. top-hat transform), run-length (RLE) and gray-level co-occurrence matrices (GLCM), the latter being the most popular of the lot [88, 199]. In a two-dimensional digital image  $I(\mathbf{x})$  of dimensions  $n \times m$ , GLCM counts the number of occurrences of a sequence of gray-level values of a defined length and along a defined direction given by offsets in the  $x$  and  $y$  directions:  $\Delta x$  and  $\Delta y$ . An example is given in Figure 1.9. Mathematically, GLCMs are defined as follows:

$$\text{GLCM}_{\Delta x, \Delta y}(i, j) = \sum_{p=1}^n \sum_{q=1}^m \begin{cases} 1, & \text{if } I(p, q) = i \text{ and } I(p + \Delta x, q + \Delta y) = j, \\ 0, & \text{otherwise.} \end{cases} \quad (1.4)$$

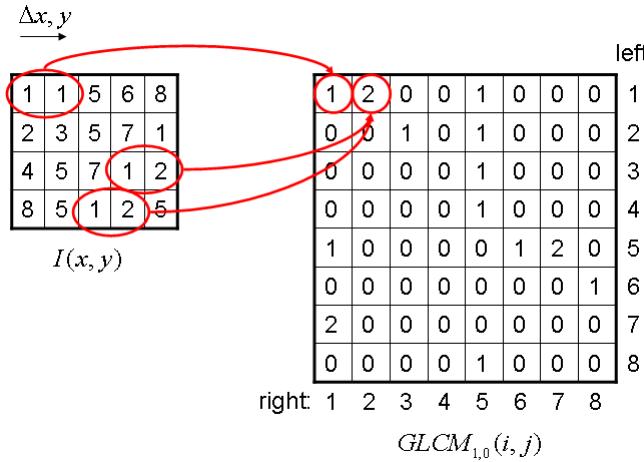


Figure 1.9: An example of GLCM for  $\Delta x = 1$  and  $\Delta y = 0$  of an image with 8 gray-levels.

From GLCM, various metrics are often computed to obtain a more useful set of features, called Haralick features that will not be detailed here. For more details see [88]. GLCM are not rotation-invariant. As an approximation, several GLCM matrices can be computed with different  $\Delta x/\Delta y$  ratios to span 180 degrees. Typically, four matrices are computed at 0, 45, 90 and 135 degrees. Another drawback is the size of the matrices, which is  $N_{gray} \times N_{gray}$ , with  $N_{gray}$  the number of gray-levels of  $I$ . For example, to analyze 12-bit images from a CT scanner, the size of each matrix is equal to  $2^{12} \times 2^{12}$  which is impractical because the number of possible co-occurrences is huge and cause problems to be calculated and stored. A reduction of  $N_{gray}$  is required which results in an important loss of relevant information. This can be handled by compressing the grayscale or by considering an alternative representation in terms of sums and differences [247].

**Spectral analysis and wavelet transform** Julesz introduced in 1981 the texton theory which was an important step towards understanding the different parameters that influence the perception of textures [107]. The idea behind is that our pre-attentive visual system uses second-order measures to discriminate among textures. As the second-order statistics determine the autocorrelation function and the Fourier transform of the autocorrelation function is the power spectrum, the *iso-second-order* textures have identical autocorrelation functions and identical power spectra. Based on the property that some image patterns (especially, periodic ones) are well described in terms of sinusoidal components and the repartition of the energy in the spectrum, the Fourier transform has been proposed for texture analysis [11, 99]. However, the latter is not appropriate for segmentation because the Fourier transform is neither shift-invariant nor rotation-invariant and thus Fourier coefficients must be evaluated over neighborhoods of varying sizes and in various directions.

An attractive solution for carrying out a more local texture analysis, which also takes into account scale, is to use the wavelet transform (WT) [6, 150]. It was mentioned above that the WT is very close to the way the human visual system analyzes image patterns. Wavelets are mathematical analysis functions that decompose signals into subbands, and then analyze each component with a resolution matching its scale (the analyzed image is iteratively subsampled by a factor of 2 in the standard dyadic discrete version). In  $\mathbb{R}^1$ , the admissibility condition for a wavelet  $\psi$  is to have a zero average:

$$\int_{\mathbb{R}} \psi(x) dx = 0. \quad (1.5)$$

This ensures that the convolution of  $\psi$  with a constant signal  $s(x) = C$  remains zero. Indeed,  $\psi$  is shaped to detect changes in  $s(x)$ . When  $\psi$  is dilated by a factor  $a$  and translated by  $b$ , we can

create a spanning set of functions  $\psi_{a,b}(x)$  as follows:

$$\psi_{a,b}(x) = \frac{1}{\sqrt{a}}\psi\left(\frac{x-b}{a}\right). \quad (1.6)$$

It will be seen in Section 3.1.2 that the family of functions  $\psi_{a,b}(x)$  may be a base or not. To obtain a wavelet transform of  $s(x)$  at the scale  $a$  and position  $b$ ,  $s$  is correlated with  $\psi$ :

$$W_\psi^s(a, b) = \int_{\mathbb{R}} s(x) \frac{1}{\sqrt{a}}\psi^*\left(\frac{x-b}{a}\right) dx, \quad (1.7)$$

with  $\psi^*$  the complex conjugate of  $\psi$ . This corresponds to a scalar product of  $s(x)$  with the family of templates

$$\mathcal{S} = \{g_1(x-b), \dots, g_J(x-b), h_J(x-b)\}_{b \in \mathbb{R}}$$

where  $h_j$  stands for a lowpass filter at iteration  $j$  and  $g_j$  a family of highpass filters with  $j = 1, \dots, J$ .  $\mathcal{S}$  is derived from the set of functions  $\psi_{a,b}(x)$ . The associated decomposition algorithm is

$$\begin{cases} G_j(x) := \langle G_j(x-b), s(x) \rangle_{l_2} \\ H_J(x) := \langle H_J(x-b), s(x) \rangle_{l_2} \end{cases} \quad (1.8)$$

where  $\langle \cdot, \cdot \rangle_{l_2}$  is the scalar product using the  $l_2$  norm.  $G_j$  contains coefficients generated by the convolution of  $s(x)$  with the highpass filters at iteration  $j$  and  $H_J$  the convolution of  $s(x)$  with the lowpass filter at the last iteration  $J$ .

Using two-dimensional isotropic wavelets  $\psi_{iso}(\mathbf{x})$ , the wavelet transform is easily extended to the second dimension as:

$$W_{\psi_{iso}}^I(a, \mathbf{b}) = \int_{\mathbb{R}^2} I(\mathbf{x}) \frac{1}{\sqrt{a}}\psi_{iso}^*\left(\frac{\mathbf{x}-\mathbf{b}}{a}\right) d\mathbf{x}. \quad (1.9)$$

The WT thus allows local multiscale analyzes and is particularly well suited for the representation of piecewise-smooth signals, as well as stochastic processes with a fractal-like behavior, which partly explains their success in biomedical imaging applications [76, 250]. When compared to the Fourier series, the approximation error based on the number of coefficients  $N_c$  decreases faster following  $1/2^{N_c}$  (wavelet bases) instead of  $1/N_c$  with the Fourier series. However, WT must be used with precautions and it will be seen in Section 3.1.2 that the sampling of  $I(\mathbf{x})$  alters the affine-invariance of the features derived from WTs in several ways. It will be shown that near affine-invariant visual features can be obtained with particular WT designs [65, 252, 255, 258].

### Color histograms

Complementary to the characterization of spatial dependencies, the distributions of the colors and intensities within the image carry out relevant information for texture and object recognition. For example, color information is enough to recognize an orange within a basket of apples. In our visual system, the recognition of colors is performed already in the retina, where three types of cones are sensitive to different wavelengths corresponding to the three colors red, green and blue [266]. In computer vision, colors are often described with statistical measures of color (or intensities in grayscale images) histograms (see Figure 1.10). Commonly used first-order texture measures derived from histograms are mean, standard deviation, mode (index of the bin with highest value), skewness and kurtosis, etc...

### 1.2.2 Trained brain: supervised machine learning

A feature set that describes the observed scene well is a prerequisite step towards the classification of the represented objects and, if applicable, a global decision taken from the observed scene and its context. In the animal world, recognition of complex objects and decision-making is processed beyond the visual cortex by cognitive schemes implemented by several interactions

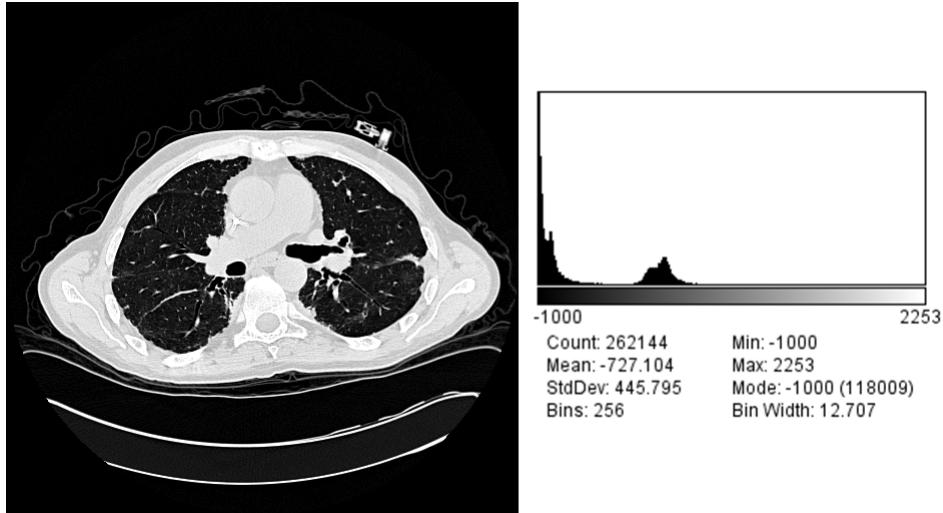


Figure 1.10: Gray–level histogram and statistical measures of an axial slice of HRCT of the lung.



Figure 1.11: Generalization: one can easily categorize the object depicted on image (b) as a “car” even without having seen it before.

between the lobes of the brain [184]. The ability of recognizing objects is usually not given by birth but is rather gained through learning and experience. Learning is defined as the aptitude to modify one’s responses to an event through repeated experiences. It is not restricted to learning by heart. Indeed, humans are able to recognize an object of a given class without having seen it before and have the ability to generalize the global shapes of the objects. For example, one will surely be able to recognize that Figure 1.11 (b) depicts a car even without having seen this specific model before but by knowing the general characteristics (features) and variations that define the class of car as illustrated by the set of images in Figure 1.11 (a).

In computer vision, once the feature space is built, algorithms have to be used to detect and create boundaries among the several classes defined by clusters of instances in the feature space. Similar to the way we learn, computers can be taught to recognize objects. One such learning technique is called *supervised learning*. During this learning process, the underlying parameters of the machine learning algorithm are optimized in order to draw decision boundaries that minimize the errors of categorization of instances with known class labels. The goal is to find the functions  $\mathcal{F}$  which best model the boundaries among the distinct classes represented in the feature space. The best functions are those that perform classification of a test set with the lowest error rate. The test set is composed of labeled instances, which have not been used to train the classifier. It simulates future unknown instances and thus allows measuring the generalization performance (GP) [262]. The objective is to minimize the error rate on the training set (empirical risk minimization) while

at the same time avoiding overfitting of the training instances (true risk minimization).

Several approaches are available to implement  $\mathcal{F}$ . Three general approaches including five classifier families are studied in this work:

- learning by density estimation with naive Bayes and k–nearest neighbor (k–NN) classifiers,
- recursive partitioning of the feature space or learning by the process of elimination with decision trees,
- nonlinear numerical approaches with the multi–layer perceptron (MLP) and kernel support vector machines (SVM).

In practice, the choice of a classifier family is a difficult problem and it is often based on the classifier happening to be available or best known to the user [19, 104]. A methodology for this is proposed in Section 4.1.1.

### Learning by density estimation

Learning by density estimation relies on the assumption that the instances belonging to the same class form homogeneous clusters in the feature space. Two approaches are studied: a probabilistic one with the naive Bayes classifier and a geometric one with k–NN.

**Naive Bayes** The naive Bayes classifier is based on a probability model and assigns the class, which has the maximum estimated posterior probability to the feature vector extracted from the ROI. The posterior probability  $P(w_j|\mathbf{v})$  of a class  $w_j$  given a feature vector  $\mathbf{v}$  is determined using Bayes’ theorem:

$$P(w_j|\mathbf{v}) = \frac{P(\mathbf{v}|w_j)P(w_j)}{P(\mathbf{v})} \quad (1.10)$$

This method is optimal when the attributes are orthogonal. However, in practice it performs well without this assumption. The simplicity of the method allows good performance with small training sets [259]. Indeed, by building probabilistic models, it is robust to outliers (i.e. feature vectors that are not representative of the class to which they belong). Moreover, it creates soft decision boundaries, which has the effect to avoid overtraining. However, the arbitrary choice of the distribution model for estimating the probabilities  $P(x)$  along with the lack of flexibility of the decision boundaries results in limited performance for complex multiclass configurations.

**k–NN** The k–nearest neighbor classifier cuts out hyper spheres in the space of instances by assigning the majority class of the  $k$  nearest instances according to a defined metric (e.g. Euclidean distance) [44]. It is asymptotically optimal and its straightforward implementation allows rapid tests for example for evaluating features. However, several shortcomings are inherent to this method. It is very sensitive to the curse of the dimensionality. Increasing the dimensionality has the effect to sparse the feature space and local homogeneous regions that represent the prototypes of the diverse classes are spread out. The classification performance strongly depends upon the used metric [259]. Moreover, a small value of  $k$  results in chaotic boundaries and makes the method very sensitive to outliers.

### Learning by the process of elimination: decision trees

Decision trees carry out recursive partitioning of the feature space by hierarchically testing features. A commonly used algorithm is the C4.5 introduced by Quinlan in 1986 [195] and consists of dividing the feature space successively by choosing primarily features with the highest information gain  $I_G$  (similar to Kullback–Leibler divergence [128]). In medicine, it is in correspondence to the approach used by clinicians to establish a diagnosis by answering successive questions. This is nevertheless only partially true when radiologists interpret HRCT images, where the information contained in images does not appear successively but at a glance.

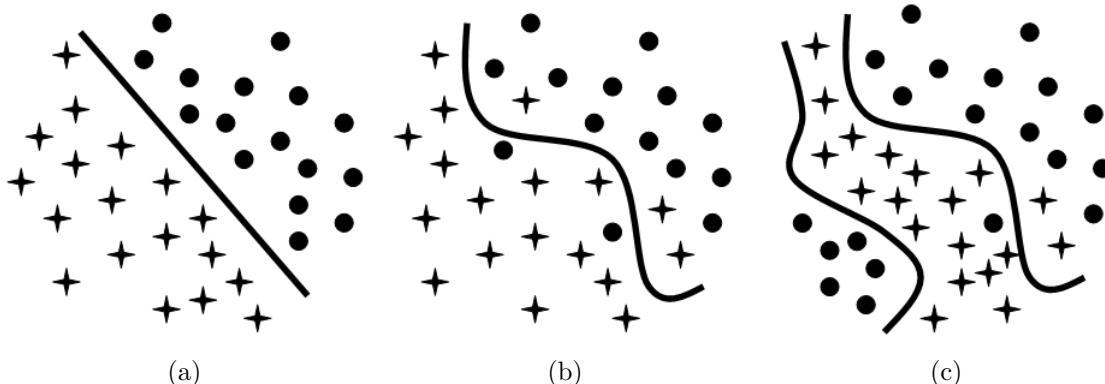


Figure 1.12: A bidimensional illustration of non-linear classification problems. (a): the two clusters of instances are linearly separable. (b) and (c): intricate decision boundaries are required to best separate the two classes.

Decision trees are robust to noisy features as only attributes with high information gain are used. However they are sensitive to the variability of data. The structure of the tree is likely to change completely when a new instance is added to the training set. Another drawback is its incapability to detect interactions between features as it treats them separately. This results in decision boundaries that are orthogonal to dimensions, which is not accurate for highly nonlinear problems. Two main parameters influencing the generalization performance require optimization:

- $N_{instances}$  : the minimum number of instances per leaf, which determines the size of the tree.
- $C_{pruning}$  : the feature confidence factor used for pruning the tree, which consists of removing branches that are deemed to provide little or no gain in statistical accuracy of the model.

### Nonlinear numerical approaches

Nonlinear numerical approaches are able to model any decision boundary  $\mathcal{F}$  and thus offer unlimited flexibility. On the one hand, they are well-suited for non-linear problems where the classes do not form localized clusters but several clusters as illustrated in Figure 1.12 (c). On the other hand, the large flexibility increases the risk of overtraining where the decision boundary fits perfectly the training set which is not necessarily representative of the complete population. In this case, regularization constraints are employed to smooth the decision boundaries.

**MLP** Multi-layer perceptrons (MLP) are inspired by the human nervous system, where information is processed through interconnected neurons [18]. The MLP is a feed-forward neural network, which means that the information propagates from input to output. The inputs are fed with values of each feature and the outputs are providing the class value. With one layer of neurons, the output is a weighted linear combination of the inputs. This network is called the linear perceptron. By adding an extra layer of neurons with nonlinear activation functions (the hidden layer), a nonlinear mapping between the input and output is possible [105]. The training phase consists of iterative optimization of the weights connecting the neurons by minimizing the mean squared error rate of classification. The learning rate  $R_{learn}$ , which controls the adjustments of the weights during the training phase must be chosen as a trade-off between error on the training set and overtraining. Another critical parameter is the number of units  $N_{hidden}$  of the hidden layer. Indeed the MLP is subject to overfitting and requires an optimal choice of the parameters for regularization. The MLP can create models with arbitrary complexity by drawing unlimited decision boundaries. It is also robust to noisy features as these will obtain a low weight after training.

**Kernel support vector machines** Kernel support vector machines (SVMs) implicitly map input feature vectors  $\mathbf{v}_i$  to a higher dimensional space by using the kernel function  $K(\mathbf{v}_i, \mathbf{v}_j) =$

$\langle \phi(\mathbf{v}_i), \phi(\mathbf{v}_j) \rangle$  which needs to be a positive definite function. For example, the Gaussian kernel is defined by:

$$K(\mathbf{v}_i, \mathbf{v}_j) = e^{-\frac{\|\mathbf{v}_i - \mathbf{v}_j\|^2}{2\sigma_K^2}} \quad (1.11)$$

with  $\sigma_K$  being the width of the Gaussian to determine. The Gaussian kernel maps each feature vector to a Gaussian (which is continuous) and thus spans a space of infinite dimension.

In the transformed space, a maximal separating hyperplane is built considering a two-class problem. Two parallel hyperplanes are constructed symmetrically on each side of the hyperplane that separates the data. The goal is to maximize the distance between the two external hyperplanes, called the margin [25, 262]. An assumption is made that the larger the margin is the better the generalization error of the classifier will be. Indeed, SVMs were developed based on the *Structural Risk Minimization* principle, which seeks to minimize an upper bound of the generalization error while most of the classifiers aims at minimizing the empirical risk, the error on the training set [42, 43]. The SVM algorithm aims at finding a decision function  $f(\mathbf{v})$ , which minimizes the functional:

$$\min C \sum_i^{N_v} \max(0, 1 - y_i f(\mathbf{v}_i))^2 + \|f\|_K \quad (1.12)$$

where  $N_v$  is the total number of feature vectors,  $\|f\|_K$  is a norm in a *Reproducing Kernel Hilbert Space*  $\mathcal{H}$  defined by the positive definite function  $K$ , which means that the functionals  $f$  are bounded.  $y_i$  is the label of  $\mathbf{v}_i$  with  $y_i \in \{-1; 1\}$  (2-class problem). The parameter  $C$  determines the cost attributed to errors and requires optimization. For the multiclass configuration, several SVM models are built using either *one versus one* or *one versus all* combinations. Finally, the majority class is attributed.

In summary, SVMs allow training generalizable, nonlinear classifiers in high-dimensional spaces using a small training set. This is enabled through the selection of a subset of vectors (called the support vectors), which characterize the true boundaries between the classes well.

## 1.3 CAD in HRCT imaging of the chest

HRCT imaging of the chest provides a three-dimensional view of the organs with submillimetric resolution in axial sections. It has become the gold standard for diagnosing interstitial lung diseases which induce diverse alterations of the lung tissue that have characteristic textural signatures. HRCT imaging of the chest thus represents an ideal candidate for developing texture-based CAD systems.

In this section, the challenges of diagnosing interstitial lung diseases are detailed. The importance of HRCT imaging of the chest in the diagnosis process of ILDs and the efforts that a correct interpretation represents are presented.

### 1.3.1 Interstitial lung diseases

Interstitial lung diseases (ILDs) can be characterized by the gradual alteration of the lung parenchyma leading to breathing dysfunction. They regroup more than 150 histological diagnoses associated with disorders of the lung parenchyma [114]. The factors and mechanisms of the disease processes vary from one disease to another and the cause of many ILDs is still unknown [114]. Physical examination of a patient affected by ILD is frequently abnormal but with unspecific findings. The diagnosis of these pathologies is established based on the complete history of the patient, a physical examination, laboratory tests, pulmonary function testing (PFT) as well as visual findings on chest X-ray.

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 [78]. The first imaging examination used is the chest radiograph because of its low cost and weak radiation exposure. It also provides a quick overview of the whole chest. However, chest radiographs are normal in more than 10% of the patients with some forms of ILD and can provide a confident

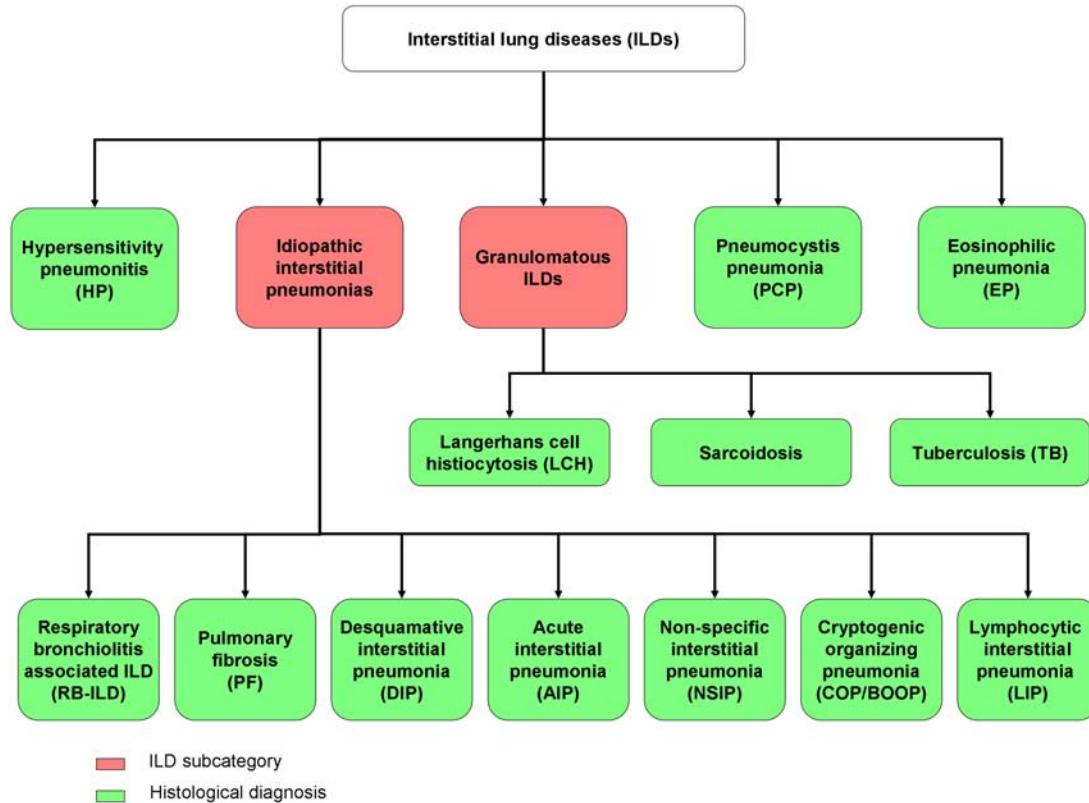


Figure 1.13: Most common histological diagnoses of ILDs according to [114] and [268].

diagnosis in only 23% of the cases with lung diseases in general [236]. When the synthesis of this information arouses suspicions toward an ILD, HRCT imaging of the chest is often required to acquire a rapid and accurate visual assessment of the lung tissue. Indeed, the three-dimensional form of HRCT data avoids superposition of organs and provides an accurate assessment of the pattern and distribution of the lung tissue with a submillimetric resolution. It quickly became the gold standard imaging protocol for the diagnosis of diffuse pulmonary parenchymal diseases.

The most common histological diagnoses of ILDs according to [114] and [268] are detailed in Figure 1.13 and the associated lung tissue patterns in HRCT are listed in Table 1.5.

### High-resolution computed tomography of the lungs

The first patient scan acquired with X-ray computed tomography (CT) was acquired in 1972 at the Atkinson Morley Hospital in England. The method utilizes an X-ray tube which rotates axially around the patient and a diametrically opposed array of detectors detects the residual radiation traversing the body. Three-dimensional arrays of pixels are reconstructed using the inverse Radon transform. The numerical value of each pixel is related to the X-ray attenuation and are expressed in Hounsfield Units (HU). HU values are obtained by computing the difference in X-ray attenuation between the observed material (e.g. bone, lung tissue, air) and the water providing an absolute measure of the density of the organs. A correspondence of the density of certain organic material and the HU value is shown in Figure 1.14.

In 1972, the first commercialized CT scanner was creating image series with relatively low resolution of an  $80 \times 80$  pixel matrix in each axial slices. Nowadays, scanners with multiple detectors and using a helical scanning mode can provide a 3D array of isotropic voxels with a submillimetric resolution. This protocol is called multidetector computed tomography (MDCT).

Table 1.4: The HRCT scanning protocol.

slice thickness	1–2 mm
spacing between slices	10–15 mm
scan time	1–2 seconds
lung shape	inspirium
contrast agent	none
axial pixel matrix	512 × 512
$x, y$ spacing	0.4–1 mm

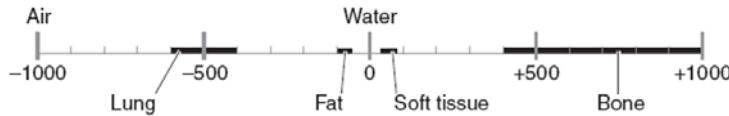


Figure 1.14: Hounsfield Units values of several organic materials.

The main drawback of this protocol is the high amount of radiation dose the patient is exposed to. To assess the visual appearance of healthy and pathological lung tissue, a submillimetric resolution is required but some areas can be skipped between the thin sections to limit the radiation exposure. This protocol is called high-resolution computed tomography (HRCT) and is the gold standard imaging protocol for diagnosing ILDs [174]. The specifications of the HRCT protocol are listed in Table 1.4. At last, HRCT is more appropriate than MRI to assess the visual appearance of the lung tissue. Indeed, MRI is only sensitive to inflammatory changes of the pulmonary parenchyma as other tissue sorts have a low density of protons [147]. A comparison between HRCT and MRI is shown in Figure 1.15.

### Lung tissue patterns associated with ILDs in HRCT

The appearance and quantification of the sorts of lung tissue patterns in HRCT are very informative for establishing the differential diagnosis of an ILDs. The Table 1.5 lists 13 common histological diagnoses of ILDs, the associated HRCT findings as well as the region of the lungs where the disease is predominant. The visual aspects of the most common lung tissue patterns are depicted in Figure 1.16. The taxonomy used to describe them often relates to texture properties. The term *fibrosis* is used in this work to describe all HRCT findings that are associated with the histological diagnosis “pulmonary fibrosis” and includes *reticulation*, *traction bronchiectasis*, *architecture distortion* and *honeycombing* [235]. As observed in Table 1.5, *ground glass* patterns are encountered in most of the ILDs and is thus non-specific. Therefore, the clinical context and other HRCT findings are required to orient the diagnosis.

**Interpretation of HRCT image series** Interpreting HRCT images of the chest represents a challenge even for trained chest radiologists and lung specialists [236, 268]. The three-dimensional form requires significant reading time, effort, and experience for a correct interpretation [66]. Most often, the interpretation process is carried out by comparing a case with similar images in textbooks such as [268] or with similar cases in personal image collections, which are most often organized by pathology. To do so, the radiologists must have a guess of the suspected disease present in the image and may miss the true pathology shown. In certain medical services (e.g. emergency radiology service), radiologists have recourse to a large diversity of imaging modalities such as conventional projection radiography, CT, MRI, functional imaging (fMRI, PET), and ultrasound applied to different organs such as the brain, colon, breast, chest, liver, kidney and the vascular and skeletal systems. They have to provide a first radiological report with ideas on the diagnosis quickly. This may result in errors by omission or confusion of diverse pathologic lung tissues [108].

Table 1.5: 13 common histological diagnoses of ILDs and associated HRCT findings.

histological diagnosis	HRCT lung tissue patterns	predominance
Hypersensitivity pneumonitis (HP)	<i>ground glass, emphysema, fibrosis</i>	diffuse
Pneumocystis pneumonia (PCP)	<i>ground glass, crazy-paving, cysts, pneumothorax</i>	central, perihilar
Eosinophilic pneumonia (EP)	<i>ground glass, consolidation, crazy-paving</i>	peripheral, apex
Langerhans cell histiocytosis (LCH)	<i>cysts, ground glass, micronodules, reticulation</i>	apex
Sarcoidosis	<i>micronodules, consolidation, macronodules, ground glass, fibrosis (end-stage)</i>	peribronchovascular, subpleural, peripheral
Tuberculosis (TB)	<i>micronodules (miliary), tree-in-bud, consolidation</i>	diffuse
Respiratory bronchiolitis associated ILD (RB-ILD)	<i>ground glass, emphysema</i>	diffuse, centrilobular
Pulmonary fibrosis (PF)	<i>fibrosis, bronchiectasis, ground glass</i>	peripheral, subpleural, basal, posterior
Desquamative interstitial pneumonia (DIP)	<i>ground glass, , emphysema, fibrosis (uncommon)</i>	subpleural, basal
Acute interstitial pneumonia (AIP)	<i>ground glass, consolidation</i>	basal, diffuse
Non-specific interstitial pneumonia (NSIP)	<i>ground glass, consolidation, reticulation, fibrosis (uncommon)</i>	peripheral, basal
Cryptogenic organizing pneumonia (COP/BOOP <sup>a</sup> )  _____ aBronchiolitis obliterans organizing pneumonia (BOOP) was formerly used and replaced by COP	<i>ground glass, consolidation (patchy), macronodules, macronodules, bronchial wall thickening, crazy-paving</i>	peribronchovascular, subpleural
Lymphocytic interstitial pneumonia (LIP)	<i>ground glass, micronodules</i>	peribronchovascular, subpleural

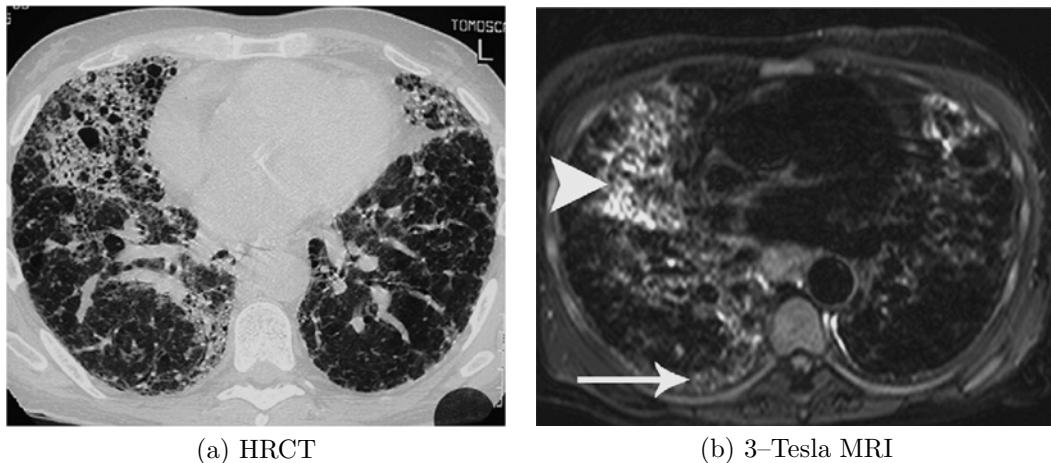


Figure 1.15: A comparison of HRCT and MRI for the visual assessment of the lung tissue (source [147]). In the MRI image (b), the signal of the infected lung-area (arrowhead) is 120% higher than fibrous tissue area (arrow).

Moreover, the context is fundamental for correct interpretation: healthy tissue, for example, may have different visual aspects depending on the age or the smoking history of the patient and *ground glass* findings are non-specific without complementary clinical parameters [177].

### 1.3.2 Texture-based CADs for lung tissue analysis in thin-section CT

Owing to the intrinsic complexity of the interpretation of HRCT image series, a real-time image-based computerized assistance appears useful for radiologists. Computerized texture analysis of the lung parenchyma in thin-section CT has been a lively research topic for more than ten years with contributions from about 24 research groups around the world. The challenging and tedious work that represents a meticulous interpretation a thin-section CT image series in a time-constraining clinical environment calls upon computerized assistance with CAD systems. Mostly four CAD scenarios were proposed in the literature:

- abnormality detection,
- categorization of the lung tissue,
- quantification of the disorders,
- retrieval of similar images or cases with known diagnosis.

The first three scenarios are related to detection-based CADs (Section 1.1.1) and the last one refers to CBIR-based CADs (see Section 1.1.2). Most often, the proposed detection-based CAD systems aim to assess a combination of these scenarios with a purpose of providing a “second opinion” to the radiologist. Texture analysis is well adapted for analyzing the lung parenchyma affected with ILDs and chronic obstructive pulmonary disease (COPD), but neither for nodule detection nor for the segmentation of the bronchial tree. Thereby, the literature review proposed in this section focuses on texture-based CAD systems and does not include studies on nodule detection or bronchus characterization in thin-section CT. Recent surveys with a wider scope on CAD systems in CT imaging of the chest are proposed by Doi *et al.* [140], Sluimer *et al.* [224] and Sonka *et al.* [231]. The performances of the various CADs are not compared in this review as strongly depending of the dataset, the number of patterns included and the validation methods.

In order to explore and compare the scopes and methods of the various contributions in the field, this section is structured as follows. A chronological overview of the techniques used is proposed in the first subsection. The scopes of the studies are detailed in Tables A.1, A.2, A.3 and

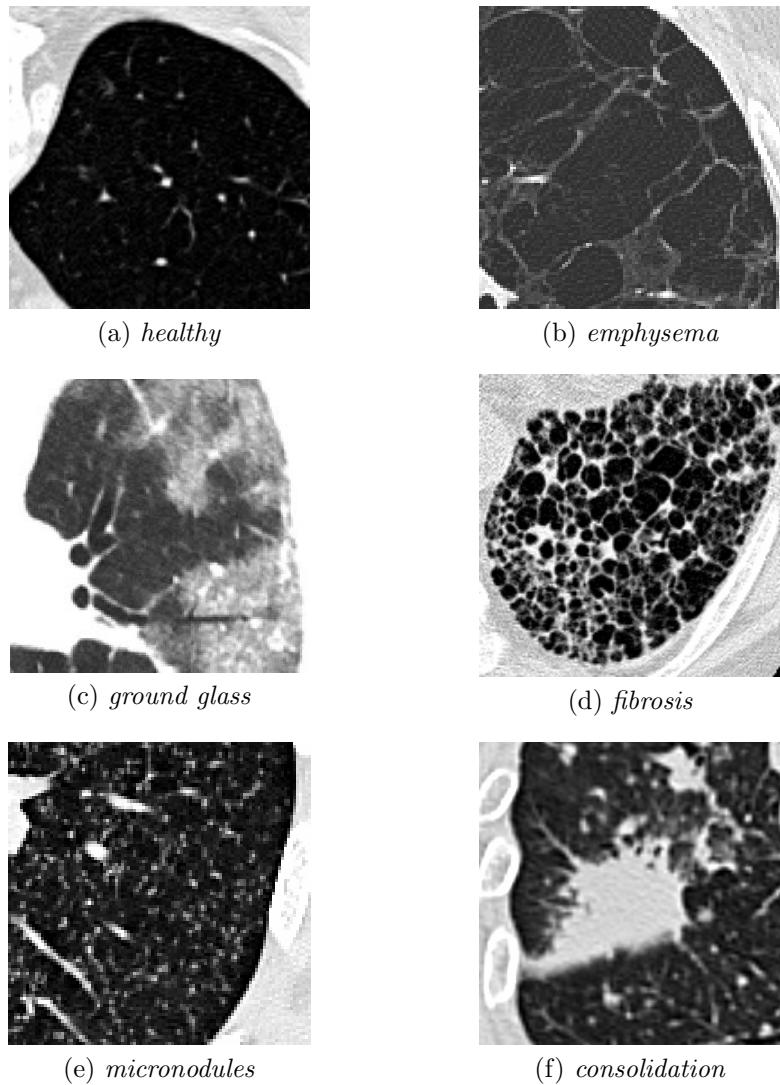


Figure 1.16: Visual aspects of the most common lung tissue patterns in HRCT of patients with ILDs.

A.4 and the methods are summarized in Tables A.5, A.6, A.7 and A.8. Some statistics and trends of the various research groups are identified and discussed in the second subsection.

### Texture-based lung parenchyma analysis in the literature

Although some studies proposed semi-automated quantification of the parenchymal disorders in the early nineties, the first fully automated texture-based CAD system for ILDs was proposed by Delorme *et al.* in 1997 [47]. 2D HRCT slices are sampled with  $5 \times 5$  blocks at three scales. For each block, GLH, GLCM and RLE features are extracted and mined with linear discriminant analysis (LDA) to be categorized into 7 classes of lung tissue. It yields an automated detection, characterization and quantification of the lung tissue. LDA may be not appropriate for the separation of the lung tissue classes in the proposed feature space, though. Moreover, the validation method used does not assess the true generalization performance as it allows to train and test the system with blocks belonging to a same patient. A few months later, Heitmann *et al.* [90, 110] proposed the usage of self-organizing maps (SOM) [122] at the pixel level to detect *ground glass* opacities. The reported performance obtained with this method are again not representative of a clinical setup because the SOM are trained and tested with identical images.

The first lung CBIR system “ASSERT” was introduced by Shyu *et al.* in 1999 [23, 220]. A selection of features from GLH, GLCM and Sobel’s edge detection based on sequential forward selection (SFS) is analyzed with a decision tree to assess the similarity between images. An evaluation of ASSERT in clinical conditions showed a positive impact of the CAD on the diagnosis outcomes, especially when used by little experienced radiologists [4]. A limitation of this system is that the user has to delineate a suspicious ROIs as query for retrieval, which does not enable automatic analysis of whole HRCT image series but is rather based on a physician-in-the-loop scheme. Moreover, the images are stored in the JPEG<sup>3</sup> format (and not DICOM), which does not allow to use the whole range of HU values.

An adaptive multiple feature method (AMFM) was developed by Uppaluri *et al.* in 1999 [85, 95, 253, 254]. The feature set consists of a selection of GLH, GLCM, RLE and fractals using correlation-based feature selection (CFS) and divergence [5] to find the most discriminative ensemble of features on a training set. These features are then classified into 6 classes of lung tissue using a naive Bayes classifier. The correlation of the features and the pulmonary function tests is studied. This method assumes that no relevant information is contained in the spacial frequencies directly. The agreement of the CAD output with the radiologists shows reliable performance but the assessment of the classical performance measures is carried out on a test set that may contain instance from patients that have been used for training yielding to a biased evaluation of the CAD. The extension of the AMFM feature set by Xu *et al.* in 3D MDCT imaging in 2005 [279–281] shows the superiority of 3D-based approaches. However, MDCT imaging is not widely used in clinical routine because of high X-ray radiation exposure and accurate CAD systems for HRCT are required.

Gabor filters showed higher performance compared to GLCM for classifying 5 classes of lung tissue with a k-NN classifier using Euclidean distance in [160]. Unfortunately, the validation method is not representative of a clinical usage as the number of each class is equal in the test set. Moreover, the random draw of ROIs to create training and testing set allow ROIs from the same patient in both sets of instances.

In 2000, Liu *et al.* proposed a CBIR-based CAD [144, 145] using Fourier analysis and SOM networks for computing the distance measure. The evaluation is in accordance with clinical conditions but is limited as based only on two classes which are *healthy* and *pathological*. Similarly to the ASSERT CBIR system, the user has to select blocks in 2D HRCT slices as query for the retrieval which entails the risk of omitting diagnostically useful events in the whole HRCT stack of images.

In 2003, Sluimer *et al.* focused on abnormality detection in HRCT image series from patients with ILDs [225]. The features are based on a filter bank of Gaussian derivatives; namely the Gaussian itself, the Laplacian and oriented first and second derivatives of Gaussians (4 scales, 6

---

<sup>3</sup>Joint Expert Picture Group (JPEG), <http://www.jpeg.org/>, as of 5 November 2009

orientations for the non-axisymmetric filters). A subset of features selected using SFS is mined using various classification algorithms such as LDA, quadratic discriminant analysis (QDA), k-NN and SVM to determine if circular blocks are *healthy* or *pathological*. The validation method reflects performance of a clinical usage but is restricted to the anterior part of the lungs.

Still in 2003, Chabat *et al.* proposed a CAD system for the COPDs [33]. GLH, GLCM and RLE features are used to classify circular blocks using a supervised Bayes classifier into *healthy* and 3 sorts of *emphysema*. The realistic validation shows that the CAD yields high specificity but rather low sensitivity for detecting and categorizing the tissues with decreased attenuation.

Mathematical morphology is used for feature extraction by Uchiyama *et al.* [246] and Doi *et al.* [66] in a CAD system for ILDs in 2003. GLH along with the top-hat transform extract features simultaneously from  $32 \times 32$  and  $64 \times 64$  blocks and classify them into 7 classes with a MLP. A 2-class configuration (*healthy* versus *all*) is also evaluated. Unfortunately the CAD performance in the multiclass configuration are evaluated on the training set and denotes an obvious overfitting of the data. This is not convenient to assess the generalization performance, especially with MLPs that are subject to overfitting as they can draw unlimited decision boundaries.

In 2004, Malone *et al.* used genetic algorithms to select features among GLH, FFT, fractal, autocorrelation measures from three block sizes [151]. Two experiments are run: the first uses manually selected blocks for evaluation whereas the second is based on contiguous blocks laid in the whole slices. The classification performance in the first experiment is much higher than in the second showing that the validation method has to be as close as possible to the clinical situation to really assess the performance of the CAD.

Prasad *et al.* used several machine learning methods with GLH, GLCM and grey-level differences features for the quantification of *emphysema*. Unsupervised and semi-supervised learning is proposed to cope with the problem of acquiring costly ground truth [187, 188]. Pearson's correlation coefficient is used to determine uncorrelated groups of features that are combined using multiple classifiers consisting of naive Bayes and C4.5. Using the group of features described above, the influences of state-of-the-art feature reduction techniques such as independent component analysis (ICA) and Principal Component Analysis (PCA) on CAD performance are investigated in [189–191]. In the same research group, Shamsheyeva *et al.* and Vo *et al.* focused on wavelet-based features for the characterization of ILDs. Quincunx wavelet frames are used in [217, 218] and are combined with empirical knowledge rules in [219]. In [264], contourlets [185] are used to carry out directional analysis of the lung tissue. The combination of the contourlet and discrete wavelet frames showed increased classification performance with *healthy*, *emphysema*, *ground glass* and *fibrosis* patterns. However, the directional analysis increases the computational complexity dramatically for little performance improvement. The cross-validation (CV) used allows for training and testing with blocks from the same patients.

Mendonça *et al.* used the earth mover's distance (EMD) [208] to assess the similarity between two-dimensional GLH and edge distributions to quantify *emphysema* into five levels (including healthy) in [157]. EMD yields unsupervised classification and does not require costly manually segmented ground truth for training. It would be interesting to evaluate this method for local classification of the lung tissues instead of the entire lung volume at the patient level as in this study.

In 2007, Zrimec *et al.* introduced prior knowledge based on the anatomical region for the detection of *honeycombing* patterns in [290, 293]. This addresses the problem of integrating anatomical knowledge for tissue characterization which is widely used by the radiologists for image interpretation and shows significant improvement of the CAD performance. For instance, *honeycombing* patterns have a peripheral and basal predominance. However, this method requires an accurate identification of the peripheral, central, apical and basal regions [291]. Then, features from GLH and GLCM are selected using CSF and classified using naive Bayes and C4.5. A comparison between the performance obtained with a 10-fold CV allowing training and testing with blocks belonging to the same patient and unseen datasets shows how the performance measures are affected by the validation method. In [292], the texture features are compared with a structure-based approach that uses a seeded region growing [274, 275]. The structure-based approach shows superior specificity whereas the texture-based features allow for better sensitivity.

Three-dimensional sum and difference histograms (SADH) [247], GLCM, RLE and GLH with adaptive binning were proposed by Zavaletta *et al.* for the characterization of lung tissue patterns associated with ILDs in MDCT imaging in [286]. K-means clustering chooses the groups of histogram bins that best discriminate the lung tissue patterns and yields an histogram signature for each pattern in [287]. The EMD distance is then used to assess the similarities between the histogram of an instance and the signatures. As soon as the signatures are computed for each class, this second method allows a fast classification of 3D volumes of interest (VOI) with high sensitivity and specificity. However, low sensitivity is obtained with *fibrosis* patterns which suggests that features characterizing the spatial frequencies are required to complete the adaptive GLH features.

An innovative three dimensional multi-scale density-based approach is proposed by Fetita *et al.* to characterize patterns associated with ILDs in MDCT imaging in 2007 [77]. According to the dominant scales and the mean density of the patterns, a graph-based hierarchical classification allows accurate predictions of 4 classes of lung tissue. A validation with a larger number of cases (10 in the study) is required to demonstrate the robustness of the method to inter-patient variations.

Lee *et al.* compared the performances of state-of-the-art classifiers fed with GLCM, gradient, GLH and RLE features with patterns associated with COPDs in [136, 137]. The SVM classifier showed superior sensitivity and specificity for the predictions of 3 classes of lung tissue among a Bayesian, a naive Bayesian and a MLP classifier. Custom shape features (circularity, aspect ratio, barycentric position and second central moment) are added to the texture features to describe *healthy* and 3 sorts of *emphysema* in [135] and showed to improve the sensitivity of classification but with a CV that allows for training and testing with blocks from an identical patient. Using the same system with slight variations, Park *et al.* showed a correlation between the pulmonary function tests (PFT) and amount of *emphysema* patterns in [180]. Lately (2009) Kim *et al.* added top-hat features to the former system and investigated several feature reduction methods in [113], namely SFS, sequential backward selection (SBS), sequential floating forward selection (SFFS), sequential floating backward selection (SFBS) and PTA( $l,r$ ) (plus  $l$ -take away  $r$ ) [193]. There was no significant difference in terms of CAD performance among the feature reduction algorithms, which is not surprising as these methods are all based on the similar iterative process of adding or removing features one by one. Again, it is important to note that the validation is still not in accordance with a clinical usage of the CAD as the five-fold stratified CV permits training and testing with instances from the same patients and the ROIs are manually chosen as being most representative of their class by the radiologists.

The influence of 45 degrees rotated wavelet frames for lung tissue classification is studied by Tolouee *et al.* in [242]. Features from the rotated wavelet frames (RWF) and the classical discrete wavelet frames (DWF) are extracted by measuring the energy in each subband. A classification with SVMs shows improvements of the CAD performance when DWF and RWF are combined. However, it is surprising that the single-handed RWF features perform significantly better than the DWF. The LORO CV with overlapping blocks introduces a large bias to the CAD performance because lung regions are used both for training and testing.

Boehm *et al.* used GLH along 3D Minkowski functionals in 3D MDCT imaging to discriminate among *healthy*, *emphysema* and *fibrosis* tissue in [20]. The performance shows that the Minkowski functionals have good discriminatory properties. However the selected lung tissue sorts are among the easiest to distinguish and this method should be tested on a large number of lung tissue patterns associated with ILDs. The validation method also leads to biased results by training and testing with VOIs from the same patient.

In 2008, Sorensen *et al.* compared local binary patterns (LBP [178]) and a filter bank based on Gaussian derivatives to categorize *healthy* and 2 sorts of *emphysema* in [234]. The classification based on k-NN with histogram intersection of a selection of features (with SFS) showed no statistically significant differences between the two groups of features. The validation method is in accordance with clinical situations. Histogram dissimilarity representations as distance measures are proposed in [232]. A 3D extension of the Gaussian and Laplacian Filters is presented in [233]. The validation is based on PFT and although high correlation is observed with some of the PFT measures it is difficult to assess the performance the proposed method for local categorization of the lung patterns.

### Trends in lung tissue analysis in thin-section CT

Clear trends in the scopes and techniques utilized in the state-of-the-art can be identified. 57 conference and journal papers from 27 research groups were categorized according to the imaging modality, studied diseases, methods used and validation in Tables A.1–A.8. Most of the studies focused on ILDs and used GLH combined with GLCM as visual features (see Figure 1.17). Detection-based CADs constitute the majority of the proposed systems. Few studies investigated the influence of integrating multimodal knowledge sources on the CAD performance. A lack of rigor is observed concerning the evaluations of the proposed systems as the validations strategies are most often not in accordance with clinical settings.

**Diseases, datasets and CAD tasks** 18 research groups focused their studies on ILDs, 7 on COPD, one on cancerous *ground glass* nodule detection and two groups did not communicate the studied diseases. Most often, the studied disease is chosen according to the available dataset which are very often in-house collections with varying quality of ground truth and small in size which does not allow reproducible results. In the 57 papers studied, a mean of 49.25 patients with a standard deviation of 37.27 are used for the validation of the CAD systems. This is rather small to have statistically significant evaluations of the techniques. To overcome the difficulty of obtaining manually acquired ground truth, some groups used unsupervised or semi-supervised learning or evaluated the correlation of the features with the PFT measures. There currently is a lack of publicly available datasets with high-quality ground truth to objectively compare the performance of the techniques such as proposed in the ImageCLEF<sup>4</sup> initiative in CBIR. Some efforts are currently going in this direction such as the lung tissue research consortium (LTRC<sup>5</sup>) at the national institutes of health (NIH). The two groups that did not specify the studied diseases were implementing CBIR systems and proposed CAD that have a too wide purpose to be really useful in clinical practice but rather proposed CBIR as a tool for creating personal image collections. Nevertheless, the ASSERT CBIR system showed to have a positive impact on the diagnosis in clinical routine, which suggests that CBIR-based CAD systems must be developed while focusing on the studied diseases. In total, 4 groups proposed a CBIR-based CAD system as diagnostic aid for the interpretation of thin-section CT whereas 24 groups proposed lung tissue categorization and detection. In Sections 4.3 and 5.1.2, we will discover that lung tissue categorization and CBIR can be complementary both on the user's side and on the algorithmic side. 21 groups used the HRCT imaging protocol and 6 groups used MDCT. MDCT enables three-dimensional isotropic texture analysis but is not widely used in clinical routine because of high X-ray radiation exposure when compared to HRCT. Thereby HRCT-based CAD systems may have a more important impact on the pulmonary health care community.

**Texture analysis** A wide range of techniques were proposed for feature extraction, reduction and classification. The distribution of the various feature extraction techniques is detailed in Figure 1.17. The heterogeneous feature group composed by GLH, GLCM and RLE was used in most of the studies starting from 1997 in [47] to nowadays in [113, 137]. Although being able to well describe the lung tissue patterns, the performance of the features derived from GLCM and RLE strongly depends of the underlying parameters (i.e. scales, directions). Often, heterogeneous sets of features that are modeling common information are used. Filtering techniques and wavelets offer an overcomplete feature set able to fit most of the texture functions in condition to efficiently derive features from the coefficients. Wavelet-based texture features are covering the whole spectrum, which truly allows to detect the important spectrum signatures of the patterns, being perfectly complementary to the measures of density using GLH.

Various classification algorithms are used. In general, the comparisons between state-of-the-art classifiers showed that the SVM classifier performs significantly better in most of the cases [53, 137]. Classifiers with linear decision boundaries (e.g. LDA) may not be optimal as clusters of instances from distinct classes are most often not linearly separable.

<sup>4</sup><http://imageclef.org/>, as of 5 November 2009

<sup>5</sup><http://www.nhlbi.nih.gov/resources/ltrc.htm>, as of 5 November 2009

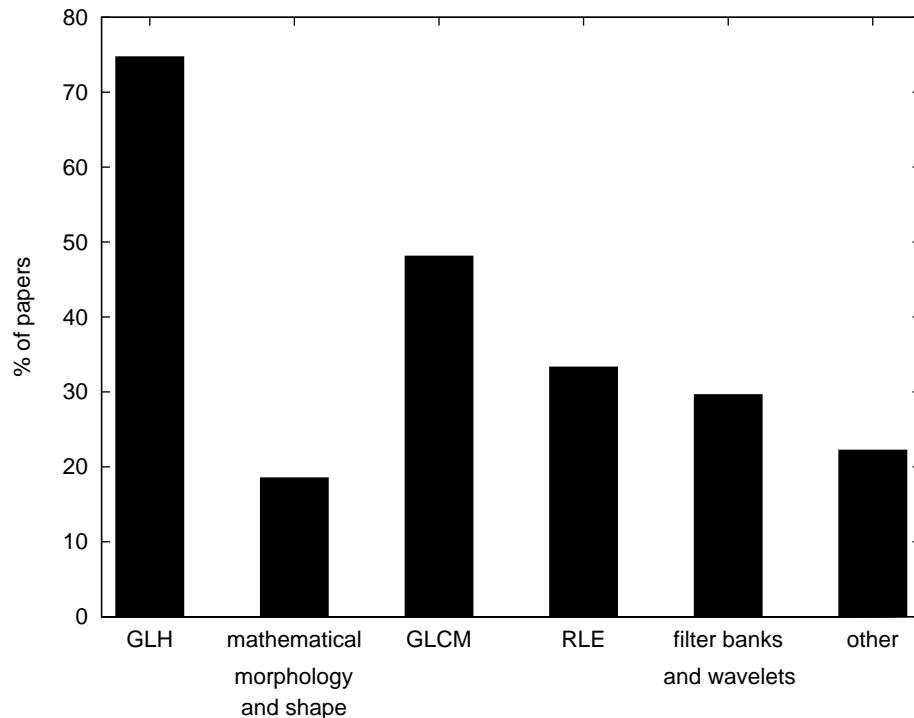


Figure 1.17: Popularity in % of the various feature extraction techniques used by 27 research groups for texture analysis of the lung tissues (57 papers).

**Multimodal approaches** Although no radiologist would interpret medical images without knowing the clinical context of the images, very few studies integrated external knowledge to enhance the performance of the CAD systems. Zrimec *et al.* showed that the knowledge of disease appearance is essential to accurately detect *fibrosis* patterns. We will discover in Section 4.2 that combining visual and clinical features allows to gain up to 10% (absolute gain) of global accuracy with a 5-class lung tissue categorization. The current state of the literature calls upon contributions towards a next generation of CAD systems based on multiple modalities.

**Validations and limitations** Validation methods to assess the CAD performance are crucial for identifying the techniques that will perform well in clinical routine. However the validation techniques used are most often not in accordance with clinical situations. First, 53% of the studies use validations methods that allow for training and testing with instances that belong to the same patient, which leads to overfitting of the learning algorithms. Indeed, the tissue patterns look very similar because of belonging to the same patient and this introduces a positive bias to the performance of CAD system as the probability of belonging to the same class is high for instances coming from the same patient (see Section 5.2.1). Second, the CAD performance are sometimes evaluated on datasets with balanced distributions of the classes, which is not representative of a clinical usage as the prevalences of each lung tissue pattern are not equal. Several CAD are evaluated on manually chosen ROIs. This creates also a positive bias as the selected ROIs are most often chosen as being the most representative of their class and does not take into account the problems encountered at the border of the lungs or near the mediastinum. At last, some studies included only a small number of lung tissue sorts which does not make sense for an automated categorization of entire image series as the radiologist will still have to search for the other disorders of the lung parenchyma. It is fine for the quantification of a certain pattern (e.g. *emphysema*, *fibrosis*), though.

These observations call upon future contributions with realistic validation methods as well as shared datasets to allow for a clear identification of the techniques that would perform well in a

clinical environment.

## 1.4 Organization and scientific contributions of this thesis

This section provides a global view of the organization of the chapters of this thesis in Section 1.4.1 and summarizes the research parts that constitutes scientific contributions in the respective domains and the associated publications in Section 1.4.2.

### 1.4.1 Thesis overview

This thesis deals with various aspects of image-based CAD systems, which are studied with a focus on texture analysis of lung tissue patterns in HRCT images of patients affected with ILDs.

This first chapter introduces the main motivations calling upon scientific contributions on image-based CAD systems. The first part explains the challenges and trends of medical imaging and provides an overview of computer-based tools based on computer vision techniques that have the potential to help the clinicians to cope with growing amounts of visual medical information. Analogies between the human visual and cognitive systems and computer vision are proposed and highlight the advantages of affine-invariant descriptions of the visual information for robust computerized interpretations. The foundations of texture analysis and supervised machine learning methods are described. In the second part, the problematic of diagnosing ILDs and interpreting the associated HRCT images is detailed. The state-of-the-art of texture-based CADs for lung tissue analysis in thin-section CT is studied to identify the trends and weaknesses of the proposed solutions. An overview of the thesis and the principal scientific contributions are listed at the end of the chapter.

Chapter 2 describes the steps of the construction of a multimedia database of ILD cases at the HUG used for the evaluation of the pattern recognition algorithms. The scope of the library of cases and the methodology elaborated for the selection, annotation and capture of the cases is detailed. The current contents of the database is described.

In Chapter 3, a multimodal feature set based on visual and clinical attributes is built in order to characterize lung tissue patterns associated with ILDs in HRCT. Research outcomes concerning the development of affine-invariant texture features based on a tailored WT are presented. The ability of the visual features to discriminate among five lung tissue patterns is evaluated on the dataset described in Chapter 2. Measures for ranking clinical attributes are proposed and compared in the second part of the chapter. The consistency of the multimodal feature space combining visual features with a selection of clinical parameters is studied in the last part of the chapter.

Chapter 4 deals with *machine learning* and *information retrieval* aspects of CAD systems and is divided into three main parts. First, five state-of-the-art classifier families are compared with a methodology based on McNemar's statistical test. Then, the influence of the integration of the clinical context of HRCT images on the classification of the lung tissue patterns is studied. The optimal scheme for fusing the visual and the clinical features is studied. In the last part of the chapter, methods enabling case-based retrieval using a similarity measure based on the volumes of segmented lung tissue as well as clinical parameters are developed.

Chapter 5 is divided in two parts. The first one details practical aspects of a hybrid detection-CBIR-based CAD system for ILDs starting from use cases to user interfaces. The second part contains the evaluation formalism used to assess the performance of the pattern recognition algorithms proposed in this thesis. Discussions concerning the potential pitfalls of commonly used CAD evaluation methods and the challenging demands of clinical environments are given at the end of the chapter.

The conclusions of the conducted research and the associated limitations are given in Chapter 6. Short, medium and long term perspectives of the work presented in this thesis are proposed in the second part of this chapter.

Additional parts were created to ease the reading of the manuscript:

- the *table of contents* can be found at Page iii,

- the *abstract* of the contents of this thesis is given at Page vii and a corresponding version in French starts at Page v,
- *acknowledgements* to all persons without whom this thesis could not have been achieved can be found at Page ix,
- the various mathematical *notations* are summarized starting at Page 119,
- a *glossary* of the abbreviations used in this thesis are listed at Page 123,
- the *list of figures* is given at Page 126,
- the *list of tables* is can be found at Page 128,
- the *index* lists selected keywords and their respective locations in the text starting at Page 150.

### 1.4.2 Scientific contributions

The main scientific contributions of this thesis are in the fields of *image processing*, *machine learning* and *information retrieval*.

The contribution in *image processing* concerns the development and evaluation of affine-invariant texture feature based on tailored WTs. The latter were built by combining several properties of WTs such as:

- shift-invariance obtained with a discrete wavelet frame transform [58],
- near-rotation-invariance and tunable scale-progression enabled by isotropic polyharmonic  $B$ -spline wavelets [59],
- near-scale-invariance achieved by using a fine scale-progression with the quincunx lattice.

The parameters of mixtures of Gaussians characterizing the distribution of the wavelet coefficients in each subband are combined with grey-level histogram bins in HU to create an affine-invariant description of lung tissue patterns associated with ILDs in HRCT.

A first contribution in *machine learning (ML)* address the problem of classification model selection with the elaboration of a methodology for comparing classifiers based on McNemar's statistical test [52, 53].

A second contribution in *ML* concerns contextual medical image analysis, which is enabled by the development of several methods for the fusion of visual and clinical information [51, 57]. The consistency of multimodal feature spaces is studied using correlation analysis and two fusion schemes are compared for the combination of the visual and the clinical features.

A contribution in *information retrieval* is constituted by the development of case-based retrieval where a multimodal inter-case similarity measure is built [60].

Some smaller contributions in the domains of *human-computer interactions*, *evaluation* of image-based CAD systems and *multimedia medical reference libraries* are achieved in this thesis.

A multimedia reference library of ILD cases was built for teaching and for the evaluation of the pattern recognition algorithms [55, 56, 263]. The methods for the selection, annotation and capture of ILD cases as well as quality assessment are the fruits of several refinement and are applicable for other disease groups.

A hybrid detection- and CBIR-based CAD system for ILDs implementing the pattern recognition techniques described above is enabled by the definition of use cases (i.e. database browsing, 3D lung tissue categorization and case-based retrieval) and the development of associated graphical user interfaces [49, 54].

At last, the pitfalls and common mistakes of the evaluation of image-based CAD systems are identified and a subsequent evaluation methodology was proposed to reproduce actual clinical conditions of the target usage of the CAD system.

## Chapter 2

# Multimedia database of interstitial lung diseases

In order to understand the underlying mechanisms of a certain group of diseases, a first requirement is to collect a sufficient number of cases that are representative of the various realizations of the studied diseases. Moreover, the quality of the acquired data is a key issue to carry out non-biased evaluations of CAD systems. With an aim to enhance the disease therapy in clinical routine, the library of cases must be as representative as possible of the hospital's population, meaning that the cases have to be chosen randomly among the whole population.

These requirements are of high importance when building image-based computerized diagnostic aid based on medical image processing [93, 98]. A high-quality multimedia collection of cases containing annotated image series and associated clinical parameters is required to ensure the success of a CAD system at the time it will be integrated in clinical routine [146].

On the one hand, the database constitutes a basis for developing computerized tools such as automatic detection of abnormal pulmonary tissues in HRCT images and retrieval of similar cases. The ground truth and a large number of cases allow to reliably evaluate and compare medical image processing algorithms with defined tasks (i.e. benchmarks). Popular datasets such as Lena, Brodatz [22] or Iris<sup>1</sup> allow to qualitatively and quantitatively evaluate a huge number of basic methods in image processing and machine learning and thereby established a *de facto* reference dataset (see Figure 2.1). The popularity of these datasets is partly due to the fact that they reflect real-life challenges thus offering more credibility of the obtained results when compared to artificial datasets. On the other hand, the database also creates opportunities for specialized studies and teaching. The cases with confirmed diagnoses constitute a knowledge database that can be used as diagnostic aid. Advanced browsing is enabled by CBIR or multimodal case-based retrieval. In summary, a high-quality multimedia library of cases is valuable for:

- teaching purposes,
- specialized descriptive studies,
- training and testing pattern recognition techniques,
- retrieving similar cases as diagnostic aid,
- comparative performance analysis of medical image processing methods (i.e. benchmarks [173]).

However, the construction of a high-quality multimedia collection of cases is extremely time-consuming and expensive and is often a bottleneck in studies on image-based CAD systems. The identification of the relevant cases, the consultation of the EHR and PACS to gather the clinical parameters and the image series, the data entry as well as the database infrastructure and maintenance involves a large number of person months with a wide range of skills from medical

---

<sup>1</sup><http://archive.ics.uci.edu/ml/datasets/Iris/>, as of 5 November 2009



(a) Lena

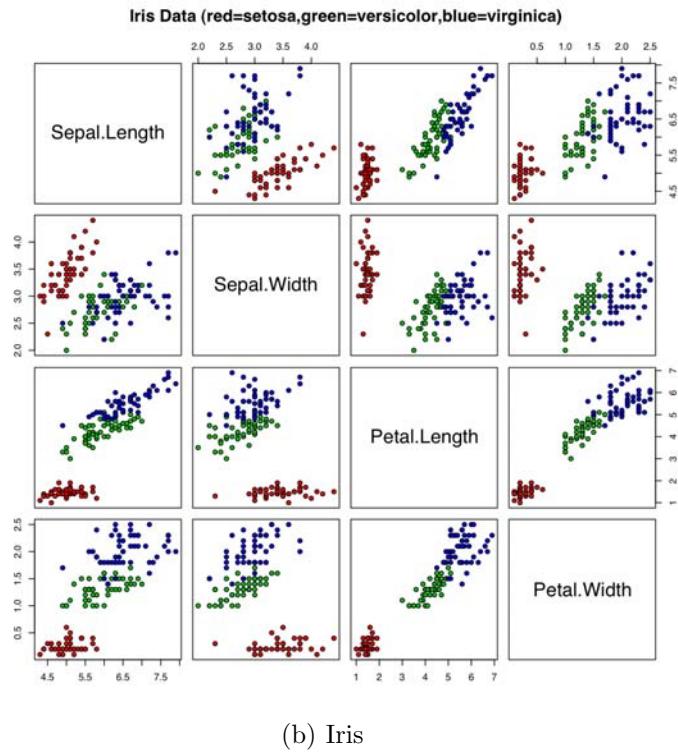


Figure 2.1: Datasets that became references for comparing image processing algorithms (Lena) and machine learning algorithms (Iris).

knowledge to information technology (IT) expertise. To assess high-quality of the data, several researchers have to be involved in the case selection process and the delineation of ROIs to cope with the inter- and intra- observer variability, the latter being particularly important in radiology [8, 9]. The consent of the ethics board has to be obtained before starting any investigations. The latter constitutes the required justification to access the content of the EHR. Depending on the studied diseases, cases are very rare and are encountered casually even in large university hospitals. Incidentally, these disease are the ones requiring reference databases to palliate the lack of experience due to their sparsity. Efforts from the European Union (EU) research programme came up with IT infrastructures to cope with the difficulty of collecting rare cases with aneurysms in the AneurIST project<sup>2</sup> [197]. Multimedia data from six clinical center within Europe is gathered. Anonymization of the patient data as well as images [204] is required as soon as the data leaves the medical institution [72].

A major observation when studying the state-of-the-art of texture-based CADs for lung tissue analysis in thin-section CT is the lack of statistical significance of the measured performance as the CAD systems are evaluated on a small number of cases (see Section 1.3.2). In this Chapter, the steps for building a multimedia database of cases with ILDs and the current content of the created database is detailed. This database was built in the context of the Talisman<sup>3</sup> project [50].

## 2.1 Existing databases of CT imaging of the lung

Efforts for building a resource for the lung imaging research community are detailed in [7, 156]. In order to test and develop lung CADs from reliable datasets for the detection of lung nodules on CT scans, the lung imaging database consortium (LIDC) is constituted of five academic institutions from across the United States. The database includes healthy and pathologic CT images with annotated nodules and primary clinical data of the patient. Expert radiologists from the five institutions agreed on the definition of nodules and the criteria for inclusion in the database. The nodules are categorized into three classes based on their diameters: “nodule  $\geq$  3 mm”, “nodule  $<$  3 mm” and “non-nodule  $\geq$  3 mm”. The annotations of up to five radiologists are available. After a blinded session of annotations, the radiologists had access to the annotations of their colleagues and had the possibility to retrospectively modify their own. The number of nodules on which all four radiologists agreed was 33.8% for the blinded session and 45.8% after review [8]. The database is publicly available and can be downloaded online from the national biomedical image archive<sup>4</sup> (NBIA). Unfortunately, the LIDC database does not contain ILD cases as it only focused on nodules in CT imaging. A small database of annotated nodules in CT imaging of the lung is also publicly available with a purpose of comparing CAD performance described in [200] and can be downloaded online<sup>5</sup>.

Similar efforts are found at the National Heart, Lung, and Blood Institute (NHLBI) but focusing on lung tissue with the creation of the lung tissue research consortium (LTRC<sup>6</sup>) [96]. The goal of LTRC is to improve the management of diffuse lung diseases through a better understanding of the biology of COPDs and fibrotic ILDs including idiopathic pulmonary fibrosis (IPF). Control cases are also enrolled. It aims at creating an open database containing histological, clinical and radiological data. The LTRC began recruitment in February 2005 and are ahead of their goal of collecting 1600 subjects with a total of 1844 enrolled as of September 30, 2008. The lung tissue patterns are described in a structured report. However, based on [96] no regions of interests were delineated in the image series to serve as ground truth for the evaluation of computerized categorization of the lung tissue or as teaching examples. Free access to the HRCT image series and associated metadata is possible after obtaining the approval of the LTRC data coordinating center based on the submission of a concept sheet describing the aim of the study using a standardized

<sup>2</sup><http://www.aneurIST.org/>, as of 5 November 2009

<sup>3</sup>TALISMAN: Texture Analysis of Lung ImageS for Medical diagnostic AssistaNce, [http://www.sim.hcuge.ch/medgift/01\\_Talisman\\_EN.htm](http://www.sim.hcuge.ch/medgift/01_Talisman_EN.htm), as of 5 November 2009

<sup>4</sup><http://imaging.cancer.gov/programsandresources/InformationSystems/LIDC/>, as of 5 November 2009

<sup>5</sup><http://www.via.cornell.edu/databases/lungdb.html>, as of 5 November 2009

<sup>6</sup><http://www.ltrcpublic.com/index.htm>, as of 8 November 2009

format. Based on the LTRC data, computerized quantification of the disease patterns was proposed as measures of the extent of pulmonary disease in [109].

A web-based teleradiology framework for acquiring cases of diffuse lung diseases with annotated regions is proposed in [209]. The Learning Medical Image Knowledge (LMIK) collaborative platform provides tools to the clinicians to delineate ROIs in HRCT imaging of the chest. The cases are then anonymized and stored in a central database that can be queried by authorized researchers for CAD evaluation purposes and by radiologists for teaching purposes. Unfortunately, no public access to the case repository is mentioned and no recent report on the LMIK activities has been found since 2003.

To our knowledge, beside these efforts, no large dataset with annotated image regions is available to be used as ground truth for the evaluation and comparison of computerized categorization of lung tissue in HRCT.

## 2.2 Scope of the database

Before any collection of cases, the scope of the database was defined in order to obtain a consistent set of cases with the aim of building computerized diagnosis aid for ILDs. In this section, the selection of the histological diagnoses to be included in the database and the associated clinical parameters are detailed.

### 2.2.1 Selection of the histological diagnoses

In collaboration with the Service of Emergency Radiology and the Service of Pneumology of the HUG, 15 histological diagnoses that are known as the most frequent causes of lung parenchymal disorders were selected [94]. The objective was to retrospectively analyze at least 150 cases representative of the 15 most frequent ILDs from the EHR at the HUG during the four years of the Talisman project. For two diagnoses, no pure case was found during the project resulting in the 13 diagnoses listed in Figure 1.13. Although healthy cases that underwent an HRCT exam are rare, the latter were included as often as possible to serve as control cases.

### 2.2.2 Selection of the clinical parameters

Based on each pathology, the most discriminative clinical parameters for the establishment of the differential diagnostic were kept. This selection process was carried out based on domain-specific literature [114, 268] along with knowledge bases of computer-based diagnostic decision support systems [81]. Discussions and remarks from lung specialists, radiologists and the medical informatics research group (Service of Medical Informatics, HUG) allowed an iterative review of the selected parameters as well as standardized units and data formats to be used. The parameters that were not available from the EHR were removed. After several modifications of the list based on the availability of the parameters in the EHR, 159 fields were used to characterize the subgroup of ILDs. The HTML (hypertext markup language) form used to capture the clinical parameters is depicted in Figures B.1–B.4. Terminology used in the HTML form is in accordance with MeSH<sup>7</sup> and SNOMED-CT<sup>8</sup> medical terminology references. As often as possible, pull-down menus were used for textual data to favor homogeneity of the data required for further computerized analysis. Units for laboratory tests and other numerical data were chosen depending on the formats used in the electronic patient record.

## 2.3 Data collection

The process of the selection of the cases, annotation of the images and data entry is the result of regular discussions involving the radiologists, the research physician and the computer scientists

---

<sup>7</sup>MeSH: Medical Subject Headings, <http://www.nlm.nih.gov/mesh/>, as of 7 November 2009

<sup>8</sup>SNOMED-CT: Systematized Nomenclature of Medicine — Clinical Terms, [http://www.nlm.nih.gov/research/umls/Snomed/snomed\\_main.html](http://www.nlm.nih.gov/research/umls/Snomed/snomed_main.html), as of 7 November 2009

during the four years of the project. The several steps of the construction of the multimedia dataset starting from the selection of the cases to the data entry are detailed in this section.

### 2.3.1 Selection of the cases at the University Hospitals of Geneva

A raw list of 1266 patients that underwent a thorax CT within a stay in the pneumology service between 2003 and 2008 was extracted from the data repository of the EHR by the helpdesk at the medical informatics service. Only cases with HRCT (without contrast agent, 1 mm slice thickness) were retrospectively analyzed. Cases from paediatrics were left aside. The diagnosis of each of the remaining cases were retraced in the EHR based on clinical history, reports and clinical examinations.

At first, the discharge summary and the pneumological consultation report<sup>9</sup> were revised to decide whether the cases can potentially contain an ILD. When the reports of the CT scans were consistent with the clinical reports, the clinical history, and the laboratory studies (pathology, pulmonary function testing, hematologic tests) the cases were nominated as candidates. Cases of which the histological diagnosis is confirmed to be one of the 13 listed in Figure 1.13 by at least one of the pathological exams (biopsy, bronchoalveolar lavage (BAL)) were selected for inclusion. When the radiographic impression was consistent with the verified diagnosis, the case was retained for the annotation sessions with the radiologists. These cases are subsequently studied from the radiological point of view during regular meetings with two attending radiologists. A very selective process retaining only cases with high confidence was required in order to gain time at the annotation sessions and concentrate on the radiological aspects only.

The time necessary to decide whether the case must be kept and annotated varied from 15 minutes to two hours, with an approximate average time of one hour per case. The selection process at the early months of the project was longer as the methodology was still not well established. It was usually quicker to reject a case than to decide to keep it as ruling out an ILD diagnosis or discovering a large number of comorbidities can be quick.

Up to now, more than 700 cases were revised and 128 were stored in the databases (approximately 18% of the cases had an ILD). Occasionally, patients that had a confirmed diagnosis of ILD but from which the HRCT examination was not corresponding to the disease episode were kept. 108 cases have an annotated HRCT image series and 20 have no images associated.

### 2.3.2 Annotations

The purpose of HRCT annotation is to establish the ground truth for lung tissue classification as well as to show examples of HRCT findings related to a studied disease for teaching purposes. The need for high-quality annotations was highlighted in [168]. Indeed, since the annotations are intended for computerized pattern recognition, the ROIs have to delineate pathologic patterns very precisely to avoid the introduction of noise in the training data (see Figure 2.2).

The possibility to visualize and delineate three-dimensional ROIs in the entire HRCT volume and to set the window level used for displaying the 16-bit DICOM image series on a computer screen was required for annotation. These specifications were fulfilled by adapting an existing graphical software originally developed for delineating hepatic tumors in CT scans at the HUG (see Figure 2.3). The radiologist opens an entire DICOM series and then draws precise ROIs in any layer of the CT volume in the axial view. 3D ROIs can be drawn by linearly interpolating the regions between 2D ROIs delineated in non-contiguous axial slices. Sagittal and coronal views are only available for visualization purposes. Depending on the spacing between slices used, anisotropy in the vertical direction prevents from delineating ROIs in sagittal and coronal views. No exhaustive annotation of every patterns contained in an entire HRCT scan is performed, only patterns that are related to the disease of the patient are delineated. A very simple text file format was developed to save or load ROIs. The coordinates of the points belonging to the contour of polygons demarcating the ROIs in 2D slices are stored. An example is depicted in Figure 2.4. The

---

<sup>9</sup>based on their availability

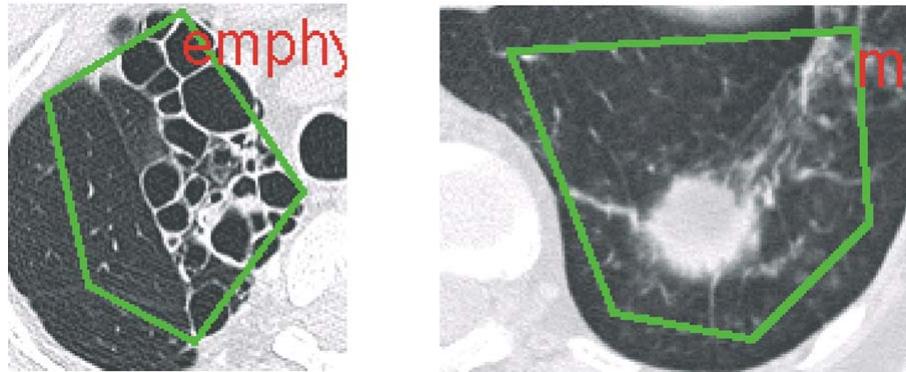


Figure 2.2: A typical example of a non-accurately delineated ROIs. The ROI contains as much healthy tissue as pathological, introducing noise in the training data (source [168]).

Table 2.1: Terminology used to describe the locations of HRCT findings.

name	description
apical	upper region
basal	lower region
diffuse	uniformly distributed
perihilar	middle region, around the mediastinum
peripheral/subpleural	lung periphery/under the pleural membrane surrounding the lobes
non-relevant	used when the pattern has no prevailing location

ROI files were also translated to the binary file format `*.seg` used in YaDiV<sup>10</sup> to facilitate the handling of 3D ROIs. These files are based on “BitCubes”, where a 1 represents a voxel that does belong to this segment and 0 otherwise. The bits are stored in an integer array where width of the slice  $xSize$  modulo 32 forms a row and the total number of integers is  $N_{rows} \times ySize \times zSize$ . It is a trade-off between memory efficiency and operation optimization. Patterns used to describe lung parenchymal disorders are not standardized among radiology communities. A detailed description of common patterns is given in [236]. The terminology used in this work was mainly derived from [114, 268] and a list of the annotated patterns can be found in Table 2.3. A first set of lung tissue sorts associated with the 15 initially selected diagnoses was defined with the radiologists and was adapted to the annotation need through the annotation sessions. Localization of the parenchymal disorders is relevant for several diseases. Thereby, the localizations of the ROIs are stored along with pattern labels in the ROI files. Table 2.1 lists the localizations used. Series with several co-morbidities, or with blur caused by breathing or movements of the patient or containing artifacts were not selected for annotations. Some images taken with contrast agent were annotated but were left aside for further analysis. When possible, healthy tissue was delineated in the studied series to provide a wide range of the aspects of normal lung parenchyma.

The average time for annotating one case was approximately of 45 minutes. 30 minutes were necessary to the two radiologists to interpret the HRCT image series and to draw coarse annotations highlighting the important events in the series. Another 10–15 minutes were required to refine and obtain accurate delineations of the lung tissue patterns as well as to capture them in the database using Java routines.

<sup>10</sup>YaDiV: Yet Another DICOM Viewer, [http://www.welfenlab.de/en/research/fields\\_of\\_research/yadiv/](http://www.welfenlab.de/en/research/fields_of_research/yadiv/), as of 5 November 2009

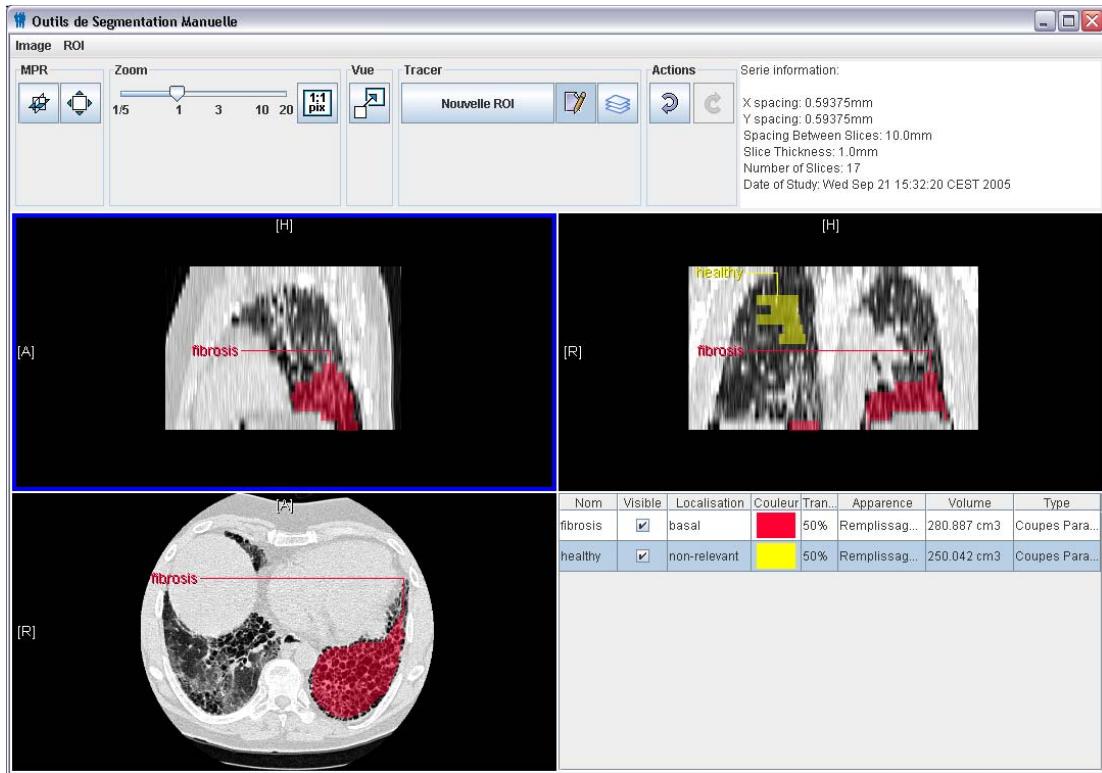


Figure 2.3: A screen shot of the graphical tool for the annotation of image regions.

```
CT-INSPIRIUM-0921.txt - Notepad
File Edit Format View Help
Study: 2.16.840.1.113669.632.20.1211.10000160766
Series: 1.2.840.113704.1.111.2700.1136961994.6
SpacingX: 0.84765625
SpacingY: 0.84765625
SpacingZ: 10.0
label: ground_glass
localisation: peripheral/subpleural
slice_number: 20
nb_points_on_contour: 67
150.96148690840627 267.1950827644733
150.96148690840627 266.1925668881204
150.96148690840627 265.1900510117675
150.96148690840627 264.1875351354146
150.96148690840627 263.1850192590616
151.2956588671906 262.18250338270866
152.2981747435435 261.5141594651401
153.30069061989644 261.1799875063558
```

Figure 2.4: An example of the text-based file format to store the ROIs in the database.

### 2.3.3 Data entry

The content of the EHR is systematically analyzed to fill the content of the HTML form depicted in Figures B.1–B.4 consistently. When multiple instances of clinical parameters (e.g. laboratory data) were available in the EHR within an interval of two weeks around the date of the HRCT image series, the instance as close as possible to HRCT examinations was retained. For each HRCT image series stored in the database, the whole set of clinical parameters was filled and cases that have several relevant image series have several instances in the database as soon as the HRCT examinations were not corresponding to the same episode of disease or if they were separated by more than two weeks.

The discharge summary and any free-text documents that contained evidence of the diagnosis were anonymized and stored in the database. The EHR at the HUG contains computer tools for the anonymization of the documents. No confidential data is stored in the database, except the patient and stay numbers, which both require authorized access to the EHR to retrieve the identity of the corresponding patient. All data is contained inside the hospital on a desktop computer dedicated to research.

Three medical doctors (MD) were successively responsible for the selection of the cases and the data entry. The final protocol described in Section 2.3.1 was developed and refined over time by the three MDs. The time necessary to capture a case was on average of 75 minutes, varying from 40 minutes to more than two hours mostly depending on the diseases. Patients with Sarcoidosis were quickly completed whereas IPF, HP and AIP were requiring more efforts to retrace the history of the patients and to gather all associated parameters. An adaptation period was necessary to the MDs to get used to the EHR at the HUG, as well as to gain experience with the various ILD diagnoses.

## 2.4 Contents

In this section, the content of the database resulting from 38 months of data entry is quantitatively described and analyzed. The first case was captured on the 21<sup>st</sup> of July 2006 and the last on the 6<sup>th</sup> of November 2009.

### 2.4.1 Numbers and statistics

128 cases are currently captured in the database. Among them, 108 have an annotated HRCT image series. 68.8% (88 cases) underwent a biopsy, 64.8% (83 cases) present a BAL. Of the 128 cases, 21.1% (27 cases) have neither a biopsy nor a BAL but had a specific test confirming the diagnosis (e.g. tuberculin skin test for TB, Kveim test for Sarcoidosis, ...). The distribution of the diagnoses is detailed in Table 2.2 and Figure 2.5. 1946 ROIs were delineated in 108 HRCT image series resulting in a total of 41.65 liters of annotated tissues. The distributions of the number of ROIs and the corresponding volumes of the lung tissue sorts are detailed in Table 2.3 and Figures 2.6 and 2.7.

### 2.4.2 Control of Quality

A retrospective control of the diagnosis of each case was carried out to ensure the consistency of the database. Since the global methodology for the selection, the annotation and the capture of the patients was refined during the 38 months of the collection of the cases, inconsistencies occurred, especially with the cases that were captured at the beginning of the project. The diagnosis of each case was retraced in the EHR with a methodology similar to the selection process. Cases with several comorbidities were removed as the visual aspect can be altered strongly (e.g. cardiac insufficiency). 21 cases were removed as their diagnosis was not reliably demonstrable or was mixed with comorbidities. The value of several clinical parameters were corrected for 13 patients. When required, a whole set of clinical parameters was re-entered. Special care was taken for clinical parameters that are subject to change rapidly over time to use the parameters that are

Table 2.2: Distribution, mean age and gender statistics of the diagnoses.

histological diagnosis	patients	image series	age (mean $\pm$ std)	female (%)
Healthy	2	2	$63.5 \pm 9.5$	100
Pulmonary fibrosis (PF)	40	39	$71.4 \pm 13.4$	50
Hypersensitivity pneumonitis (HP)	24	19	$65.3 \pm 17.1$	80.3
Tuberculosis (TB)	15	12	$41.1 \pm 17.6$	40
Pneumocystis pneumonia (PCP)	8	4	$59.6 \pm 20.4$	12.5
Cryptogenic organizing pneumonia (COP/BOOP)	8	3	$45 \pm 24.6$	50
Eosinophilic pneumonia (EP)	1	1	33	100
Sarcoidosis	20	18	$48.5 \pm 17.3$	30
Acute interstitial pneumonia (AIP)	4	4	$65.3 \pm 5.5$	50
Desquamative interstitial pneumonia (DIP)	1	1	46	100
Respiratory bronchiolitis associated ILD (RB-ILD)	1	1	54	100
Non-specific interstitial pneumonia (NSIP)	2	2	$61.5 \pm 12.5$	50
Langerhans cell histiocytosis (LCH)	1	1	24	0
Lymphocytic interstitial pneumonia (LIP)	1	1	32	0
total	128	108	$59 \pm 20.2$	36.7

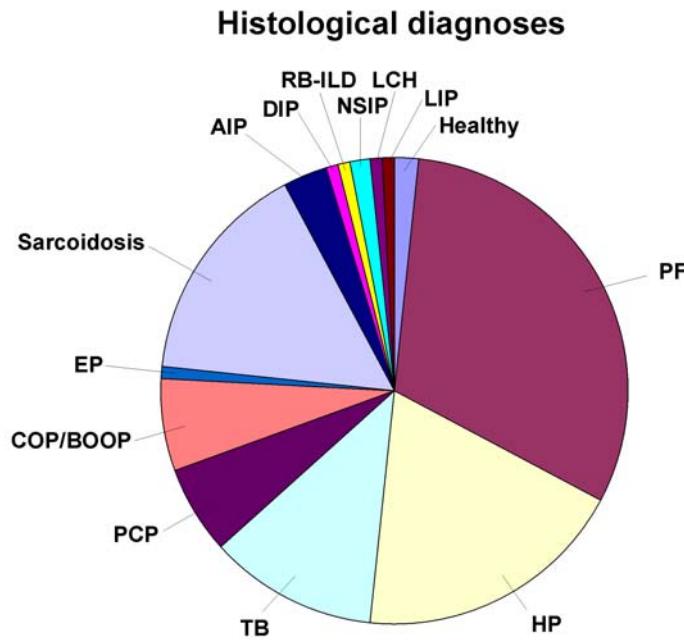


Figure 2.5: Distribution of the histological diagnoses. The database contains 128 cases of which 84.37% (108) have an annotated image series.

Table 2.3: Distribution of the lung tissue patterns. Note that the total number of image series is not equal to the sum of the number of series per pattern as each image series may contain several lung tissue sorts.

label	volume (liters)	image series	ROIs	mean volume per ROI ( $10^{-2}$ liters)
<i>healthy</i>	5.12	7	100	5.12
<i>fibrosis</i>	8.45	38	473	1.79
<i>ground glass</i>	4.91	37	427	1.15
<i>micronodules</i>	16.06	16	297	5.41
<i>consolidation</i>	0.69	14	196	0.35
<i>reticulation</i>	1.88	10	131	1.43
<i>emphysema</i>	1.15	5	66	1.75
<i>bronchiectasis</i>	0.08	8	44	0.18
<i>macronodules</i>	0.18	7	37	0.49
<i>bronchial wall thickening</i>	0.01	1	15	0.08
<i>cysts</i>	0.15	3	11	1.40
others	2.95	14	149	—
total	41.65	108	1946	—

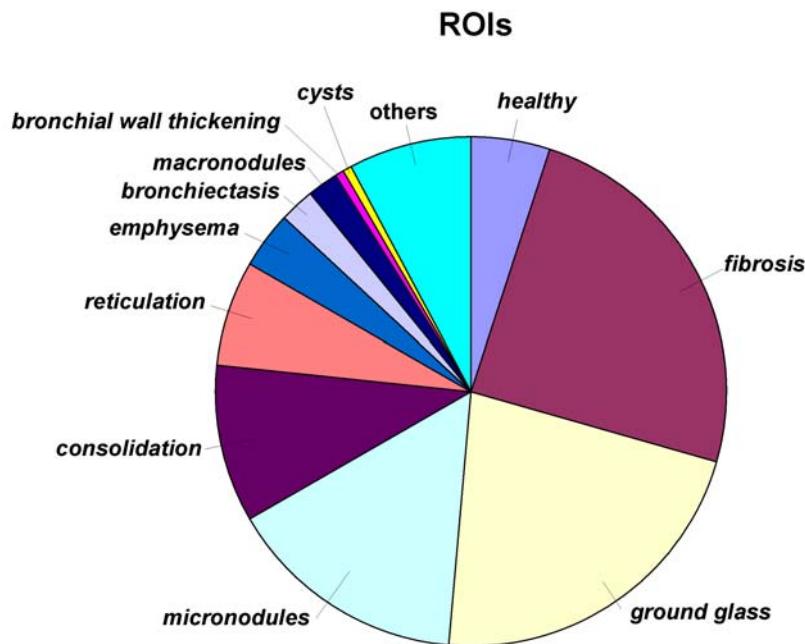


Figure 2.6: Distribution of the various lung tissue patterns in terms of hand-drawn ROIs. The total number of ROIs is 1946.

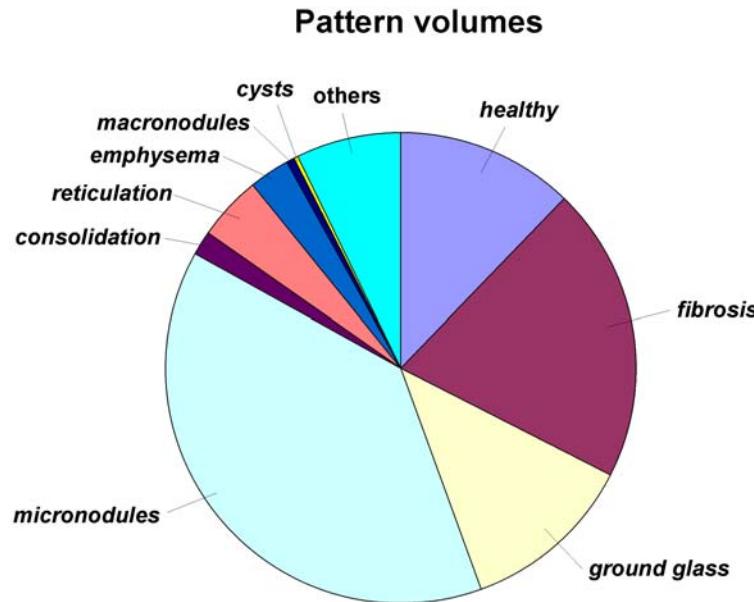


Figure 2.7: Distribution of the various lung tissue patterns in terms of volumes of annotated tissue. The total volume is 41.65 liters.

corresponding to the most acute manifestation of the disease and also as closest as possible to the HRCT examination.

## 2.5 Discussions

In this chapter, a multimedia database built at the HUG is described. The structured multimedia information allows using the stored data to carry out research on an image-based diagnosis aid tool for ILDs. So far, no similar datasets are publicly available for comparing pattern recognition techniques on ILDs, which did not allow to identify the best algorithms proposed in the literature to be further integrated into a CAD system used in a clinical environment. The database is also designed for online consultation for teaching using a web-based interfaces that is detailed in Section 5.1.1.

The distribution of the encountered diagnoses described in Table 2.2 and Figure 2.5 is in accordance with the expected frequency in Europe [241]. A male predominance is observed as only 36.7% of the cases are female, which was also observed in [40, 114, 241]. The mean age of the patients with the respective diseases is shown in Table 2.2, and is in accordance with the expected values in the literature [114]. For instance, the distribution of the ages of patients with LCH, Sarcoidosis and PF shown in Figure 2.8 are similar to the mean age of the corresponding diagnoses in Table 2.2. For instance, PF is found at patients over the age 50 [265] and the mean age of cases of PF is 71.4 years in Table 2.2. The distribution of the diagnoses is very unbalanced where the most represented disease PF contains 40 patients and five diseases have only one case. The number of healthy control cases is also very low and are constituted by former ILD cases that have a histological proof of having recovered. These cases cannot truly be considered as control cases as the former ILD episode is related to abnormal values of some of the clinical parameters.

The total time for completing the whole process of selecting, annotating the image series and capturing a case was on average of three hours. The 7 most represented types of lung tissue<sup>11</sup> in terms of number of ROIs allow describing a wide range of visual findings in HRCT image series associated with ILDs [268]. The distributions of annotated volumes of the various classes of lung

<sup>11</sup>i.e. healthy, emphysema, ground glass, fibrosis, micronodules, consolidation and reticulation.

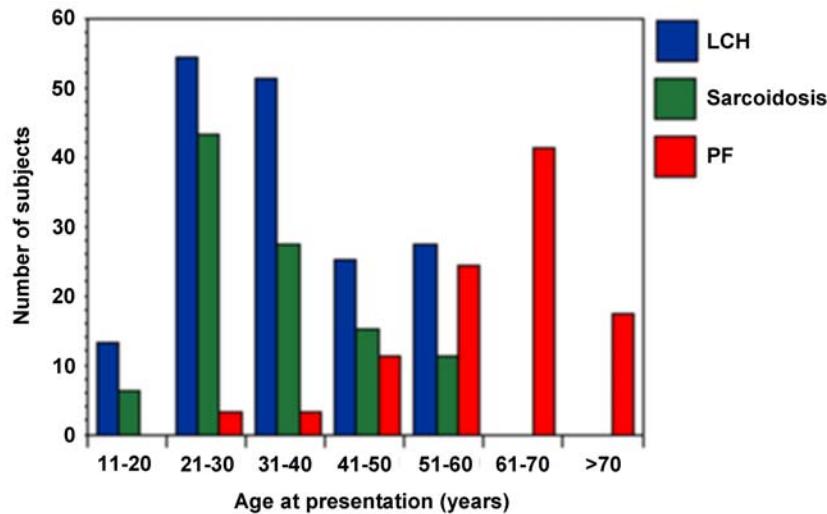


Figure 2.8: Age at onset presentation in LCH, sarcoidosis and PF (source [114]).

tissue in Figure 2.7 are not directly related to the distributions of the ROIs. This can be explained by observing the ratios of mean volume per ROI in Table 2.3. Based on the type of lung tissue, the size of the ROIs varies significantly. For instance, *micronodules* patterns found in cases of Miliary tuberculosis are diffusely distributed in the whole lung, which allowed to delineate very large regions. This is not the case for classes such as *macronodules* or *consolidation* for which the alterations of the lung parenchyma are narrowly localized. The annotation of the latter is comparatively more time-consuming to acquire identical volumes of annotated tissue.

The annotations of the lung image regions are successively carried out by two radiologists with 15 to 20 years of experience. However, no assessment of the quality of the annotations by measuring inter-agreement measurement is carried out, which showed to be an important source of errors in the interpretation process of HRCT image series showing diffuse parenchymal lung diseases in [9]. Nevertheless, the two radiologists have to agree on the delineated region and, thanks to the retrospective analysis, the annotations of the cases were based on the description of the image regions in the radiological report written by several radiologists from 2003 to 2008 at the HUG and enabled cross-control of the interpretation of the image series.

Limitations occur in the representation of the variability of the class *healthy* as normal tissue was annotated in patients that have had an ILD in their medical history. Thereby, beyond continuing the collection of cases, future work is needed to add healthy cases in order to have a rich representation of healthy lung tissue and associated clinical parameters. This is challenging because healthy subjects that undergo an HRCT examination are very rare due to high cost and radiation dose delivery.

# Chapter 3

## Features

Features in the context of pattern recognition and computer vision are defined as “individual measurable heuristic properties of the phenomena being observed” in Wikipedia<sup>1</sup>. The very wide concept of *feature* ranges from continuous measures of an observed signal (low-level) to concepts with high semantics such as evoked feelings (high-level). The purpose of features is to provide a compact representation of an observed phenomenon or scene that is further processed by cognitive processes such as a classification algorithm in machine learning (see Section 1.2). Features can be associated with the concept of a variable in statistics which can be continuous, categorical or binary.

To efficiently describe the content of medical images, we followed the concepts used by radiologists to interpret images. The latter can be classified into two groups: visual features and clinical parameters. The visual features comprise all the information contained in the image and can be color descriptors, edge detectors, power spectrums, object detectors, etc... The clinical parameters denote the context of the image and regroup the clinical parameters related to a certain disease (e.g. age, gender, laboratory results, etc...). It is certain that the design of an image-based system must rely on visual features to integrate the content of the images. Nevertheless, the clinical context of medical images is also essential for a correct interpretation and a radiologist would never interpret an image without knowing at least the age and the gender of the patient (see Section 4.2). Yet, few image-based CAD systems integrate the clinical context of the images and computerized classification of lung tissue patterns associated with ILDs in HRCT imaging is no exception to this as shown in Section 1.3.2, which calls upon a next generation of CAD systems integrating multiple sources of information. In this chapter, we introduce the visual and clinical features used in the proposed CAD system for ILDs.

This chapter is organized as follows. In Section 3.1, the theoretical basis of the visual features is introduced in Sections 3.1.1 and 3.1.2. Evaluations of the visual features are proposed in Section 3.1.3 and are discussed in Section 3.1.4. In Section 3.2, methods are proposed to create a consistent multimodal feature space from clinical features stored in the EHR combined with the visual features described in Section 3.1.

### 3.1 Visual features

In Sections 1.1 and 1.2, we have seen that the performance of image-based CAD strongly depends on the visual features used. The latter constitute the feature space modeling visual information contained in medical images. We found in Section 1.2 that affine-invariant features were desirable to recognize objects that exists in three-dimensions and are projected to two-dimensional slices because they correspond to our visual perception of textures. In this section, we aim at developing an affine-invariant set of features able to discriminate among several texture patterns found in HRCT of the lung from patients affected with ILDs.

---

<sup>1</sup>[http://en.wikipedia.org/wiki/Features\\_\(pattern\\_recognition\)](http://en.wikipedia.org/wiki/Features_(pattern_recognition)), as of 5 November 2009

Most of the texture features used in the state-of-the-art are modeling the same information: the spatial periodicity and scales contained in the images (GLCM, Gabor filters, wavelets, LBP, etc...). The question is then: which one characterizes best the lung tissue patterns and is the most adaptable to the various needs of lung tissue analysis in HRCT imaging?

Although not being the most popular in the state-of-the-art of CAD in HRCT imaging of the chest, filtering techniques have several advantages because they provide continuous responses to the image in contrast to GLCM, which have binary responses to a given sequence of pixels and thus are less adaptive to the heterogeneous data that medical images are. Secondly, filtering allows to seek for special events in the images (such as edge or ridge detection) by modeling the shape of the filters either in the spatial or in the frequency domain. Wavelet transforms (WT) yield filter banks that have the desirable property of being multiscale and thus allow to cover the frequency domain (scale-invariance). It will be shown in the next sections that affine invariance can be approximate using particular designs of WTs. Therefore, our choice is to investigate the ability of wavelet-based texture features to discriminate among the several classes of healthy and pathological lung tissue. The grey-level histograms are modeling a complementary information corresponding to the density of the structures in CT imaging and are thus utilized along with wavelet transforms to characterize the lung tissue.

### 3.1.1 Gray-level histograms

In Section 1.2.1, we have seen that grayscale values can contain relevant information for the characterization of objects and textures and is complementary to texture features. In HRCT imaging, scanners deliver 12-bit DICOM images with voxel values in HU in the interval [-1024;3071]. These values correspond univoquely to densities of the anatomic organs and thus allow the identification of lung tissue components [89]. To extract this information, gray-level histograms (GLH) with a number  $N_{bins}$  of bins are built. The value of each bin is directly used as a visual feature. The air component value given by the number of pixels with value less than -1000 HU is computed as an additional feature [55]. Preliminary results showed that using the values of the bins directly allows better categorization of the lung tissue patterns compared to statistical measures of the distributions (i.e. mean, variance, etc...). Features from histograms showed to be significantly different for 6 lung tissue patterns from 38 patients using a Mann–Whitney U test in [237]. An evaluation of the histogram features for lung tissue classification is carried out in Section 3.1.3, and the results are discussed in Section 3.1.4.

### 3.1.2 Wavelet transforms for lung tissue analysis

In Section 1.2.1, we have seen that the WT was a good candidate to obtain affine-invariant texture features. Wavelet transforms constitute a vast domain and have been applied to various domains such as signal compression, denoising, pattern recognition and many more. They have received a particular attention in biomedical applications [76, 250, 260]. Although being able to detect events in signals, each WT contains all the information to perfectly reconstruct the signal and a reduction of the information is needed to obtain compact feature representations. In this section we show how we can create efficient texture features from tailored wavelet transforms.

#### Affine invariance

The WT has several shortcomings for texture analysis in the standard classical form that are partly due to the sampling of digital images. It will be further shown how the standard WT can be modified slightly to obtain near-affine-invariant texture features.

**Translation invariance: wavelet bases versus redundant wavelet frames** The WT exists in two flavors: bases or frames. Wavelet bases provide a one-to-one decomposition (that may be orthogonal or not), which makes them ideally suited for image compression. They have a fast implementation, but their main drawback is their lack of translation invariance. For example, in the standard dyadic discrete wavelet transform (basis), multiple scales are obtained by downsampling

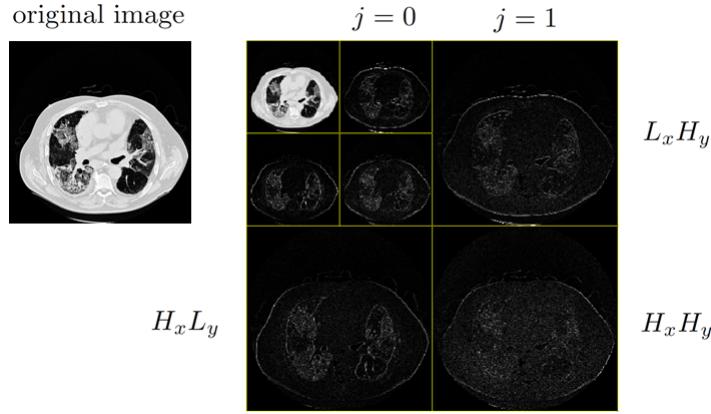


Figure 3.1: On the right, two iterations of a standard separable dyadic wavelet decomposition of the original image on the left. At each iteration  $j$ , the original image is downsampled by a factor of two to obtain a multiscale representation.

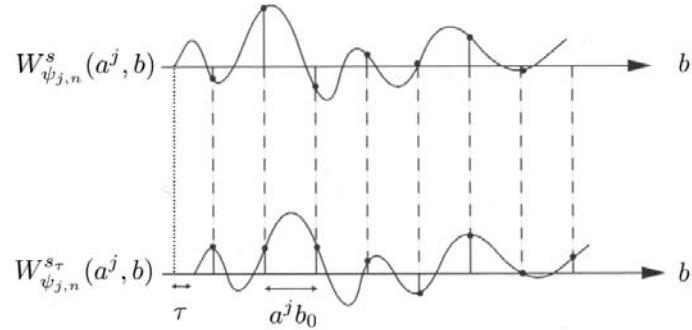


Figure 3.2: The WT is not shift-invariant any more when sampled on an uniform sampling grid because the grid step  $a^j b_0$  may differ from any continuous translation  $\tau$  of the signal  $s(x)$  (source [150]).

the image  $I$  by a factor of 2 at each iteration  $j$ , which implements the pyramidal image decomposition (see example in Figure 3.1). Unfortunately, although bases provide translation-invariant representations in the continuous domain, sampling the translation factor  $b$  over the uniform grid of digital images removes this property. The image  $I$  may be translated by a continuous factor  $\tau$  which is not a priori equal to the grid interval  $a^j b_0$  at scale  $a^j$  (see Figure 3.2 for an example in 1D). The obtained wavelet coefficients  $W_{\psi_{j,n}}^{I_\tau}(a^j, b)$  are very different from  $W_{\psi_{j,n}}^I(a^j, b)$ . To obtain translation-invariant features, a solution is to keep the original resolution of  $I$  and to oversample  $\psi$  so that the minimum shift  $\tau_{min}$  is equal to  $a^j b_0$ , and all values of  $\tau$  are multiples of  $a^j b_0$ . Figure 3.3 illustrates two iterations of the discrete wavelet frame (DWF) decomposition. The DWF decomposition yields highly redundant image representations which offer more flexibility for image analysis and performs quite well for texture analysis. It was first proposed by Unser in 1995 [248] and has been widely used thereafter [131, 141, 142, 198, 261]; also with biomedical images [37]. When compared to the WT, the translation invariance of DWF tends to decrease the variability of the estimated texture features thereby improving classification performance. The ability of DWF in describing lung tissue patterns is evaluated in Section 3.1.3.

**Rotation and scale invariance: isotropic polyharmonic  $B$ -spline wavelets and the quincunx lattice** The combination of isotropic polyharmonic  $B$ -Spline wavelets along with the quincunx lattice offers several desirable properties both for scale-invariance and rotation-

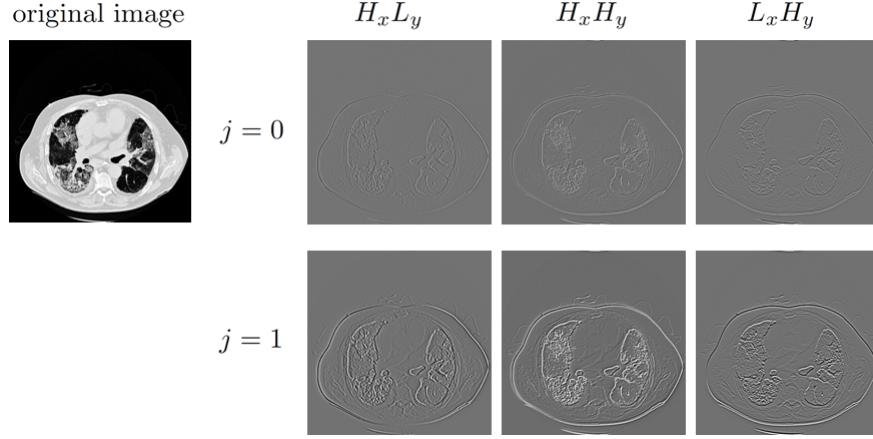


Figure 3.3: On the right, two iterations of the DWF of the original image on the left. At each iteration  $j$ , the filter is upsampled by a factor of two to obtain a multiscale representation.

invariance [255].

When the analyzed patterns do not have *a priori* orientation (e.g. lung tissue patterns in axial slices of HRCT), Marr *et al.* suggested that the Laplacian of Gaussian  $\psi_{\Delta_G}$  was optimal for the detection of zero-crossings according to the mammalian early visual system [154, 258]. This was supported by the following hypotheses:

- the orientation of the detector should be aligned with the local dominant orientation, which is the orientation of the underlying segment of zero-crossings,
- this orientation is also the one at which the zero-crossing has the maximum slope,
- these lines of zero-crossing are corresponding to the orientation-independent differential operator: the Laplacian  $\Delta = \frac{\partial^2}{\partial x_1^2} + \frac{\partial^2}{\partial x_2^2}$ .

Indeed,  $\psi_{\Delta_G}$  is able to “face” image edges in any direction as being isotropic. Moreover, we have seen in Section 1.2.1 that it offers an appropriate trade-off between spatial localization and bandwidth which allows to locally examine a portion of the spectrum of the image. Now that we are convinced that  $\psi_{\Delta_G}$  will be our ideal filter for texture analysis, the question is: how to best implement a multiscale version of  $\psi_{\Delta_G}$  on the cartesian grid imposed by digital images?

In 2005, while looking for an isotropic localization operator that makes polyharmonic  $B$ -splines converge to a Gaussian, Van De Ville *et al.* introduced a new family of wavelets: the isotropic polyharmonic  $B$ -spline wavelets. The latter are derived from elementary  $m$ -harmonic cardinal  $B$ -splines, where the second moment is well-defined. For  $N = 2, 3$  dimensions, the isotropic polyharmonic  $B$ -spline is defined in the Fourier domain as:

$$\hat{\beta}_\gamma(\omega) = \frac{V_\gamma(e^{j\omega})}{\|\omega\|^\gamma} \quad (3.1)$$

where  $V_\gamma(e^{j\omega}) = V_2(e^{j\omega})^{\gamma/2}$  and

$$V_2(e^{j\omega}) = 4 \sum_{k=1}^N \sin^2\left(\frac{\omega_k}{2}\right) - \frac{8}{3} \sum_{k=1}^{N-1} \sum_{l=k+1}^N \sin^2\left(\frac{\omega_k}{2}\right) \sin^2\left(\frac{\omega_l}{2}\right). \quad (3.2)$$

Explicitly, the two-dimensional case is:

$$\hat{\beta}_\gamma(\omega) = \left( \frac{4 \sin^2(\omega_1/2) + 4 \sin^2(\omega_2/2)}{\omega_1^2 + \omega_2^2} - \frac{8}{3} \frac{\sin^2(\omega_1/2) \sin^2(\omega_2/2)}{\omega_1^2 + \omega_2^2} \right)^{\frac{\gamma}{2}}. \quad (3.3)$$

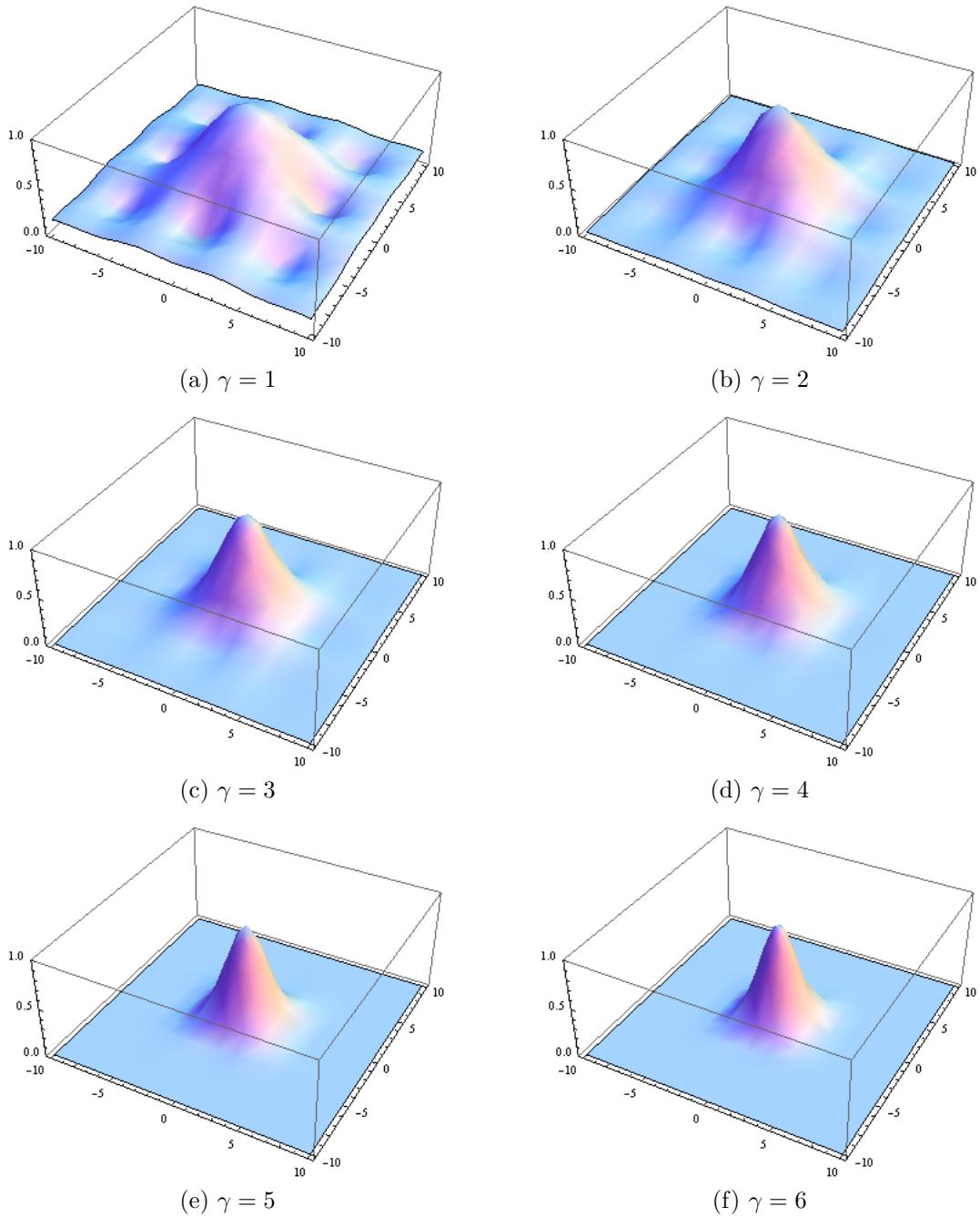


Figure 3.4: Shapes of the 2D isotropic polyharmonic  $B$ -splines  $\hat{\beta}_\gamma(\omega)$  of various orders in the Fourier domain.

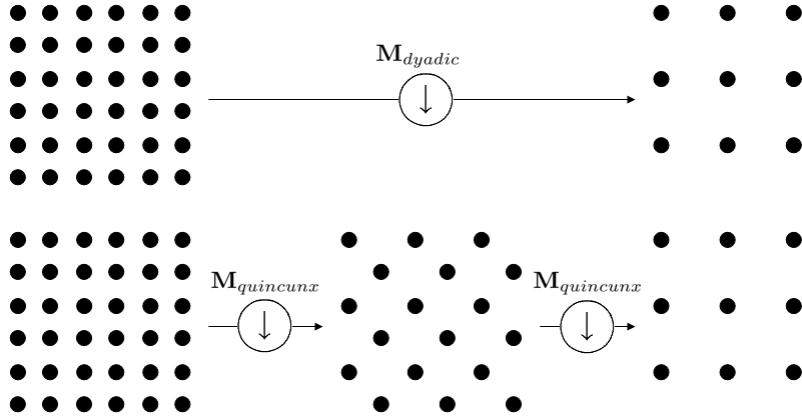


Figure 3.5: Dyadic and quincunx lattices.

Their shapes in the Fourier domain are depicted in Figure 3.4 for  $\gamma = 1, \dots, 6$ . The three-dimensional extension is given by:

$$\hat{\beta}_\gamma(\omega) = \left( \frac{4 \sin^2(\omega_1/2) + 4 \sin^2(\omega_2/2) + 4 \sin^2(\omega_3/2)}{\omega_1^2 + \omega_2^2 + \omega_3^2} - \frac{8 \sin^2(\omega_1/2) \sin^2(\omega_2/2) + \sin^2(\omega_1/2) \sin^2(\omega_3/2) + \sin^2(\omega_2/2) \sin^2(\omega_3/2)}{3(\omega_1^2 + \omega_2^2 + \omega_3^2)} \right)^{\frac{\gamma}{2}}. \quad (3.4)$$

Van de Ville *et al.* demonstrated that these wavelets behave as a  $\gamma$ th-order Laplacian  $\Delta$  for low-frequencies. Indeed, the isotropic polyharmonic  $B$ -spline  $\beta_\gamma$  is an interesting candidate to approximate Gaussian-like basis functions as it converges to the following Gaussian when the order increases:

$$\beta_\gamma(\mathbf{x}) \approx \frac{6}{\pi\gamma} \exp\left(-\frac{6\|\mathbf{x}\|^2}{\gamma}\right) \quad (3.5)$$

with  $\mathbf{x} \in \mathbb{R}^2$ , and a standard deviation  $\sigma = \sqrt{\gamma/12}$ . The convergence is fast: the normalized squared differences between  $\beta_\gamma$  and its corresponding Gaussian function is inferior to 5% for  $\gamma \geq 3$ . Thus, isotropic polyharmonic  $B$ -spline wavelets implement a multiscale smoothed version of the Laplacian from which the initial scale can be tuned through  $\gamma$ . This wavelet  $\psi_\gamma$ , at the first decomposition level, can be characterized as

$$\psi_\gamma(\mathbf{M}^{-1}\mathbf{x}) = \Delta^{\frac{\gamma}{2}} \{\phi\}(\mathbf{x}), \quad (3.6)$$

where  $\phi$  is an appropriate smoothing (low-pass) function (e.g. Gaussian-like) and  $\mathbf{M}$  is the subsampling matrix<sup>2</sup>.  $\gamma$  tunes the iterate of the Laplacian operator (comparable to the traditional vanishing moments). Large values of  $\gamma$  reduce the energy of the wavelet coefficients but increase the ringing effect [75]. At last,  $\psi_\gamma$  yield symmetric filters that have a linear phase response, where the delay (shift) is predictable and no phase distortion is introduced. Thus  $\psi_\gamma$  allow to preserve the shift-invariance provided by the redundant frame transform.

The choice of the subsampling scheme defined by  $\mathbf{M}$  has itself an important influence both on scale and rotation invariance. Mainly two subsampling schemes are found in the literature for the two-dimensional case: dyadic and quincunx, the former being by far most widely used. The quincunx subsampling removes one sample over two while the dyadic one removes three samples over four as depicted in Figure 3.5 and corresponds to a similarity transform in 2D<sup>3</sup>. With particular wavelets, dyadic subsampling enable separability of the WT which allows computational efficiency

<sup>2</sup>In the wavelet frame decomposition,  $\mathbf{M}$  corresponds to the upsampling of the wavelet.

<sup>3</sup>It is important to note that the extension of the quincunx subsampling in more than two dimensions does not correspond to a similarity transform for two-channel filter banks (i.e. wavelets) [256].

Table 3.1: Desirable properties of the combination of isotropic polyharmonic  $B$ -spline wavelets and the quincunx lattice along with a redundant frame transform.

	translation invariance	scale invariance	rotation invariance
isotropic polyharmonic $B$ -spline wavelets	symmetric filters: no phase distortion	initialization of the scale progression tunable with $\gamma$	near isotropic: implement a multiscale smoothed version of $\Delta$
quincunx lattice	—	finer scale progression compared to dyadic: $\sqrt{2}$ instead of 2	preserves isotropy
redundant frame transform	no image downsampling	multiscale	—

because wavelet coefficients within each subband can be obtained by successive one-dimensional convolutions along the columns and the rows of the image. However, the dyadic lattice has two major drawbacks to perform texture analysis. On the one hand, the scale progression is coarse as images are downsampled by a factor of 2 (in each dimension) between two decomposition levels. Relevant information might be padded out when having major energy contained in a narrow subband located between two successive levels of the dyadic transform. For example, subtle changes in the scale of lung tissue patterns (e.g. *micronodules* versus bronchovascular structures) might be neglected by the dyadic scale progression [58]. On the other hand, although separability allows fast WTs, this process tends to favor the vertical and the horizontal directions, and produces a so-called “diagonal” wavelet component [36], which does not have a straightforward directional interpretation.

The scale–progression of the quincunx subsampling scheme defined by  $\mathbf{M}_{\text{quincunx}} = [1 \ 1; 1 \ -1]$  is slower compared to dyadic decomposition, with an equivalent one–dimensional downsampling factor of  $\sqrt{2}$  instead of 2 for dyadic [243]. Compared to the dyadic separable case,  $\mathbf{M}_{\text{quincunx}}$  preserves isotropy and the wavelet space is spanned by only one wavelet subband per decomposition level (versus three for separable dyadic) [162]. This leads to a direct and easy interpretation of the subbands; the small number of subbands also breeds small features spaces, which are preferable for classification.

The desirable properties of the isotropic polyharmonic  $B$ -spline wavelets combined with the quincunx lattice along with a redundant frame transform are summarized in Table 3.1. This combination yields isotropic quincunx wavelet frames (QWF) that are near–affine–invariant. QWF are evaluated and compared to DWF for the classification of the lung tissue patterns in Section 3.1.3.

### Features from the wavelet coefficients

For the classification of the lung tissue patterns, a subset of features able to characterize the coefficients of the wavelet filterbanks is required. The wavelet frame transforms yield a number of wavelet coefficients  $N_c$  that is directly proportional to the number of pixels  $N_{pix}$  of the input image, the number of iterations  $J$  and the number of subbands per iteration  $N_{subband}$ :

$$N_c = N_{pix} \times J \times N_{subband}.$$

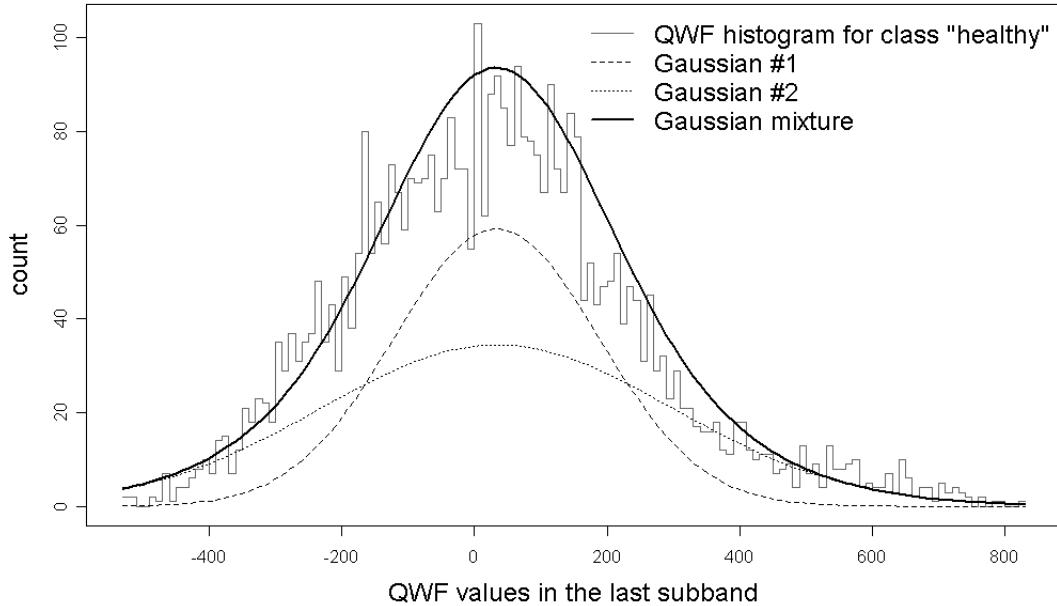


Figure 3.6: Mixture of two Gaussians ( $\mu_{1,2} = \mu$ ,  $\sigma_{1,2}$ ) to model the distribution of wavelet coefficients within one subband.

For example, 4 iterations of DWF (3 subbands per iteration) of an HRCT slice of  $512 \times 512$  would lead to more than 3 million coefficients. No classification algorithm can learn from such high-dimensional feature spaces and compact feature representations are required. In the literature, localized energy measurements were often used to characterize the coefficients in the subbands [131, 248, 257]. As an alternative, Portilla *et al.* showed that Gaussian mixture models (GMM) are modeling well the distributions of the wavelet coefficients through the successive subbands as well as in a spatial neighborhood within the same subband [186]. Based on this assumption, the distributions of the wavelet coefficients in each subband were characterized through the parameters of GMMs. Under the hypothesis that the global mean of the coefficient values is close to zero according to (1.5), the parameters of a mixture of two Gaussians with same means  $\mu_{1,2} = \mu$  and two standard-deviations  $\sigma_{1,2}$  are estimated using the expectation–maximization (EM) algorithm [48]. In Figure 3.6, an example showing the coefficients in the eighth subband of QWF from *healthy* patterns suggests that the mixture of two Gaussians with fixed means allows a reasonable fit of the distributions of the wavelet coefficients.

### 3.1.3 Evaluation of the visual features

The visual features are evaluated and compared through several experiments in order to identify groups of visual features (GLH and texture) that allows the best discrimination among the lung tissue patterns associated with ILDs for further integration in the CAD. Please note that the validation method, the dataset used and the classification algorithm varied over time and are detailed for each experiment.

#### GLH

The normalized histograms of the most frequent lung tissue patterns and their respective mean, standard deviation along with third and fourth moments, respectively skewness and kurtosis are depicted in Figure 3.7. The optimal number of GLH bins  $N_{bins}$  is investigated in Figure 3.8. With a leave-one-ROI-out (LORO) cross-validation (see Section 5.2.1), 40 bins allow the best classification accuracy  $A^{LORO}$  using a k-NN classifier with optimized number of neighbors  $k$  and the dataset described in Table 3.2.

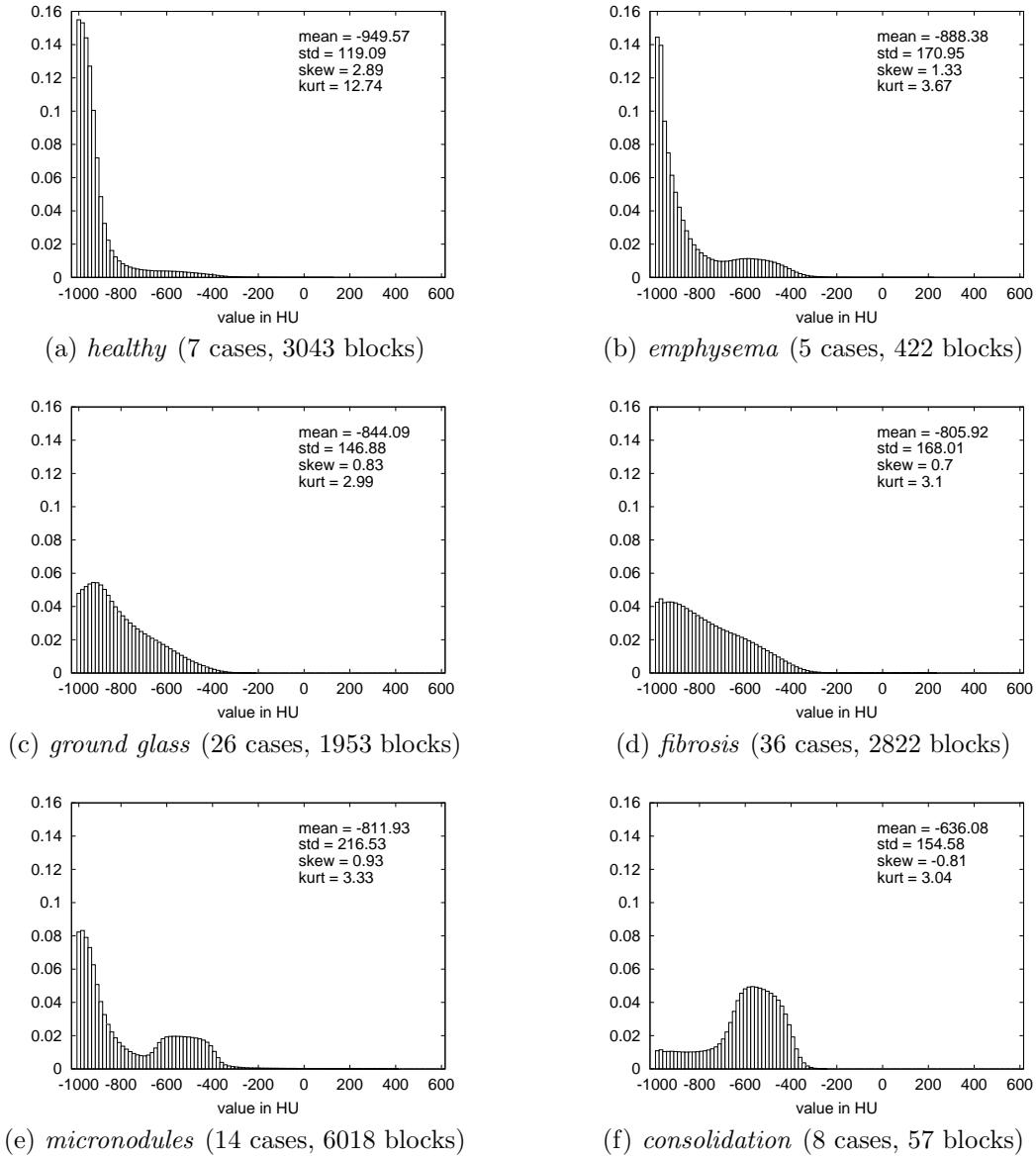


Figure 3.7: Gray-level histograms of the most frequent lung tissue patterns. Each histogram is obtained from the average of histograms computed over  $32 \times 32$  blocks.

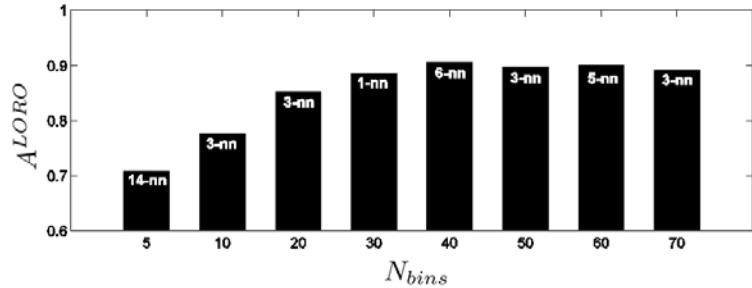


Figure 3.8: Classification accuracies for varying number of GLH bins, using the best number of nearest neighbors  $k$  for each.

Table 3.2: Dataset used to determine the optimal  $N_{bins}$  and to evaluate DWF features.

class	<i>healthy</i>	<i>emphysema</i>	<i>ground glass</i>	<i>fibrosis</i>	<i>micronodules</i>
ROIs	77	72	113	64	155
patients	10	5	13	11	5

## DWF

The ability of the DWF combined with GLH to discriminate among five lung tissue patterns is evaluated [58]. A family of  $B$ -splines of third degree is used as wavelet basis in a separable transform based on the dyadic lattice for upsampling the filters.  $B$ -spline wavelets are compactly-supported smooth piecewise polynomials from which the degree  $\alpha$  modulates their scale-space properties [249]. This is convenient for analyzing medical images where the varying scales of the objects require adjustable wavelets. The symmetric  $B$ -spline is expressed in the Fourier domain as:

$$\hat{\beta}^\alpha(\omega) = \left( \frac{\sin(\omega/2)}{(\omega/2)} \right)^{\alpha+1} \quad (3.7)$$

The influence of the degree  $\alpha$  on classification accuracy is studied in the next subsection.

The coefficients  $H_j(\mathbf{x})$  resulting from the convolutions with lowpass filters  $h_j$  are kept for each iteration  $j$  in order to investigate continuous components of the lung tissue patterns at different scales. Moreover the  $l_2$ -norm of composite diagonal coefficients  $C_j(\mathbf{x})$  are computed for each iteration as follows:

$$C_j(\mathbf{x}) = \sqrt{(G_x H_y)_j^2(\mathbf{x}) + (G_y H_x)_j^2(\mathbf{x})} \quad (3.8)$$

where  $(G_x H_y)_j(\mathbf{x})$  and  $(G_y H_x)_j(\mathbf{x})$  are the coefficients resulting from the convolution with the highpass filter on  $x$  and with the lowpass filter on  $y$ , and vice versa. The norm of both is computed because we believe that no directionality is contained in lung tissue textures. To extract higher-frequency features at smaller scales, the input images are upsampled by a factor of  $2^{N_{upsample}}$ . The mean  $\mu$  and the variance  $\sigma$  of the coefficients  $G_j(\mathbf{x})$ ,  $C_j(\mathbf{x})$  and  $H_j(\mathbf{x})$  are computed on the whole ROI for each iteration  $j$  (scale) to create the feature vector

$$\begin{aligned} & (\mu, \sigma(G_1(\mathbf{x}_R)) \quad \mu, \sigma(C_1(\mathbf{x}_R)) \quad \mu, \sigma(H_1(\mathbf{x}_R)) \quad \dots \\ & \dots \quad \mu, \sigma(G_J(\mathbf{x}_R)) \quad \mu, \sigma(C_J(\mathbf{x}_R)) \quad \mu, \sigma(H_J(\mathbf{x}_R)) \quad ) \end{aligned} \quad (3.9)$$

where  $\mathbf{x}_R$  denotes the points belonging to the ROI.

The relevant scales of the DWF are investigated by performing a multi-class classification of the five patterns detailed in Table 3.2 with a 1-nearest neighbor using the Euclidean distance computed between normalized feature vectors. Global accuracies  $A^{LORO}$  obtained with a LORO cross-validation using only coefficients from the first iteration of the DWF are shown in Figure 3.9. Since the original images are resampled by a factor  $2^{N_{upsample}}$ , the sampling factor is inversely proportional to the scales contained in the patterns.

In order to quantify the global discriminative properties of the DWF feature combined with GLH, additional features from GLH consisting of 40 bins in  $[-1050; 1500]$  HU and a feature that counts the number of air pixels with value below  $-1000$  HU are concatenated into the feature vector described in (3.9). A k-NN classifier predicts the class of each global ROI. Each feature is normalized using a linear mapping between 0 and 1 from each realization to give an equal weight to each of them. No weighting is used for the combination of heterogeneous features. The confusion matrix shows the classification results using a LORO cross-validation in Table 3.3 where the original images are upsampled by 4 ( $N_{upsample} = 2$  allowed best  $A^{LORO}$  in Figure 3.9). The associated performance measures are listed in Table 3.4. The comparison of accuracies ( $A^{LORO}$ ) of DWF, GLH plus percentage of air pixels, and the combination is illustrated in Figure 3.10.

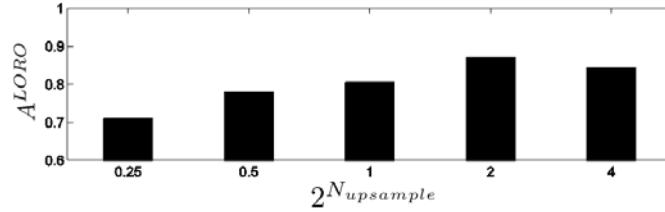


Figure 3.9: Overall accuracies at iterative scales. The sampling factor  $2^{N_{upsample}}$  is inversely proportional to the studied scale.

Table 3.3: Confusion matrix of the lung tissue pattern represented by full hand-drawn ROIs using DWF combined with GLH in %.  $N_{ROI}$  denotes the number of ROIs used for evaluation.

	<i>healthy</i>	<i>emphysema</i>	<i>ground glass</i>	<i>fibrosis</i>	<i>micronodules</i>	$N_{ROI}$
<i>healthy</i>	<b>92.2</b>	0	1.3	0	6.5	77
<i>emphysema</i>	0	<b>100</b>	0	0	0	72
<i>ground glass</i>	0.9	4.4	<b>86.7</b>	0	8	113
<i>fibrosis</i>	0	0	6.2	<b>93.8</b>	0	64
<i>micronodules</i>	3.9	0	0.6	2.6	<b>92.9</b>	155

Table 3.4: Performance measures associated with the confusion matrix in Table 3.3. See Section 5.2.2 for details about the performance measures.

	recall	precision	F-measure	accuracy
<i>healthy</i>	91	92.2	91.6	97.3
<i>emphysema</i>	93.5	100	96.6	98.9
<i>ground glass</i>	94.2	86.7	90.3	95.6
<i>fibrosis</i>	93.8	93.8	93.8	98.3
<i>micronodules</i>	91.1	92.9	92	94.8

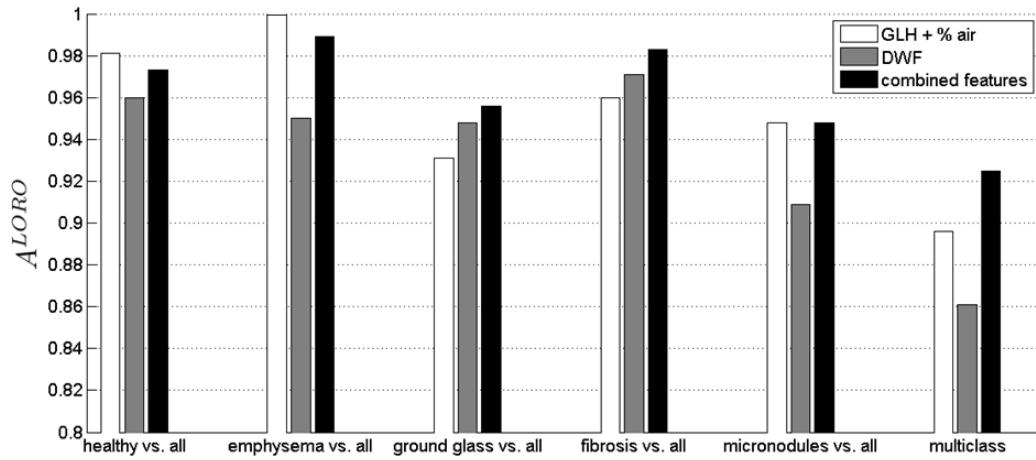


Figure 3.10: Comparison of accuracies ( $A^LORO$ ) using GLH plus air percentage, DWF and the combination of all features applied to each pattern versus all and to the multiclass configuration.

Table 3.5: Dataset used for the comparison between DWF and QWF features.

class	<i>healthy</i>	<i>emphysema</i>	<i>ground glass</i>	<i>fibrosis</i>	<i>micronodules</i>	<i>macronodules</i>
ROIs	113	93	148	312	155	22
patients	11	6	14	28	5	5

The algorithm for the extraction of the texture features and GLH from DICOM files was implemented as an *ImageJ*<sup>4</sup> plugin in Java, and the k-NN classification was carried out using the function *knn.cv* of the package *class* of the free data mining software *R*<sup>5</sup>.

### DWF versus QWF

In this section, the abilities of DWF and QWF for classifying the lung tissue patterns are compared [59]. In order to build the feature space for further classification of the ROIs, several measures are computed from the original image as well as from the DWF/QWF wavelet coefficients of each subband. On the original images, values of pixels belonging to the ROIs are categorized using GLH with 22 bins in [-1050; 600] HU. The distributions of the wavelet coefficients in each subband are characterized through the parameters of mixtures of two Gaussians with fixed means  $\mu_{1,2} = \mu$  and the standard-deviations  $\sigma_{1,2}$  as described in Section 3.1.2. The feature vector thus consists of 24 features for 8 iterations of QWF. Various orders of the isotropic polyharmonic *B*-spline wavelets  $\beta_\gamma$  are compared with  $\gamma = 2, 3, 4$ . To compare performances, 4 levels of DWFs were performed using *B*-spline wavelets of degree  $\alpha = 1, 2, 3$ . The equivalent order of derivatives  $\gamma$  corresponds to  $\alpha + 1$  (see equations (3.3) and (3.7)). Using parameters of a mixture of two Gaussians for each subband, the feature vector contains 36 measures of the DWF coefficients. In both cases, the low-pass filtered images are left aside.

The validation is organized as follows: feature vectors from 843 ROIs containing healthy and five pathological lung tissue sorts are extracted. The dataset used in this experiment is detailed in Table 3.5. Then, 674 instances (80%) are randomly drawn from the full dataset and used to train and optimize the parameters of SVMs, which have shown to be effective to categorize texture in wavelet feature spaces in [141] and in particular lung tissue as shown in Section 4.1. The remaining 169 instances are used for testing. The global experimentation is repeated 30 times and means of the global classification accuracies on the testing set  $A_{mean}^{test}$  along with means of class-specific accuracies  $A_{mean}^{geom}$  (geometric mean) are computed. The optimization of the parameters of SVMs follows the methodology detailed in Section 4.1.1. Pairwise comparisons of classification accuracies using DWF versus QWF for several degrees and orders ( $\alpha + 1, \gamma$ ) are shown in Table 3.6.

### Blockwise classification with QWF

This section describes the various steps of the blockwise classification with QWF and GLH features. In a first step, a semi-automatic segmentation of the lung volumes is used to locate the lung parenchyma to be further analyzed. Then, the segmented regions are entirely categorized using QWF and GLH in overlapping blocks classified using SVMs to create a three-dimensional map of five lung tissue patterns.

**Semi-automatic segmentation of the lung volumes** Segmentation of the lung volumes is a required preliminary step to lung tissue categorization [123]. The result of this step is a binary mask  $M_{lung}$  that indicates the regions to be analyzed by the texture analysis routines. Since the geometries and shapes of the lungs are subject to large variations among the cases, semi-automatic segmentation based on region growing and mathematical morphology is used. The region growing routine contained in YaDiV is used. Starting from a seed point  $p(x, y, z)$  defined by the user, each

<sup>4</sup><http://rsb.info.nih.gov/ij/>, as of 5 November 2009

<sup>5</sup><http://www.r-project.org/>, as of 5 November 2009

Table 3.6: Mean accuracies in % with experiments repeated 30 times. Isotropic polyharmonic  $B$ -spline wavelets (QWF) with order  $\gamma = 3$  allowed a mean of 94.3% of correct predictions among the six lung tissue classes with high precision (geometric mean = 89%). Best accuracies are marked in bold.

class		$\alpha = 1, \gamma = 2$	$\alpha = 2, \gamma = 3$	$\alpha = 3, \gamma = 4$
<i>healthy</i>	DWF	91.1	93.6	92.5
	QWF	95.9	<b>98.1</b>	92.4
<i>emphysema</i>	DWF	97.2	98.7	97.7
	QWF	<b>100</b>	<b>100</b>	99.7
<i>ground glass</i>	DWF	84.2	88.3	86.3
	QWF	85.7	<b>89</b>	87.7
<i>fibrosis</i>	DWF	95.8	95.2	<b>96.5</b>
	QWF	<b>96.5</b>	96.3	94.5
<i>micronodules</i>	DWF	89.8	93.3	88.8
	QWF	94.1	<b>95.2</b>	91.7
<i>macronodules</i>	DWF	40.3	48	46.9
	QWF	54.2	<b>55.5</b>	48.5
$A_{\text{mean}}^{\text{geom}}$	DWF	83.1	86.2	84.8
	QWF	87.8	<b>89</b>	85.7
$A_{\text{mean}}^{\text{test}}$	DWF	$90.6 \pm 2.6$	$92.5 \pm 1.4$	$91.4 \pm 2.2$
	QWF	$93.3 \pm 1.6$	<b><math>94.3 \pm 1.6</math></b>	$92 \pm 1.9$

26-connected neighbor is added to the region  $M_{lung}$  if the summed value of its own neighbors differs of less than a given variance defined by the user. The resulting binary mask  $M_{lung}$  describes the global lung regions well but contains many holes where the region growing algorithm was stopped by denser regions such as vessels or consolidations of the lung parenchyma. To fill these holes, a closing operation is applied to  $M_{lung}$  using a spherical structuring element. Two parameters require attention from the user: the radius  $r$  of the structuring element in millimeters and  $N_{op}$  which defines the number of closing operations. To assess satisfying segmentation results, the user can tune the parameters  $r$  and  $N_{op}$  and, if required, add manual corrections using the YaDiV interface where a tab was created for the closing operation (see Figure 5.12). An example of the segmentation is depicted in Figure 3.11.

**3D lung tissue categorization** In order to automatically categorize every pixel of  $M_{lung}$ , each 2D slice is divided into overlapping blocks. Preliminary results using block sizes of  $\{8 \times 8; 16 \times 16; 24 \times 24; 32 \times 32; 40 \times 40; 48 \times 48; 56 \times 56; 64 \times 64\}$  showed that blocks of size  $32 \times 32$  have the best trade-off between classification performance and localization. For each block, 22 histogram bins  $bin_j$  of GLH in  $[-1050; 600]$  and the number of air pixels  $pix_{air}$  are concatenated into one hybrid feature vector  $\mathbf{v}$  along with GMM parameters of 8 iterations of QWF using  $\beta_\gamma$  of order  $\gamma = 3$  as follows (see also Figure 3.12):

$$\mathbf{v} = (bin_0 \quad \dots \quad bin_{21} \quad pix_{air} \quad \mu^0 \quad \sigma_1^0 \quad \sigma_2^0 \quad \dots \quad \mu^7 \quad \sigma_1^7 \quad \sigma_2^7) \quad (3.10)$$

The low-pass filtered images of the QWF transform are left aside. A SVM classifier learns from the space spanned by  $\mathbf{v}$  to find the decision boundaries among five classes of lung tissue using a *one versus one* approach. The optimal cost of the errors  $C$  and the width of the Gaussian kernel  $\sigma_K$  are found using a grid search with  $C \in [0; 100]$  and  $\sigma_K \in [10^{-2}; 10^2]$ . For each case, the whole lung volume is segmented using a distance between the centers of the blocks equal to 4 pixels, leading to an overlap of 87.5% (see Figure 3.13). The confusion matrix resulting from a leave-one-patient-out (LOPO) cross-validation with 69 patients is shown in Table 3.7. The associated performance measures are listed in Table 3.8. Note that patients may contain several sorts of lung tissue.

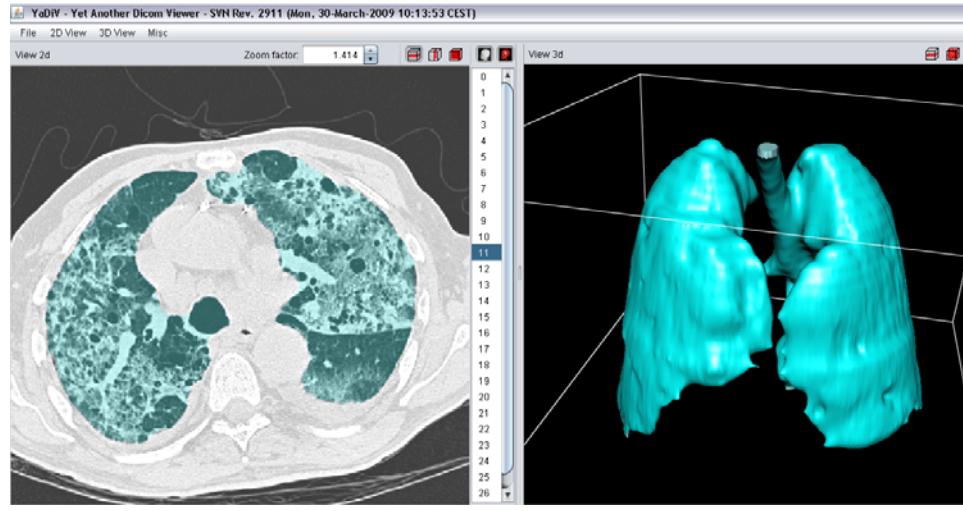


Figure 3.11: An example of the segmentation of the lung volumes using a modified version of YaDiV.

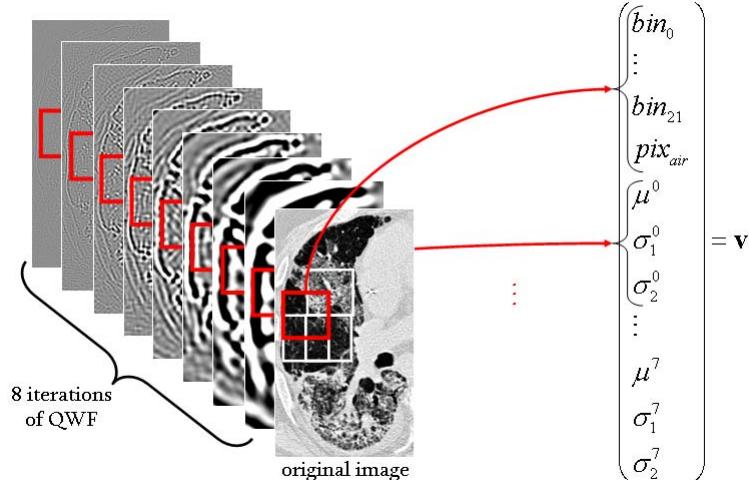


Figure 3.12: Construction of the feature vector  $\mathbf{v}$  for each block.

Table 3.7: Confusion matrix of the blockwise classification of lung tissue patterns using QWF and GLH features and a LOPO cross-validation in %. Global arithmetic and geometric means of 75.1% and 74.7% are obtained respectively.  $N_{vox}$  denotes the number of manually segmented voxels used for evaluation and  $N_{cases}$  the number of patients.

	<i>healthy</i>	<i>emphysema</i>	<i>ground glass</i>	<i>fibrosis</i>	<i>micronodules</i>	$N_{vox}$	$N_{cases}$
<i>healthy</i>	<b>78.1</b>	2.8	0.7	0.2	18.1	63'914	7
<i>emphysema</i>	0.9	<b>70.1</b>	0	4.7	24.2	61'578	5
<i>ground glass</i>	4.6	1.6	<b>76</b>	14.7	3.1	644'814	21
<i>fibrosis</i>	2.3	1.9	17	<b>73.5</b>	5.3	860'474	28
<i>micronodules</i>	13.7	1.8	2.2	6.7	<b>75.7</b>	1'436'055	10

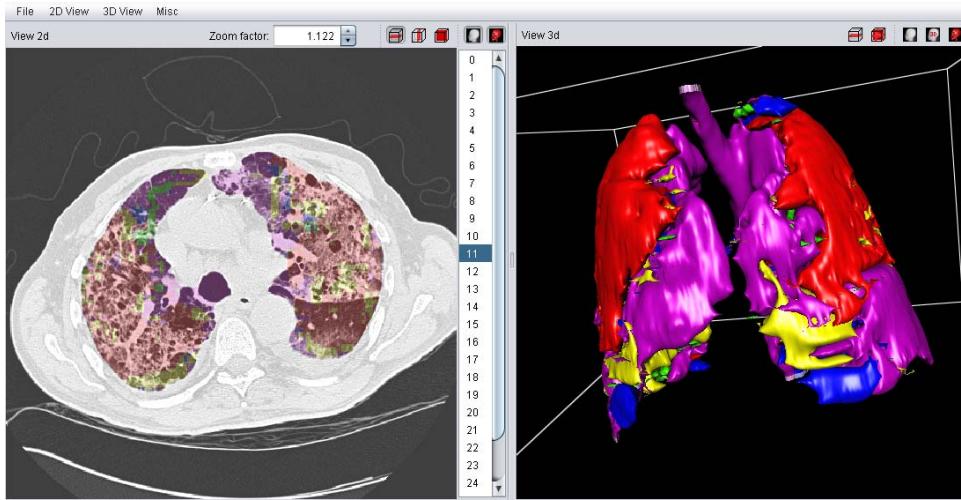


Figure 3.13: Automated segmentation of the lung tissue patterns of a patient affected with pulmonary fibrosis. The 3D segmented regions are displayed to the clinician using YaDiV. Green: *healthy* (0.1 liters), blue: *emphysema* (0.39 l), yellow: *ground glass* (0.53 l), red: *fibrosis* (1.91 l), pink: *micronodules* (1.77 l).

Table 3.8: Performance measures of the blockwise classification of the lung tissue patterns using QWF and GLH features.

	recall	precision	F-measure	accuracy
<i>healthy</i>	78.4	78.2	78.3	91.3
<i>emphysema</i>	89.6	70.2	78.7	92.4
<i>ground glass</i>	79.2	76	77.6	91.2
<i>fibrosis</i>	73.6	73.5	73.6	89.4
<i>micronodules</i>	59.9	75.6	66.8	85

### 3.1.4 Discussions on the visual features

In this section, the results presented in Section 3.1.3 are discussed. Global remarks and limitations of the studies are given at the end of this section.

#### GLH of the lung tissue patterns

Gray-level histograms in Figure 3.7 show high variability of distributions among the six patterns. *Healthy* patterns are constituted of soft tissue with HU values in the range [-1050;-300] and have the lowest mean value of -950 HU. Surprisingly, even *emphysema* patterns have a higher mean value of -888. This can be explained by looking at Figure 3.7 (b) where a relatively high number of voxels have values in [-700;400] which corresponds to increased attenuations surrounding large regions of air. Distribution of the *ground glass* voxels is close to those of *fibrosis* voxels which highlights the need of texture features that characterize the spatial organization of the voxels to separate the two patterns. The shape of the histogram characterizing *micronodule* voxels in Figure 3.7 (e) is the result of the combination of healthy tissue with small nodules of density similar to *consolidation* voxels. *consolidation* voxels are corresponding to a very high mean (around -600 HU) as well as a weak air component.

According to the notable variation among the gray-level distributions, features such as histogram bins, mean, variance, skewness and kurtosis of the distribution along with air components may have strong discriminative potential for the classification of the patterns. The investigations on the best number of histogram bins  $N_{bins}$  in Figure 3.8 shows that 40 bins allow the best classification accuracy  $A^{LORO}$ . However, in order to be further combined with texture features for classification, a reduced number is desirable to face the curse of dimensionality of feature spaces [102]. An accurate description of the gray-levels can be obtained in reducing the number of bins to 22 corresponding to pixel values in [-1050;600] because the bins outside this interval are very sparsely populated. [-1050;600] corresponds to an extended interval of the lung HU values that includes pathological tissues of high density (e.g. calcified nodules [152]). Because the majority of the information is contained in low HU values, a non-linear binning with bins of exponential width was tested but did not lead to better results. A reason for this is that the information that allows to differentiate among the patterns is mostly contained by HU values in [-700;-200] (see Figure 3.7).

It is important to note that although the shapes of the grey-level distributions are very distinctive in Figure 3.7, they are averaged over a large number of cases and thus corresponds to the center of mass of each class in the GLH feature space. However, large intra-class variations are observed as depicted in Figure 3.14, which makes the classification task not trivial. Additional features along with machine learning algorithms able to draw intricate decision boundaries are required for an accurate recognition of the lung tissue patterns. This is especially true with *healthy* tissue [3]. The blockwise analysis also tends to increase the intra-class variability.

#### DWF

The results in Figure 3.9 show that smaller scales allow better discrimination of patterns. Despite the fact that upsampling images without carrying out interpolation introduces high-frequency artifacts, the best overall accuracy is reached using upsampled original images with  $N_{upsample} = 2$ . However, we believe that each pattern contains specific scales and the effect of the scale has to be studied for each of them. QWF are expected to partly solve this problem by providing a finer scale progression with a factor of  $\sqrt{2}$  instead of 2 at each iteration (see next subsection). Indeed, relevant information can be hidden between two consecutive scales.

The confusion matrix in Table 3.3 shows how the combined features can separate each pattern. All 72 ROIs showing *emphysema*, which is characterized by the destruction of lung tissue (air), are 98.8% correctly classified with a precision of 100%. In this case, features such as percentage of air pixels are much more relevant compared to frequential analysis. *ground glass* has the lowest precision and is confused 9 times with *micronodules*. This can be explained by the lack of intermediate scales with the DWF, where the size of small nodules in *micronodules* is not correctly matched

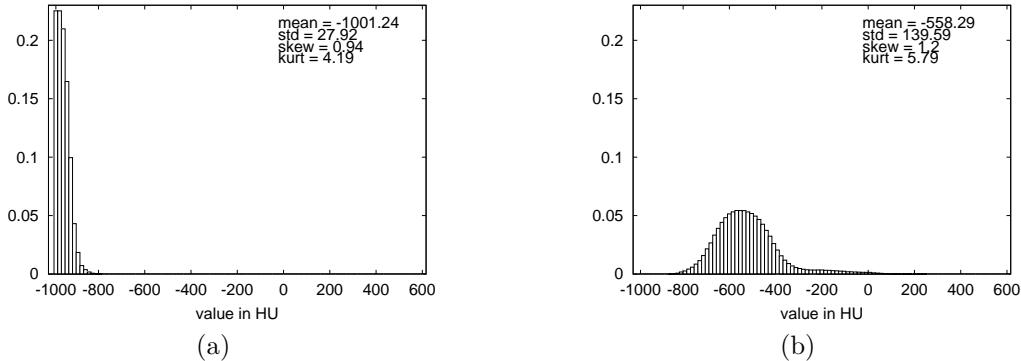


Figure 3.14: Intra-class variations for class *healthy*. The histograms of the two blocks shows two outliers that both mark the boundaries of the class. The block in (b) contains several bronchovascular structures which increase the global density.

by the scaled templates. These observations seem to be validated by the respective accuracies in Figure 3.10 where DWF is not accurate to classify *micronodules*. For patterns *ground glass* and *fibrosis*, the DWF shows superior discrimination performance compared to GLH. This can be explained by the fact that fibrosis is characterized by sharp transitions between high density tissue and small air bubbles that yield a specific spectrum signature. The combination of GLH and DWF features shows improvement in classification accuracy for most of the patterns except *healthy* and *emphysema*, and is particularly effective when carrying out multiclass classification.

### DWF versus QWF

Pairwise comparisons shown in Table 3.6 indicate that QWF outperform DWF in 88.9% of the comparisons (16 among 18) for all degrees/orders. This global increase in performance is primarily due to the better isotropy properties of the non-separable wavelet transform, thanks to their close connection to the Laplacian. The favored directions of the separable transform lead to noisy features creating non-homogeneous clusters of instances belonging to the same class in the feature space, which decreases global classification performance due to a lack of regularization.

Although having influence on global accuracy as well, the finer scale progression allowed by the quincunx subsampling scheme increases the precision of the classification; e.g., by reducing confusions between patterns with well-defined object sizes, such as micro- and macro-nodules. Indeed, the highest improvements using QWF are achieved for classes *healthy*, *micronodules* and *macronodules* in which the size of objects is important.

The reduced number of subbands with QWF (one for QWF and three for DWF) is also a very important clue to avoid the curse of dimensionality. QWF features allow to characterize twice more scales with even a reduced number of features when compared to DWF. In this study the DWF requires 36 features ( $4 \times 3 \times 3$ ) for characterizing 4 scales when the QWF use no more than 24 features ( $8 \times 3$ ) to characterize 8 scales. This highlights that albeit separable transforms being faster because the convolutions are carried out in one dimension, they yield several subbands per scale, which is less adapted for multidimensional signal analysis where the amount of information increases exponentially.

Degree/order  $\alpha = 2$  and  $\gamma = 3$  allow best classification for all patterns except for *fibrosis* although being very close to the best accuracy as well. A first reason for this is that the scale-progression is well initialized and fits best the characteristic structures of each patterns leading to increased classification accuracy. These degrees/orders are best trade-offs between spatial and spectral localization. Secondly, a sufficiently high order is required to obtain isotropy with QWF. When looking at the shapes of the templates  $\hat{\beta}_\gamma(\omega)$ , in the Fourier domain in Figure 3.4, one can easily remark that  $\beta_\gamma$  begins to have acceptable isotropic properties starting from  $\gamma = 3$  and lower values favor the  $x$  and  $y$  directions.

## Blockwise classification with QWF

The results obtained with the two components of the blockwise classification of the lung tissue patterns are discussed in this section.

**Segmentation of the lung volumes** Our experience with the segmentation of 69 lung volumes shows that the 3D region growing associated with closing allows an almost fully-automatic segmentation. However, the trachea is included as lung tissue in most cases (see Figure 3.11). Manual corrections are required when the closing operation cannot fill large regions of consolidated tissue.

**Automatic 3D lung tissue categorization with QWF** The automatic segmentation of the lung tissue is a crucial step for the success of the CAD. The accuracies obtained in Table 3.8 show that the SVM classifier can learn efficiently from the hybrid feature space. However, the recurrent confusion between *healthy* and *micronodules* patterns in Table 3.7 and Figure 3.13 suggests that the decision boundaries are not trivial in some cases. Indeed, the small nodules in *micronodules* are mixed with the bronchovascular structures contained in *healthy* tissue. This is a major limitation of the two-dimensional QWF used in this work. The use of three-dimensional isotropic polyharmonic *B*-spline wavelets as explicitly formulated in (3.4) in near-isotropic MDCT image series may be able to overcome the confusions between *healthy* and *micronodules* as the bronchovascular structures are cylindric whereas small nodules are roughly round. Unfortunately, the HRCT protocol used at the HUG does not allow for a true three-dimensional analysis as the spacing between the slices is 10 millimeters and yielding extremely anisotropic volumetric images but limiting the radiation dose delivered to the patients when compared to MDCT protocol. Using the clinical context of the images such as the age of the patient shows to allow clarifications between visually similar patterns in Section 4.2. Indeed, *micronodules* in a 20-year-old subject are very visually similar to *healthy* tissue surrounded by vessels of a 80-year-old man. The clinical context allows significant improvements particularly for the characterization of *micronodules* in Section 4.2.

Table 3.7 also shows recurrent confusions between *fibrosis* and *ground glass*. This may be partially explained by the fact that *fibrosis* patterns are most often accompanied with small regions of *ground glass* because of the re-distribution of the perfusion to the functional tissue remaining. This has the effect to overload the healthy tissue which thus has the visual appearance of *ground glass* because of increased attenuation. However, during the annotations sessions, the label *fibrosis* was assigned to the whole ROI leading to classification errors when the system correctly detects the small *ground glass* regions.

The distribution of the classes given by  $N_{vox}$  in Table 3.7 is very unbalanced, which renders the classification task more difficult [127]. This is mostly due to the way the radiologists delineate regions during the annotation session (see Section 2.4, Figure 2.7). In case of unbalanced classes, SVM classifiers with asymmetric margins can be used to favor minority classes as suggested in [42]. The most represented pattern *micronodules* is the most difficult to classify and obtains the worst performance in Table 3.8 which has the effect to decrease the global performance of the CAD system and requires further improvements. At the border of the lungs, missclassifications occur due to the response of the wavelets to the sharp change of intensity. A solution to this is to use a symmetry of the lung tissue using the tangent to the lung border as axis. To remove noise in the blockwise classification, a 3D averaging of the outputs of the classifier may avoid small isolated regions. The anatomical regions of the lung have not the same prior probabilities of having abnormal tissue. For instance, *fibrosis* is prevailing at the periphery of the lung bases. Integrating the anatomical regions for improving the classification performance requires to be able to automatically segment the lung regions [291] and showed to significantly improve classification accuracy in [290, 293].

## Global remarks on the visual features

Wavelets and filter banks are not the most popular feature extraction techniques for the characterization of parenchymal textures in the literature (see Figure 1.17). Although a direct comparison of the performance reported does not make sense as the sorts of tissue included, the validation

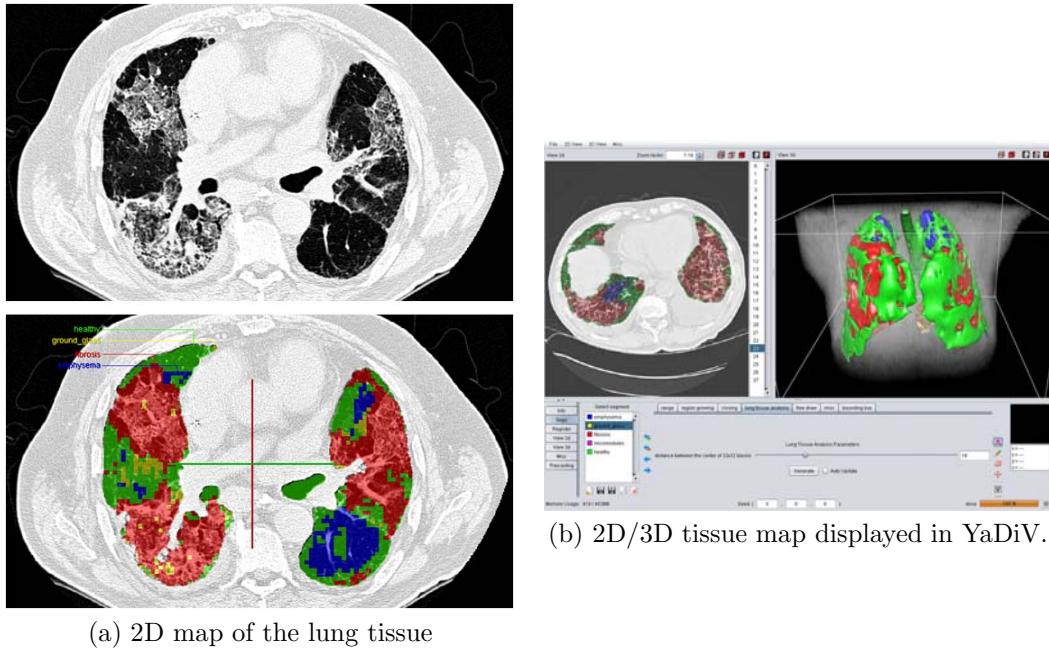


Figure 3.15: 2D and 3D maps of the lung tissue as diagnostic aid tools for the radiologists.

methods and the dataset used vary a lot among the studies, we believe that wavelet-based texture features are more adapted to analyze texture patterns in medical images. Indeed, filtering techniques and wavelets offer an overcomplete feature set able to fit most of the texture functions in condition to efficiently derive features from the coefficients. Wavelet-based texture features are covering the whole spectrum, which truly allows to detect the important spectrum signatures of the patterns, being perfectly complementary to the measures of density using GLH.

Among all the experiments run in this chapter, the combination of QWF with GLH showed to best discriminate the lung tissue patterns. However, no comparison with directional analysis is carried out. Preliminary tests partially using the methods described in [257] did not lead to better results which suggests that isotropic analysis is adapted to lung tissue analysis that may not contain prevailing orientation. Nevertheless, this is contradictory to the results obtained in [242] and [264], and further investigations are required to learn about the orientations in the lung tissue patterns.

As already highlighted in the literature, we observed that the dataset, the selected classes of lung tissue and the validation methods have all a strong influence on the measured performance of the CAD. In Table 3.6 for instance, the class *macronodules* decreases the global performance significantly. In addition, it is important to note that the performance shown in Table 3.4 is biased by the LORO cross-validation which allows training and testing with ROIs that belong to a same patient. The evaluation used to compare QWF and DWF in Table 3.6 is subject to the same bias as ROIs from the same patient can be used in the training and the testing sets. However, the hand-drawn ROIs used as instances do not overlap in these two studies, which limits the bias as the classification algorithms are never trained and tested with the same data. Another limitation of these two validations is that classification of entire ROIs does not allow automatic detection of pathological lung tissue in a whole HRCT image series as it requires a prior selection of a region to be analyzed. This biases again the performance measures because it corresponds to the way the radiologists annotate ROIs and not to the true repartition of the pathological tissues. The performance shown in Tables 3.7 and 3.8 has to be considered as the baseline performance in accordance with actual clinical situations as the LOPO cross-validation simulates true clinical routine usage of the CAD (see Section 5.2.1). Moreover, the blockwise classification allows to detect abnormalities in HRCT images series and provides a three-dimensional map of the lung tissue (see Figure 3.15) that indicates the suspicious regions to the unexperienced radiologist.

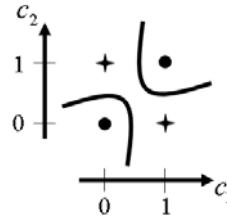


Figure 3.16: Binary features increase the risk of XOR configurations leading to highly non-linear decision boundaries.

## 3.2 Clinical and multimodal features

In this section, the construction of the clinical feature space is described. The choice of the clinical parameters based on their relevance for diagnosing ILDs is described in Section 2.2. The purpose of the clinical parameters are multiple. In Section 4.2, a multimodal feature space consisting of the clinical features combined with the visual features is used to investigate the influence of the clinical context in lung tissue classification. Then, in Section 4.3, the clinical parameters are combined with visual features to build a similarity measure for case-based retrieval.

This section is structured as follows. In Section 3.2.1 the methods for extracting features ready to be used in a classification algorithm from the clinical parameters collected from the EHR and stored in a MySQL database are described. A feature selection based on two ranking measures is proposed. Section 3.2.2 investigates the consistency of the multimodal feature space and shows the results of the feature ranking. The results are discussed in Section 3.2.3.

### 3.2.1 Extraction: from MySQL variables to a consistent feature space

The clinical parameters entered in the MySQL database as described in Section 2.3.3 are not directly usable for data-mining. Pre-processing steps are required to build a consistent feature space. Clinical features  $c_n$  have to be selected according to their relevance for the classification of the lung tissue.

#### Preprocessing

Nominal variables are divided into several 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. 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.

#### Feature ranking

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 4.2.2), dependencies between clinical attributes and classes of lung tissue must be investigated before any fusion with the visual features. Due to missing values, binarization or irrelevance according to the studied diseases, some features may 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 (see Figure 3.16). Moreover, it is preferable to keep as few features as possible to limit the curse of dimensionality. A feature ranking is thus required to build an effective set of attributes.

Table 3.9: Dataset used to rank the clinical parameters and to evaluate the influence of the clinical context on lung tissue categorization. Each patient may contain several kinds of lung tissue patterns.

class	<i>healthy</i>	<i>emphysema</i>	<i>ground glass</i>	<i>fibrosis</i>	<i>micronodules</i>
ROIs	63	58	148	312	155
patients	5	4	14	28	5

Two measures are compared for 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 [195]. 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:

$$I_G(Y|X) = H(Y) - H(Y|X) \quad (3.11)$$

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:

$$H(Y) = - \sum_{y \in \mathcal{Y}} p(y) \log p(y) \quad (3.12)$$

$$H(Y|X) = - \sum_{x \in \mathcal{X}, y \in \mathcal{Y}} p(x, y) \log p(y|x) \quad (3.13)$$

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

$$I_{G_{ratio}}(Y|X) = \frac{I_G(Y|X)}{- \sum_{i=1}^l \frac{|T_i|}{|T|} \log(\frac{|T_i|}{|T|})} \quad (3.14)$$

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 [183, 196]. 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 based on a feature vector concatenating all visual features  $t_m$  with the studied clinical feature  $c_n$ :

$$\mathbf{v} = (t_1 \dots t_M \ c_n) \quad (3.15)$$

### 3.2.2 Evaluation

In this section, the feature extraction methods and the ranking measures introduced in Section 3.2.1 are applied to a dataset consisting of 48 patients described in Table 3.9.

#### Clinical feature space $\mathbf{c}_n$

After having gathered binary and continuous variables, the created clinical feature space contains 72 attributes (63 binary and 9 continuous). There are a majority of binary parameters as all categorical parameters were divided into a number of binary parameters equal to the number of categories. The *unknown* values were substituted by the mean of the binary attribute over all the patients. 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 mean filling rate of the retained attributes is 88.7%.

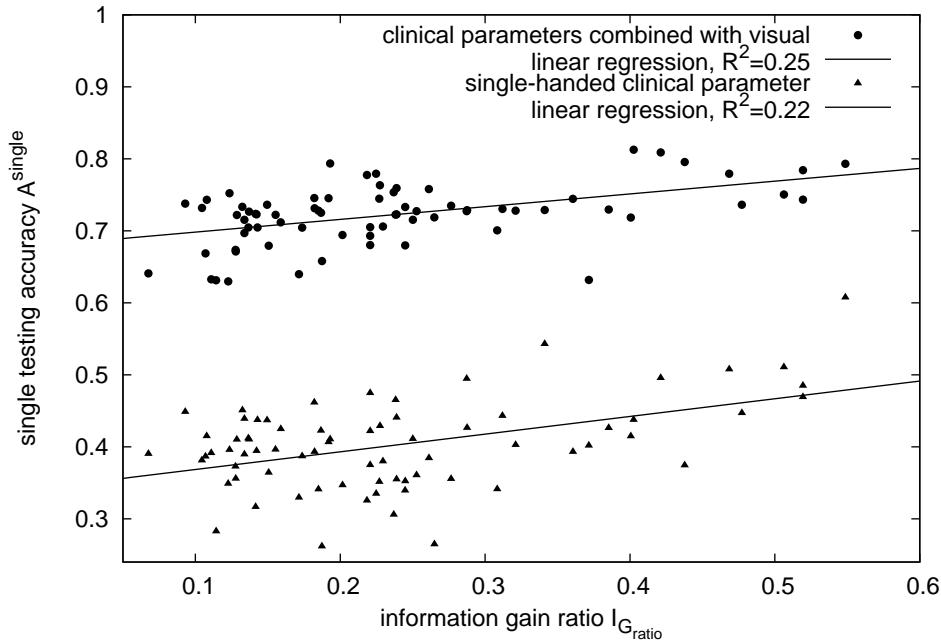


Figure 3.17: Correlation of  $I_{G_{ratio}}$  and  $A^{single}$ .

### Ranking and consistency of the multimodal feature space

The correlation of  $I_{G_{ratio}}$  and  $A^{single}$  is evaluated in Figure 3.17.  $A^{single}$  is first obtained using visual features  $t_m$  and a LOPO cross-validation using the full two-dimensional ROIs to train an SVM classifier. Secondly,  $A^{single}$  is evaluated for each single-handed clinical parameter  $c_n$ . In both cases, the training set is used both for grid search for optimal SVM 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  $\sigma_K$  of the Gaussian kernel. A grid search is carried out within the intervals  $C \in [1; 100]$  and  $\sigma_K \in [10^{-2}; 10^2]$ . For every coordinate of the grid, another 10-fold crossvalidation is carried out on the training set. Optimal parameters ( $C^{opt}$ ,  $\sigma_K^{opt}$ ) that allowed best mean accuracy  $A^{cv}$  are used to train the final model on the entire training set. A preliminary coarse grid search was performed to locate regions of the space with high  $A^{cv}$  values.

Table 3.10 lists the first 20 clinical attributes with highest  $A^{single}$  value. The multimodal feature vector  $\mathbf{v}$  is obtained by concatenating the visual features  $t_m$  consisting of QWF and GLH as described in Section 3.1.3 and the 20 first clinical features  $c_n$  with the highest  $A^{single}$  value:

$$\mathbf{v} = (t_1 \dots t_M \ c_1 \dots c_N) \quad (3.16)$$

with  $m \in [1; M]$  and  $n \in [1; N]$ . The consistency of the multimodal feature space spanned by  $\mathbf{v}$  is studied using a correlation matrix in Figure 3.18.

### 3.2.3 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 Figure 3.18.

#### Measures for ranking

Figure 3.17 shows that  $I_{G_{ratio}}$  is little correlated to both  $A^{single}$  and testing accuracy obtained with each single-handed clinical parameter. Pearson's coefficient of regression  $R^2$  is below 0.25 for

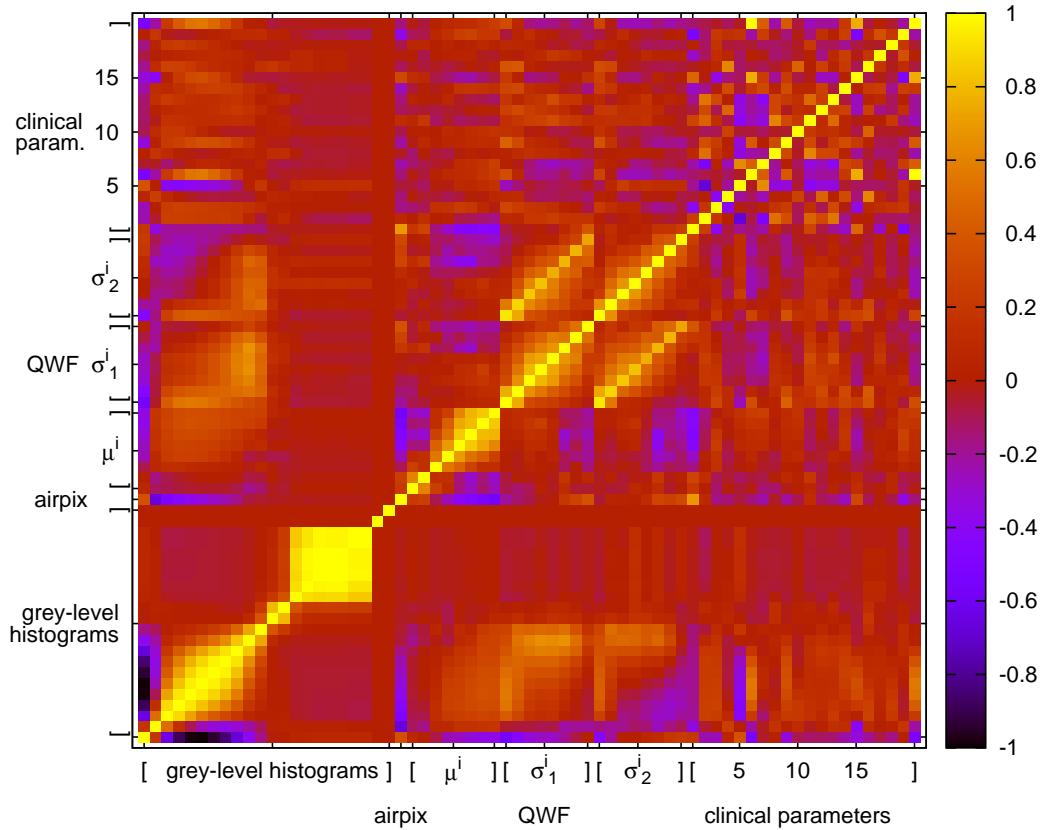


Figure 3.18: Correlation matrix of the multimodal feature space. Indexes of the clinical parameters corresponds to their rank described in Table 3.10.

Table 3.10: 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_medicalHTA	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

both comparisons. Even if  $A^{single}$  is averaged over 30 experiments, the values obtained still have high variance according to the draws 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 [97], 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 [64] 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.

### 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 an important influence on the quality of clinical data. At last, the relevance of the parameters according to the studied diseases is of course primordial.

As observed in Section 4.2.2, the age has an important influence on the visual aspect of lung tissue (see Figure 4.8) and this is confirmed by finding it at the 2<sup>nd</sup> rank in Table 3.10. 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 hematocrit level. This phenomenon is indeed commonly observed in cases affected with COPD [34], characterized by HRCT images showing *emphysema* patterns. The latter observation

is firmly confirmed by looking at the correlation matrix in Figure 3.18, where the first clinical parameter (*laboratory\_hematocrit*) is strongly anticorrelated with the means  $\mu_i$  of the QWF and highly correlated to *airpix*. The means  $\mu_i$  of the QWF have high values for inhomogeneous patterns, where *emphysema* patterns are very homogeneous due to absence of lung tissue. Coherently, the 3<sup>rd</sup> rank is occupied by the parameter *laboratory\_hemoglobin*, which is also involved in the transport of oxygen. Hemoglobin is the protein contained in red blood cells that is responsible for delivery of oxygen to the tissue. 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 5<sup>th</sup> and 7<sup>th</sup> ranks are not surprising as they usually highlight the presence of a host illness. Finding the parameter *medication\_cordarone* at the 8<sup>th</sup> rank is in accordance with the well-known side effect of the cordarone drug creating pulmonary fibrosis when used over long periods.

### Consistency of the multimodal feature space

The study of the correlation of the multimodal feature space is carried out in Figure 3.18. A first look at the correlation matrix shows that clinical features have little correlation with the visual features. This is confirmed as the mean correlation  $\rho_{mean}$  of the two groups is equal to 0.0143. Several homogeneous groups can be identified within the visual features. The first histogram bin representing pixel values within  $[-1050; -975[$  HU is of course highly correlated with *airpix* and is anticorrelated with bins in range  $[-900; -450[$ . This is partly due to patterns with low-density tissue (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 tissue with HU 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  $\mu^i$  and the standard-deviations  $\sigma_{1,2}^i$ .  $\rho_{mean}$  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 [143].



# Chapter 4

## Classification and retrieval

In Chapter 3, a multimodal feature set was presented. Features modeling visual and text-based clinical features proved to create a consistent feature space adapted to describe lung tissue patterns associated with ILDs in HRCT imaging. Although feature extraction is a crucial step for the success of the CAD system, particular care must also be employed to draw adequate decisions or to assess similarity between instances according to the values of the features [19]. In the case of a detection-based CAD system, classification algorithms are required to predict the class of an unseen instance (see Section 1.1.1). With CBIR-based CAD, the retrieval engine needs a distance measure in order to evaluate the similarity between the two instances represented by feature vectors (see Section 1.1.2).

In the first part of this chapter, a classification framework is proposed to adequately mine the multimodal feature set described in Chapter 3. The goal is to categorize full HRCT image series from patients affected with ILDs and to provide a three-dimensional map of the lung tissue. In a first step, the choice of the appropriate model for mining the visual features is carried out using a novel methodology based on McNemar's statistical test in Section 4.1. Based on the chosen model, the strategy required to fuse visual and clinical attributes is investigated in Section 4.2. In the second part of the chapter, methods based on the three-dimensional map of lung tissue as well as on clinical attributes are proposed to carry out case-based retrieval in Section 4.3.

### 4.1 Choice of the model based on visual features

A large variety of classification algorithms exists in the literature [79, 104, 273]. Each of them was developed in the context of a particular application, which led to machine learning algorithms with various skills and limitations. Most often, it is difficult to predict which one will perform best with a given classification problem and a given dataset. In this section, a comparison of five state-of-the-art classification algorithms applied to lung tissue categorization is carried out. The theoretical basis of each classifier family is detailed in Section 1.2.2. The supervised learning approach is used as it corresponds best to the radiologists' approach consisting of learning and gaining experience through training sessions and clinical routine. The methodology utilized to compare the performance of each classifier family and the results are described in Section 4.1.1. The comparison of the classifier performances are discussed in Section 4.1.2.

#### 4.1.1 Methodology for comparing classifier families

The full dataset is divided into two equal parts: 50% for training and 50% for testing. Training means both search for optimal parameters and creation of the model (i.e. adjustments of the decision boundary). The methodology is explained in Figure 4.1.

The dataset used for the comparison is detailed in Table 4.1. 843 ROIs from healthy and five pathological lung tissue patterns are selected for training and testing the classifiers. The selected patterns are *healthy*, *emphysema*, *ground glass*, *fibrosis*, *micronodules* and *macronodules*.

Table 4.1: Distribution of the ROIs per class of lung tissue pattern used to compare the classifier performances.

class	<i>healthy</i>	<i>emphysema</i>	<i>ground glass</i>	<i>fibrosis</i>	<i>micronodules</i>	<i>macronodules</i>
ROIs	113	93	148	312	155	22
patients	11	6	14	28	5	5

Global experimentation is performed 50 times, which results in 50 divisions of the dataset

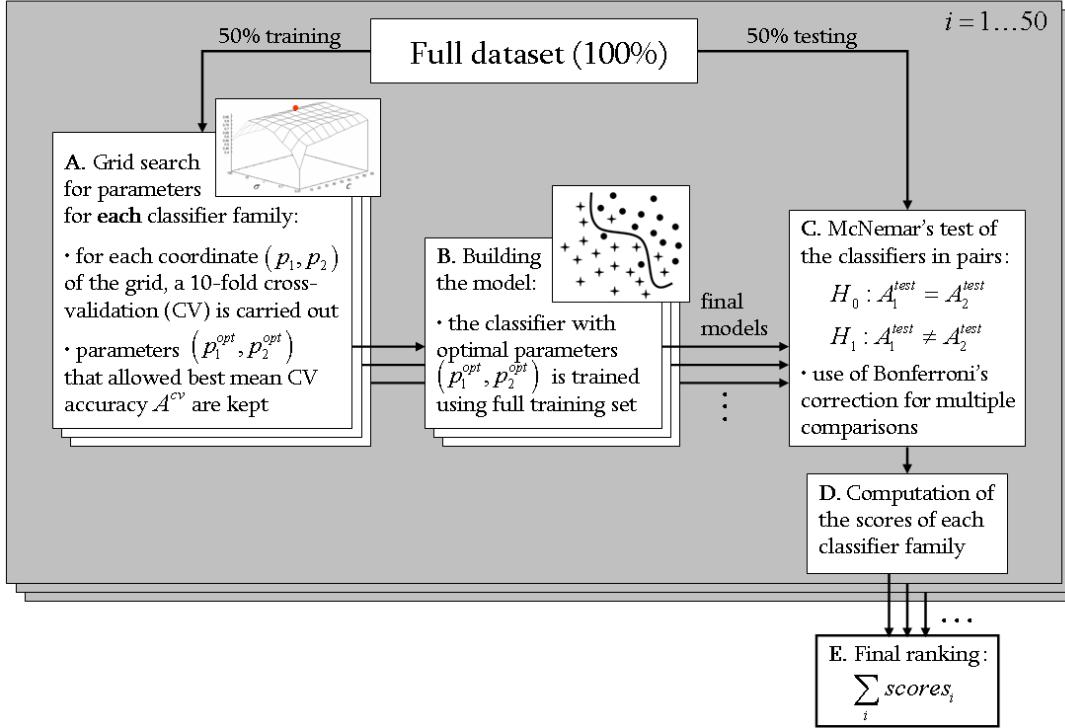


Figure 4.1: Methodology for comparing the classifier performance.

Distributions of the classes are highly imbalanced as the largest class *fibrosis* contains 312 ROIs and the smallest class *macronodules* only 22 ROIs. There is a mean of 140.5 ROIs per class. The entire ROI constitute the instance to be classified (the ROIs are not cut into blocks for classification).

Classifier implementations were taken from the open source Java library *Weka* [79, 273]. The *LIBSVM* library is used for the SVMs' C-support vector machine classification [35].

### Grid search for optimal parameters

In order to determine the optimal parameters  $p_i$ , a grid search is performed for each classifier family. When required, exponential grid steps were used for coarse search. For every coordinate of the grid, a 10-fold CV is carried out on the training set. Optimal parameters  $p_i^{opt}$  that allowed best mean CV accuracy  $A^{cv}$  are used to train the final model on the entire training set. Optimized parameters are detailed in Table 4.2. An example of grid search for best  $A^{cv}$  is shown in Figure 4.2 where the cost  $C$  and the  $\sigma_K$  value of the Gaussian kernel of the SVMs are optimized. A preliminary coarse grid search is performed to locate regions of the space with high  $A^{cv}$  values.

Table 4.2: Grid search for optimal parameters  $p_i^{opt}$ . The values for the number of hidden layer units  $N_{hidden}$  of the MLP are chosen as {none, number of classes, (number of attributes + number of classes)/2, number of attributes + number of classes}.

classifier family	parameters	ranges	step
Naive Bayes	—	—	—
k–NN	$k$	[0; 100]	linear
C4.5	$N_{instances}, C_{pruning}$	[0; 5], [0.02; 0.24]	lin, lin
MLP	$R_{learn}, N_{hidden}$	$[10^{-10}; 10^5]$ , {0, 6, 22, 45}	log, —
SVMs	$C, \sigma_K$	[1; 100], $[10^{-2}; 10^2]$	lin, log

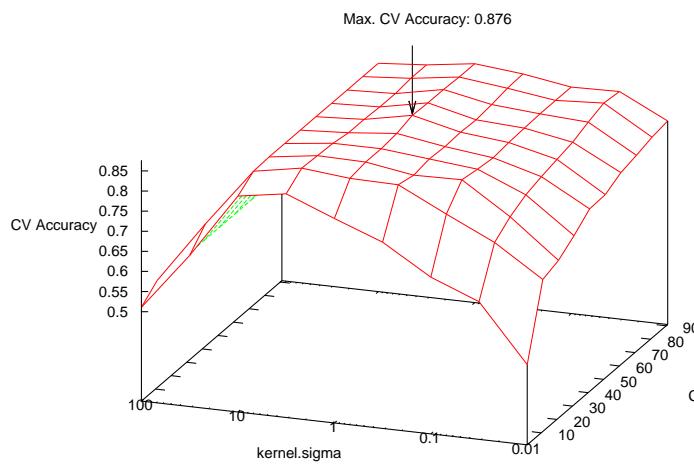


Figure 4.2: Grid search for optimal SVMs parameters  $C$  and  $\sigma_K$ .

Table 4.3: Class-specific accuracies for each classifier family. Best performances are highlighted in bold. SVMs reached three times the best accuracy and is no more than 2% behind the best performance over all classes.

	<i>healthy</i>	<i>emphysema</i>	<i>ground glass</i>	<i>fibrosis</i>	<i>micronodules</i>	<i>macronodules</i>
Naive Bayes	0.9161	0.8753	<b>0.7873</b>	0.7454	0.3778	0.3076
k-NN	0.9104	<b>0.9978</b>	0.7369	0.8926	<b>0.91</b>	0.3605
C4.5	0.7568	0.9419	0.6803	0.8821	0.7555	0.3054
MLP	0.7290	0.9707	0.6756	0.8751	0.8035	0.2461
SVMs	<b>0.9242</b>	0.9874	0.7731	<b>0.9218</b>	0.8907	<b>0.4250</b>

Table 4.4: Class-specific accuracies for each classifier family with a 2-class configuration. Naive Bayes performs well for classifying healthy tissue.

	<i>healthy</i>	<i>pathological</i>
Naive Bayes	<b>0.922</b>	0.8087
k-NN	0.8923	0.9764
C4.5	0.6985	0.9675
MLP	0.711	0.958
SVMs	0.8535	<b>0.9818</b>

### Ranking of the classifier families

Instances of the test set are classified by each classifier family and McNemar's test is applied to the classifiers in pairs with the hypotheses:

$$H_0 : A_1^{test} = A_2^{test}$$

$$H_1 : A_1^{test} \neq A_2^{test}$$

with  $A_{1,2}^{test}$  the testing accuracy of the classifiers 1,2 computed as the number of correctly classified instances divided by the total number of instances in the test set. Compared to other statistical tests for comparing supervised classification learning algorithms, McNemar's test showed to be the only test with acceptable type I error rate in [63]. Type I errors correspond to a false detection of difference in performance between two algorithms. Bonferroni's correction for multiple comparisons is used to adjust the threshold of the test. When  $H_0$  is rejected and  $A_1^{test}$  is greater than  $A_2^{test}$ , the score of the classifier 1 is incremented. When  $H_0$  is accepted, 0.5 is added to the scores of both classifiers. The global experimentation is repeated 50 times and a final ranking based on the total of the scores is performed. As the distribution of the classes is highly imbalanced, the geometric mean<sup>1</sup>  $A_{mean}^{geom}$  of each class-specific accuracy  $A_l$  on the test set is computed for every classifier. Two classification configurations are investigated. First, classifiers are evaluated on a multiclass configuration using all 6 classes of lung tissue. Secondly, a 2-class configuration opposing *healthy* versus *pathological* tissues is investigated. In this configuration, the classes *emphysema*, *ground glass*, *fibrosis*, *micronodules* and *macronodules* are grouped together to form the class *pathological* containing 730 ROIs versus 113 for the class *healthy* (86% *pathological*). Final rankings, mean testing accuracies  $A_{mean}^{test}$  and mean geometric accuracies  $A_{mean}^{geom}$  are shown in Figures 4.3, 4.4, 4.5 and 4.6. The class-specific accuracies achieved by each classifier family are presented in Tables 4.3 and 4.4.

---

<sup>1</sup>see Section 5.2.2 for details on the performance measures

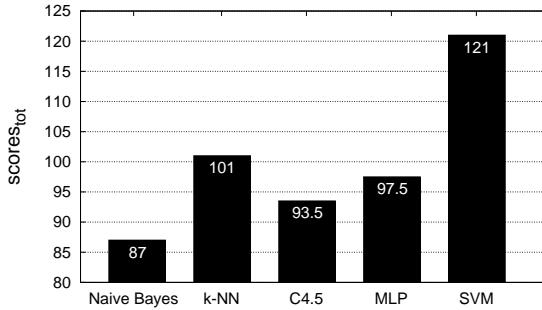


Figure 4.3: Final ranking based on the total of the scores of the classifiers with the 6-class configuration. SVMs reached the best score with 121.

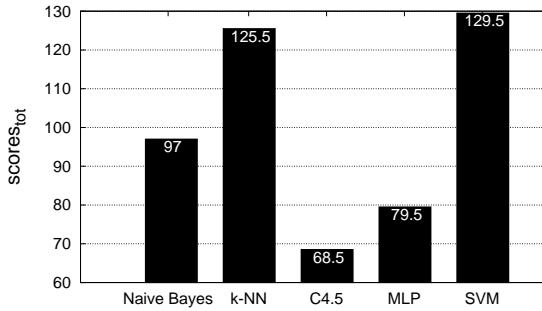


Figure 4.5: Final ranking with the 2-class configuration. Again, SVMs reached the best score with 129.5.

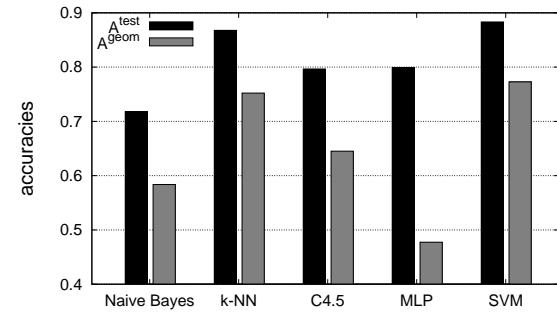


Figure 4.4: Overall mean testing accuracies  $A_{mean}^{test}$  and  $A_{mean}^{geom}$  with 6 classes. SVMs reached best accuracies with  $A_{mean}^{test} = 88.3\%$  and  $A_{mean}^{geom} = 77.3\%$ .

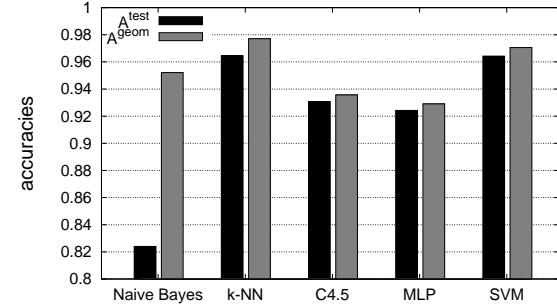


Figure 4.6: Overall mean testing accuracies with 2 classes. k-NN reached best accuracies with  $A_{mean}^{test} = 96.7\%$  and  $A_{mean}^{geom} = 97.7\%$  and is closely followed by SVMs.

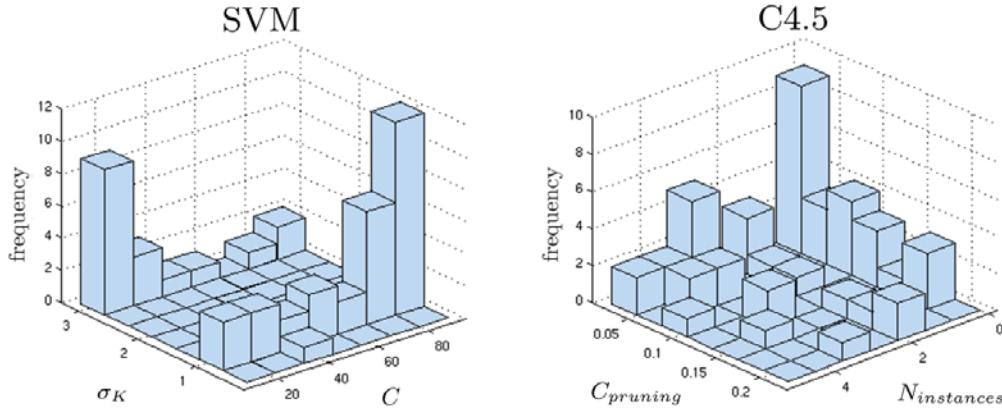


Figure 4.7: Bivariate histograms of the optimal parameters  $(p_1^{opt}, p_2^{opt})$  for SVMs and C4.5.

### Stability of the classifier parameters

Optimal parameters of each classifier family were stored for every 50 : 50 division. In order to study the stability, histograms of the values of  $(p_1^{opt}, p_2^{opt})$  are built for SVMs and C4.5 as shown in Figure 4.7.

#### 4.1.2 Discussion on classifier performances

All scores shown in Figures 4.3 and 4.5 show strong variations among the performances of the classifiers. Moreover, the variations can be decreased by the use of Bonferroni's correction, which makes the tests more permissive (i.e. McNemar's test rejects  $H_0$  more easily). Two classifier families reach highest scores: k-NN and SVMs. These performances are confirmed by their respective accuracies in Figures 4.4 and 4.6. Overall scores and mean testing accuracies  $A_{mean}^{test}$  show to be complementary metrics. For example, with the 6-class configuration the MLP reaches high global accuracy of 79.9% with a low score of 97.5. Those discordances can be understood by looking at the mean geometric accuracies  $A_{mean}^{geom}$ . The latter is very low for MLP with a value of 47.7%, which indicates that the MLP has a very low class-specific accuracy, and thus a low precision for each class, which is not suitable for the characterization of lung tissue where the prevalence of each tissue sort is unbalanced. Therefore, the SVMs that reached best score and global accuracy is able to classify tissue of each class accurately, even from those that are little represented. Beyond the fact that the k-NN classifier reached a slightly lower score and global accuracy compared to SVMs with 6 classes, one problem occurs with this classifier. The optimal number  $k$  of nearest neighbors for each of the 50 training/testing splits was 1. This can be explained by the fact that for some classes, the number of patients is low and thus many ROIs are extracted from the same image series. Training and testing with images from the same image series can result in a biased classification as images are similar as they belong to the same patient. Two such instances are artificially close in the feature space and will facilitate the classification task while attributing the class of the closest neighbor, which probably belongs to the same image series. In that sense, the k-NN classifier carries out overfitting of the training instances, which is not suitable for classifying ROIs from new ILD cases.

The complementarity of the classifiers is studied in Tables 4.3 and 4.4. Naive Bayes performs surprisingly well for classifying *healthy* tissue in the the 2-class configuration. However, the low accuracy achieved on the majority class *pathological* suggests that naive Bayes tends to favor the class *healthy*. Again, the competition between k-NN and SVMs is tight with an advantage for SVMs. For the 6-class configuration, SVMs reached three times the best accuracy and is no more than 2% behind the best performance over all classes. On the difficult class *macronodules*, SVMs outperforms all other classifier families by more than 6% of accuracy.

Distributions of the optimal parameters  $(p_1^{opt}, p_2^{opt})$  represented in Figure 4.7 show distinct be-

havior for SVMs and C4.5. Coupled parameters are more uniformly distributed for C4.5 compared to SVMs:  $\sigma_K$  of the Gaussian kernel of SVMs is characterized by a bimodal distribution. This means that two values of  $\sigma_K$  allow a convenient mapping of the feature space to higher dimensions for accurate separation of the classes. These values affect the optimal cost value  $C$ . Indeed, the organization of the classes in the transformed space is fixed by the value of  $\sigma_K$ , which requires a corresponding readjustment of the cost  $C$  that allows best trade-off between regularization and accuracy with the training examples. The most frequent pair of values of C4.5 occurs 9 times over 50, while the second most frequent pair occurs 5 times. For the SVMs, the most frequent pair occurs 12 times over 50 while the second most frequent pair occurs 9 times. In that sense, the SVMs classifier offers more stability. The stability has an important influence on the generalization performance: a classifier that frequently obtained identical pairs of optimal parameters has a high probability to be optimal for classifying new data.

The outcomes of the comparison are validated by other studies where the SVMs classifier outperformed other classification algorithms. In [141], a SVMs classifier allowed best classification results using DWF-based texture features on the Brodatz database [22]. Another study showed that the SVMs classifier was better adapted than MLP or naive Bayes for categorizing lung tissue in HRCT images of patient affected with COPD in [137].

## 4.2 Clinical context integration for medical image analysis

When analyzing an image, one interprets its content according to a given context. This is particularly true when analyzing medical images. Although fundamental in almost every medical field, the context is rarely used in computer vision applications. In CBIR for example, it became quickly clear that visual data alone can only achieve a limited retrieval quality when used for diagnosis aid [166] and that visual and other clinical data contain complementary information. However, the integration of the clinical context for classification or CBIR implicates several challenges. On the one hand, collecting contextual information beside images is usually time-consuming, requiring the help of a specialist for finding the relevant information and converting it from text to structured data to be used as a feature. For the majority of the common parameters, the value can be extracted automatically from the EHR [239]. On the other hand, a high-level of knowledge of the application domain is required to find relevant contextual parameters [245]. The selection of parameters for contextual medical image analysis has to be carried out based on the 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 same patient. Integrating information from multiple modalities consists of two major steps [277]. First, 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  $m_i$  is represented by a set of features  $\mathbf{v}_i$ . 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 [84, 230]:

- early fusion, where features  $\mathbf{v}_i$  from each modality are concatenated into one vector  $\mathbf{v} = (\mathbf{v}_1 \dots \mathbf{v}_{N_{mod}})$  to create one unique feature space,
- late fusion, where several classifiers  $\{q_1, \dots, q_{N_{mod}}\}$  are built on each modality  $\{\mathbf{v}_1, \dots, \mathbf{v}_{N_{mod}}\}$  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  $\mathbf{v}$  is equal to the sum of the dimensions of the subspaces  $\{\mathbf{v}_1, \dots, \mathbf{v}_{N_{mod}}\}$ . 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 and multivariate feature selection methods are required [14, 116].

Combining classifiers has been a very active domain over the past ten years [115, 132]. Wide interest for the latter mainly relies on the assumption that the heterogeneity of classifiers  $\{q_1, \dots, q_{N_{mod}}\}$  allows self-correction of the errors leading to better results [64]. Late fusion is a special case of classifier combination where heterogeneity occurs in the input spaces. It allows to face the *curse of dimensionality* by dividing the feature space without neglecting information contained in features that would be discarded by classical feature selection methods [223].

This section is structured as follows. Section 4.2.1 reviews the literature in contextual image analysis. Section 4.2.2 points out the importance of the clinical context for a correct interpretation of HRCT images. In Section 4.2.3, two fusion techniques are proposed to carry out multimodal classification of the lung tissue patterns. Visual and clinical features are combined using early and late fusion. The two approaches are compared and discussed.

### 4.2.1 Related work in contextual image analysis

Studies in contextual image analysis are discussed in this section. The idea of integrating the context to improve image analysis was used both in the CBIR and general medical pattern recognition communities. Both early and late fusion schemes are found in the literature.

#### Context in CBIR

Context has been used in CBIR where information from textual annotations of images was fused with image features (color histogram and Gabor filters) using early fusion to compare the performances of image retrieval using pure content, pure context and the combination of both in a collection of images from a Dutch newspaper in [270]. It is shown that the combination of content and context outperforms each method separately.

In [29], a CBIR system combined visual with textual statistics again directly in the feature vector space representation (early fusion) to retrieve images from the Internet. Textual statistics are extracted from the content of the HTML document and the visual content of the images is extracted with color histograms and histograms of the dominant orientations. The weighting of the modalities is carried out using relevance feedback. The maximum improvement was achieved when both visual and textual were used in the relevance feedback framework.

Inter-media medical image retrieval was carried out in [130] using textual, color and texture features semantically parsed and described with the Unified Medical Language System (UMLS). The visual and textual information was combined in the calculation of the similarity measure. Best mean average precision (MAP) is obtained with the mixed retrieval. The results show that the use of explicit knowledge with UMLS allows to improve the retrieval performance for text only and this approach showed to be less effective on visual features.

Investigation of the effectiveness of combining text and images for retrieval including medical image retrieval is one of the main goals of the CBIR benchmarking campaign ImageCLEF<sup>2</sup> [92, 170].

#### Context in medical pattern recognition

Combined decisions of classifiers constructed on sequentially selected sets of features were tested on four datasets including medical data in [223]. 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 [38]. 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 and the associated prediction performance was performed.

---

<sup>2</sup><http://imageclef.org/>, as of 5 November 2009

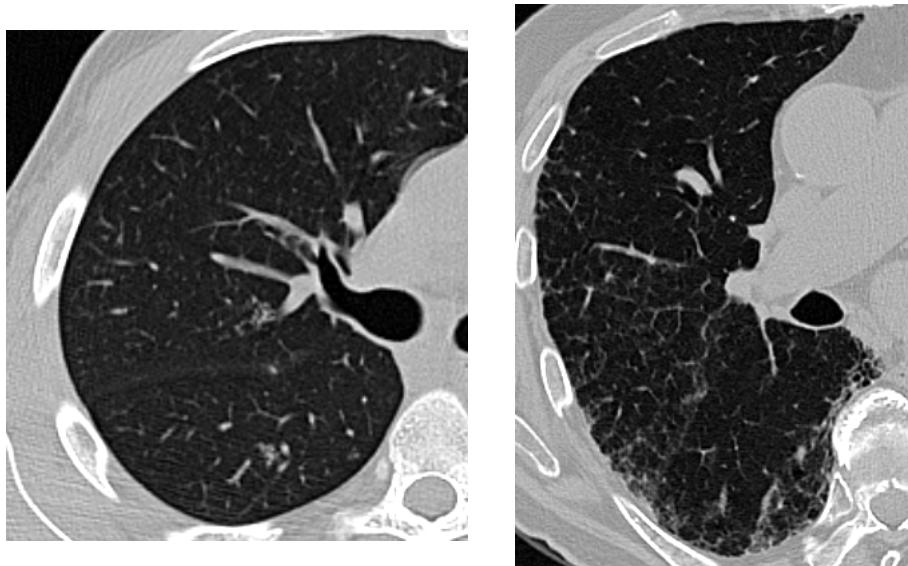


Figure 4.8: Influence of the age on the visual aspect of the lung tissue. Healthy tissue from a 25-year-old man on the left, and from an 88-year-old man on the right. Both images have identical window level settings.

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 [1]. 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 [293]. The locations of the patterns showed to significantly improve detection performance but require an accurate segmentation of the anatomy of the lung.

As observed in Section 1.3.2, many image-based diagnostic aid systems for ILDs were developed for the categorization of lung tissue patterns in HRCT data [220, 246, 253] and some of them showed to be effective in clinical routine [4]. 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 QWF and GLH was investigated in Chapter 3 and support vector machines (SVMs) showed to be optimal for the categorization of lung tissue in Section 4.1. In Section 4.2.3, we study the influence of the integration of the clinical context of HRCT images on classification performance of 2D hand-drawn ROIs in axial slices from patients affected with an ILD. Early and late fusion strategies are studied.

#### 4.2.2 The clinical context of HRCT images from patients with ILDs

Radiologists do never interpret HRCT images without taking into account the clinical context. Several clinical parameters — in particular the age of the patient (see Fig. 4.8) — have a major influence on the visual aspects (densities) of HRCT images of the chest [51, 57, 161, 210]. For example, discovering fibrotic findings in a lung belonging to an 80-year-old patient is not as surprising as finding them in a lung of a 25-years-old young person. In Figure 4.8, 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. A comprehensive description of the clinical parameters involved in the diagnostic process of the ILDs can be found in Section 2.2.2. To accurately analyze HRCT images an image-based computerized diagnostic aid system for ILDs must integrate the clinical context of the images.

### 4.2.3 Modality fusion

Integrating information from multiple sources for pattern recognition is not straightforward and require precautions to achieve best performance. The trivial way of fusing information from various modalities is to create one large feature space by concatenating the features from every modality (early fusion). The redundancy and synergies among groups of features must be studied and the increased number of features associated with multiple sources of information requires special attention to avoid the *curse of dimensionality*. A solution to manage this is to split the feature spaces into subgroups that are mined by multiple classifier systems (late fusion). The latter has been an active research domain during the last decade. In particular, the International Workshops on Multiple Classifier Systems<sup>3</sup> (MCS) endorsed by the International Association for Pattern Recognition (IAPR) brought several contributions on the problematic of combining classifiers. Several other fusion scheme were proposed in the literature. For instance, non-homogeneous transformations of the feature space were proposed in [289] using separated kernels for each modality.

In this section, two fusion strategies are proposed, evaluated and compared: early and late fusion.

#### Early fusion

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

$$\mathbf{v} = (t_1 \dots t_M \ c_1 \dots c_N) \quad (4.1)$$

with  $t_m, m \in [1; M]$  the visual features (GLH and QWF) and  $c_n, n \in [1; N]$  the clinical attributes (see Section 3.2). Using all 72 clinical attributes, the maximum dimensionality of the multimodal feature space reaches 119 with 56 continuous and 63 binary features.  $\mathbf{v}$  is used as input of SVMs which directly output the predicted class using *one versus one* multiclass approach.

#### Late fusion: multiple classifier systems

Two SVMs classifiers  $q_t$  and  $q_c$  are trained using visual features  $\mathbf{t}$  and clinical attributes  $\mathbf{c}$  respectively. Attributes in  $\mathbf{t}$  and  $\mathbf{c}$  are normalized (within each group) in order to give equal importance to each of them. Both SVMs deliver probabilities  $p_t$  and  $p_c$  using pairwise coupling [276]. For each class  $w_j$ , probabilities are multiplied to compute final probability  $p(w_j)$

$$p(w_j) = p_t(w_j) \cdot p_c(w_j). \quad (4.2)$$

The final predicted class  $w_j$  is given by  $\underset{j}{\operatorname{argmax}} p(w_j)$ . Using the product of probabilities for predicting the final class assumes that the modalities  $\mathbf{t}$  and  $\mathbf{c}$  are conditionally statistically independent [15, 115].

Other combination strategies of the classifier's outputs are found in the literature. The mean of the probabilities has been used although it does not make sense in theory of statistics. A majority vote was also proposed several times, but it is not applicable to a combination of two classifiers and special cases need to be treated when the number of modalities  $N_{mod}$  is even. Using the product of probabilities has the advantage of attributing more importance to the classifiers that have higher probability, which favors classifiers that actually have more confidence. The late fusion scheme is summarized in Figure 4.9.

### 4.2.4 Evaluation: 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

---

<sup>3</sup><http://www.diee.unica.it/mcs/>, as of 5 November 2009

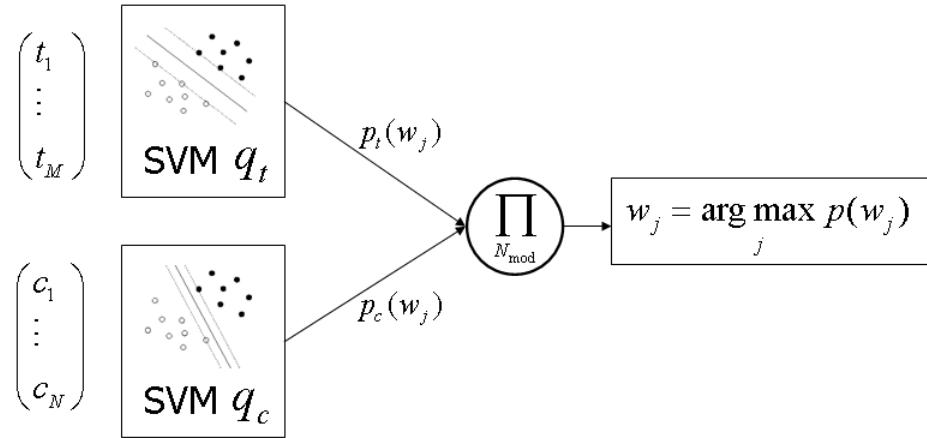


Figure 4.9: Classification scheme for late fusion. Two expert SVMs classifiers  $q_t$  and  $q_c$  output probabilities  $p_{\text{mod}}(w_j)$  which are multiplied to obtain the final probability of each class  $w_j$ .

shown to be effective to categorize lung tissue patterns from visual features in Section 4.1 and are adapted to mine clinical parameters as shown in [42]. Two methods for combining visual and clinical attributes are compared: early versus late fusion. Comparisons can be found in the literature for emotion recognition from face and body gestural in [84] and for video retrieval where visual and textual information are fused in [230].

### Principal component analysis

The effect of reducing the dimensionality of feature spaces using principal component analysis (PCA) is investigated in this section. First, PCA is applied only on the clinical feature set  $c_n$  to be used with combined SVMs. Second, PCA is applied on the whole concatenated feature set. The number of principal components  $N_{\text{PCA}}$  kept is chosen according to [182]:

$$N_{\text{PCA}} > 1 + 2\sqrt{\frac{N_{\text{features}} - 1}{N_{\text{instances}} - 1}} \quad (4.3)$$

with  $N_{\text{features}}$  the number of features and  $N_{\text{instances}}$  the number of instances. In both cases,  $N_{\text{PCA}} = 2$  is chosen based on (4.3).

### Material and methods

736 ROIs from healthy and four pathological lung tissue patterns belonging to 48 patients with filled clinical parameters were selected for this study. The dataset used for the comparison is identical to the one used in Section 3.2 to build the multimodal feature space and is detailed in Table 3.9. 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 and PCA transform is taken from the open source Java library *Weka*<sup>4</sup> using a wrapper for *LIBSVM*<sup>5</sup>. The image feature extraction and the optimization of SVMs is implemented in Java. Quincunx Wavelet frames are implemented in Java [255].

### Validation

In order to test the influence of clinical parameters on classification accuracy of the 2D ROIs, a LOPO cross-validation was used. The training set is used both for grid search for optimal param-

<sup>4</sup><http://www.cs.waikato.ac.nz/ml/weka/>, as of 5 November 2009

<sup>5</sup><http://www.csie.ntu.edu.tw/~cjlin/libsvm/>, as of 5 November 2009

Table 4.5: Averaged accuracies obtained with the various techniques. Best performances are highlighted in bold.

	visual features	clinical features	concatenated features (early fusion)	combined SVMs (late fusion)	combined SVMs, PCA on clinical features	PCA on concatenated features
<i>healthy</i>	0.46	0.01	0.19	0.43	<b>0.48</b>	0.22
<i>emphysema</i>	0.64	0.08	0.26	0.43	0.47	<b>0.78</b>
<i>ground glass</i>	0.57	0.56	0.6	<b>0.71</b>	0.62	0.5
<i>fibrosis</i>	0.91	0.77	<b>0.95</b>	<b>0.95</b>	<b>0.95</b>	0.87
<i>micronodules</i>	0.68	0.8	<b>0.86</b>	0.83	0.45	0.21
global	0.74	0.61	0.74	<b>0.79</b>	0.72	0.58

eters and adjustment of the maximum–margin hyperplane of the SVMs. Optimized parameters of the SVMs are the cost of the errors  $C$  and the width  $\sigma_K$  of the Gaussian kernel. A grid search is carried out within the intervals  $C \in [1; 100]$  and  $\sigma_K \in [10^{-2}; 10^2]$ . For every coordinate of the grid, a 10-fold CV is carried out on the training set. Optimal parameters  $(C^{opt}, \sigma_K^{opt})$  that allowed best mean CV accuracy  $A^{cv}$  are used to train the final model on the entire training set.

In order to study the optimal number  $N_{clin}$  of clinical attributes to be used, mean classification accuracies over the 48 patients are computed for each  $N_{clin} \in [1; 72]$ , with clinical attributes ordered by  $A^{single}$  values (see Section 3.2.3). Mean classification accuracies  $A_{mean}^{test}$  according to  $N_{clin}$  obtained with the test set using visual features only, clinical features only and combined features with early and late fusion are shown in Figures 4.10, 4.11, 4.12, 4.13, 4.14 for each class. Global accuracies of each method are summarized in Figure 4.15. Mean accuracies over  $N_{clin}$  values as well as classification based on PCA are contained in Table 4.5.

#### 4.2.5 Discussion on the fusion schemes

Influence of the clinical context of HRCT images on lung tissue classification accuracy is studied in Figures 4.10, 4.11, 4.12, 4.13, 4.14 and 4.15. As baseline performance, the accuracy achieved by using visual features only is considered, which has a mean global value of 74%. The last line of Table 4.5 shows that integrating the clinical context of the images allows for significant global improvements of the classification accuracies. On average, 5% is gained (absolute gain) in global accuracy using combined SVMs compared to using visual features only. Classification accuracies of combined modalities (all combination strategies taken together) are always superior to single modalities as observed in Table 4.5.

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. 4.14 for the class *micronodules* and for low  $N_{clin}$  values in Fig. 4.10 for class *healthy*. However, concatenating all features in a single vector has the drawback that less informative attributes of one modality scatter homogeneous clusters of instances in the feature space of the other modality. Figures 4.10, 4.11 and 4.15 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 them as well as with the visual ones, which partly explains the irregular shape of the curves. Separating visual and clinical features for mining using late fusion avoid interactions between the feature groups which show more stable performance compared to early fusion. This is particularly true when clinical attributes carry little information as it is the case for classes *healthy*. The conditional statistical independence required for using the product of probabilities in Eq. (4.2) is admissible as the mean correlation value  $\rho_{mean}$  of each feature pair  $(t_m, c_n)$  is equal to 0.0143 (see Figure 3.18). The late fusion scheme divides the multimodal feature space into smaller feature subspaces that have the advantage to reduce the computational complexity for solving the quadratic problem of finding the maximum margin hyperplane of SVMs in Eq. 1.12. Moreover the subspaces can be processed

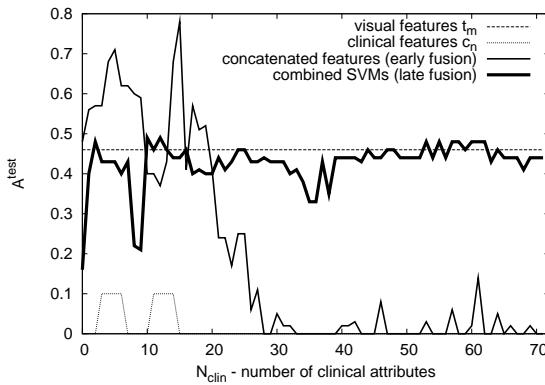
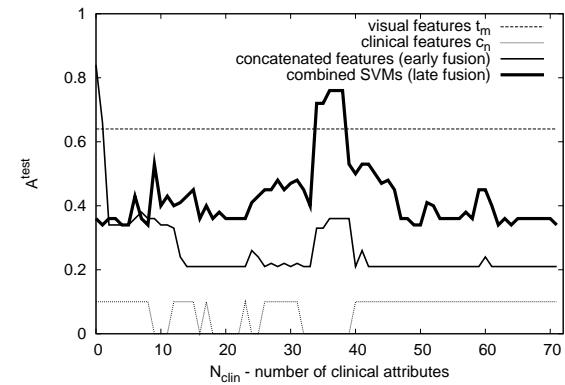
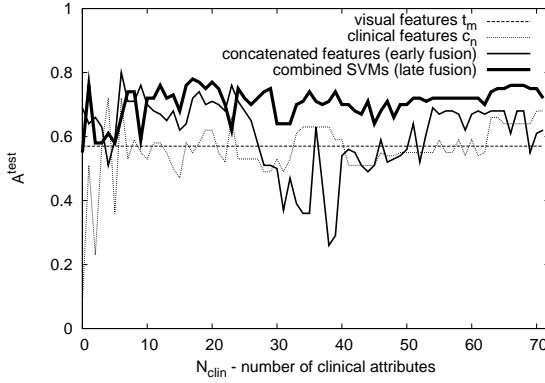
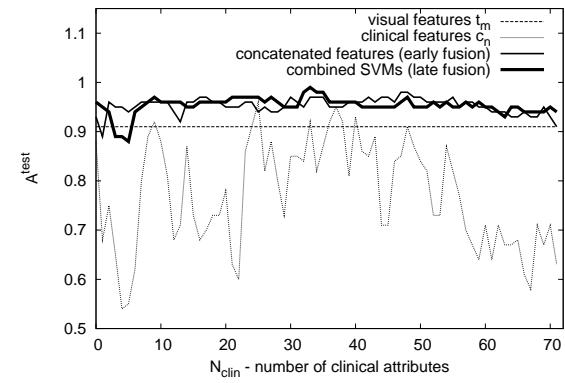
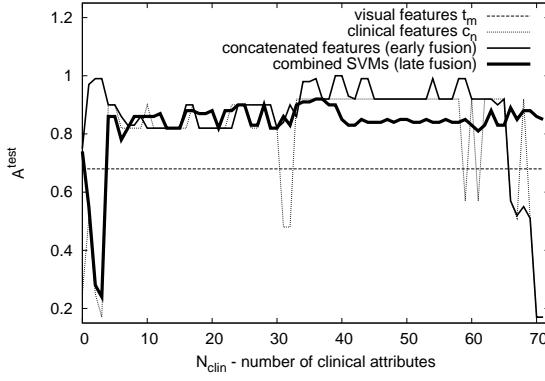
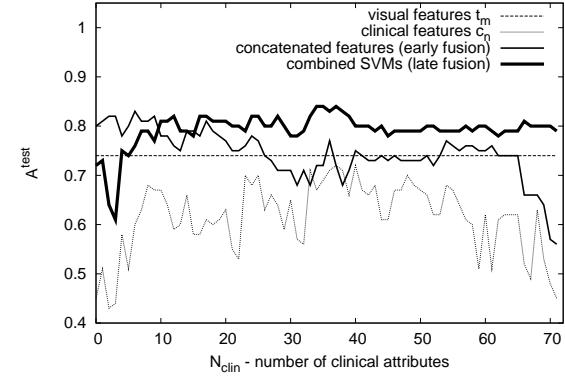
Figure 4.10: *healthy*Figure 4.11: *emphysema*Figure 4.12: *ground glass*Figure 4.13: *fibrosis*Figure 4.14: *micronodules*Figure 4.15: *multiclass*

Figure 4.16: Testing accuracies  $A_{\text{mean}}^{\text{test}}$  with a varying number of clinical parameters  $N_{\text{clin}}$  for the diverse techniques.

in parallel to reduce the training time of the SVMs. The combined SVMs (late fusion) show high robustness towards the number  $N_{clin}$  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  $N_{clin} = 35$ . Although the PCA transform allows best results for classes where clinical attributes have low discriminatory power (see Table 4.5), it does not improve the global accuracy.

A limitation of this study occurs as some of the ROIs in the class *healthy* were delineated in patients that have had an ILD in their medical history, leading to abnormal clinical parameters. This partly explains the very low accuracies obtained for *healthy* patterns when the classification is based on clinical parameters. Further validation with true healthy control subjects has to be carried out to assess the benefits of using the clinical parameters for the classification of *healthy* lung tissue.

The substitution with the average value for the clinical parameters that had an *unknown* value in Section 3.2.1 is questionable. Indeed, the ideal solution would be to teach the SVMs classifier that the value is unknown and that the feature should not be used for the classification. This is implemented by placing the feature value right on the decision boundary. By substituting *unknown* by the mean, the instance is positioned somewhere in between the classes, which may not be on the decision boundary, thus having an influence on the classification outcome.

At last, we believe that negative synergies still occur among features in the subspaces using a late fusion scheme. Groups of features with positive synergy [14], 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 of the performance according to  $N_{clin}$  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 [116, 117]. Alternatively, a visual approach based on Bayesian networks is proposed in [38]. GLH and texture features from GLCM, RLE and grey-level differences are divided into uncorrelated groups using Pearson's coefficient of correlation  $R^2$  for the categorization of COPD in [191]. In this study, the visual and the clinical features have very low mean correlation  $\rho_{mean} = 0.0143$  (see Section 3.2.3). The correlation among the features can be visualized in the correlation matrix of the multimodal feature space in Figure 3.18. At last, identifying feature groups with high consistency should also include medical knowledge.

### 4.3 Case-based retrieval

Searching for external information to fill in the lack of knowledge and experience towards a problem to solve corresponds to the daily routine of the majority of the physicians. The most common external sources of information used are:

- the Internet,
- literature databases accessed through web portals such as PubMed<sup>6</sup>, BioMed Central<sup>7</sup> or UpToDate<sup>8</sup>,
- textbooks owned by the department,
- personal collections of cases.

The Internet is a very vast knowledge base, but the quality of the information is questionable. Standardized quality control was proposed by the health on the net foundation (HON) by introducing certification codes that are visible on the web pages (HONcode<sup>9</sup>). However, searching for information on the Internet, scientific literature or textbooks can be time-consuming and often

---

<sup>6</sup><http://www.ncbi.nlm.nih.gov/sites/entrez?db=pubmed>, as of 5 November 2009

<sup>7</sup><http://www.biomedcentral.com/>, as of 6 November 2009

<sup>8</sup><http://www.utdol.com/home/index.html>, as of 5 November 2009

<sup>9</sup>Health On the Net, <http://www.hon.ch/>, as of 5 November 2009

yields theoretical solutions that are not directly applicable to a given medical problem. Problem-based learning is complementary to the theoretical knowledge and provides practical guidelines. The latter was successfully applied to the medical domain [120, 228]. In order to store and organize the practical experience gained with solved problems, physicians and radiologists often create their own collections of cases [165]. The latter can be digital and several initiatives such as Casimage<sup>10</sup>, MedTing<sup>11</sup>, MyPACS<sup>12</sup> or MIRC<sup>13</sup> offer disk storage and software tools to organize, anonymize and share collection of cases. These multimedia collection of cases can be used for teaching enabling the possibility to gain practical experience and to compare the patient under investigations to cases with verified diagnoses and follow-up [206].

Most often, the access to these collections of cases is text-based where the user can search by diagnosis, organ, modality, age, etc... This is often of limited use to the radiologists who are actually looking for similar images and may want to retrieve cases with various diagnoses to discover a potential match with the patient under investigation. Assessing visual similarity is thus required and CBIR was proposed several times [145, 170, 213, 220, 272]. However, radiologists are most often looking for similar cases as they are considering the image within the context of a patient with a personal history, findings on the physical examination, laboratory tests, etc.

Thereby, a possible extension to CBIR is to carry out case-based retrieval based on visual and clinical similarity. Moreover, in terms of retrieval performance, CBIR based on the visual data alone can only achieve a limited retrieval quality when used for diagnosis aid [166]. The case has to be taken as a whole with visual information from images and other clinical data that contain complementary information required in the process.

In this section, three-dimensional case-based retrieval is proposed to retrieve similar cases from the database described in Chapter 2.

### 4.3.1 3D multimodal retrieval

Case-based retrieval is enabled by the automated categorization of the lung tissue described in Section 3.1.3. The three-dimensional map of the lung tissue obtained with the blockwise classification of the lung regions yields a semantically-related basis for the comparison of the cases. As a first approach, the percentages of the respective volumes  $v_i$  of the five classes of lung tissue are used to assess the visual similarity between HRCT image series from two patients. The respective volumes of lung tissue are semantically related to the ILDs as each histological diagnosis is associated to a given combination of HRCT findings. This allows to reduce the semantic gap between the user's intentions and the visual features, which is often a bottleneck in CBIR [226]. The Euclidean distance is computed from the percentages of the five volumes of tissue as follows:

$$d_{vol} = \sqrt{v_h^2 + v_e^2 + v_g^2 + v_f^2 + v_m^2} \quad (4.4)$$

with  $v_h$  corresponding to *healthy* tissue,  $v_e$  to *emphysema*,  $v_g$  to *ground glass*,  $v_f$  to *fibrosis* to  $v_m$  for *micronodules*.

44 clinical parameters with two levels of importance are used to assess the “meta-similarity” between the cases. The levels of importance are defined by a physician according to the relevance for establishing the diagnosis of eight common ILDs. 3 clinical parameters of first importance include age, gender and smoking history. Another 41 parameters of second importance included physical findings, medical history, and laboratory results. The parameters associated with biopsy outcomes were not included as the goal of the CAD is to provide quick information to the radiologists before any biopsy.

The multimodal distance measure  $d_M$  between two cases is computed as a linear combination of three modalities:

$$d_M = a_1 d_{vol} + a_2 d_{param1} + a_3 d_{param2}, \quad (4.5)$$

<sup>10</sup><http://pubimage.hcuge.ch/>, as of 5 November 2009

<sup>11</sup><http://medting.com/>, as of 5 November 2009

<sup>12</sup><http://www.mypacs.net/>, as of 5 November 2009

<sup>13</sup>Medical Imaging Resource Center, <http://mirc.rsna.org/>, as of 5 November 2009

Table 4.6: Mean precisions based on the diagnosis of the retrieved cases. The values of the weight  $a_i$  that allowed best global precisions show the respective importances of the modalities.

	$P@1$	$P@5$	$P@10$	$P@N_r$	$N_r$
PF	79.2	58.3	51.7	42.7	24
COP/BOOP	60	20	18	20	5
TB	71.4	48.6	34.3	42.9	7
PCP	25	20	10	25	4
HP	54.5	40	39.1	38	11
AIP	66.7	33.3	25.5	27.2	9
Sarcoidosis	100	66.6	52.2	56.8	9
average/total weights $a_{1,2,3}$ in (4.5)	59.4 8;1;39	39.7 6;9;38	34.2 8;5;48	32.4 10;4;41	69

with  $a_j$  being the weights of each modality.  $d_{vol}$  is the Euclidean distance in terms of percentages of the volumes of segmented tissue according to (4.4) and  $d_{param1,2}$  the euclidean distance in terms of clinical parameters of importance 1 and 2, respectively.  $d_{vol}$  and  $d_{param1,2}$  are normalized before being combined in (4.5).

### 4.3.2 Evaluation

To assess case-based retrieval performance, mean precisions at rank 1, 5, 10 and at rank equal to the number of instances of the diagnosis  $N_r$  are computed using a LOPO cross-validation with 69 cases (see Table 4.6). A grid search for optimal weights of the modalities in (4.5) is carried out with  $a_j \in [0 : 50[$ .

### 4.3.3 Discussions on the retrieval performance

Visual information retrieval in 3D data sets has rarely been reported in the medical domain, and even less when based on textures whereas it does exist for non-medical shape-based retrieval [238]. The retrieval precisions presented in Table 4.6 are currently fairly low to be used in clinical routine but show the feasibility of indexing ILD cases using the volumes of automatically segmented lung tissue as well as clinical parameters. It is important to note that the link between visual similarity of two HRCT scans and their associated diagnoses is not straightforward. The values of the weights  $a_{1,2,3}$  that allowed best performance reflect the importance of the each modality. High values obtained for  $a_3$  shows the unexpectedly high importance of the clinical parameters of secondary priority. High variations of the precision can be explained by the fact that the number of cases is still fairly small, particularly for COP/BOOP and PCP. Further developments are required to improve the assessment of the visual similarity: local values of the low-level feature vector  $\vec{v}$  containing the QWF and GLH features can be used directly as a modality in (4.5), under the condition to overcome the difficulty in setting up a standardized localization system for the lung anatomy. This was proposed by Zrimc *et al.* in [291] where a 3D model of the human lung with lung region characterization is robustly built. A simple model was also proposed in [144] for CBIR. Learning to rank was also proposed in the literature where the distance measure is learned using a set of training examples. It allowed to improve the retrieval performance of text-based queries on the Internet in [24]. To improve the retrieval performance of the ILD cases, SVMs regression can be used to learn the similarity measure. A higher number of cases is required, though.

# Chapter 5

# Applications and evaluation

In the previous chapters, methods for feature extraction, classification and retrieval were proposed and evaluated on a multimedia dataset built from clinical routine at the HUG. Although constituting the foundations of the CAD system for ILDs, these techniques are not directly usable in clinical routine and efforts are required to successfully bring them to the end users. Ergonomic interfaces and realistic evaluation settings must be put in place in order to take benefit of the full potential of the proposed methods.

In the first part of this chapter, use cases are defined to serve as a basis to develop a rich Internet application (RIA) that implements the methods previously proposed for lung tissue categorization and case-based retrieval. In the second part, the evaluation strategies are detailed and revisited to ensure their accordance to real clinical settings.

## 5.1 Use cases, visualization and software

Up until recently, film-based radiological images allowed to visualize organs only via 2D projections of three-dimensional organs. Nowadays, the digital form of medical visual information enables advanced visualization techniques. Three-dimensional colored objects with motion correspond to our visual perception of the world. Thereby, several DICOM viewers integrated tools for representing four-dimensional color visual information [212]. A notable example is the open source software OsiriX<sup>1</sup> [207] based on the visualization toolkit<sup>2</sup> (VTK) [216]. Attempts using virtual reality was proposed in [82]. Transfer functions mapping density values to colors allow realistic rendering of bone structures and organs (see Figure 5.1). The visualization tools aim at representing the most information at once in order to reduce the reading time and to assess global views of organs synthesizing visual information to ease the diagnosis workup.

The definition of use cases adapted to the needs of the radiologists in a particular context is also an important clue for the success of the CAD. In this thesis, three use cases are defined to optimize diagnostic aid for ILDs:

- navigation through the multimedia database records for teaching and learning purposes,
- automated 3D categorization of the lung tissue in HRCT image series,
- case-based retrieval.

The second and the third use cases are strongly interdependent, yielding a hybrid detection-CBIR-based CAD system (see Section 4.3). The three use cases were implemented in a RIA based on Adobe FLEX<sup>3</sup> and Java Applet<sup>4</sup> using Java3D<sup>5</sup>. The goal was to develop a fully web-based application, so it can be used by either students or clinicians and from various places inside

<sup>1</sup><http://www.osirix-viewer.com/>, as of 5 November 2009

<sup>2</sup><http://www.vtk.org/>, as of 5 November 2009

<sup>3</sup><http://www.adobe.com/fr/products/flex/>, as of 5 November 2009

<sup>4</sup><http://java.sun.com/applets/>, as of 5 November 2009

<sup>5</sup><http://java3d.j3d.org/download.html>, as of 5 November 2009

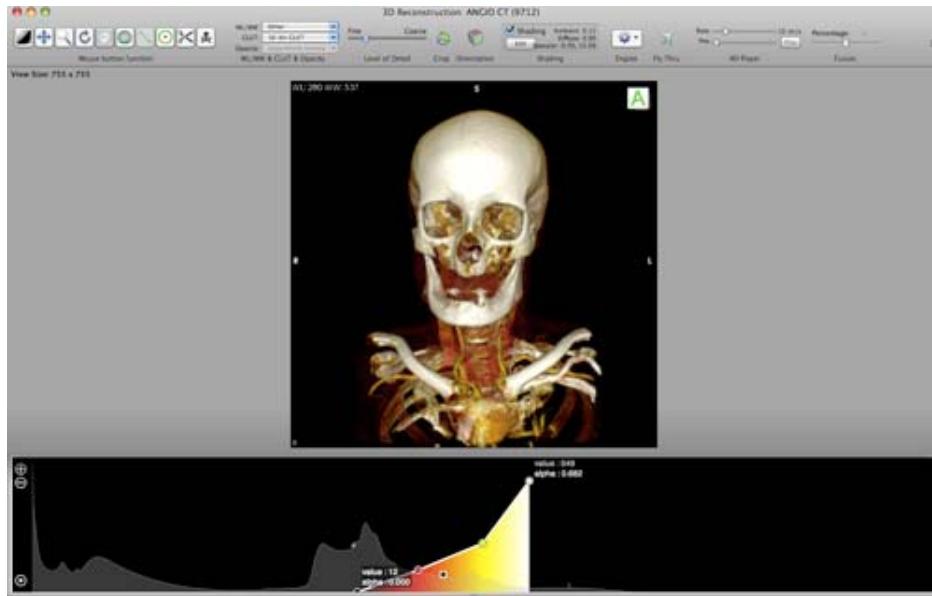


Figure 5.1: CT scan of the head: a transfer function maps HU values to colors for realistic rendering of the bone structures and organs (source: <http://www.osiriximaging.com/2008/12/how-to-create-a-fly-through-in-osirix-written-by-esther-pulley/>, as of 17 November 2009).

the hospitals. All software to be installed on clinical desktops at the HUG has to go through a tedious administrative task. As soon as a Java virtual machine and Java3D are installed, the RIA application can be accessed directly by typing the corresponding uniform resource locator (URL) in a web browser. Once logged in, the user can choose between browsing the multimedia database using exact text search and analyzing a new case along with retrieving similar patients (see Figure 5.2).

This section is organized as follows. The use case for browsing the database is detailed in Section 5.1.1. The workflow for the analysis of lung tissue of a new case combined with the case-based retrieval is described in Section 5.1.2.

### 5.1.1 Multimedia database browsing

The workflow for the navigation through the database records is depicted in Figure 5.3. This use case is mostly intended for teaching where students, interns and little experienced radiologists can consult the typical clinical and radiological manifestations of ILD diagnoses but also realize their variabilities through real cases. This use case is recommended to be used along with textbooks and literature such as [114, 268] that describe the fundamental mechanisms of the studied histological diagnosis.

The user can perform a text-based search by defining the target value of each clinical parameter as shown in Figure 5.4. *And* and *Or* combination rules are available to favor either precision or recall of the retrieved information. Based on the query, an exact text search is performed. The retrieved patients are listed in a dynamic page that allows to preview the clinical parameters organized in three levels of importance (see Figure 5.5). The importance level of each clinical parameter vary based on the diagnosis of the case in order to show the most relevant parameters in the second level. A pop-up window showing a preview of the HRCT image series can be obtained by clicking on the miniature image displayed on the right side of each case. As soon as the user wants to access full case data based on the previewed information, a “Detail” button can be clicked to open a new page that displays every available clinical parameter using a tree-like structure (see

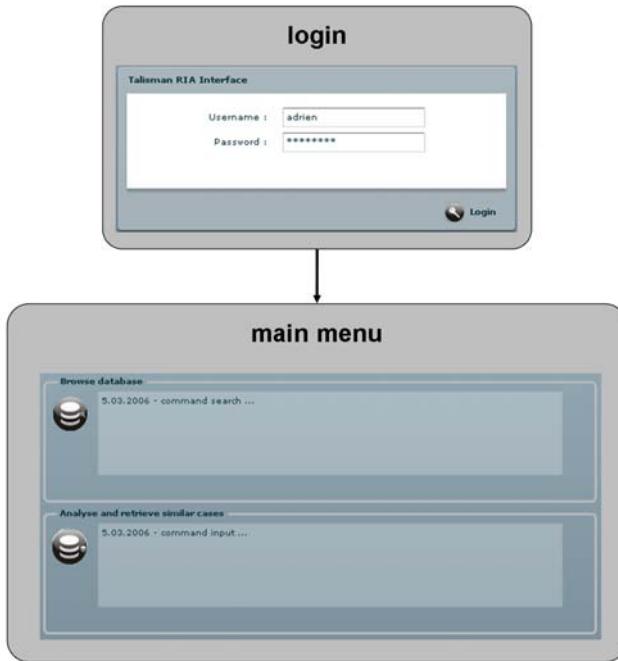


Figure 5.2: Welcome menus of the RIA. Once logged in, the user can choose between navigating through database records (see Figure 5.3) and analyzing a new case along with retrieving similar patients (Figure 5.9).

Figure 5.6). On the right panel of this page, the HRCT image series are listed by the date of the study. By double-clicking on an image series, the YaDiV Java Applet is launched and opens the requested image series along with the associated annotated regions (see Figure 5.7). 2D and 3D views are available to navigate through the image series and visualize the delineated pathological lung tissues. The user can switch between clinical parameters and HRCT images by clicking either on the “Patient data” tab or the “YaDiV” tab. In order to perform a new search or to return to the main menu, the button showing a backward arrow on the top left of each page can be used.

### 5.1.2 3D lung tissue categorization and case-based retrieval

The second use case implements the analysis of the image series of a new case and case-based retrieval. The main components of the hybrid approach are detailed in Figure 5.8. In a first step, the clinician can run the three-dimensional categorization of the lung tissue on the undiagnosed incoming HRCT image series in order to obtain a 3D map of the lung tissue. On the one hand, this map highlights diagnostically useful events in the image series and thus reduce the risk of omission of important lesions of the lung tissue. On the other hand, it provides first insights of the potential histological diagnosis of the patient under investigation by providing the respective volumes of five lung tissue sorts<sup>6</sup> that are associated with most of the ILDs. In a second step, the clinician can retrieve similar cases from the multimedia database based both on the respective volumes of the previously segmented lung tissue sorts and on the value of the clinical parameters (see Section 4.3). The range of the diagnoses of the retrieved cases provides information to the clinician with detailed examples. The workflow for the categorization of the lung tissue and case-based retrieval is detailed in Figure 5.9. Starting from the page for submitting the query, the user can either categorize lung tissue and fill clinical parameters to retrieve similar cases based both on visual features and text or skip the image analysis step to retrieve cases based only on the clinical

<sup>6</sup>healthy, emphysema, ground glass, fibrosis and micronodules

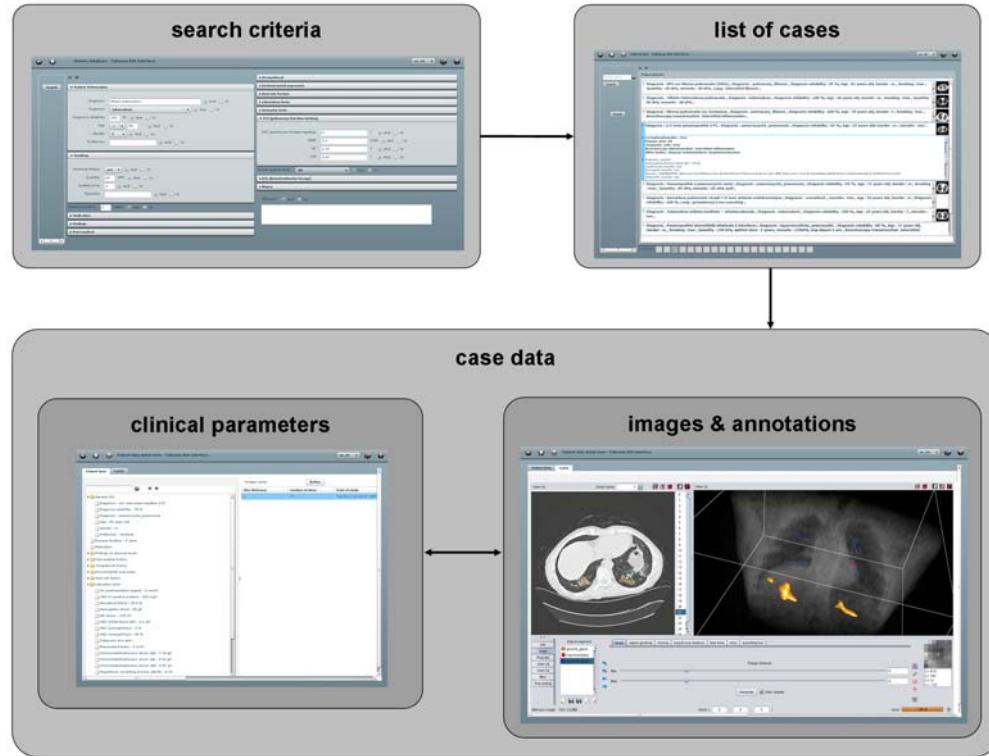


Figure 5.3: Workflow of the RIA for the navigation in the database using exact text search.

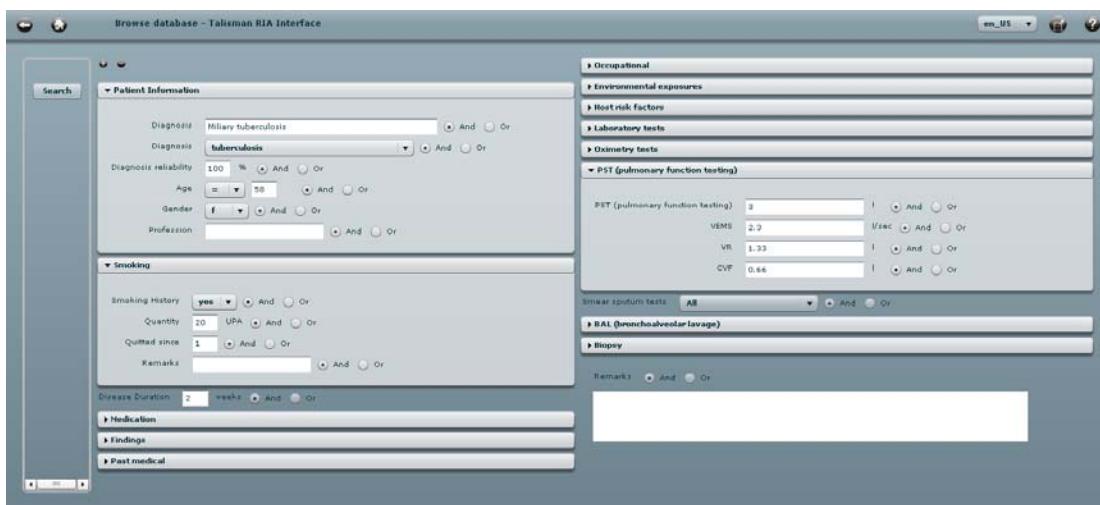


Figure 5.4: Form containing the clinical parameters for querying ILD cases using exact text search. *And* and *Or* combination rules are available.

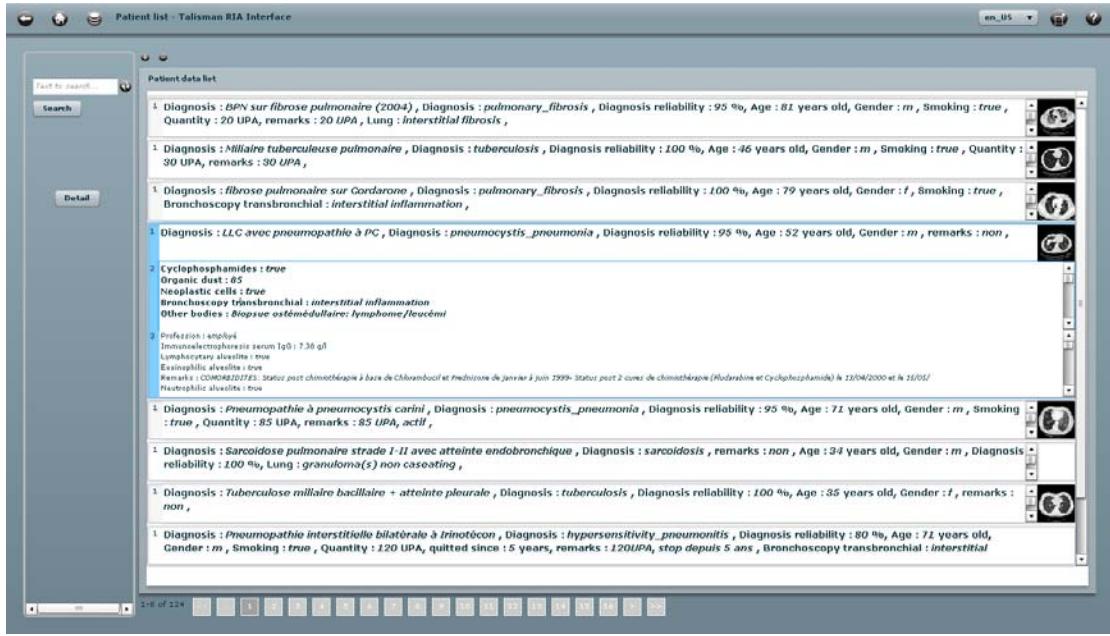


Figure 5.5: List of cases returned by an exact test search. The user can preview the clinical parameters of the patients with three levels of importance. Levels of importance are varying based on the diagnosis.

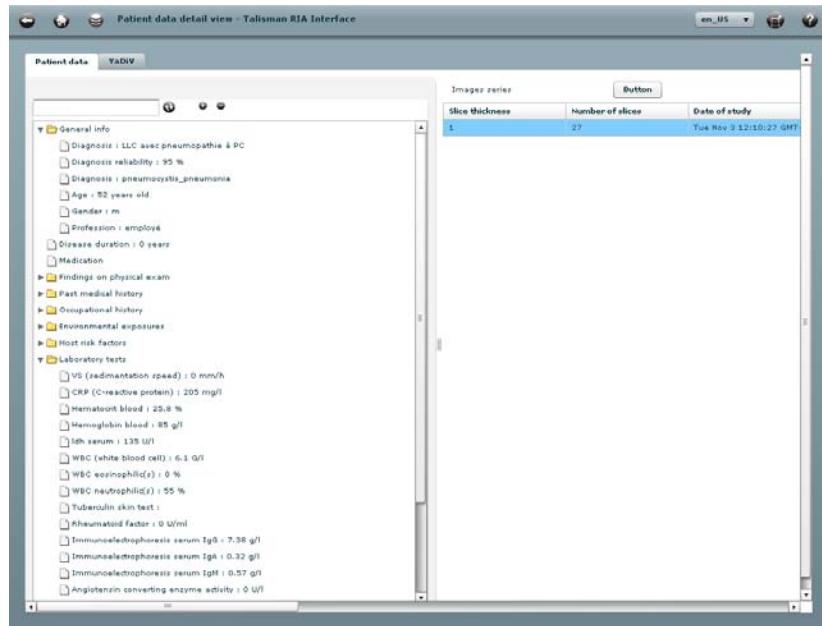


Figure 5.6: Detailed view of a selected case: clinical parameters. This page is reached from the button “Detail” of Figure 5.5.

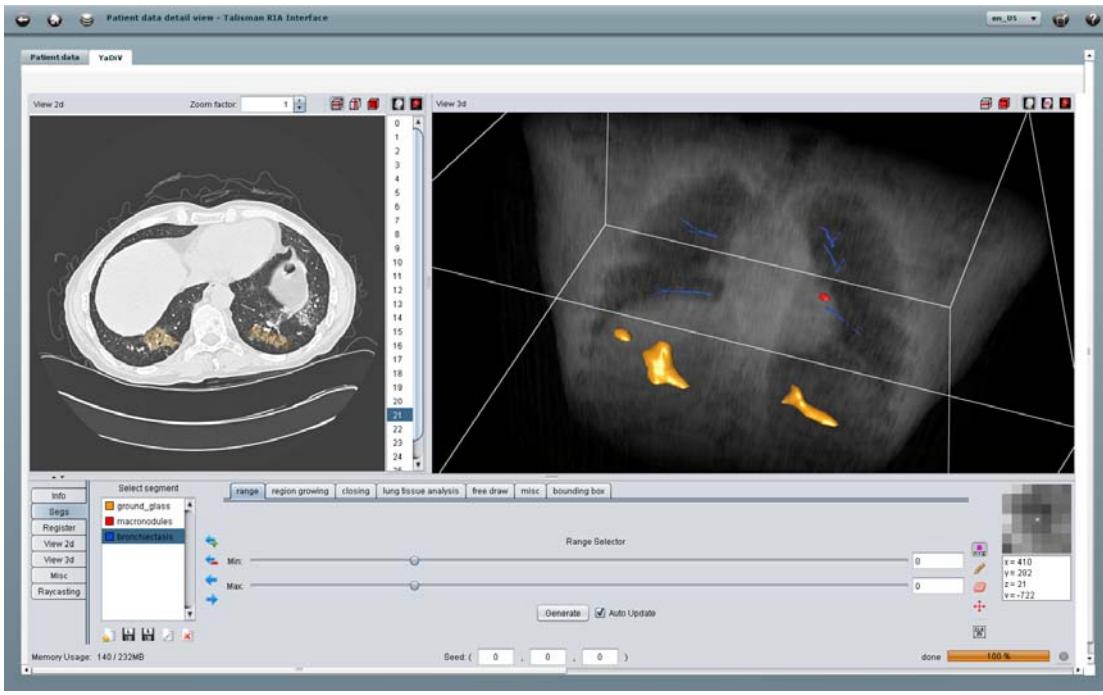


Figure 5.7: Detailed view of a selected case: image series and annotations. 2D and 3D views are available to visualize the ROIs. YaDiV was adapted and embedded in a Java Applet.

parameters.

#### Similar case retrieval

Starting from the Flex page for the specification of the clinical parameters (top left image in Figure 5.9 and Figure 5.10) for similar case retrieval, the user has two choices:

- analysis of a new image series,
- similar case retrieval.

By clicking on the button showing a 3D map of lung tissue (see Figure 5.10), the YaDiV Java Applet is launched and the user can perform the 3D categorization of the lung tissue by opening an HRCT image series, segmenting the lung volumes and running the blockwise classification. The user can submit the query for the retrieval of similar cases by clicking on the button showing a pie chart (see Figure 5.10). If an image series was previously categorized, the query will be based both on visual and clinical features. Alternately, the retrieval will be based on the clinical parameters only.

The ranked list of retrieved cases shown in Figure 5.11 contains the same functionalities as the list of cases used in the use case for browsing the database in Figure 5.5. A pie chart was added to display the similarity measure in percent for each case. Detailed views of the selected case identical to the one used in the “browse” use case (see Figures 5.6 and 5.7) can be obtained by clicking on the “Detail” button.

#### 3D categorization of lung tissue

The use case described in this section implements the categorization of the lung tissue detailed in Section 3.1.3. It composed of the segmentation of the lung volumes in a first step, and of the blockwise classification of the lung tissue in a second step.

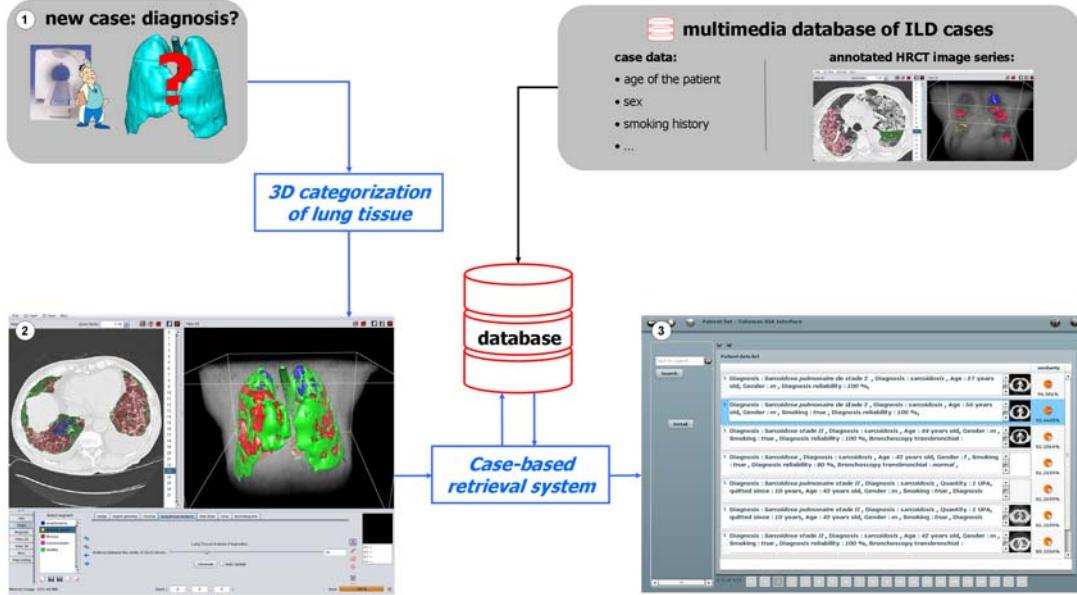


Figure 5.8: Main components of the hybrid (detection- and CBIR-based) CAD system.

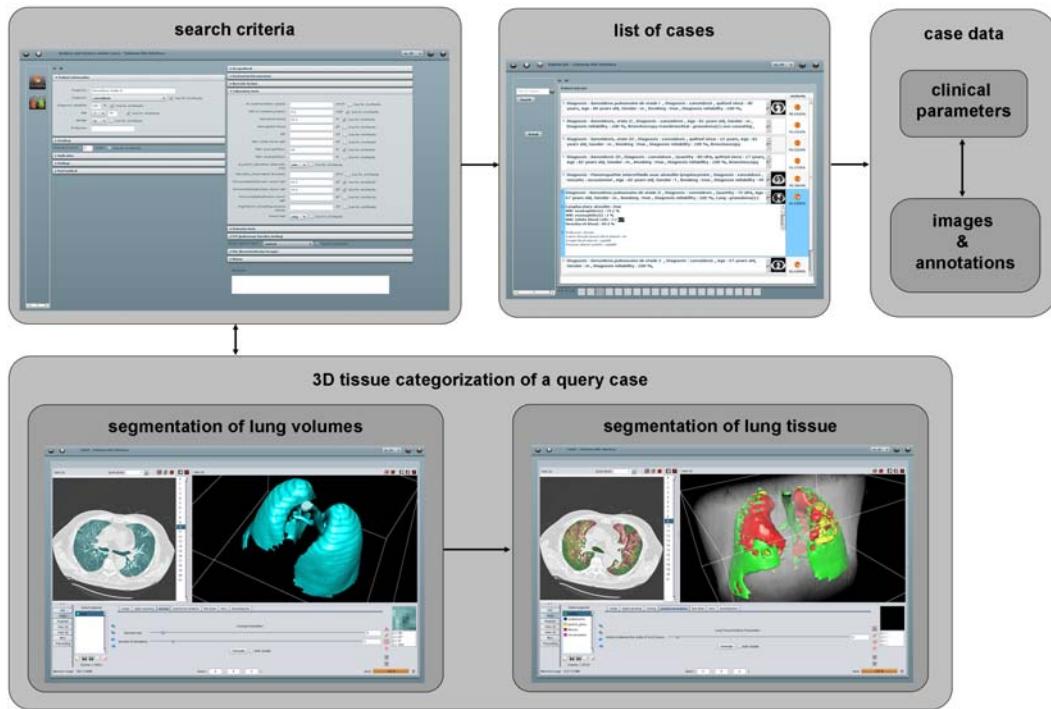


Figure 5.9: Workflow of the RIA for the lung tissue categorization of a new case and the retrieval of similar patients. The query can be based both on specified clinical parameters and on the volumes of the segmented tissue sorts.

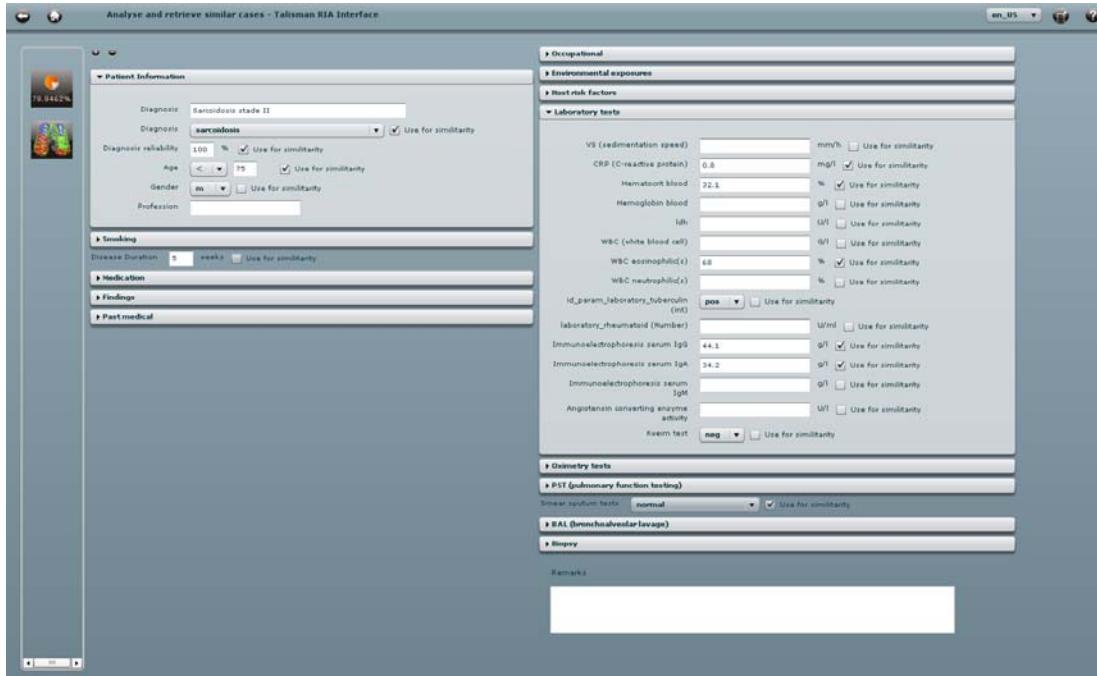


Figure 5.10: Selection of the clinical parameters for case-based retrieval. The multimodal Euclidean distance described in Section 4.3.1 is used (see Eq. 4.5). The visual modality is included only if a query image series was previously analyzed.

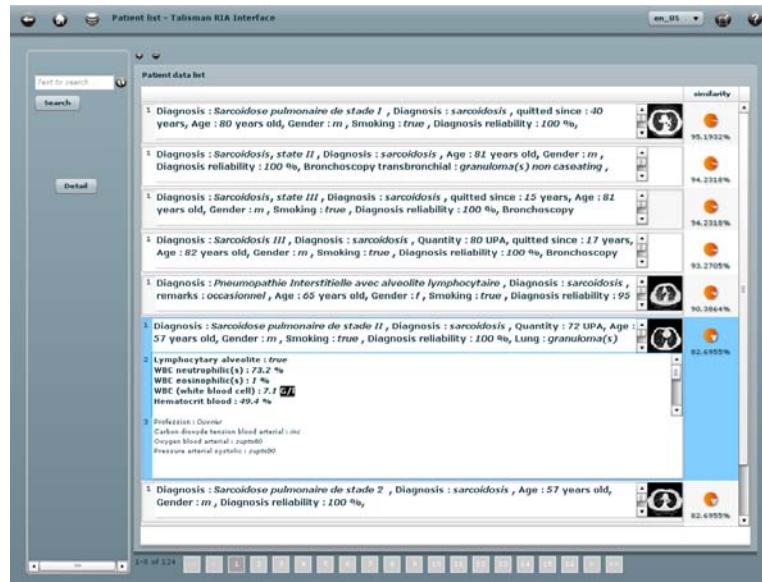


Figure 5.11: List of cases returned by the multimodal query. The similarity measure is displayed using pie charts. Similarly to the “browse” use case, detailed views of a selected case are available by clicking the “Detail” button to access pages depicted in Figures 5.6 and 5.7.

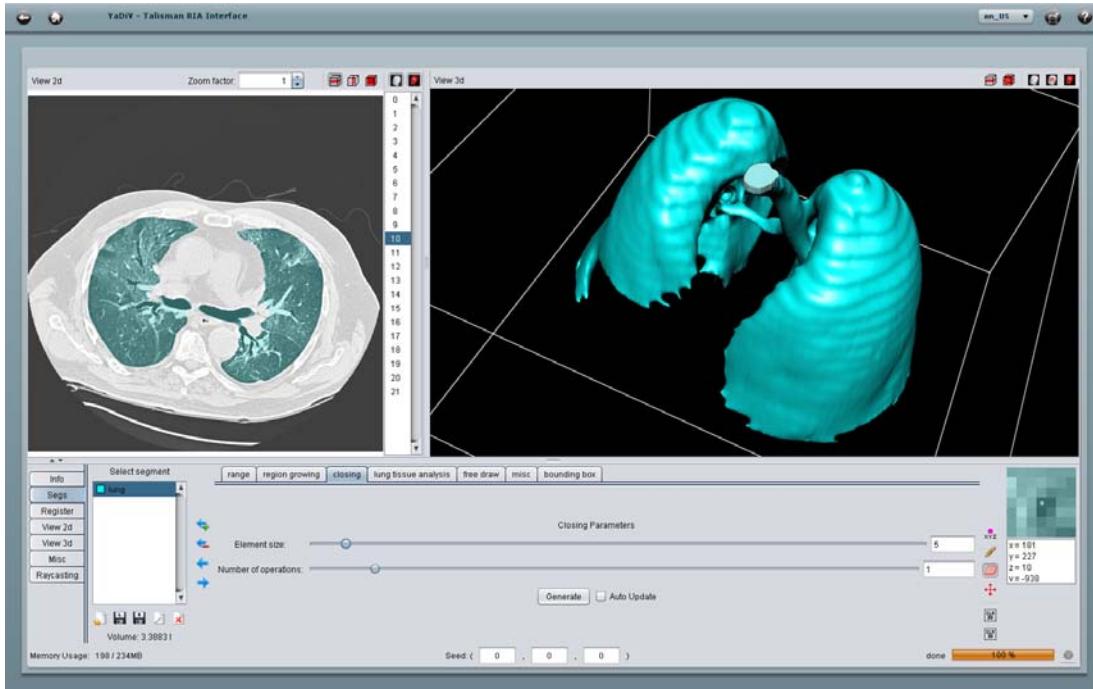


Figure 5.12: Segmentation of the lung volumes using YaDiV. A “closing” tab was added to YaDiV to perform a closing operation after the region growing.

**Segmentation of lung volumes** The segmentation of the lung volumes is a prerequisite step for the categorization of the lung tissue. It consists of a 3D region growing followed by a closing operation. More details can be found in Section 3.1.3. The 3D region growing routines are already available in the original distribution of YaDiV. A tab was added for the closing operation (see Figure 5.12). When required, the user can manually edit the lung mask. The lung mask can be visualized in 2D and 3D to verify the completeness of the segmented volume.

**Blockwise classification of the lung tissue** As soon as the user is satisfied with the segmentation of the lung volumes, the routines for the categorization of the lung tissue can be run from the “lung tissue analysis” tab that was added to YaDiV. The distance separating the centers of  $32 \times 32$  blocks can be specified. The “Generate” button runs the feature extraction and classification of each block belonging to the lung mask. Thanks to the wavelet frame transform, the values of the wavelet coefficients can be accessed through every scale using identical  $(x, y, z)$  coordinates (see Section 3.1.3 and Figure 3.12). For each block, the corresponding feature vector is classified using previously trained SVMs. The result of the categorization is displayed using a three-dimensional map of the lung tissue that can be visualized using the 2D and 3D views of YaDiV as depicted in Figure 5.13. The volumes in liters of each tissue sort can be obtained by clicking on the corresponding segment.

## 5.2 Evaluation

Although almost all CAD systems described in Section 1.3.2 provide an estimation of their retrieval or classification efficiency, it remains very problematic to compare them objectively [167]. Indeed, the diversity of the addressed tasks and the database used for validation render the comparison of the various systems very difficult. Moreover, the validation methods used to assess the performance of the CADs proposed in the literature have also a large variability over the studies and the performance measures are strongly related to the latter.

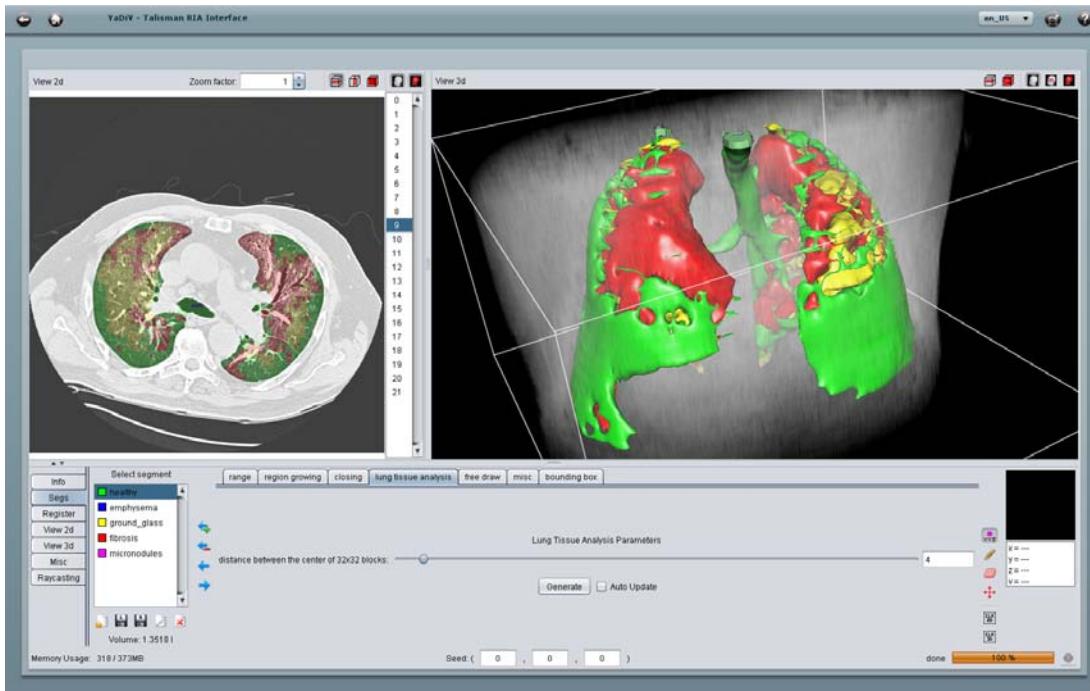


Figure 5.13: Segmentation of the lung tissue using YaDiV. A “lung tissue analysis” tab was created to run the blockwise feature extraction and classification. The distance between the centers of the blocks can be tuned.

In this section, the evaluation methods are discussed with both quantitative and qualitative considerations of the performance assessment. The need for simulating clinical situations as reliably as possible to ensure the success of the CAD system at the time it will be integrated in clinical routine and used by clinicians under time constraints is highlighted. This section is organized as follows. Section 5.2.1 lists and discusses the validation strategies used in this work and in the literature. Section 5.2.2 details the metrics used to assess the performance measures of image-based CAD systems. The Section 5.2.3 enumerates and discusses common causes of error and bias found in this work and in the state-of-the-art. At last, a discussion on qualitative considerations is given in Section 5.2.4.

### 5.2.1 Validation strategies

The validation of the classification and retrieval algorithms is a critical step to assess the performance of the proposed methods. The latter must be able simulate the conditions in which the system is intended to be used. In this work, the targeted usage is clearly the clinical routine and the validation methods must reproduce the clinical conditions where the diagnosis and the lung tissue patterns are predicted for one unknown incoming patient based on training information derived from ILD cases with verified diagnosis and annotated HRCT image series that constitutes the so-called “ground truth”. This ground truth is obtained from the multimedia database described in Chapter 2.

In machine learning, it is crucial to be able to evaluate the GP (or true risk) of the algorithms (Section 1.2.2). The GP cannot be directly computed because an infinite number of instances that form the clusters of classes would be required. However, the error obtained on a training set called “empirical risk” is known to provide an unbiased estimation of the GP [262]. The challenge then is to obtain an unbiased estimation of the empirical risk using an appropriate validation method. In this work, the latter is estimated using one of the four types of validations:

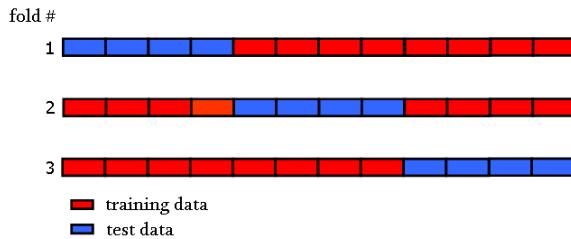


Figure 5.14: Cross-validation. In this example, 12 instances are split into 3 groups. Over  $N_{folds}$  folds, the instances of one group will serve  $N_{folds} - 1$  times for training and 1 time for testing.

- repeated random drawing to split the data into training and testing sets,
- $N_{fold}$ -fold cross-validation (CV),
- leave-one-ROI-out (LORO) CV,
- leave-one-patient-out (LOPO) CV.

Random drawing consists of splitting data into two parts and use one for training and the remaining instances for testing. The two parts can be unbalanced (e.g. 80% of the samples for training and 20% for testing). The performance measures are averaged over repeated drawings to obtain an estimation of the performance with limited bias. The other three validation strategies are based on CV and are detailed and discussed below.

### Cross-validation

The cross-validation (CV) consists of partitioning samples of data into  $N_{folds}$  complementary subsets (see Figure 5.14) called “folds”. Aiming at assessing the GP, one of the subset is left aside for further testing of the algorithms and the remaining subsets are used to train the learning algorithms. The former subset is called the “testing set” and simulates unknown instances and the latter is called the “training set”. To reduce variance of the performance measures, this experiment is repeated as many times as  $N_{folds}$  and the global performance measures are obtained by averaging the performances of each fold. A particular case of the cross-validation is the leave-one-out (LOO) where  $N_{folds}$  is equal to the number of instances  $N_{instances}$ . LOO is appropriate to evaluate the performance of the algorithm on relatively small datasets and is computationally intensive. CV showed to provide an estimation of the empirical risk with a tunable bias-variance trade-off with  $N_{folds}$  [121].

**Leave-one-ROI-out cross-validation** In the context of an image-based CAD system, the instances are constituted of ROIs (e.g. hand-drawn, square or circular blocks, entire image, ...). Thereby, a LOO CV consists of leaving aside one ROI for testing and use all the remaining ROIs for training (see Figure 5.15 (a)). This can be extremely computationally intensive as the number of models to train is equal to the number of instances. For example, the number of instances obtained by cutting all hand-drawn ROIs of the 5 most represented lung tissue patterns of the multimedia dataset described in Table 2.3 into  $32 \times 32$  blocks is superior to 15'000. Running the whole LORO CV requires to train more than 15'000 classifiers using very large training sets of  $N_{instances} - 1$ .

Another problem occurs when splitting the training and testing sets by ROIs. Most often in medical pattern recognition, the number of instances is larger than the number of patients as several ROIs are drawn in the same image series. It allows to assess the intra-patient variance of the classes. However, the LORO CV permits to train and test with ROIs that belong the same patient, which has the effect of introducing a positive bias to the performance measures. The visual aspects of two ROIs of the same patient are indeed very similar because of individual anatomy and image settings that are independent of the classification problematic. The probability that

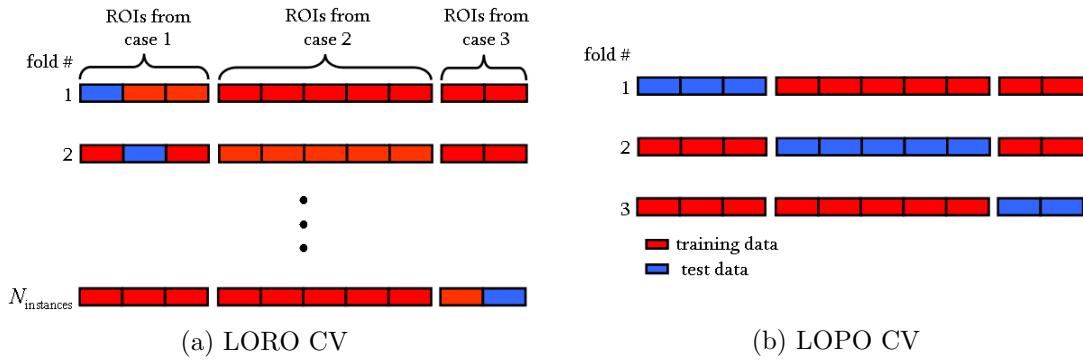


Figure 5.15: LORO and LOPO validation schemes.

the two ROIs from the same patients belongs to the same class is very high because both ROIs are associated with the same predominant disease. Thereby, the classification task is simplified as the classifier will learn visual aspects that are related to the particular patient and not to the disease, leading to overfitting. The bias is even stronger when the ROIs consist of overlapping blocks where groups of voxels are used both for training and testing.

**Leave-one-patient-out cross-validation** A solution to reduce the estimation bias of the empirical risk is to build the folds of the CV based on the patients (see Figure 5.15 (b)). The leave-one-patient-out (LOPO) ensures that all ROIs belonging to the same patient are contained in the same fold and thus does not allow to train and test with identical patients [69]. This situation corresponds to the clinical routine where the CAD system is trained using the entire database and unseen ROIs from an unknown patient are classified.

When compared to random drawing and  $N_{folds}$ -fold CV, the LOPO has the advantage of LOO which eliminates the variance of the performance measures as exactly the same experience is carried out for each run (no random draw is carried out with LOO). This is very desirable when searching for optimal parameters where the variance introduced by a random draw of the folds can lead to an inappropriate choice of parameters.

At last, the computational cost is affordable with  $N_{folds}$  equals to the number of cases  $N_{cases}$ . The number of classifiers to train is equal to  $N_{cases} - 1$ .

### 5.2.2 Metrics

Mostly depending on the research communities, complementary metrics are used in the literature to assess the performance of CAD systems. In this section, the performance metrics used throughout this work are made explicit and discussed.

## Confusion matrix and derived measures

The confusion matrix (sometimes also called “table of contingency”) categorizes the predictions of a classifier in order to visualize and count the occurrences of the correct predictions and the confusions among the classes (see Table 5.1). It contains all the results of a classification run of a given dataset with a given model. From the confusion matrix four class-specific measures can be derived:

- true positive (TP): number of positive instances classified as positive,
  - true negative (TN): number of negative instances classified as negative,
  - false positive (FP): number of negative instances classified as positive,
  - false negative (FN): number of positive instances classified as negative.

Table 5.1:  $N_{class}$ -confusion matrix.

predicted	actual		
	class 1	$\dots$	class $N_{class}$
class 1	$a_{11}$	$\dots$	$a_{1N_{class}}$
$\vdots$	$\vdots$	$\ddots$	$\vdots$
class $N_{class}$	$a_{N_{class}1}$	$\dots$	$a_{N_{class}N_{class}}$

The latter can be computed for each class  $c_l$  from the elements  $a_{ij}$  of the confusion matrix as follows:

$$TP_l = a_{ll}, \quad (5.1)$$

$$TN_l = \sum_{i=1}^{N_{class}} \sum_{j=1}^{N_{class}} \begin{cases} a_{ij}, & \text{if } i \neq l \text{ and } j \neq l, \\ 0, & \text{otherwise.} \end{cases}, \quad (5.2)$$

$$FP_l = \sum_{j=1}^{N_{class}} \begin{cases} a_{ij}, & \text{if } i = l \text{ and } j \neq l, \\ 0, & \text{otherwise.} \end{cases}, \quad (5.3)$$

$$FN_l = \sum_{i=1}^{N_{class}} \begin{cases} a_{ij}, & \text{if } i \neq l \text{ and } j = l, \\ 0, & \text{otherwise.} \end{cases}. \quad (5.4)$$

With  $N_{class}$  the number of classes. From these measures, several performance measures can be derived such as the precision, recall, accuracy and F-measure.

**Precision and recall** The precision measures the correctness of the predictions for the class  $l$ :

$$\text{precision}_l = \frac{TP_l}{TP_l + FP_l}. \quad (5.5)$$

The recall is complementary to precision as it assess the completeness of the predictions for the class  $l$ :

$$\text{recall}_l = \frac{TP_l}{TP_l + FN_l}. \quad (5.6)$$

Precision will be favored for CAD systems where each prediction needs to be correct whereas systems with high recall ensure to detect every instance of the considered class (e.g. important for the detection of malignant lung nodules).

**Accuracy** Whereas specialized CAD systems require either high precision or recall, most of the systems aim at maximizing the two measures together. The accuracy  $A_l$  combines the two by computing the rate of correctly classified instances for the class  $l$ :

$$A_l = \frac{TP_l + TN_l}{TP_l + TN_l + FP_l + FN_l}. \quad (5.7)$$

For multiclass systems, either the arithmetic or the geometric mean of  $A_l$  can be used to assess the global multiclass performance. The arithmetic mean  $A$  of the class-specific accuracies computes the rate of correctly classified instances over all classes:

$$A = \frac{\sum_{l=1}^{N_{class}} TP_l + TN_l}{\sum_{l=1}^{N_{class}} TP_l + TN_l + FP_l + FN_l}. \quad (5.8)$$

$A$  allows to assess the global performance of all predictions for all classes but is of limited use when the distribution of the classes is highly imbalanced. For example, consider the two-class

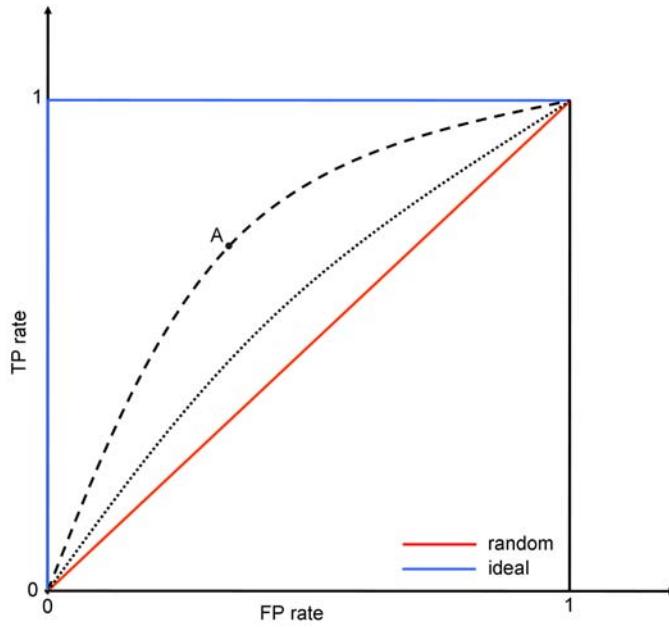


Figure 5.16: ROC curves. The point A denotes a working point of a CAD system and corresponds to a prediction of all instances of the dataset (i.e. full confusion matrix).

configuration with a dataset of 90% and 10% of negative and positive instances, respectively. Based on this dataset, a classifier always predicting the negative class will lead to  $A=90\%$  but is useless in practice. Alternately, the geometric mean  $A^{geom}$  of each class-specific accuracy  $A_l$  gives the same importance to each class, even if the classes are imbalanced [127]:

$$A^{geom} = \sqrt[N_{class}]{\prod_{i=1}^{N_{class}} A_l}. \quad (5.9)$$

**F-measure** F-measure (also called “F-score”) is the weighted harmonic mean of precision and recall rates:

$$F_l = \frac{2 \cdot \text{precision}_l \cdot \text{recall}_l}{\text{precision}_l + \text{recall}_l}. \quad (5.10)$$

$F_l$  yields a class-specific performance based both on precision and recall. The F-measure is useful to assess the performance of *one versus all* configurations.

### ROC curves

As mentioned above, depending on the CAD task, either the precision or the recall will be favored. Most often, the trade-off between the two can be tuned with the underlying algorithm parameters of the CAD (e.g. the cost of the errors with SVM classifiers). In order to assess the behavior of the precision-recall trade-off, the receiver operating characteristics curves (ROC) represent the TP rate functions of the FP rate (see Figure 5.16). Clearly, a system with best performance will maximize the area under the curve. Based on the ROC curve, the user can choose the working point that provides an appropriate precision-recall trade-off. However, ROC curves are class-specific and several *one versus all* configuration have to be used for multiclass configurations, which is not appropriate to visualize the global performance of a multiclass CAD system.

### 5.2.3 Causes of error and bias

Clinical environments are extremely demanding in terms of robustness (high variability of data to be analyzed), usability and efficiency of the CAD system. The clinicians are expecting to access accurate information simply and rapidly [31]. Recently, the Food and Drug Administration (FDA<sup>7</sup>) of the United States issued two draft guidance documents<sup>8</sup> concerning the premarket approval process for CAD devices that contain valuable recommendations for evaluating image-based CAD systems. Hazards of the validation methods were discussed in Section 5.2.1. In this section, potential pitfalls of the validation methods concerning the influences of the dataset and optimization techniques are identified.

#### Influence of the dataset

The dataset used for the validation of the CAD is of course of primary importance to obtain statistically significant performance assessment. An inappropriate dataset will lead to very biased results. For instance, in the context of CBIR, Müller *et al.* showed that validating the same CBIR system on various selected subsets of the Corel Photo CDs leads to very different performance values [167]. The testing database has to be representative of the target population and the target disease, condition or abnormality for which the CAD system is intended (see Chapter 2). It is important to have a collection of cases showing the full range of the variability of the considered classes or diseases. This is especially true for the normal cases as the variability among healthy subjects is very large.

To objectively compare the performance of CAD systems, benchmarking with standardized databases was proposed in CBIR with ImageCLEF [91], in pulmonary image analysis at the MICCAI 2009 conference<sup>9</sup> with VolCANO<sup>10</sup> for the characterization of lung nodules and EXACT<sup>11</sup> for the segmentation of the airways. Yet, no challenge was proposed for lung tissue categorization in HRCT imaging, which is probably due to the lack of databases of ILD cases for comparing the methods.

**Influence of the annotation** As already mentioned in Section 2.3.2, particular care is required when carrying out annotations to be used as ground truth for computerized analysis. A measure of the performance of the radiologists is desirable to assess the quality of the ground truth itself. The intra and inter-observer variations allow to objectively evaluate the radiologists as well as to be aware of the specific challenges of interpreting the considered images [9]. For example, common causes of error when interpreting HRCT images of the lung are to mix partial volume with *ground glass* or nodules with bronchovascular structures [90].

As soon as high-quality ground truth is obtained, bias can be introduced when the validation is carried out only with typical visual aspects of the patterns. In most of the studies in the state of the art as well as in this work, the validation is based on manually selected ROIs (hand-drawn or blocks). This can potentially bias the performance measures as the typical instances are usually easier to classify. In lung tissue characterization in HRCT, problems occur for the classification of blocks that are close to the mediastinum or adjacent to the lung border where high gradient values are found because the radiologists are usually delineating ROIs in the central regions of the lung parenchyma (see Figure 5.17). The system is thus evaluated on facilitated instances, which introduce a positive bias in the performance measures. A negative bias was found in this work that is highlighted in Table 3.7 with the blockwise classification where recurrent confusions are found between *ground glass* and *fibrosis*. Most often, small regions of *ground glass* are found in the

---

<sup>7</sup><http://www.fda.gov/>, as of 14 November 2009

<sup>8</sup>The documents are available online at

<http://www.fda.gov/MedicalDevices/DeviceRegulationandGuidance/GuidanceDocuments/ucm187249.htm> and <http://www.fda.gov/MedicalDevices/DeviceRegulationandGuidance/GuidanceDocuments/ucm187277.htm>, as of 14 November 2009.

<sup>9</sup>Medical Image Computing and Computer Assisted Intervention, <http://www.miccai2009.org/>, as of 15 November 2009

<sup>10</sup>Volume Change Analysis of NOdules, <http://www.via.cornell.edu/challenge/> as of 15 November 2009

<sup>11</sup>EXtraction of Airways from CT, <http://image.diku.dk/exact/> as of 15 November 2009

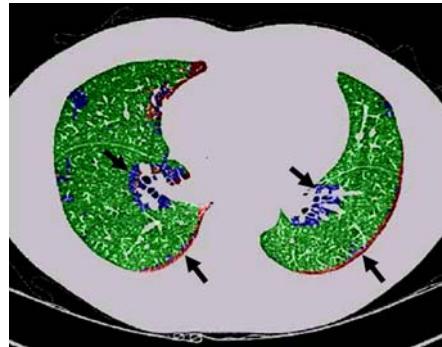


Figure 5.17: Common missclassifications near the lung border and the mediastinum (source [113]).

remaining functional tissue surrounding fibrotic patterns and were encircled with *fibrosis* regions, which induce classification errors.

### Optimization

Optimization is required to find optimal values of the underlying parameters of the CAD. It aims at minimizing the empirical risk but the final performance claimed must be measured with an untouched testing set (see Section 4.1.1). Test data, used once before, does not constitute independent data because the CAD algorithm may have become trained to that data, either implicitly or explicitly. Although rather straightforward, it is important to note that these guidelines are not always respected in the state-of-the-art.

#### 5.2.4 Qualitative evaluation

Although fundamental to be aware of the limitations of the CAD, the standalone quantitative evaluation is not directly related to the usability of the system in a clinical environment [81]. The performance level of automated computer diagnosis is equal to the performance achieved by computers. However, the performance level of the CAD system is equal to the performance achieved by the physician who makes the final decision by using the computer output as a second opinion. The latter is of course related to the standalone performance, but also to the allowance offered by the CAD as well as the user-friendliness of the GUIs. For instance, with image-based CAD for ILDs in HRCT, the choice of the classes of lung tissue is crucial for an optimal usage in clinical routine. While aiming at providing an exhaustive analysis of the HRCT image series as diagnostic aid for ILDs, the set of classes must be able to characterize most of the ILDs to avoid that the clinicians still have to go through the whole image series to seek for the patterns that were not included in the CAD. Normal versus abnormal classification configuration should be favored to detection of one single specific pattern (e.g. *ground glass*, *emphysema*, *honeycombing*, ...).

At last, regular meetings and discussions with radiologists are strongly recommended starting from the beginning of the design of a CAD system to have continuous feedback on the functionalities, user interfaces and of course to gain experience providing insights of solutions to the considered problems and challenges. A methodology must be put in place to conduct a clinical evaluation aiming at quantitatively measure the benefits of the CAD system by comparing the precision and recall measures of the radiologists with and without the CAD [4].

# Chapter 6

# Conclusions and Perspectives

At the end of the early days of computer vision, multidisciplinary research efforts are required to enable image-based computer-aided diagnosis to clinicians. It is the role of computer scientists to make the first step towards the clinicians to demonstrate the full potential of their emerging technologies for clinical benefits. Close collaborations with health care professionals are necessary to adapt and further develop the existing state-of-the-art techniques in computer vision to the very challenging demands of clinical environments.

In this thesis, image analysis and retrieval methods are proposed to cope with the exponential growth of the amount and variability of medical visual information. A focus on texture analysis of HRCT imaging of the chest of patients with ILD allowed to study medical image-based CAD systems from theoretical aspects of image processing and machine learning to ergonomics of user interfaces through the construction of a multimedia case collection and CAD performance assessment. The main achievements representing contributions in the fields of *image processing*, *machine learning* and *information retrieval* as well as the respective limitations are recalled in Section 6.1. Based on the proposed achievements, perspectives of computer-aided analysis and management of medical visual information are discussed in Section 6.2.

## 6.1 Achievements and limitations

At the crossing of computer vision, medicine, artificial intelligence, information retrieval, data management and human-computer interaction, the main contributions and underlying limitations proposed in this thesis are summarized in this section.

### 6.1.1 Literature analysis

57 papers on texture-based CADs for lung tissue analysis in thin-section CT from 27 research groups from 1997 to 2009 were systematically categorized and reviewed in Section 1.3.2. The trends and weaknesses were identified and showed a lack of systems integrating non-visual data calling upon contributions on multimodal CAD systems fusing visual and clinical data.

### 6.1.2 Multimedia library

A multimedia collection of 128 ILD cases (108 with image series) was created to provide a reference dataset for the evaluation of pattern recognition techniques and teaching purposes. A methodology for selecting, annotating and capturing patients with verified diagnoses of ILD and quality assessment was established. A limitation is the lack of healthy cases. Consequently, most of the *healthy* regions of the lung parenchyma were annotated in formerly diseased cases and further investigations are required to ensure that the annotated regions of healthy tissue are representative of the large variety of normal lung tissue in HRCT.

### 6.1.3 Features

A multimodal feature set for the categorization of lung tissue patterns associated with ILDs in HRCT was developed and validated in Chapter 3. The consistency of the heterogeneous feature space was studied using correlation analysis. Further work is necessary to study the interactions among the subgroup of features in order to define features spaces with maximal positive synergy.

#### Visual features

An affine-invariant texture feature set based on a tailored WT was developed. It is the result of the combination of isotropic polyharmonic  $B$ -spline wavelets with a redundant frame transform using a quincunx lattice. The characterization of the wavelets coefficients using parameters of a mixture of Gaussian combined with grey-level histogram bins in Hounsfield Units showed to efficiently describe five lung tissue patterns. Limitations occur as no directional analysis is carried out. The assumption was that no prevailing directions are contained in lung tissue patterns projected on 2D axial slices in HRCT.

A  $N$ -dimensional extension of the visual features is desirable for the characterization of organs in volumetric medical images (see Section 6.2.2). The three-dimensional expression of an isotropic polyharmonic  $B$ -spline in the Fourier domain is given Section 3.1.2. The custom WT transform is consequently easily extendable to three dimensions with the limitation that the extension of the quincunx lattice will no more yield inter-scale similarity transforms [256]. It is important to note that 3D wavelet transforms are not appropriate for analyzing HRCT image series because filtering along the  $z$ -axis with a very low axial resolution (20 to 50 slices with 10 mm distance) would lead to coarse blurring of the relevant information.

#### Clinical attributes

A set of clinical features was built from MySQL fields of the database. Two measures for ranking heterogeneous clinical parameters based on the information gain ratio and the generalization performance were compared in Section 3.2. The results of the comparison were qualitatively evaluated based on the medical significance of the ranking. The substitution of the missing values of the clinical parameters with the average value may arbitrarily favor one of the class and the associated impact on the generalization performance has to be investigated.

### 6.1.4 Machine learning

A methodology for comparing the performances of classification models based on McNemar's test was built in Section 4.1.1. It assess the global and class-specific accuracies and studies the stability of the hyperparameters of the models. This methodology can be applied to any classification task.

Contextual medical image analysis was proposed in Section 4.2. The optimal scheme for fusing visual and clinical parameters was studied. Although fundamental for medical image analysis, the clinical context of images is rarely used for improving the performance of image-based CADs in the literature. To our knowledge, it was never used for the categorization of lung tissue patterns associated with ILDs in HRCT so far. In this work, the combination of two SVMs classifiers (late fusion) allowed an average absolute gain of 5% of accuracy for the classification of five lung tissue sorts.

### 6.1.5 Information retrieval

Case-based retrieval based on a multimodal inter-case similarity was developed and described in Section 4.3. The clinical and visual similarities are combined using a weighted Euclidean distance. No low-level local features are used for inter-case visual comparison, which can potentially be a bottleneck to the retrieval performance.

### 6.1.6 Application

A hybrid detection–CBIR–based CAD system was developed and integrates the texture analysis framework along with case–based retrieval (see Section 5.1). Three use cases adapted to the needs of the radiologists were developed including:

- multimedia database browsing,
- automated 3D categorization of the lung tissue in HRCT image series,
- case–based retrieval.

Ergonomic interfaces were implemented in a rich Internet application accessible from a web browser. The major limitations of the CAD system are:

- no possibility to store the cases analyzed by a user to enrich the multimedia database,
- high computational time for the categorization of the lung tissue ( $\sim 1$  minute per slice),
- the clinical parameters are not used for the classification of the lung tissue.

### 6.1.7 Evaluation of image–based CAD systems

The pitfalls and common mistakes of the evaluation of image–based CAD systems were identified in Section 5.2 and a subsequent evaluation methodology was proposed to reproduce actual clinical conditions of the target usage of the CAD system. However, the performance of human observers using the CAD system in a true clinical environment was not carried out. The robustness of the lung tissue classification framework towards the image parameters (i.e. slice thickness, manufacturer, ...) was not assessed.

## 6.2 Perspectives

Short, medium and long term future research directions are proposed in this section, which is structured as follows. Section 6.2.1 contains a couple of practical solutions that may improve the classification and retrieval performances of the hybrid CAD system. In Section 6.2.2, the need for further developments required to extend the proposed retrieval techniques to more dimensions is highlighted. Section 6.2.3 briefly reviews the literature of content–based indexing of the PACS. At last, general conclusions concerning future clinician–computer interactions are given in Section 6.3.

### 6.2.1 Potential improvements and extensions of the CAD system

Future work concerning the visual features, the classification algorithms, the multimodal fusion and the case–based retrieval are proposed in this section. Some other domains where the affine–invariant texture analysis framework can potentially perform well are listed at the end of the section.

#### Monogenic directional texture analysis

A possible directional extension of the proposed texture analysis framework is to use the monogenic extension of the polyharmonic wavelets using a complexified Riesz transform [251, 285]. The complex wavelet domain yields directional analysis using steerability. For instance, directional texture features can be obtained by measuring local orientations, energy, coherency and many more using the structure tensor [257]. The complex wavelet transform is available in a frame flavor and is extendable to  $N$  dimensions.

### Imbalanced class distributions

SVM with asymmetric margins was proposed by Cohen *et al.* in 2006 to cope with imbalanced class distributions [42]. In two-class problems, errors from positive and negative example are not equally penalized using two cost values  $C_+$  and  $C_-$ . In order to optimize and regularize the classification performance for every regions of the lung parenchyma, a uniform distribution of the lung tissue classes can be used for training with class-specific  $C_n$  values using *one versus all* configurations initialized based on the respective prevalence of the lung tissue sorts.

### Kernel modality fusion

The challenge when fusing modalities at the early level is to obtain a coherent feature space, which strongly depends on the transformation implicitly applied by the kernel. Consequently, multiple kernels with adjustable feature-specific parameters (e.g. Gaussian kernel with varying  $\sigma_K$  for each feature or group of feature) can potentially build optimal feature spaces for multimodal classification problems [10]. Fusing modalities with multiple kernels is very computationally intensive but can be formulated as a semi-definite program, which rapidly converges to the optimum [133].

### Local texture analysis and learning to rank for similar case retrieval and literature retrieval

Currently, the visual similarity used for case-based retrieval is not based on local comparisons of the texture measures. Local features can potentially improve the importance of the visual modality in Eq. 4.5. To do so, a localization system based on the anatomy of the lung is required, which was proposed by Zrimec *et al.* in [291].

Supervised learning with regression-SVMs can be used to predict the inter-case similarity measure to improve the generalization performance when compared to the weighted Euclidean distance [24, 229]. In this case, the multimodal fusion can still be performed either using early or late fusion.

At last, based on the images of the query cases, a search for similar images in the literature can provide relevant additional peer-reviewed information for treatment planning [70].

### Grid computing

In the context of the KnowARC project<sup>1</sup>, parts of algorithms were parallelized to speed up the following computationally intensive tasks [171]:

- feature extraction over the entire database (split by image series),
- categorization of the lung tissue of an image series (separate by slices),
- cross-validation (split by folds),
- grid search for SVM hyperparameters (distribution of the parameter values).

Virtualization techniques allowed to create virtual computing nodes inside the hospital yielding a secure solution to speed up image analysis tasks of the CAD system [175].

### Other potential application domains of the texture analysis framework

The texture analysis framework can potentially provide satisfying results on other biomedical data such as, for instance:

- characterization of skin lesions in dermoscopic images [271],
- differentiation of dementia of patients with Alzheimer's disease in MRI images [118],
- categorization of the liver tissue in ultrasonic images [100, 162],

---

<sup>1</sup>[http://www.sim.hcuge.ch/medgift/01\\_KnowARC\\_EN.htm](http://www.sim.hcuge.ch/medgift/01_KnowARC_EN.htm), as of 21 November 2009

- and many more [103, 106, 163, 283]

Most often, the characterization of biomedical tissue is related to texture analysis [30, 138], which opens a wide variety of potential applications of the texture analysis framework proposed in this work. However, at least some modifications concerning the GLH (based on HU) and tuning of the parameters of QWF (i.e.  $\gamma$ , number of iterations) are required.

### 6.2.2 Classification and retrieval in $N$ dimensions

In modern hospitals, the majority of the medical images are produced in more than two dimensions. The commonly used imaging modalities are CT and MRI. Both are three-dimensional and the combination of these imaging modalities with functional imaging (i.e. PET, fMRI) as well as their temporal extension (i.e. videos) lead to multidimensional data. Although many image processing and analysis algorithms were initially developed in two dimensions, some of them can be extended to additional dimensions.

In the literature, the most common higher-order texture measures (GLCM, RLE, autocorrelation function as well as multi-resolution filtering such as multiscale Gabor filters and gradients, fractal analysis and wavelet transforms) were already extended to three dimensions and evaluated at least on small test data sets.

GLCM and run-length encoding being the most populars of the lot, were extended to three-dimensions several times [124, 129, 278]. Volumetric textures were characterized using GLCM in MRI data in [39, 125, 149]. Three-dimensional GLCM, run-length and fractal features are used together to identify normal and pathological lung parenchyma in MDCT images in [279, 280]. Another 3D hybrid feature set is proposed in [148] where first-order statistics, GLCM, gradient features and multiscale steerable Gabor filters are combined to segment prostatic adenocarcinoma in high-resolution MRI. Feature selection on 3D multi-resolution filtering allows for a superior classification performance compared to GLCM with various 3D texture datasets in [201]. Although several studies showed that GLCM and run-length encoding can be outperformed by the WT for 2D texture categorization [61, 198], few research groups have extended it to 3D for texture analysis, yet. The simplest way to extend it to 3D is to carry out separable WTs along all the three directions. The latter was used in [103] to characterize the volumetric texture of the hippocampus in MRI images to quantify epilepsy. In [13], texture features based on a non-redundant 3D wavelet transform are used to discriminate among cerebro-spinal fluid, gray matter and white matter. However, there is no *a priori* reason for carrying out the transforms along  $x$ ,  $y$  and  $z$  axes and directional and/or isotropic transforms are preferable for analyzing textures in medical images [59] (see Section 3.1.2).

Motion analysis in videos can be considered a multi-dimensional texture analysis problem as proposed in [215]. Spatio-temporal textures are characterized in video sequences with GLCM in [21] and multi-resolution analysis based on three-dimensional Gabor filters in [41]. A separable 3D wavelet transform (along  $x$ ,  $y$ , and time axes) is used in [227] to index video sequences for retrieval. In moving 3D data (or 4D datasets) very little has so far been published in terms of texture analysis.

Globally, multi-resolution analysis using filter banks (Gabor filters, wavelets) is generally more appropriate for characterizing  $N$ -dimensional textures in comparison with GLCM and run-length encoding as it corresponds better to our perception of textures. Following this direction, research efforts are needed to develop  $N$ -dimensional affine-invariant wavelet transforms with tunable parameters and able to manage anisotropy inherent in medical image series as well as in video sequences.

### 6.2.3 Towards indexing the whole PACS

As soon as visual similarity measures can be built in  $N$ -dimensions, indexing the whole PACS is tempting to enable access to all image series and the associated metadata within the hospital. Whereas concepts, insights and even prototypes addressing the methodology for the indexing the PACS [71, 134, 194] as well as practical implementations for a particular imaging modality

or organ [83, 213] was proposed several times in the literature, to our knowledge, no practical implementation able to deal with the heterogeneity of the medical visual information was proposed yet.

### 6.3 Better than clinicians?

As mentioned in the introduction, computer vision is still far behind the formidable skills of human vision for the interpretation of visual scenes in terms of recognition time, accuracy and robustness as well as context integration. Consequently, clinicians will still be the main actors within the health care loop for a long time. Expert computerized systems have the potential of being of precious help to little experienced radiologists and non-specialist. Thanks to machine learning techniques, modern CAD systems have the capacity of learning from unstructured data and thus are able to gain experience by extracting knowledge from data repositories. Expert CAD systems proved to be incontrovertible tools also in high-risk professional domains from aeronautics to architecture. Information retrieval technologies enable access to up-to-date multimedia information, which is absolutely complementary to human skills and does not alter the workflow of the clinicians. In that perspective, clinicians and computers are expected to team up in a very close future.

## **Appendix A**

### **CAD in HRCT of the lungs**

Table A.1: Studies in computerized lung tissue analysis: research groups and scopes (1997–2001)

study	research group	tissue patterns	diseases	2D/3D	imaging protocol	slice thickness	clinical context	anatomical features
Delorme <i>et al.</i> 1997 [47]	German Cancer Research Center, Germany	healthy, emphysema, ground glass, fibrosis, vessel, bronchus, cysts	ILDs	2D	HRCT	1 mm	—	—
Heitmann <i>et al.</i> 1997 [90], Kauczor <i>et al.</i> 2000 [110]	Johannes Gutenberg University of Mainz, Germany	ground glass	ILDs	2D	HRCT	1 mm	—	—
Shyu <i>et al.</i> 1999 [23, 220]	Purdue University, USA	healthy, emphysema, ground glass, fibrosis, micronodules, macronodules, cysts	ILDs	2D	HRCT	—	—	—
Uppaluri <i>et al.</i> 1999 [253, 254], Guo <i>et al.</i> 2002 [85], Hoffman <i>et al.</i> 2003 [95]	University of Iowa, USA	healthy, emphysema, ground glass, fibrosis, nodular, bronchovascular normal, ground glass,	ILDs	2D	HRCT, MDCT	1 mm, 3 mm	—	—
Mitani <i>et al.</i> 2000 [160]	Yamaguchi University, Japan	consolidation, reticular, nodular	ILDs	2D	HRCT	—	—	—
Liu <i>et al.</i> 2000 [145], 2001 [144]	National Tsing Hua University, Taiwan	pathological	—	2D	HRCT	—	—	yes

Table A.2: Studies in computerized lung tissue analysis: research groups and scopes (2003–2004)

study	research group	tissue patterns	diseases	2D/3D	imaging protocol	slice thickness	clinical context	anatomical features
Sluimer <i>et al.</i> 2003 [225]	University Medical Center Utrecht, The Netherlands	<i>healthy,</i> <i>pathological</i>	ILDs	2D	HRCT	1–1.25 mm	–	–
Chabat <i>et al.</i> 2003 [33]	Imperial College London, UK	<i>healthy,</i> <i>emphysema</i> (3 sorts)	COPD	2D	HRCT	1.5 mm	–	–
Uchiyama <i>et al.</i> 2003 [246], Doi <i>et al.</i> 2005 [66]	University of Chicago, USA	<i>healthy,</i> <i>emphysema,</i> <i>ground glass,</i> <i>fibrosis,</i> <i>micronodules,</i> <i>consolidation,</i> <i>reticular</i>	ILDs	2D	HRCT	1–3 mm	–	–
Malone <i>et al.</i> 2004 [151]	University of Bristol, UK	<i>healthy,</i> <i>emphysema,</i> <i>fibrosis, body</i>	ILDs	2D	HRCT	2 mm, 6 mm, 8 mm	–	–
Yu <i>et al.</i> 2004 [284]	National Chung Cheng University, Taiwan	<i>pathological</i>	–	2D	CT	–	–	–
Prasad <i>et al.</i> 2004 [187, 188, 190], 2008 [189], 2009 [191]	Sowmya's group, University of New South Wales, Australia	<i>emphysema</i>	COPD	2D	HRCT	–	–	–
Shamsheyeva <i>et al.</i> 2004 [217, 218], 2008 [219], Vo <i>et al.</i> 2009 [264]	Sowmya's group, University of New South Wales, Australia	<i>healthy,</i> <i>emphysema,</i> <i>ground glass,</i> <i>fibrosis</i>	ILDs	2D	HRCT	1 mm, 2 mm	yes	–

Table A.3: Studies in computerized lung tissue analysis: research groups and scopes (2005–2007)

study	research group	tissue patterns	diseases	2D/3D	imaging protocol	slice thickness	clinical context	anatomical features
Mendonça <i>et al.</i> 2005 [157]	GE Research Center, USA	<i>normal, emphysema</i> (4 levels)	COPD	2D	HRCT	1 mm	—	—
Kim <i>et al.</i> 2005 [112]	Seoul National University College of Medicine, Korea	<i>healthy, ground glass</i>	cancer detection	2D	HRCT	1 mm, 2 mm, 5 mm	—	—
Müller <i>et al.</i> 2005 [168, 169]	University Hospitals of Geneva, Switzerland	<i>healthy, emphysema, ground glass, fibrosis, micronodules, macronodules, interstitial syndrome</i>	ILDs	2D	HRCT	—	—	—
Wong <i>et al.</i> 2005 [274], 2006 [275], Zrimec <i>et al.</i> 2007 [290, 292, 293]	Zrimec's group, University of New South Wales, Australia	<i>healthy, fibrosis</i>	ILDs	2D	HRCT	1 mm	—	yes
Xu <i>et al.</i> 2005 [279], 2006 [280, 281]	University of Iowa, USA	<i>healthy (2 sorts), emphysema</i> (2 sorts), <i>ground glass, fibrosis</i>	COPD, ILDs	3D	MDCT	1.3 mm	—	—
Saita <i>et al.</i> 2007 [211]	University of Tokushima, Japan	<i>emphysema</i>	COPD	3D	MDCT	1 mm	—	—
Caban <i>et al.</i> 2007 [26]	National Institute of Health, USA	<i>healthy, fibrosis</i>	ILDs	2D	HRCT	—	—	—

Table A.4: Studies in computerized lung tissue analysis: research groups and scopes (2007-2009)

study	research group	tissue patterns	diseases	2D/3D	imaging protocol	slice thickness	clinical context	anatomical features
Zavaletta <i>et al.</i> 2007 [286, 287]	Mayo Clinic, USA	<i>healthy, emphysema, ground glass, fibrosis, reticular</i>	ILDs	3D	MDCT	1.25 mm	—	—
Fetita <i>et al.</i> 2007 [77]	Groupe des Ecoles des Télécommunications, France	<i>healthy, emphysema, ground glass, fibrosis</i>	ILDs	3D	MDCT	—	—	—
Lee <i>et al.</i> 2007 [135, 136], Park <i>et al.</i> 2008 [180], Kim <i>et al.</i> 2009 [113, 137]	Seoul National University, Korea	<i>healthy, emphysema (3 sorts)</i>	COPD	2D	HRCT	1 mm	—	—
Depeursinge <i>et al.</i> 2007 [58], 2008 [59], 2009 [53, 57, 60]	University Hospitals of Geneva, Switzerland	<i>healthy, emphysema, ground glass, fibrosis, micronodules, macronodules</i>	ILDs	2D, 3D	HRCT	1 mm	yes	—
Tolouee <i>et al.</i> 2008 [242]	K.N.Tooi University of Technology, Iran	<i>healthy, ground glass, fibrosis, reticular</i>	ILDs	2D	HRCT	1 mm	—	—
Boehm <i>et al.</i> 2008 [20]	University of Munich, Germany	<i>healthy, emphysema, fibrosis</i>	ILDs	3D	MDCT	0.6 mm	—	—
Sorensen <i>et al.</i> 2008 [234], 2009 [232, 233]	University of Copenhagen, Denmark	<i>healthy, emphysema (2 sorts)</i>	COPD	2D, 3D	HRCT	1.25 mm	—	—

Table A.5: Studies in computerized lung tissue analysis: methods (1997–2001)

study	visual features	feature reduction	instances	CAD task	validation	dataset
Delorme <i>et al.</i> 1997 [47]	GLH, GLCM, RLE	stepwise discriminant analysis, factor analysis	5 × 5 blocks at three image scales	LDA	7-class tissue categorization	ROI random drawing 1022 ROIs, 10 patients
Heitmann <i>et al.</i> 1997 [90], Kauczor <i>et al.</i> 2000 [110]	no features – image pixels	–	5 × 5, 9 × 9 blocks	2-stage SOM	ground glass detection	manual evaluation 84 patients
Shyu <i>et al.</i> 1999 [23, 220]	GLH, GLCM, Sobel operator, custom shape features	SFS	hand-drawn ROIs	decision tree	CBIR	LOPO 302 labelled images, 78 patients
Uppaluri <i>et al.</i> 1999 [253, 254], Guo <i>et al.</i> 2002 [85], Hoffman <i>et al.</i> 2003 [95]	GLH, GLCM, RLE, fractal (AMFM)	CFS, divergence feature method [5]	31 × 31 blocks	Naive Bayes	6-class tissue categorization	agreement with radiologists, ROI random drawing 899 ROIs, 72 patients
Mitani <i>et al.</i> 2000 [160]	Gabor filters	–	15 × 15, 23 × 23, 31 × 31, 39 × 39 blocks	k-NN with Euclidean distance	5-class tissue categorization	ROI random drawing 500 ROIs, 11 patients
Liu <i>et al.</i> 2000 [145], 2001 [144]	Fourier analysis	–	manually selected square blocks as query, 2D image as result	SOM	CBIR	patient random drawing 2260 ROIs, 119 patients

Table A.6: Studies in computerized lung tissue analysis: methods (2003–2004)

study	visual features	feature reduction	instances	classifier	CAD task	validation	dataset
Sluimer <i>et al.</i> 2003 [225]	Gaussian derivatives filter bank	SFS	circular ROIs, radius = 40	LDA, QDA, k-NN, SVM	healthy vs pathological tissue categorization	4-patient-fold CV	657 ROIs, 116 patients
Chabat <i>et al.</i> 2003 [33]	GLH, GLCM, RLE	—	circular ROIs, radius = 22	Bayes	4-class tissue categorization	patient random drawing	2420 ROIs, 33 patients
Uchiyama <i>et al.</i> 2003 [246], Doi <i>et al.</i> 2005 [66]	GLH, top-hat transform	—	32 × 32, 64 × 64 blocks	MLP	normal vs all, 7-class tissue categorization	identical training and testing sets	315 ROIs, 105 patients
Malone <i>et al.</i> 2004 [151]	GLH, Fourier analysis, fractal, auto-correlation	genetic algorithm	4 × 4, 8 × 8, 16 × 16 blocks	Naive Bayes, SVM	4-class tissue categorization	3-patient-fold CV	102 ROIs, 32 patients
Yu <i>et al.</i> 2004 [284]	GLCM, custom shape features	—	images	k-NN with Euclidean distance	CBIR	image random drawing	500 images
Prasad <i>et al.</i> 2004 [187, 188, 190], 2008 [189], 2009 [191]	GLH, GLCM, RLE, gray-level differences	ICA, PCA, SFS, Information gain $I_G$	pixels	semi-supervised, classifier combination with Naive Bayes, C4.5, SVM	emphysema detection	manual pixel drawing	60 patients with fully-annotated scans
Shamsheyeva <i>et al.</i> 2004 [217, 218], 2008 [219], Vo <i>et al.</i> 2009 [264]	GLH, quincunx wavelet frames, directional wavelet frames	—	pixels, 32 × 32 blocks	SVM with anisotropic Gaussian kernel	4-class tissue categorization	pixel random drawing, 10-fold CV	177 ROIs, 66 patients

Table A.7: Studies in computerized lung tissue analysis: methods (2005–2007)

study	visual features	feature reduction	instances	classifier	CAD task	validation	dataset
Mendonça <i>et al.</i> 2005 [157]	GLH, Canny edge detection	—	patients	k-NN with EMD	5-class patient classification	LPOO	81 patients
Kim <i>et al.</i> 2005 [112]	GLH, GLCM, surface curvature [181]	—	30 × 30 blocks	MLP	healthy vs ground glass tissue	5-fold CV 17734 ROI blocks, 14 patients	
Müller <i>et al.</i> 2005 [168, 169]	GLH, GLCM, RLE, Gabor filters	—	16 × 16, 32 × 32 blocks	Naïve Bayes, k-NN, C4.5, SVM	healthy vs pathological tissue	CV	112 ROIs, 12 patients
Wong <i>et al.</i> 2005 [274], 2006 [275], Zrimec <i>et al.</i> 2007 [290, 292, 293]	CFS, consistency feature selection	7 × 7, 15 × 15 blocks	Naïve Bayes, C4.5, seeded region growing	healthy vs fibrosis tissue	10-fold CV, separated test set	18407 ROI blocks, 8 patients	
Xu <i>et al.</i> 2005 [279], 2006 [280, 281]	GLH, GLCM, RLE, fractal (3D AMFM)	31 × 31 isotropic 3D blocks	Naïve Bayes, SVM	4-class, 5-class tissue categorization	LORO, 10-fold CV	1865 VOIs, 34 patients	
Saita <i>et al.</i> 2007 [211]	GLH	—	pixels	—	emphysema detection	only test set (no training)	—
Caban <i>et al.</i> 2007 [26]	GLH, GLCM, RLE	4 × 4, 8 × 8, 16 × 16 blocks	SVM	healthy vs fibrosis tissue categorization	ROI random drawing		

Table A.8: Studies in computerized lung tissue analysis: methods (2007–2009)

study	visual features	feature reduction	instances	classifier	CAD task	validation	dataset
Zavaletta <i>et al.</i> 2007 [286, 287]	GLH with adaptive binning using K-means clustering, GLCM, RLE, SADH	—	15 × 15 × 15 3D blocks	k-NN with EMD or Mahalanobis	4-class and 5-class lung tissue categorization	14 scans training, 4 scans testing	909 VOIs, 14 patients with 4 fully-annotated scans
Retita <i>et al.</i> 2007 [77]	custom 3D morphological filter	—	pixels	graph-based hierarchical classification	4-class tissue categorization	only test set (no training)	10 patients with fully-annotated scans
Lee <i>et al.</i> 2007 [135, 136], Park <i>et al.</i> 2008 [180], Kim <i>et al.</i> 2009 [113, 137]	GLH, GLCM, RLE, shape, top-hat transform	SFS, SBS, $\text{PTA}(l,r)$ , SFBS, SFFS	16 × 16, 32 × 32, 64 × 64 blocks	Naive Bayes, Bayes, MLP, SVM	4-class tissue categorization	5-fold CV, 5-patient-fold CV	265 ROIs, 92 patients
Depenninge <i>et al.</i> 2007 [58], 2008 [59], 2009 [53, 57, 60]	GLH, isotropic quincunx wavelet frames	PCA	hand-drawn ROIs, 32 × 32 blocks	Naive Bayes, k-NN, C4.5, MLP, SVM	healthy vs pathological, 5-class and 6-class lung tissue categorization, CBIR	LORO, Lopo, ROI random drawing	14'315 ROI blocks, 69 patients
Tolouee <i>et al.</i> 2008 [242]	discrete wavelet frames and rotated wavelet frames	—	31 × 31 blocks	SVM	4-class tissue categorization	LORO	435 ROI blocks, 17 patients
Boehm <i>et al.</i> 2008 [20]	GLH, 3D Minkowski functionals	—	40 × 40 × 40 3D blocks	LDA	3-class tissue categorization	LORO	275 VOIs, 21 patients
Sorensen <i>et al.</i> 2008 [234], 2009 [232, 233]	GLH, LBP, Gaussian derivatives filter bank (2D and 3D)	—	31 × 31, 41 × 41, 51 × 51 blocks	k-NN and histogram dissimilarity measure, LDA and QDA	3-class tissue categorization	LORO	168 ROIs, 25 patients



## Appendix B

# HTML form for clinical data

### New ILD case

#### 1.Values for table "confidential data":

1.1 Initials:

1.2 Stay number:

1.3 Patient number:

1.4 Birthdate (YYYY-MM-DD):

1.5 Medical service:

1.6 ICD 10 code:

1.7 Stay dates (YYYY-MM-DD): a) arrival:  b) departure:

#### 2.Values for table "patient data":

2.1 Diagnosis:  reliability:  %

2.2 Diagnosis category:

2.3 Age:  years old

2.4 Gender:  F  M

2.5 Profession:

2.6 Smoking history:  yes  no  unknown a) quantity:  UPA  
b) quitted since:  years  
c) remarks:

2.7 Disease duration:  weeks

2.8 Medication: a) cordarone:  yes  no  unknown, 1. quantity:   
2. duration:  days

b) cyclines:  yes  no  unknown, 1. quantity:   
2. duration:  days

c) cyclophosphamides:  yes  no  unknown, 1. quantity:   
2. duration:  days

d) nitrofurantoines:  yes  no  unknown, i. quantity:   
ii. duration:  days

e) methotrexate:  yes  no  unknown, i. quantity:   
ii. duration:  days

f) busulfan:  yes  no  unknown, i. quantity:   
ii. duration:  days

g) gold salts:  yes  no  unknown, i. quantity:   
ii. duration:  days

h) bleomycin:  yes  no  unknown, i. quantity:   
ii. duration:  days

i) corticosteroid:  yes  no  unknown, i. quantity:   
ii. duration:  days

j) other drugs:  yes  no  unknown, 1. quantity:   
2. duration:  days  
3. type:

Figure B.1: HTML form used to capture the clinical parameters (page 1/4).

2.9 Findings on physical exam:

a) generals:

1. fever:	<input type="radio"/> yes <input type="radio"/> no <input type="radio"/> unknown
2. cyanosis of mucous membrane(s):	<input type="radio"/> yes <input type="radio"/> no <input type="radio"/> unknown
3. anorexia:	<input type="radio"/> yes <input type="radio"/> no <input type="radio"/> unknown
4. finger(s) clubbed:	<input type="radio"/> yes <input type="radio"/> no <input type="radio"/> unknown
5. Raynaud's phenomenon:	<input type="radio"/> yes <input type="radio"/> no <input type="radio"/> unknown
6. myalgia:	<input type="radio"/> yes <input type="radio"/> no <input type="radio"/> unknown
7. arthralgia:	<input type="radio"/> yes <input type="radio"/> no <input type="radio"/> unknown
8. lymph node(s) enlarged:	<input type="radio"/> yes <input type="radio"/> no <input type="radio"/> unknown, localisation: <input type="text"/>

b) respiratory:

1. cough:	<input type="radio"/> yes <input type="radio"/> no <input type="radio"/> unknown
2. tachypnea:	<input type="radio"/> yes <input type="radio"/> no <input type="radio"/> unknown
3. breathing intercostal retraction inspiratory:	<input type="radio"/> yes <input type="radio"/> no <input type="radio"/> unknown
4. breathing accessory muscle(s) used:	<input type="radio"/> yes <input type="radio"/> no <input type="radio"/> unknown
5. lung(s) percussion and/or auscultation abnormal:	<input type="radio"/> yes <input type="radio"/> no <input type="radio"/> unknown
6. sputum production:	<input type="radio"/> yes <input type="radio"/> no <input type="radio"/> unknown, type: <input type="text"/>
7. chestpercussion diaphragm low and/or movement decreased bilateral:	<input type="radio"/> yes <input type="radio"/> no <input type="radio"/> unknown

c) cardiac:

1. tachycardia:	<input type="radio"/> yes <input type="radio"/> no <input type="radio"/> unknown
2. leg(s) edema bilateral slight or moderate:	<input type="radio"/> yes <input type="radio"/> no <input type="radio"/> unknown
3. pressure arterial systolic:	<input type="radio"/> <90 <input type="radio"/> >90 <input type="radio"/> unknown

d) abdominals:

1. liver enlarged slight:	<input type="radio"/> yes <input type="radio"/> no <input type="radio"/> unknown
2. splenomegaly present:	<input type="radio"/> yes <input type="radio"/> no <input type="radio"/> unknown

e) ORL anomalies:

1. <input type="radio"/> yes <input type="radio"/> no <input type="radio"/> unknown, type: <input type="text"/>
---

f) visual anomalies:

1. <input type="radio"/> yes <input type="radio"/> no <input type="radio"/> unknown, type: <input type="text"/>
---

g) skin lesion(s):

1. <input type="radio"/> yes <input type="radio"/> no <input type="radio"/> unknown, type: <input type="text"/>
---

h) articular deformations:

1. <input type="radio"/> yes <input type="radio"/> no <input type="radio"/> unknown
---

2.10 Past medical history:

a) lymphoma or leukemia:

1. <input type="radio"/> yes <input type="radio"/> no <input type="radio"/> unknown
---

b) neoplasm malignant:

1. <input type="radio"/> yes <input type="radio"/> no <input type="radio"/> unknown
---

c) respiratory infection recent:

1. <input type="radio"/> yes <input type="radio"/> no <input type="radio"/> unknown
---

d) pneumonia recurrent:

1. <input type="radio"/> yes <input type="radio"/> no <input type="radio"/> unknown
---

e) asthma:

1. <input type="radio"/> yes <input type="radio"/> no <input type="radio"/> unknown
---

f) allergy:

1. <input type="radio"/> yes <input type="radio"/> no <input type="radio"/> unknown
---

g) wheezing:

1. <input type="radio"/> yes <input type="radio"/> no <input type="radio"/> unknown
---

h) drug hypersensitivity:

1. <input type="radio"/> yes <input type="radio"/> no <input type="radio"/> unknown
---

i) dyspnea acute recurrent attack(s):

1. <input type="radio"/> yes <input type="radio"/> no <input type="radio"/> unknown
---

j) lupus erythematosus systemic:

1. <input type="radio"/> yes <input type="radio"/> no <input type="radio"/> unknown
---

k) progressive systemic sclerosis:

1. <input type="radio"/> yes <input type="radio"/> no <input type="radio"/> unknown
---

l) rheumatoid arthritis:

1. <input type="radio"/> yes <input type="radio"/> no <input type="radio"/> unknown
---

m) tuberculosis contact:

1. <input type="radio"/> yes <input type="radio"/> no <input type="radio"/> unknown
---

n) tuberculosis history:

1. <input type="radio"/> yes <input type="radio"/> no <input type="radio"/> unknown
---

o) arthritis:

1. <input type="radio"/> yes <input type="radio"/> no <input type="radio"/> unknown
---

p) renal insufficiency:

1. <input type="radio"/> yes <input type="radio"/> no <input type="radio"/> unknown
---

q) HTA:

1. <input type="radio"/> yes <input type="radio"/> no <input type="radio"/> unknown
---

r) angina:

1. <input type="radio"/> yes <input type="radio"/> no <input type="radio"/> unknown
---

s) subOAP:

1. <input type="radio"/> yes <input type="radio"/> no <input type="radio"/> unknown
---

t) weight loss greater than 10%:

1. <input type="radio"/> yes <input type="radio"/> no <input type="radio"/> unknown
---

Abbreviations:

HTA: arterial hypertension  
subOAP: acute pulmonary edema

Figure B.2: HTML form used to capture the clinical parameters (page 2/4).

2.11 Occupational history: a) drug-addiction:  yes  no  unknown  
 b) activities and leisures:

2.12 Environmental exposures: a) moulds:  yes  no  unknown  
 b) birds:  yes  no  unknown  
 c) other animals:  yes  no  unknown  
 d) organic dust:   
 e) mineral particles: 1. silicosis:  yes  no  unknown  
 2. siderite:  yes  no  unknown  
 3. asbestose:  yes  no  unknown  
 f) other exposures:

2.13 Host risk factors: a) HIV:  yes  no  unknown  
 b) chemotherapy:  yes  no  unknown  
 c) radicotherapy:  yes  no  unknown  
 d) hemopathy:  yes  no  unknown  
 e) transplant received:  yes  no  unknown  
 f) diabetes:  yes  no  unknown  
 g) alcohol:  yes  no  unknown  
 h) malnutrition:  yes  no  unknown

2.14 Laboratory tests: a) VS:  mm/h  
 b) CRP:  mg/l  
 c) hematocrit blood:  %  
 d) hemoglobin blood:  g/l  
 e) ldh serum:  U/l  
 f) WBC:  G/l  
 g) WBC eosinophilic(s):  %  
 h) WBC neutrophilic(s):  %  
 i) tuberculin skin test:  positive  negative  unknown  
 j) rheumatoid factor:  U/ml  
 k) immunolectrophoresis serum IgG:  g/l  
 l) immunolectrophoresis serum IgA:  g/l  
 m) immunolectrophoresis serum IgM:  g/l  
 n) angiotensin converting enzyme activity:  U/l  
 o) kveim test:  positive  negative  unknown

Abbreviations: VS: sedimentation speed  
 CRP: C-reactive protein  
 WBC: white blood cell

2.15 Oximetry tests: a) oxygen blood arterial:  <80  >80  unknown  
 b) carbon dioxide tension blood arterial:  increased  decreased  unknown

2.16 PST (pulmonary function testing): a) CPT:  l  
 b) VEMS:  l/sec  
 c) VR:  l  
 d) CVF:  l

Abbreviations: CPT: total pulmonary capacity  
 VEMS: maximal expired volume in one second  
 VR: residual volume  
 CVF: functional vital capacity

Figure B.3: HTML form used to capture the clinical parameters (page 3/4).

2.17 Smear sputum tests:

2.18 BAL (bronchoalveolar lavage): a) opportunist agents:  yes  no  unknown, type:

b) neoplastic cells:  yes  no  unknown

c) lipoproteic alveolar liquid:  yes  no  unknown

d) mineral particles:  yes  no  unknown

e) siderophages:  yes  no  unknown

f) cells of Langerhans:  yes  no  unknown

g) total hypercellularity:  yes  no  unknown

h) lymphocitary alveolite:  yes  no  unknown

i) neutrophilic alveolite:  yes  no  unknown

j) eosinophilic alveolite:  yes  no  unknown

2.19 Biopsy: a) lung:

b) bronchoscopy transbronchial:

c) surgical:

d) other bodies:

2.20 Remarks:

Figure B.4: HTML form used to capture the clinical parameters (page 4/4).

# Notation

The mathematical notation and symbols used throughout the thesis are listed in this chapter. The section and the page number of the first apparition of the notation is indicated in brackets.

$\hat{\cdot}$  denotes the Fourier transform.

$A$  is the arithmetic mean of the class-specific accuracies  $A_l$  (Section 5.2.2, page 95, Eq. 5.8)

$A^{cv}$  is the accuracy obtained while averaging the accuracies of all the folds of the cross-validation (Section 4.1.1, page 68)

$A_i^{test}$  is the testing accuracy of the classifier  $i$  computed as the number of correctly classified instances divided by the total number of instances in the test set (Section 4.1.1, page 70)

$A_{mean}^{test}$  is the mean of  $A_i^{test}$  over the global experiment (Section 4.1.1, page 70)

$A_l$  is the accuracy specific to the class  $l$  (Section 5.2.2, page 95, Eq. 5.7)

$A^{geom}$  is the geometric mean the class-specific accuracies  $A_l$  (Section 5.2.2, page 96, Eq. 5.9)

$A_{mean}^{geom}$  is the mean of  $A^{geom}$  over the global experiment (Section 4.1.1, page 70)

$A^{LORO}$  is the accuracy obtained with a LORO cross-validation (Section 3.1.3, page 48)

$A^{single}$  is the single testing accuracy (Section 3.2.1, page 61)

$a$  is the scaling factor in Eq. 1.6 (Section 1.2.1, page 11)

$a_j$  is weight of the modality  $j$  in the multimodal distance measure (Section 4.3.1, Eq. 4.5, page 82)

$\alpha$  is the degree of  $B$ -spline wavelets (Section 3.1.3, page 50)

$B$  is the bandwidth of a continuous signal (Section 1.2.1, page 7)

$b$  is the translation factor in Eq. 1.6 (Section 1.2.1, page 11)

$\beta_\gamma$  is an isotropic polyharmonic  $B$ -spline function (Section 3.1.2, page 46)

$c_n$  denote the clinical features indexed by  $n$  (Section 3.2.1, page 60)

$C_{pruning}$  is the feature confidence factor for pruning the tree in C4.5 (Section 1.2.2, page 15)

$C$  is the cost of the errors in the soft margin formulation of the SVM (Section 1.2.2, Eq. 1.12, page 16)

$\Delta$  is the Laplacian operator (Section 3.1.2, page 44)

$\gamma$  is the order of the isotropic polyharmonic  $B$ -spline  $\beta_\gamma$  (Section 3.1.2, page 46)

$\mathcal{F}$  is the set of classification function which models the boundaries among the distinct classes of lung tissue patterns represented in the feature space (Section 1.2.2, page 13)

$f(\mathbf{v})$  is the decision function of a classifier (Section 1.2.2, page 16)

$f_s$  is the sampling frequency (Section 1.2.1, page 7)

$F_l$  is the F-measure of the class  $l$  (Section 5.2.2, page 96, Eq. 5.10)

$G_j$  contains the coefficients generated by the convolution of  $s(x)$  with the highpass filters  $g_j$  at iteration  $j$  (Section 1.2.1, page 12)

$g_j$  stands for a highpass filter at iteration  $j$  of the WT (Section 1.2.1, page 12)

$\mathcal{H}$  is the reproducing kernel Hilbert space (Section 1.2.2, page 16)

$H_{0:1}$  are the hypotheses for McNemar's test used to compare the classifier families (Section 4.1.1, page 70)

$I_G$  is the information gain (Section 3.2.1, page 61, defined in Eq. 3.11)

$I_{G_{ratio}}$  is the information gain ratio (Section 3.2.1, page 61, defined in Eq. 3.14)

$I(\mathbf{x})$  is a two-dimensional image indexed by the coordinate vector  $\mathbf{x} = \begin{pmatrix} x \\ y \end{pmatrix}$  (Section 1.2.1, page 10)

$j$  is the scale index of the WT (Section 1.2.1, page 12)

$J$  is the number of iterations of the WT (Section 1.2.1, page 12)

$k$  denotes the number of neighbors to be used in the k-NN algorithm (Section 1.2.2, page 14)

$K(\mathbf{v}_i, \mathbf{v}_j)$  denotes the kernel function (introduced in Section 1.2.2, page 16)

$\|\cdot\|_K$  is a norm in  $\mathcal{H}$  defined by the kernel function  $K$  (Section 1.2.2, Eq. 1.12, page 16)

$H_J$  contains the coefficients generated by the convolution of  $s(x)$  with the lowpass filter  $h_J$  at the last iteration  $J$  (Section 1.2.1, page 12)

$h_j$  stands for a lowpass filter at iteration  $j$  of the WT (Section 1.2.1, page 12)

$H(Y)$  is the entropy of the variable  $Y$  (Section 3.2.1, page 61, defined in Eq. 3.12)

$H(Y|X)$  is the conditional entropy of the variable  $Y$  when the value of  $X$  is known (Section 3.2.1, page 61, defined in Eq. 3.13)

$\mathbf{l}_2$  stands for the  $\mathbf{l}_2$  norm (Section 1.2.1, page 12)

$\mathbf{M}$  is a subsampling matrix in Eq. 3.6 (Section 3.1.2, page 46)

$\mathbf{M}_{quincunx}$  is the quincunx subsampling matrix (Section 3.1.2, page 47)

$M_{lung}$  is a binary mask marking the lung tissue region (Section 3.1.3, page 52)

$\mu$  is used for the mean of a distribution (Section 3.1.2, page 48)

$N_c$  is the number of wavelet (or Fourier) coefficients (Section 1.2.1, page 12)

$N_{cases}$  is the number of cases (Section 5.2.1, page 94)

$N_{class}$  denotes the number of classes (Section 5.2.2, page 95)

$N_{clin}$  denotes the number of clinical parameters (Section 4.2.4, page 78)

$N_{bins}$  is the number of bins of gray-level histograms (Section 3.1.3, page 48)

$N_{features}$  is the number of features (Section 4.2.4, page 77)

$N_{folds}$  denotes the number of folds of a cross-validation (Section 5.2.1, page 93)

$N_{gray}$  is the number of gray-levels contained in an image  $I(\mathbf{x})$  (Section 1.2.1, page 11)

- $N_{hidden}$  is the number of hidden units of the MLP (Section 1.2.2, page 15)
- $N_{instances}$  is the minimum number of instances per leaf in C4.5 decision tree algorithm (Section 1.2.2, page 15)
- $N_{mod}$  is the number of modalities (Section 4.2.3, page 76)
- $N_{PCA}$  is the number of principal component in Eq. 4.3 (Section 4.2.4, page 77)
- $N_{upsample}$  upsamples the images by  $2^{N_{upsample}}$  (Section 3.1.3, page 50)
- $N_v$  is the total number of feature vectors in Eq. 1.12, page 16
- $P(x)$  is the probability of  $x$  (Section 1.2.2, page 14)
- $p_i$  is a parameter of a classification algorithm (Section 4.1.1, page 68)
- $R^2$  is the coefficient of regression of Pearson (Section 3.2.3, page 62)
- $R_{learn}$  is learning rate of the MLP (Section 1.2.2, page 15)
- $\rho_{mean}$  denotes the mean correlation between two groups of features (Section 3.2.3, page 65)
- $s(x)$  is a 1D signal indexed by the variable  $x$  (Section 1.2.1, page 11)
- $\mathcal{S}$  is a family of templates that yield a WT of  $s(x)$  when taking the scalar product of  $s$  with each element of  $\mathcal{S}$  (Section 1.2.1, page 12)
- $\sigma$  is used for the standard-deviation of a distribution (Section 3.1.2, page 48)
- $\sigma_K$  is the width of the Gaussian kernel (Section 1.2.2, Eq. 1.11, page 16)
- $t_m$  denote the visual features indexed by  $m$  (Section 3.2.1, page 61)
- $\mathbf{v}$  is the feature vector corresponding to an instance expressed in the feature space (Section 1.2.2, page 14)
- $v_i$  is the volume percentage of the class  $i$  of lung tissue (Section 4.3.1, Eq. 4.4, page 81)
- $w_j$  denotes the class of lung tissue indexed by  $j$  (introduced in Section 1.2.2, page 14)
- $y_i$  is the class label of  $\mathbf{v}_i$  (Section 1.2.2, Eq. 1.12, page 16)
- $\phi$  is a smoothing function in Eq. 3.6 (Section 3.1.2, page 46)
- $\psi$  denotes a function that satisfies wavelet admissibility conditions as defined in Eq. 1.5 (Section 1.2.1, page 11)
- $\psi_{iso}(\mathbf{x})$  is an isotropic wavelet (Section 1.2.1, page 12)
- $\psi_{fovea}$  is the psychophysical linespread function that models receptive fields of retinal ganglion cells in the foveal region (Section 1.2.1, page 9).
- $\psi_{DG}$  is the difference of two Gaussians (Section 1.2.1, Eq. 1.2, page 9)
- $\psi_{\Delta_G}$  is the two-dimensional Laplacian of a Gaussian (Section 1.2.1, Eq. 1.3, page 10)



# Glossary

This chapter contains a list of the most common abbreviations in the field of computer vision and all the abbreviations used in this thesis, to avoid confusion. The location of the first apparition of the abbreviation is indicated in brackets.

**AIP** Acute Interstitial Pneumonia (Section 1.5, page 19)

**AMFM** Adaptive Multiple Feature Method (Section 1.3.2, page 22)

**ASSERT** Automated Search and Selection Engine with Retrieval Tools (Section 1.3.2, page 22)

**BAL** Bronchoalveolar Lavage (Section 2.3.1, page 33)

**BOOP** Bronchiolitis Obliterans Organizing Pneumonia (Section 1.5, page 19)

**CAD** Computer–Aided Diagnosis (Section 1.1, page 2)

**CBIR** Content–Based Image Retrieval (Section 1.1, page 2)

**CFS** Correlation–based Feature Selection (Section 1.3.2, page 22)

**COP** Cryptogenic Organizing Pneumonia (Section 1.5, page 19)

**COPD** Chronic Obstructive Pulmonary Disease (Section 1.3.2, page 20)

**CT** Computed Tomography (Section 1.1, page 2)

**CV** Cross–Validation (Section 5.2.1, page 93)

**DICOM** Digital Imaging and COmmunications in Medicine (Section 1.1, page 2)

**DIP** Desquamative Interstitial Pneumonia (Section 1.5, page 19)

**DWF** Discrete Wavelet Frame (Section 3.1.2, page 43)

**ECG** ElectroCardioGram (Section 1.1, page 2)

**EEG** ElectroEncephaloGram (Section 1.1, page 2)

**EHR** Electronic Health Record (Section 1, page 1)

**EM** Expectation–Maximization (Section 3.1.2, page 48)

**EMD** Earth’s Mover Distance (Section 1.3.2, page 23)

**EMG** ElectroMyoGram (Section 1.1, page 2)

**EP** Eosinophilic Pneumonia (Section 1.5, page 19)

**EU** European Union (Section 2, page 31)

**FDA** Food and Drug Administration (Section 5.2.3, page 97)

**fMRI** functional Magnetic Resonance Imaging (Section 1.1, page 2)

**FN** False Negative (Section 5.2.2, page 94, Eq. 5.4)

**FP** False Positive (Section 5.2.2, page 94, Eq. 5.3)

**GLCM** Gray–Level Co–occurrence Matrices (Section 1.2.1, page 10)

**GLH** Gray–level histograms (Section 3.1.1, page 42)

**GMM** Gaussian Mixture Models (Section 3.1.2, page 48)

**GP** Generalization Performance (Section 1.2.2, page 13)

**GUI** Graphical User Interface (Section 1.1.1, page 3)

**HON** Health On the Net (Section 4.3, page 80)

**HP** Hypersensitivity Pneumonitis (Section 1.5, page 19)

**HRCT** High–Resolution Computed Tomography (Section 1.3.1, page 18)

**HTML** HyperText Markup Language (Section 2.2.2, page 32)

**HU** Hounsfield Units (Section 1.3.1, page 17)

**HUG** Hôpitaux Universitaires de Genève (Section 1, page 1)

**IAPR** International Association for Pattern Recognition (Section 4.2.3, page 76)

**ICA** Independent Component Analysis (Section 1.3.2, page 23)

**ILD** Interstitial Lung Disease (Section 1.3.1, page 16)

**ImageCLEF** Image Cross Language Evaluation Forum (Section 1.3.2, page 25)

**IPF** Idiopathic Pulmonary Fibrosis (Section 2.1, page 31)

**IT** Information Technology (Section 2, page 31)

**JPEG** Joint Expert Picture Group (Section 1.3.2, page 22)

**k–NN** k–Nearest Neighbor (Section 1.2.2, page 14)

**LBP** Local Binary Patterns (Section 1.3.2, page 24)

**LCH** Langerhans Cell Histiocytosis (Section 1.5, page 19)

**LDA** Linear Discriminant Analysis (Section 1.3.2, page 22)

**LGN** Lateral Geniculate Nucleus (Section 1.2, page 5)

**LIDC** Lung Imaging Database Consortium (Section 2.1, page 31)

**LIP** Lymphocytic Interstitial Pneumonia (Section 1.5, page 19)

**LMIK** Learning Medical Image Knowledge (Section 2.1, page 32)

**LOPO** Leave–One–Patient–Out cross–validation (Section 5.2.1, page 93)

**LOO** Leave–One–Out cross–validation (Section 5.2.1, page 93)

**LORO** Leave–One–ROI–Out cross–validation (Section 5.2.1, page 93)

**LTRC** Lung Tissue Research Consortium (Section 2.1, page 31)

- MAP** Mean Average Precision (Section 4.2.1, page 74)
- MCS** Multiple Classifier Systems (Section 4.2.3, page 76)
- MD** Medical Doctor (Section 2.3.3, page 36)
- MDCT** MultiDetector Computed Tomography (Section 1.3.1, page 17)
- MEG** MagnetoEncephaloGram (Section 1.1, page 2)
- MeSH** Medical Subject Headings (Section 2.2.2, page 32)
- ML** Machine Learning (Section 1.4.2, page 28)
- MLP** Multi-Layer Perceptron (Section 1.2.2, page 14)
- MRI** Magnetic Resonance Imaging (Section 1.1, page 2)
- NBIA** National Biomedical Image Archive (Section 2.1, page 31)
- NHLBI** National Heart, Lung, and Blood Institute (Section 2.1, page 31)
- NIH** National Institutes of Health (Section 1.3.2, page 25)
- NSIP** Non-Specific Interstitial Pneumonia (Section 1.5, page 19)
- PACS** Picture Archiving and Communication Systems (Section 1, page 1)
- PCA** Principal Component Analysis (Section 1.3.2, page 23)
- PCP** Pneumocystis Pneumonia (Section 1.5, page 19)
- PET** Positron Emission Tomography (Section 1.1, page 2)
- PF** Pulmonary Fibrosis (Section 1.5, page 19)
- PFT** Pulmonary Function Tests (Section 1.3.2, page 24)
- PTA( $l,r$ )** Plus  $l$ -Take Away  $r$  (Section 1.3.2, page 24)
- QDA** Quadratic Discriminant Analysis (Section 1.3.2, page 23)
- QMR** Quick Medical Reference (Section 1.1, page 2)
- QWF** Quincunx Wavelet Frame (Section 3.1.2, page 47)
- RB-ILD** Respiratory Bronchiolitis associated ILD (Section 1.5, page 19)
- RIA** Rich Internet Application (Section 5.1, page 83)
- RLE** Run-Length Encoding (Section 1.2.1, page 10)
- ROC** Receiver Operating Characteristics (Section 5.2.2, page 96)
- ROI** Region Of Interest (Section 1.1.1, page 3)
- RSNA** Radiological Society of North America (Section 1.1, page 2)
- RWF** Rotated Wavelet Frames (Section 1.3.2, page 24)
- SADH** Sum And Difference Histogram (Section 1.3.2, page 24)
- SBS** Sequential Backward Selection (Section 1.3.2, page 24)
- SFS** Sequential Forward Selection (Section 1.3.2, page 22)

**SFFS** Sequential Floating Forward Selection (Section 1.3.2, page 24)

**SFBS** Sequential Floating Backward Selection (Section 1.3.2, page 24)

**SOM** Self–Organizing Maps (Section 1.3.2, page 22)

**SNOMED–CT** Systematized Nomenclature of Medicine – Clinical Terms (Section 2.2.2, page 32)

**SPECT** Single Photon Emission Computed Tomography (Section 1.1, page 2)

**SVM** Support Vector Machines (Section 1.2.2, page 14)

**TALISMAN** Texture Analysis of Lung ImageS for Medical diagnostic AssistaNce (Section 2, page 31)

**TB** Tuberculosis (Section 1.5, page 19)

**TN** True Negative (Section 5.2.2, page 94, Eq. 5.2)

**TP** True Positive (Section 5.2.2, page 94, Eq. 5.1)

**UMLS** Unified Medical Language System (Section 4.2.1, page 74)

**URL** Uniform Resource Locator (Section 5.1, page 84)

**US** Ultrasound Imaging (Section 1.1, page 2)

**VOI** Volume Of Interest (Section 1.3.2, page 24)

**VTK** Visualization ToolKit (Section 5.1, page 83)

**WT** Wavelet Transform (Section 1.2.1, page 11)

**XOR** eXclusive OR (Section 3.2.1, page 60)

**YaDiV** Yet Another DIcom Viewer (Section 2.3.2, page 34)

# List of Figures

1.1	Average number of images produced per day at the HUG during the last decade . . . . .	1
1.2	General scheme of CAD for event detection and quantification in medical image series . . . . .	3
1.3	General scheme of CBIR-based CAD . . . . .	4
1.4	Visual pathways in the human visual system . . . . .	5
1.5	Organization of cone cells in the foveal region . . . . .	7
1.6	Affine transformations of 3D objects and textures . . . . .	8
1.7	Shift invariance of the human visual system . . . . .	9
1.8	Shapes of $\psi_{fovea}$ . . . . .	9
1.9	An example of GLCM . . . . .	11
1.10	Gray-level histogram of an axial slice of HRCT of the lung . . . . .	13
1.11	Learning and recognition: generalization . . . . .	13
1.12	Non-linear decision boundaries . . . . .	15
1.13	Most common histological diagnoses of ILDs . . . . .	17
1.14	Hounsfield Units values of several organic materials . . . . .	18
1.15	A comparison of HRCT and MRI for the visual assessment of the lung tissue . . . . .	20
1.16	Visual aspects of the most common lung tissue patterns in HRCT of patients with ILDs . . . . .	21
1.17	Popularity of the various feature extraction techniques used by 27 research groups for texture analysis of the lung tissues . . . . .	26
2.1	Popular datasets for evaluation of image processing and machine learning algorithms . . . . .	30
2.2	A typical example of a non-accurately delineated ROIs . . . . .	34
2.3	A screen shot of the graphical tool for the annotation of image regions . . . . .	35
2.4	An example of the text-based file format to store the ROIs in the database . . . . .	35
2.5	Distribution of the histological diagnoses . . . . .	37
2.6	Distribution of the various lung tissue patterns in terms of hand-drawn ROIs . . . . .	38
2.7	Distribution of the various lung tissue patterns in terms of volumes of annotated tissue . . . . .	39
2.8	Age at onset presentation in LCH, sarcoidosis and PF . . . . .	40
3.1	Two iterations of a standard separable dyadic wavelet decomposition . . . . .	43
3.2	Loss of the shift-invariance property of the WT on an uniform sampling grid . . . . .	43
3.3	Two iterations of DWF . . . . .	44
3.4	Shapes of the 2D isotropic polyharmonic $B$ -spline $\hat{\beta}_\gamma(\omega)$ of various orders in the Fourier domain . . . . .	45
3.5	Dyadic and quincunx lattices . . . . .	46
3.6	GMM to model the distribution of wavelet coefficients within one subband . . . . .	48
3.7	Gray-level histograms of the most frequent lung tissue patterns . . . . .	49
3.8	Classification accuracies for varying number of GLH bins . . . . .	49
3.9	Overall accuracies at iterative scales . . . . .	51
3.10	Comparison of accuracies using GLH plus air percentage, the DWF and the combination of all features . . . . .	51
3.11	An example of the segmentation of the lung volumes using YaDiV . . . . .	54

3.12	Construction of the feature vector $\mathbf{v}$	54
3.13	An example of the three-dimensional categorization of the lung tissues	55
3.14	Intra-class variations for class <i>healthy</i>	57
3.15	2D and 3D maps of the lung tissue as diagnostic aid tools for the radiologists	59
3.16	XOR configurations leading to highly non-linear decision boundaries	60
3.17	Correlation of $I_{Gratio}$ and $A^{single}$	62
3.18	Correlation matrix of the multimodal feature space	63
4.1	Methodology for comparing the classifiers performance	68
4.2	Grid search for optimal SVMs parameters $C$ and $\sigma_K$	69
4.3	McNemar-based classifiers' scores with the 6-class configuration	71
4.4	Accuracies of the classifiers with the 6-class configuration	71
4.5	McNemar-based classifiers' scores with the 2-class configuration	71
4.6	Accuracies of the classifiers with the 2-class configuration	71
4.7	Stability of the classifier parameters	72
4.8	Influence of the age on the visual aspect of the lung tissue	75
4.9	Classification scheme for late fusion	77
4.10	Accuracies for class <i>healthy</i>	79
4.11	Accuracies for class <i>emphysema</i>	79
4.12	Accuracies for class <i>ground glass</i>	79
4.13	Accuracies for class <i>fibrosis</i>	79
4.14	Accuracies for class <i>micronodules</i>	79
4.15	Accuracies for class the multiclass configuration	79
4.16	Testing accuracies $A_{mean}^{test}$ with a varying number of clinical parameters $N_{clin}$	79
5.1	Transfer function of a CT scan of the head	84
5.2	Welcome menus of the RIA	85
5.3	Workflow of the RIA for the navigation in the database using exact text search	86
5.4	Form containing the clinical parameters for querying ILD cases using exact text search	86
5.5	List of cases returned by an exact test search	87
5.6	Detailed view of a selected case: clinical parameters	87
5.7	Detailed view of a selected case: image series and annotations	88
5.8	Main components of the hybrid CAD system	89
5.9	Workflow of the RIA for the lung tissue categorization of a new case and the retrieval of similar patients	89
5.10	Selection of the clinical parameters for case-based retrieval	90
5.11	List of cases returned by the multimodal query	90
5.12	Segmentation of the lung volumes using YaDiV	91
5.13	Segmentation of the lung tissue using YaDiV	92
5.14	Cross-validation	93
5.15	LORO and LOPO validation schemes	94
5.16	ROC curves	96
5.17	Common missclassifications near the lung border and the mediastinum	98
B.1	HTML form used to capture the clinical parameters (page 1/4)	115
B.2	HTML form used to capture the clinical parameters (page 2/4)	116
B.3	HTML form used to capture the clinical parameters (page 3/4)	117
B.4	HTML form used to capture the clinical parameters (page 4/4)	118

# List of Tables

1.1	Most common signals and imaging techniques in medicine . . . . .	2
1.2	Number of papers presented at the Annual Meeting of RSNA related to CAD from year 2000 to 2005 . . . . .	3
1.3	Analogies between computer vision and human vision . . . . .	6
1.4	HRCT scanning protocol . . . . .	18
1.5	13 common histological diagnoses of ILDs and associated HRCT findings . . . . .	19
2.1	Terminology used to describe the locations of HRCT findings . . . . .	34
2.2	Distribution, mean age and gender statistics of the diagnoses . . . . .	37
2.3	Distribution of the lung tissue patterns . . . . .	38
3.1	Desirable properties of the combination of isotropic polyharmonic $B$ -spline wavelets and the quincunx lattice along with a redundant frame transform . . . . .	47
3.2	Dataset used to determine the optimal $N_{bins}$ and to evaluate DWF features . . . . .	50
3.3	Confusion matrix of the lung tissue ROIs using DWF combined with GLH . . . . .	51
3.4	Performance measures associated with the confusion matrix in Table 3.3 . . . . .	51
3.5	Dataset used for the comparison between DWF and QWF features . . . . .	52
3.6	Mean accuracies in % with experiments repeated 30 times. Isotropic polyharmonic $B$ -spline wavelets (QWF) with order $\gamma = 3$ allowed a mean of 94.3% of correct predictions among the six lung tissue classes with high precision (geometric mean = 89%). Best accuracies are marked in bold. . . . .	53
3.7	Confusion matrix of the blockwise classification of lung tissue patterns using QWF and GLH features . . . . .	54
3.8	Performance measures of the blockwise classification of the lung tissue patterns using QWF and GLH features . . . . .	55
3.9	Dataset used to rank the clinical parameters and to evaluate the influence of the clinical context on lung tissue categorization . . . . .	61
3.10	List of the first 20 clinical attributes with highest $A^{single}$ when combined with visual features . . . . .	64
4.1	Dataset used to compare the classifier performances . . . . .	68
4.2	Grid search for optimal parameters $p_i^{opt}$ . . . . .	69
4.3	Class-specific accuracies for each classifier family . . . . .	70
4.4	Class-specific accuracies for each classifier family with a 2-class configuration . . . . .	70
4.5	Averaged accuracies obtained with the various techniques . . . . .	78
4.6	Mean precisions based on the diagnosis of the retrieved cases . . . . .	82
5.1	Confusion matrix . . . . .	95
A.1	Studies in computerized lung tissue analysis: research groups and scopes (1997–2001)	106
A.2	Studies in computerized lung tissue analysis: research groups and scopes (2003–2004)	107
A.3	Studies in computerized lung tissue analysis: research groups and scopes (2005–2007)	108
A.4	Studies in computerized lung tissue analysis: research groups and scopes (2007–2009)	109

A.5 Studies in computerized lung tissue analysis: methods (1997–2001) . . . . .	110
A.6 Studies in computerized lung tissue analysis: methods (2003–2004) . . . . .	111
A.7 Studies in computerized lung tissue analysis: methods (2005–2007) . . . . .	112
A.8 Studies in computerized lung tissue analysis: methods (2007–2009) . . . . .	113

# Bibliography

- [1] H. Abe, K. Ashizawa, F. Li, N. Matsuyama, A. Fukushima, J. Shiraishi, H. MacMahon, and K. Doi. Artificial neural networks (ANNs) for differential diagnosis of interstitial lung disease: results of a simulation test with actual clinical cases. *Academic Radiology*, 11(1):29–37, June 2004.
- [2] H. Abe, H. MacMahon, R. Engelmann, Q. Li, J. Shiraishi, S. Katsuragawa, M. Aoyama, T. Ishida, K. Ashizawa, C. E. Metz, and K. Doi. Computer-aided diagnosis in chest radiography: Results of large-scale observer tests at the 1996–2001 RSNA scientific assemblies. *RadioGraphics*, 23(1):255–265, January 2003.
- [3] H. Adams, M. S. Bernard, and K. McConnochie. An appraisal of CT pulmonary density mapping in normal subjects. *Clinical Radiology*, 43(4):238–242, April 1991.
- [4] A. M. Aisen, L. S. Broderick, H. Winer-Muram, C. E. Brodley, A. C. Kak, C. Pavlopoulou, J. Dy, C.-R. Shyu, and A. Marchiori. Automated storage and retrieval of thin-section CT images to assist diagnosis: System description and preliminary assessment. *Radiology*, 228(1):265–270, July 2003.
- [5] H. C. Andrews. *Introduction to mathematical techniques in pattern recognition*. John Wiley & Sons Inc, September 1972.
- [6] S. Arivazhagan and L. Ganesan. Texture classification using wavelet transform. *Pattern Recognition Letters*, 24(9-10):1513–1521, June 2003.
- [7] S. G. Armato, G. McLennan, M. F. McNitt-Gray, C. R. Meyer, D. Yankelevitz, D. R. Aberle, C. I. Henschke, E. A. Hoffman, E. A. Kazerooni, H. MacMahon, A. P. Reeves, B. Y. Croft, and L. P. Clarke. Lung image database consortium: Developing a resource for the medical imaging research community. *Radiology*, 232(3):739–748, September 2004.
- [8] S. G. Armato, M. F. McNitt-Gray, A. P. Reeves, C. R. Meyer, G. McLennan, D. R. Aberle, E. A. Kazerooni, H. MacMahon, E. J. R. van Beek, D. Yankelevitz, E. A. Hoffman, C. I. Henschke, R. Y. Roberts, M. S. Brown, R. M. Engelmann, R. C. Pais, C. W. Piker, D. Qing, M. Kocherginsky, B. Y. Croft, and L. P. Clarke. The lung image database consortium (LIDC): An evaluation of radiologist variability in the identification of lung nodules on CT scans. *Academic Radiology*, 14(11):1409–1421, November 2007.
- [9] Z. A. Aziz, A. U. Wells, D. M. Hansell, G. A. Bain, S. J. Copley, S. R. Desai, S. M. Ellis, F. V. Gleeson, S. Grubnic, A. G. Nicholson, S. P. Padley, K. S. Pointon, J. H. Reynolds, R. J. Robertson, and M. B. Rubens. HRCT diagnosis of diffuse parenchymal lung disease: inter-observer variation. *Thorax*, 59(6):506–511, June 2004.
- [10] F. R. Bach, G. R. G. Lanckriet, and M. I. Jordan. Multiple kernel learning, conic duality, and the SMO algorithm. In *ICML '04: Proceedings of the twenty-first international conference on Machine learning*, page 6, New York, NY, USA, July 2004. ACM.
- [11] R. Bajcsy. Computer description of textured surfaces. In *Proceedings of the 3rd International Joint Conference on Artificial Intelligence*, pages 572–579, August 1973.

- [12] C. Ballester and M. González. Affine invariant texture segmentation and shape from texture by variational methods. *Journal of Mathematical Imaging and Vision*, 9(2):141–171, September 1998.
- [13] V. Barra and J.-Y. Boire. Tissue segmentation on MR images of the brain by possibilistic clustering on a 3D wavelet representation. *Journal of Magnetic Resonance Imaging*, 11(3):167–278, March 2000.
- [14] A. J. Bell. The co-information lattice. In *Proc. 4th International Symposium on Independent Component Analysis and Blind Signal Separation (ICA2003)*, pages 921–926, Nara, Japan, April 2003. Springer.
- [15] R. Benmokhtar and B. Huet. Classifier fusion: Combination methods for semantic indexing in video content. In *ICANN 2006: International Conference on Artificial Neural Networks*, pages 65–74, Athens, Greece, September 2006. Springer.
- [16] I. Biedermann. Recognition-by-components: A theory of human image understanding. *Psychological Review*, 94(2):115–147, April 1987.
- [17] R. L. Birdwell, P. Bandodkar, and D. M. Ikeda. Computer-aided detection with screening mammography in a university hospital setting. *Radiology*, 236(2):451–457, August 2005.
- [18] C. M. Bishop. *Neural networks for pattern recognition*. Clarendon Press, Oxford, 1995.
- [19] C. M. Bishop. *Pattern Recognition and Machine Learning*. Springer, August 2006.
- [20] H. Boehm, C. Fink, U. Attenberger, C. Becker, J. Behr, and M. Reiser. Automated classification of normal and pathologic pulmonary tissue by topological texture features extracted from multi-detector CT in 3D. *European Radiology*, 18(12):2745–2755, December 2008.
- [21] P. Bouthemy and R. Fablet. Motion characterization from temporal cooccurrences of local motion-based measures for video indexing. In *Fourteenth International Conference on Pattern Recognition*, volume 1, pages 905–908, August 1998.
- [22] P. Brodatz. *Textures: A Photographic Album for Artists and Designers*. Dover Publications, New York, December 1966.
- [23] C. Brodley, A. Kak, C. Shyu, J. Dy, L. Broderick, and A. M. Aisen. Content-based retrieval from medical image databases: A synergy of human interaction, machine learning and computer vision. In *Proceedings of the 10th National Conference on Artificial Intelligence*, pages 760–767, Orlando, FL, USA, July 1999.
- [24] C. Burges, T. Shaked, E. Renshaw, A. Lazier, M. Deeds, N. Hamilton, and G. Hullender. Learning to rank using gradient descent. In *ICML '05: Proceedings of the 22nd international conference on Machine learning*, pages 89–96, New York, NY, USA, August 2005. ACM.
- [25] C. J. C. Burges. A tutorial on support vector machines for pattern recognition. *Data Mining and Knowledge Discovery*, 2(2):121–167, June 1998.
- [26] J. J. Caban, J. Yao, N. A. Avila, J. R. Fontana, and V. C. Manganiello. Texture-based computer-aided diagnosis system for lung fibrosis. In M. L. Giger and N. Karssemeijer, editors, *Medical Imaging 2007: Computer-Aided Diagnosis*, volume 6514, page 651439. SPIE, February 2007.
- [27] F. W. Campbell and R. W. Gubish. Optical quality of the human eye. *The Journal of Physiology*, 186(3):558–578, October 1966.
- [28] E. C. Caritá, E. Seraphim, M. O. Honda, and P. Mazzoncini de Azevedo-Marques. Implementation and evaluation of a medical image management system with content-based retrieval support. *Radiologia Brasileira*, 41(5):331–336, October 2008.

- [29] M. L. Cascia, S. Sethi, and S. Sclaroff. Combining textual and visual cues for content-based image retrieval on the world wide web. In *Content-Based Access of Image and Video Libraries, 1998. Proceedings. IEEE Workshop on*, pages 24–28, Washington, DC, USA, June 1998. IEEE Computer Society.
- [30] G. Castellano, L. Bonilha, L. M. Li, and F. Cendes. Texture analysis of medical images. *Clinical Radiology*, 59(12):1061–1069, December 2004.
- [31] R. A. Castellino. Computer aided detection (CAD): an overview. *Cancer Imaging*, 5(1):17–19, August 2005.
- [32] F. Chabat, D. M. Hansell, and G.-Z. Yang. Computerized decision support in medical imaging. *Engineering in Medicine and Biology Magazine, IEEE*, 19(5):89–96, September 2000.
- [33] F. Chabat, G.-Z. Yang, and D. M. Hansell. Obstructive lung diseases: Texture classification for differentiation at CT. *Radiology*, 228(3):871–877, September 2003.
- [34] A. Chambellan, E. Chailleux, and T. Similowski. Prognostic value of the hematocrit in patients with severe COPD receiving long-term oxygen therapy. *Chest*, 128(3):1201–1208, September 2005.
- [35] C. C. Chang and C. J. Lin. *LIBSVM: a library for support vector machines*, 2001.
- [36] C.-L. Chang, A. Maleki, and B. Girod. Adaptive wavelet transform for image compression via directional quincunx lifting. In *Multimedia Signal Processing, 2005 IEEE 7th Workshop on*, pages 1–4, November 2005.
- [37] A. Chebira. *Adaptive multiresolution frame classification of biomedical images*. PhD thesis, Carnegie Mellon University, 2008.
- [38] Q. Chen, G. Li, T.-Y. Leong, and C.-K. Heng. Predicting coronary artery disease with medical profile and gene polymorphisms data. In K. A. Kuhn, J. R. Warren, and T.-Y. Leong, editors, *Medinfo 2007: Proceedings of the 12th World Congress on Health (Medical) Informatics*, pages 1219–1224, Brisbane, Australia, August 2007. IOS Press.
- [39] W. Chen, M. L. Giger, H. Li, U. Bick, and G. M. Newstead. Volumetric texture analysis of breast lesions on contrast-enhanced magnetic resonance images. *Magnetic Resonance in Medicine*, 58(3):562–571, September 2007.
- [40] M. Chiyo, Y. Sekine, T. Iwata, K. Tatsumi, K. Yasufuku, A. Iyoda, M. Otsuji, S. Yoshida, K. Shibuya, T. Iizasa, Y. Saitoh, and T. Fujisawa. Impact of interstitial lung disease on surgical morbidity and mortality for lung cancer: analyses of short-term and long-term outcomes. *Journal of Thoracic and Cardiovascular Surgery*, 126(4):1141–1146, October 2003.
- [41] O. Chomat and J. L. Crowley. Probabilistic recognition of activity using local appearance. In *Conference on Computer Vision and Pattern Recognition (CCVP 1999)*, volume 2, pages 104–109, June 1999.
- [42] G. Cohen, M. Hilario, H. Sax, S. Hugonnet, and A. Geissbuhler. Learning from imbalanced data in surveillance of nosocomial infection. *Artificial Intelligence in Medicine*, 37(1):7–18, May 2006.
- [43] C. Cortes and V. Vapnik. Support-vector networks. *Machine Learning*, 20(3):273–297, sep 1995.
- [44] T. Cover and P. Hart. Nearest neighbor pattern classification. *IEEE Transactions on Information Theory*, 13(1):21–27, January 1967.

- [45] F. H. C. Crick, D. C. Marr, and T. Poggio. An information processing approach to understanding the visual cortex. Technical report, Massachusetts Institute of Technology, 1980.
- [46] C. A. Curcio, K. R. Sloan, R. E. Kalina, and A. E. Hendrickson. Human photoreceptor topography. *The Journal of Comparative Neurology*, 292(4):497–523, February 1990.
- [47] S. Delorme, M.-A. Keller-Reichenbecher, I. Zuna, W. Schlegel, and G. Van Kaick. Usual interstitial pneumonia: Quantitative assessment of high-resolution computed tomography findings by computer-assisted texture-based image analysis. *Investigative Radiology*, 32(9):566–574, September 1997.
- [48] A. P. Dempster, N. M. Laird, and D. B. Rubin. Maximum likelihood from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society. Series B (Methodological)*, 39(1):1–38, 1977.
- [49] A. Depeursinge, J. Iavindrasana, H. Andriambololoniaaina, P.-A. Poletti, A. Platon, A. Geissbuhler, and H. Müller. A framework for diagnosing interstitial lung diseases in HRCT : the TALISMAN project. *Swiss Medical Informatics*, 64:17–20, 2008.
- [50] A. Depeursinge, J. Iavindrasana, G. Cohen, A. Platon, P.-A. Poletti, and H. Müller. Computer-aided diagnostic for interstitial lung diseases in HRCT: the talisman project. In *Swiss Conference on Medical Informatics*, Sierre, Switzerland, June 2008.
- [51] A. Depeursinge, J. Iavindrasana, G. Cohen, A. Platon, P.-A. Poletti, and H. Müller. Lung tissue classification in HRCT data integrating the clinical context. In *21th IEEE Symposium on Computer-Based Medical Systems (CBMS)*, pages 542–547, Jyväskylä, Finland, June 2008.
- [52] A. Depeursinge, J. Iavindrasana, A. Hidki, G. Cohen, A. Geissbuhler, A. Platon, P.-A. Poletti, and H. Müller. A classification framework for lung tissue categorization. In K. P. Andriole and K. M. Siddiqui, editors, *Medical Imaging 2008: PACS and Imaging Informatics*, volume 6919, page 69190C, San Diego, CA, USA, 2008. SPIE.
- [53] A. Depeursinge, J. Iavindrasana, A. Hidki, G. Cohen, A. Geissbuhler, A. Platon, P.-A. Poletti, and H. Müller. Comparative performance analysis of state-of-the-art classification algorithms applied to lung tissue categorization. *Journal of Digital Imaging*, 23(1):18–30, 2010.
- [54] A. Depeursinge, H. Müller, A. Hidki, A. Platon, and P.-A. Poletti. Computerized assistance for diagnosing interstitial lung disease in emergency radiology. In *Proceedings of the annual congress of the Swiss society of radiology*, Basel, Switzerland, June 2007.
- [55] A. Depeursinge, H. Müller, A. Hidki, P.-A. Poletti, A. Platon, and A. Geissbuhler. Image-based diagnostic aid for interstitial lung disease with secondary data integration. In M. L. Giger and N. Karssemeijer, editors, *Medical Imaging 2007: Computer-Aided Diagnosis*, volume 6514, page 65143P. SPIE, February 2007.
- [56] A. Depeursinge, H. Müller, A. Hidki, P.-A. Poletti, T. Rochat, and A. Geissbuhler. Building a library of annotated pulmonary CT cases for diagnostic aid. In *Swiss conference on medical informatics (SSIM 2006)*, Basel, Switzerland, April 2006.
- [57] A. Depeursinge, D. Racoceanu, J. Iavindrasana, G. Cohen, A. Platon, P.-A. Poletti, and H. Müller. Fusing visual and clinical information for lung tissue classification in HRCT data. *Journal of Artificial Intelligence in Medicine*, 2010–to appear.
- [58] A. Depeursinge, D. Sage, A. Hidki, A. Platon, P.-A. Poletti, M. Unser, and H. Müller. Lung tissue classification using Wavelet frames. In *Engineering in Medicine and Biology Society, 2007. EMBS 2007. 29th Annual International Conference of the IEEE*, pages 6259–6262, Lyon, France, August 2007. IEEE Computer Society.

- [59] A. Depeursinge, D. Van De Ville, M. Unser, and H. Müller. Lung tissue analysis using isotropic polyharmonic B-spline wavelets. In *MICCAI 2008 Workshop on Pulmonary Image Analysis*, pages 125–134, New York, USA, September 2008.
- [60] A. Depeursinge, A. Vargas, A. Platon, A. Geissbuhler, P.-A. Poletti, and H. Müller. 3D case-based retrieval for interstitial lung diseases. In B. Caputo and H. Müller, editors, *MCBR-CDS 2009: Medical Content-based Retrieval for Clinical Decision Support*, MICCAI 2009, the 12th International Conference on Medical Image Computing and Computer Assisted Intervention, pages 18–25. Springer Lecture Notes in Computer Science, September 2010.
- [61] L. Dettori and L. Semler. A comparison of wavelet, ridgelet, and curvelet-based texture classification algorithms in computed tomography. *Computers in Biology and Medicine*, 37(4):486–498, April 2007. Wavelet-based Algorithms for Medical Problems.
- [62] E. Diamant. Modeling human-like intelligent image processing: An information processing perspective and approach. *Signal Processing: Image Communication*, 22(6):583–590, September 2007.
- [63] T. G. Dietterich. Approximate statistical test for comparing supervised classification learning algorithms. *Neural Computation*, 10(7):1895–1923, October 1998.
- [64] T. G. Dietterich. Ensemble methods in machine learning. In *MCS '00: Proceedings of the First International Workshop on Multiple Classifier Systems*, pages 1–15, London, UK, June 2000. Springer-Verlag.
- [65] M. N. Do and M. Vetterli. Rotation invariant texture characterization and retrieval using steerable wavelet-domain hidden markov models. *IEEE Transactions on Multimedia*, 4(4):517–527, December 2002.
- [66] K. Doi. Current status and future potential of computer-aided diagnosis in medical imaging. *British Journal of Radiology*, 78:3–19, 2005.
- [67] K. Doi. Computer-aided diagnosis in medical imaging: Historical review, current status and future potential. *Computerized Medical Imaging and Graphics*, 31(4–5):198–211, March 2007.
- [68] J. S. Duncan and N. Ayache. Medical image analysis: Progress over two decades and the challenges ahead. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 22(1):85–106, January 2000.
- [69] M. Dundar, G. Fung, L. Bogoni, M. Macari, A. Megibow, and B. Rao. A methodology for training and validating a CAD system and potential pitfalls. *International Congress Series*, 1268:1010–1014, June 2004. CARS 2004 – Computer Assisted Radiology and Surgery. Proceedings of the 18th International Congress and Exhibition.
- [70] I. Eggel and H. Müller. Combination of visual similarity retrieval & textual retrieval from medical documents. In *The Medical Informatics Europe Conference (MIE 2009)*, pages 841–845, Sarajevo, Bosnia-Herzegovina, August 2009.
- [71] E. El-Kwae, H. Xu, and M. R. Kabuka. Content-based retrieval in picture archiving and communication systems. *Journal of Digital Imaging*, 13(2):70–81, May 2000.
- [72] B. Elger, J. Iavindrasana, L. Lo Iacono, H. Müller, N. Roduit, P. Summers, and J. Wright. Health data depersonalisation for prospective research in the life sciences. *Computer Methods and Programs in Biomedicine*, 2010 – submitted.
- [73] R. L. Engle. Attempts to use computers as diagnostic aids in medical decision making: a thirty-year experience. *Perspectives in biology and medicine*, 35(2):207–219, February 1992.

- [74] C. Enroth-Cugell and J. G. Robson. The contrast sensitivity of retinal ganglion cells of the cat. *The Journal of Physiology*, 187(3):517–552, December 1966.
- [75] M. P. Feilner, D. Van De Ville, and M. Unser. An orthogonal family of quincunx wavelets with continuously adjustable order. *IEEE Transactions on Image Processing*, 14(4):499–510, April 2005.
- [76] M. Fernández and A. Mavilio. Texture analysis of medical images using the wavelet transform. In *MEDICAL PHYSICS: Sixth Mexican Symposium on Medical Physics*, volume 630, pages 164–168. AIP, October 2002.
- [77] C. Fetita, K.-C. Chang-Chien, P.-Y. Brillet, F. Prêteux, and P. Grenier. Diffuse parenchymal lung diseases: 3D automated detection in MDCT. In S. LNCS, editor, *Proceedings of MICCAI 2007*, volume 4791, pages 825–833, November 2007.
- [78] K. R. Flaherty, T. E. King, G. Raghu, J. P. Lynch, T. V. Colby, W. D. Travis, B. H. Gross, E. A. Kazerooni, G. B. Toews, Q. Long, S. Murray, V. N. Lama, S. E. Gay, and F. J. Martinez. Idiopathic interstitial pneumonia: what is the effect of a multidisciplinary approach to diagnosis? *American Journal of Respiratory and Critical Care Medicine*, 170(8):904–910, October 2004.
- [79] E. Frank, M. A. Hall, G. Holmes, R. Kirkby, B. Pfahringer, I. H. Witten, and L. Trigg. Weka - a machine learning workbench for data mining. In O. Maimon and L. Rokach, editors, *The Data Mining and Knowledge Discovery Handbook*, pages 1305–1314. Springer, 2005.
- [80] T. W. Freer and M. J. Ulisseyy. Screening mammography with computer-aided detection: Prospective study of 12,860 patients in a community breast center. *Radiology*, 220(3):781–786, September 2001.
- [81] C. P. Friedman, A. S. Elstein, F. M. Wolf, G. C. Murphy, T. M. Franz, P. S. Heckerling, P. L. Fine, T. M. Miller, and V. Abraham. Enhancement of clinician’s diagnostic reasoning by computer-based consultation. *Journal of the American Medical Association*, 282(19):1851–1856, November 1999.
- [82] L. Gallo, G. De Pietro, A. Coronato, and I. Marra. Toward a natural interface to virtual medical imaging environments. In *AVI ’08: Proceedings of the working conference on Advanced visual interfaces*, pages 429–432, New York, NY, USA, May 2008. ACM.
- [83] S. Gordon, G. Zimmerman, and H. Greenspan. Image segmentation of uterine cervix images for indexing in PACS. In *CBMS ’04: Proceedings of the 17th IEEE Symposium on Computer-Based Medical Systems*, pages 298–304, Washington, DC, USA, June 2004. IEEE Computer Society.
- [84] H. Gunes and M. Piccardi. Affect recognition from face and body: early fusion vs. late fusion. In *2005 IEEE International Conference on Systems, Man and Cybernetics*, volume 4, pages 3437–3443, Big Island, Hawaii, October 2005. IEEE Computer Society.
- [85] J. Guo, J. M. Reinhardt, H. Kitaoka, L. Zhang, M. Sonka, G. McLennan, and E. A. Hoffman. Integrated system for CT-based assessment of parenchymal lung disease. In *Biomedical Imaging, 2002. Proceedings. 2002 IEEE International Symposium on*, pages 871–874. IEEE Computer Society, June 2002.
- [86] A. Guyton and J. Hall. *Textbook of Medical Physiology*. Elsevier/Saunders, Philadelphia, 11th edition edition, June 2005.
- [87] R. M. Haralick. Statistical and structural approaches to texture. *Proceedings of the IEEE*, 67(5):786–804, May 1979.
- [88] R. M. Haralick, K. Shanmugam, and I. Dinstein. Textural features for image classification. *IEEE Transactions on Systems, Man and Cybernetics*, 3(6):610–621, November 1973.

- [89] P. G. Hartley, J. R. Galvin, G. W. Hunninghake, J. A. Merchant, S. J. Yagla, S. B. Speakman, and D. A. Schwartz. High-resolution CT-derived measures of lung density are valid indexes of interstitial lung disease. *Journal of Applied Physiology*, 76(1):271–277, January 1994.
- [90] K. R. Heitmann, H.-U. Kauczor, P. Mildenberger, T. Uthmann, J. Perl, and M. Thelen. Automatic detection of ground glass opacities on lung HRCT using multiple neural networks. *European Radiology*, 7(9):1463–1472, November 1997.
- [91] W. Hersh, J. Jensen, H. Müller, P. Gorman, and P. Ruch. Trans-atlantic collaboration for evaluating image retrieval systems in ImageCLEF biomedical image retrieval task. In *Workshop on NSF/FP6 cooperations at the 2005 eChallenges conference*, Ljubljana, Slovenia, October 2005.
- [92] W. Hersh, J. Kalpathy-Cramer, and J. Jensen. Medical image retrieval and automated annotation: OHSU at ImageCLEF 2006. In C. Peters, P. Clough, F. C. Gey, J. Karlgren, B. Magnini, D. W. Oard, M. de Rijke, and M. Stempfhuber, editors, *CLEF*, volume 4730 of *Lecture Notes in Computer Science*, pages 660–669. Springer, September 2007.
- [93] W. Hersh, H. Müller, J. Kalpathy-Cramer, and E. Kim. Consolidating the ImageCLEF medical task test collection: 2005-2007. In *MUSCLE/ImageCLEF workshop*, pages 31–39, Budapest, Hungary, May 2007.
- [94] A. Hidki, H. Müller, A. Depeursinge, P.-A. Poletti, and A. Geissbuhler. 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.
- [95] E. A. Hoffman, J. M. Reinhardt, M. Sonka, B. A. Simon, J. Guo, O. Saba, D. Chon, S. Samrahan, H. Shikata, J. Tschirren, K. Palagy, K. C. Beck, and G. McLennan. Characterization of the interstitial lung disease via density-based and texture-based analysis of computed tomography images of lung structure and function. *Academic Radiology*, 10(10):1104–1118, October 2003.
- [96] D. R. Holmes III, B. J. Bartholmai, R. A. Karwoski, V. Zavaletta, and R. A. Robb. The lung tissue research consortium: An extensive open database containing histological, clinical, and radiological data to study chronic lung disease. *Insight journal*, July 2006.
- [97] S. J. Hong. Use of contextual information for feature ranking and discretization. *IEEE Transactions on Knowledge and Data Engineering*, 9(5):718–730, September 1997.
- [98] A. Horsch, M. Prinz, S. Schneider, O. Sipilä, K. Spinnler, J.-P. Vallée, I. Verdonck-de Leeuw, R. Vogl, T. Wittenberg, and G. Zahlmann. Establishing an international reference image database for research and development in medical image processing. *Methods of Information in Medicine*, 43(4):409–412, February 2004.
- [99] T.-I. Hsu, A. D. Calway, and R. Wilson. Texture analysis using the multiresolution Fourier transform. In *Scandinavian Conference on Image Analysis*, pages 823–830, May 1993.
- [100] Y. Huang, L. Wang, and C. Li. Texture analysis of ultrasonic liver image based on wavelet transform and probabilistic neural network. In *BMEI '08: Proceedings of the 2008 International Conference on BioMedical Engineering and Informatics*, pages 248–252, Washington, DC, USA, May 2008. IEEE Computer Society.
- [101] D. H. Hubel and T. N. Wiesel. Receptive fields of single neurones in the cat's striate cortex. *The Journal of Physiology*, 148(3):574–591, October 1959.
- [102] P. Indyk and R. Motwani. Approximate nearest neighbors: towards removing the curse of dimensionality. In *STOC '98: Proceedings of the thirtieth annual ACM symposium on Theory of computing*, pages 604–613, New York, NY, USA, May 1998. ACM.

- [103] K. Jafari-Khouzani, H. Soltanian-Zadeh, K. Elisevich, and S. Patel. Comparison of 2D and 3D wavelet features for TLE lateralization. In A. A. Amini and A. Manduca, editors, *Medical Imaging 2004: Physiology, Function, and Structure from Medical Images*, volume 5369, pages 593–601. SPIE, February 2004.
- [104] A. K. Jain, R. P. W. Duin, and J. Mao. Statistical pattern recognition: A review. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 22(1):4–37, January 2000.
- [105] A. K. Jain, J. Mao, and K. M. Mohiuddin. Artificial neural networks: a tutorial. *Computer*, 29(3):31–44, March 1996.
- [106] Q. Ji, J. Engel, and E. Craine. Texture analysis for classification of cervix lesions. *IEEE Transactions on Medical Imaging*, 19(11):1144–1149, November 2000.
- [107] B. Julesz. Textons, the elements of texture perception, and their interactions. *Nature*, 290(5802):91–97, March 1981.
- [108] R. Kakinuma, H. Ohmatsu, M. Kaneko, K. Eguchi, T. Naruke, K. Nagai, Y. Nishiwaki, A. Suzuki, and N. Moriyama. Detection failures in spiral CT screening for lung cancer: Analysis of CT findings. *Radiology*, 212(1):61–66, July 1999.
- [109] R. A. Karwoski, B. Bartholmai, V. A. Zavaletta, D. Holmes, and R. A. Robb. Processing of CT images for analysis of diffuse lung disease in the lung tissue research consortium. In X. P. Hu and A. V. Clough, editors, *Medical Imaging 2008: Physiology, Function, and Structure from Medical Images*, volume 6916, page 691614. SPIE, February 2008.
- [110] H.-U. Kauczor, K. R. Heitmann, C. P. Heussel, D. Marwede, T. Uthmann, and M. Thelen. Automatic detection and quantification of ground-glass opacities on high-resolution CT using multiple neural networks: Comparison with a density mask. *American Journal of Roentgenology*, 175(5):1329–1334, November 2000.
- [111] D. Keysers, J. Dahmen, H. Ney, B. B. Wein, and T. M. Lehmann. A statistical framework for model-based image retrieval in medical applications. *Journal of Electronic Imaging*, 12(1):59–68, January 2003.
- [112] K. G. Kim, J. M. Goo, J. H. Kim, H. J. Lee, B. G. Min, K. T. Bae, and J.-G. Im. Computer-aided diagnosis of localized ground-glass opacity in the lung at CT: Initial experience. *Radiology*, 237(2):657–661, September 2005.
- [113] N. Kim, J. B. Seo, Y. Lee, J. G. Lee, S. S. Kim, and S. H. Kang. Development of an automatic classification system for differentiation of obstructive lung disease using HRCT. *Journal of Digital Imaging*, 22(2):136–148, April 2009.
- [114] T. E. King. *Approach to the adult with interstitial lung disease*. UpToDate, Waltham, MA, Denise S. Basow edition, 2008.
- [115] J. Kittler, M. Hatef, R. P. W. Duin, and J. Matas. On combining classifiers. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 20(3):226–239, March 1998.
- [116] J. Kludas, E. Bruno, and S. Marchand-Maillet. Information fusion in multimedia information retrieval. In *Proceedings of 5th international Workshop on Adaptive Multimedia Retrieval (AMR)*, volume 4918, pages 147–159, Paris, France, June 2008. ACM.
- [117] J. Kludas, E. Bruno, and S. Marchand-Maillet. Can feature information interaction help for information fusion in multimedia problems? *Multimedia Tools and Applications*, 42(1):57–71, March 2009.
- [118] N. Kodama, Y. Kawase, and K. Okamoto. Application of texture analysis to differentiation of dementia with lewy bodies from alzheimers disease on magnetic resonance images. In *World Congress on Medical Physics and Biomedical Engineering 2006*, pages 1444–1446. Springer, August 2007.

- [119] J. J. Koenderink. The structure of images. *Biological Cybernetics*, 50(5):363–370, August 1984.
- [120] M. M. Koh, Gerald Choon-Huat, P. Khoo, Hoon Eng, M. M. Wong, Mee Lian, and M. P. Koh, David. The effects of problem-based learning during medical school on physician competency: a systematic review. *Canadian Medical Association Journal*, 178(1):34–41, January 2008.
- [121] R. Kohavi. A study of cross-validation and bootstrap for accuracy estimation and model selection. In *Proceedings of the International Joint Conference on Artificial Intelligence*, volume 2, pages 1137–1145, August 1995.
- [122] T. Kohonen, editor. *Self-organizing maps*. Springer-Verlag New York, Inc., Secaucus, NJ, USA, 1997.
- [123] P. Korfiatis, S. Skiadopoulos, P. Sakellaropoulos, C. Kalogeropoulou, and L. Costaridou. Automated 3D segmentation of lung fields in thin slice CT exploiting wavelet preprocessing. *Computer Analysis of Images and Patterns*, 4673:237–244, August 2007.
- [124] V. Kovalev and M. Petrou. Multidimensional co-occurrence matrices for object recognition and matching. *Graphical Models in Image Processing*, 58(3):187–197, May 1996.
- [125] V. A. Kovalev, F. Kruggel, H. J. Gertz, and D. Y. von Cramon. Three-dimensional texture analysis of MRI brain datasets. *IEEE Transactions on Medical Imaging*, 20(5):424–433, May 2001.
- [126] R. P. Kruger, W. B. Thompson, and A. F. Turner. Computer diagnosis of pneumoconiosis. *IEEE Transactions on Systems, Man, and Cybernetics*, SMC-4(1):40–49, January 1974.
- [127] M. Kubat and S. Matwin. Addressing the curse of imbalanced training sets: one-sided selection. In *Proceedings of the 14th International Conference on Machine Learning*, pages 179–186. Morgan Kaufmann, 1997.
- [128] S. Kullback and R. A. Leibler. On information and sufficiency. *The Annals of Mathematical Statistics*, 22(1):79–86, March 1951.
- [129] A. S. Kurani, D.-H. Xu, J. Furst, and D. S. Raicu. Co-occurrence matrices for volumetric data. In *The 7th IASTED International Conference on Computer Graphics and Imaging – CGIM 2004*, Kauai, Hawaii, USA, August 2004.
- [130] C. Lacoste, J.-P. Chevallet, J.-H. Lim, D. T. H. Le, W. Xiong, D. Racoceanu, R. Teodorescu, and N. Vuillenemot. Inter-media concept-based medical image indexing and retrieval with UMLS at IPAL. In C. Peters, P. Clough, F. C. Gey, J. Karlsgren, B. Magnini, D. W. Oard, M. de Rijke, and M. Stempfhuber, editors, *CLEF*, volume 4730 of *Lecture Notes in Computer Science*, pages 694–701. Springer, September 2007.
- [131] A. Laine and J. Fan. Frame representations for texture segmentation. *IEEE Transactions on Image Processing*, 5(5):771–780, May 1996.
- [132] L. Lam. Classifier combinations: Implementations and theoretical issues. In *MCS '00: Proceedings of the First International Workshop on Multiple Classifier Systems*, pages 77–86, London, UK, June 2000. Springer-Verlag.
- [133] G. R. G. Lanckriet, N. Cristianini, P. Bartlett, L. E. Ghaoui, and M. I. Jordan. Learning the kernel matrix with semidefinite programming. *Journal of Machine Learning Research*, 5:27–72, December 2004.
- [134] C. Le Bozec, E. Zapletal, M.-C. Jaulent, D. Heudes, and P. Degoulet. Towards content-based image retrieval in HIS-integrated PACS. In *Proceedings of the Annual Symposium of the American Society for Medical Informatics (AMIA)*, pages 477–481, Los Angeles, CA, USA, November 2000.

- [135] Y. Lee, N. Kim, J. B. Seo, J. G. Lee, and S. H. Kang. The performance improvement of automatic classification among obstructive lung diseases on the basis of the features of shape analysis, in addition to texture analysis at HRCT. In J. P. W. Pluim and J. M. Reinhardt, editors, *Medical Imaging 2007: Image Processing*, volume 6512, page 65124F. SPIE, February 2007.
- [136] Y. Lee, J. B. Seo, B. Kang, D. Kim, J. G. Lee, S. S. Kim, N. Kim, and S. H. Kang. Performance comparison of classifiers for differentiation among obstructive lung diseases based on features of texture analysis at HRCT. In J. P. W. Pluim and J. M. Reinhardt, editors, *Medical Imaging 2007: Image Processing*, volume 6512, page 651249. SPIE, February 2007.
- [137] Y. Lee, J. B. Seo, J. G. Lee, S. S. Kim, N. Kim, and S. H. Kang. Performance testing of several classifiers for differentiating obstructive lung diseases based on texture analysis at high-resolution computerized tomography (HRCT). *Computer Methods and Programs in Biomedicine*, 2(93):206–215, February 2009.
- [138] R. A. Lerski, K. Straughan, L. R. Schad, D. Boyce, S. Blüml, and I. Zuna. VIII. MR image texture analysis—an approach to tissue characterization. *Magnetic Resonance Imaging*, 11(6):873–887, January 1993.
- [139] Q. Li, F. Li, J. Shiraishi, S. Katsuragawa, S. Sone, and K. Doi. Investigation of new psychophysical measures for evaluation of similar images on thoracic computed tomography for distinction between benign and malignant nodules. *Medical Physics*, 30(10):2584–2593, October 2003.
- [140] Q. Li, F. Li, K. Suzuki, J. Shiraishi, H. Abe, R. Engelmann, Y. Nie, H. MacMahon, and K. Doi. Computer-aided diagnosis in thoracic CT. *Seminars in ultrasound, CT, and MR, Semin Ultrasound CT MR*, 26(5):357–363, October 2005.
- [141] S. Li, J. T. Kwok, H. Zhu, and Y. Wang. Texture classification using the support vector machines. *Pattern Recognition*, 36(12):2883–2893, December 2003.
- [142] S. Liapis and G. Tziritas. Color and texture image retrieval using chromaticity histograms and wavelet frames. *IEEE Transactions on Multimedia*, 6(5):676–686, October 2004.
- [143] A. H. Limper. Chemotherapy-induced lung disease. *Clinics in Chest Medicine*, 25(1):53–64, March 2004.
- [144] C.-T. Liu, P.-L. Tai, A. Y.-J. Chen, C.-H. Peng, T. Lee, and J.-S. Wang. A content-based CT lung retrieval system for assisting differential diagnosis images collection. In *Proceedings of the second International Conference on Multimedia and Exposition (ICME'2001)*, pages 241–244, Tokyo, Japan, August 2001. IEEE Computer Society, IEEE Computer Society.
- [145] C.-T. Liu, P.-L. Tai, A. Y.-J. Chen, C.-H. Peng, and J.-S. Wang. A content-based medical teaching file assistant for CT lung image retrieval. In *Proceedings of the IEEE International Conference on Electronics, Circuits, Systems (ICECS2000)*, pages 361–365, Jouneih–Kaslik, Lebanon, December 2000.
- [146] H. J. Lowe, B. G. Buchanan, G. F. Cooper, and J. K. Vreis. Building a medical multimedia database system to integrate clinical information: An application of high-performance computing and communications technology. *Bulletin of the Medical Library Association*, 83(1):57–64, January 1995.
- [147] G. Lutterbey, C. Grohé, J. Gieseke, M. von Falkenhausen, N. Morakkabati, M. P. Wattjes, R. Manka, D. Trog, and H. H. Schild. Initial experience with lung-MRI at 3.0T: Comparison with CT and clinical data in the evaluation of interstitial lung disease activity. *European journal of radiology*, 61(2):256–261, February 2007.

- [148] A. Madabhushi, M. D. Feldman, D. N. Metaxas, D. Chute, and J. E. Tomaszewski. A novel stochastic combination of 3D texture features for automated segmentation of prostatic adenocarcinoma from high resolution MRI. In *Medical Image Computing and Computer-Assisted Intervention - MICCAI 2003, 6th International Conference*, pages 581–591, November 2003.
- [149] D. Mahmoud-Ghoneim, G. Toussaint, J.-M. Constans, and J. D. de Certaines. Three dimensional texture analysis in MRI: a preliminary evaluation in gliomas. *Magnetic Resonance Imaging*, 21(9):983–987, November 2003.
- [150] S. G. Mallat. *A Wavelet Tour of Signal Processing*. Academic Press, San Diego, 2nd edition edition, 1999.
- [151] J. Malone, J. Rossiter, S. Prabhu, and P. Goddard. Identification of disease in CT of the lung using texture-based image analysis. In *2004. Conference Record of the Thirty-Eight Asilomar Conference on Signals, Systems and Computers*, volume 2, pages 1620–1624, November 2004.
- [152] E. M. Marom. CT of the solitary pulmonary nodule – a commentary. *American Journal of Roentgenology*, 190(5):1154–1155, May 2008.
- [153] D. Marr. *Vision: a computational investigation into the human representation and processing of visual information*. W. H. Freeman, San Francisco, 1982.
- [154] D. Marr and E. Hildreth. Theory of edge detection. *Proceedings Royal Society of London*, 207(1167):187–217, February 1980.
- [155] J. G. Mazoue. Diagnosis without doctors. *Journal of Medicine and Philosophy*, 15(6):559–579, December 1990.
- [156] M. F. McNitt-Gray, S. G. Armato, C. R. Meyer, A. P. Reeves, G. McLennan, R. C. Pais, J. Freymann, M. S. Brown, R. M. Engelmann, P. H. Bland, G. E. Laderach, C. Piker, J. Guo, Z. Towfic, D. Qing, D. F. Yankelevitz, D. R. Aberle, E. J. R. van Beek, H. MacMahon, E. A. Kazerooni, B. Y. Croft, and L. P. Clarke. The lung image database consortium (LIDC) data collection process for nodule detection and annotation. *Academic Radiology*, 14(12):1464–1474, December 2007.
- [157] P. R. S. Mendonça, D. R. Padfield, J. C. Ross, J. V. Miller, S. Dutta, and S. M. Gautham. Quantification of emphysema severity by histogram analysis of CT scans. In *Medical Image Computing and Computer-Assisted Intervention - MICCAI 2005*, volume 3749, pages 738–744. Springer Berlin / Heidelberg, September 2005.
- [158] P. H. Meyers, C. M. Nice, H. C. Becker, W. J. Nettleton, J. W. Sweeney, and G. R. Meckstroth. Automated computer analysis of radiographic images. *Radiology*, 83:1029–1034, December 1964.
- [159] R. A. Miller. Medical diagnostic decision support systems—past, present, and future: a threaded bibliography and brief commentary. *Journal of the American Medical Informatics Association*, 1(1):8–27, February 1994.
- [160] Y. Mitani, H. Hirayama, H. Yasuda, S. Kido, Y. Hamamoto, K. Ueda, and N. Matsunaga. A gabor filter-based classification for diffuse lung opacities in thin-section computed tomography images. In *Fourth International Conference on Knowledge-Based Intelligent Engineering Systems and Allied Technologies, 2000. Proceedings*, volume 2, pages 780–783. IEEE Computer Society, August 2000.
- [161] F. Mitsunobu, T. Mifune, K. Ashida, Y. Hosaki, H. Tsugeno, M. Okamoto, S. Harada, S. Takata, and Y. Tanizaki. Influence of age and disease severity on high resolution CT lung densitometry in asthma. *Thorax*, 56(11):851–856, November 2001.

- [162] A. Mojsilovic, M. Popovic, S. Markovic, and M. Krstic. Characterization of visually similar diffuse disease from B-scan liver images using nonseparable wavelet transform. *IEEE Transactions on Medical Imaging*, 17(4):541–549, August 1998.
- [163] N. R. Mudigonda, R. Rangayyan, and J. E. L. Desautels. Gradient and texture analysis for the classification of mammographic masses. *IEEE Transactions on Medical Imaging*, 19(10):1032–1043, October 2000.
- [164] H. Müller, P. Clough, W. Hersh, and A. Geissbuhler. Variations of relevance assessments for medical image retrieval. In *Adaptive Multimedia Retrieval (AMR)*, volume 4398 of *Springer Lecture Notes in Computer Science (LNCS)*, pages 233–247, Geneva, Switzerland, June 2007.
- [165] H. Müller, C. Despont-Gros, W. Hersh, J. Jensen, C. Lovis, and A. Geissbuhler. Health care professionals’ image use and search behaviour. In *Proceedings of the Medical Informatics Europe Conference (MIE 2006)*, IOS Press, Studies in Health Technology and Informatics, pages 24–32, Maastricht, The Netherlands, August 2006.
- [166] H. Müller and J. Kalpathy-Cramer. Analyzing the content out of context — features and gaps in medical image retrieval. *International Journal on Healthcare Information Systems and Informatics*, 4(1):88–98, 2009.
- [167] H. Müller, S. Marchand-Maillet, and T. Pun. The truth about Corel – Evaluation in image retrieval. In M. S. Lew, N. Sebe, and J. P. Eakins, editors, *Proceedings of the International Conference on the Challenge of Image and Video Retrieval (CIVR 2002)*, volume 2383 of *Lecture Notes in Computer Science (LNCS)*, pages 38–49, London, England, July 2002. Springer-Verlag.
- [168] H. Müller, S. Marquis, G. Cohen, and A. Geissbuhler. Lung CT analysis and retrieval as a diagnostic aid. In R. Engelbrecht, A. Geissbuhler, C. Lovis, and G. Mihalas, editors, *Proceedings of the Medical Informatics Europe Conference (MIE 2005)*, volume 116, pages 453–458, Geneva, Switzerland, August 2005.
- [169] H. Müller, S. Marquis, G. Cohen, P.-A. Poletti, C. Lovis, and A. Geissbuhler. Automatic abnormal region detection in lung CT images for visual retrieval. *Swiss Medical Informatics*, 57:2–6, 2005.
- [170] H. Müller, N. Michoux, D. Bandon, and A. Geissbuhler. A review of content-based image retrieval systems in medicine—clinical benefits and future directions. *International Journal of Medical Informatics*, 73(1):1–23, February 2004.
- [171] H. Müller, M. Pitkanen, X. Zhou, A. Depeursinge, J. Iavindrasana, and A. Geissbuhler. Knowarc: Enabling Grid networks for the biomedical research community. In *Healthgrid 2007*, pages 261–268, Geneva, Switzerland, April 2007.
- [172] H. Müller, A. Rosset, A. Garcia, J.-P. Vallée, and A. Geissbuhler. Benefits from content-based visual data access in radiology. *RadioGraphics*, 25(3):849–858, May 2005.
- [173] H. Müller, A. Rosset, J.-P. Vallée, F. Terrier, and A. Geissbuhler. A reference data set for the evaluation of medical image retrieval systems. *Computerized Medical Imaging and Graphics*, 28(6):295–305, September 2004.
- [174] N. L. Müller. Clinical value of high-resolution CT in chronic diffuse lung disease. *American Journal of Roentgenology*, 157(6):1163–1170, December 1991.
- [175] M. Niinimaki, X. Zhou, A. Depeursinge, and H. Müller. Plug-in grid: A virtualized grid cluster. In *MICCAI workshop on HealthGrids*, pages 74–83, London, UK, September 2009.
- [176] R. M. Nishikawa. Current status and future directions of computer-aided diagnosis in mammography. *Computerized Medical Imaging and Graphics*, 31(4–5):224–235, June 2007.

- [177] K. Nowers, J. D. Rasband, G. Berges, and M. Gosselin. Approach to ground-glass opacification of the lung. *Seminars in Ultrasound, CT, and MRI*, 23(4):302–323, August 2002. Diffuse Lung Disease.
- [178] T. Ojala, M. Pietikainen, and T. Maenpaa. Multiresolution gray-scale and rotation invariant texture classification with local binary patterns. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 24(7):971–987, July 2002.
- [179] M. C. Oliveira, W. Cirne, and P. M. de Azevedo Marques. Towards applying content-based image retrieval in the clinical routine. *Future Generation Computer Systems*, 23(3):466–474, March 2007.
- [180] Y. S. Park, J. B. Seo, N. Kim, E. J. Chae, Y. M. Oh, S. D. Lee, Y. Lee, and S. H. Kang. Texture-based quantification of pulmonary emphysema on high-resolution computed tomography: Comparison with density-based quantification and correlation with pulmonary function test. *Investigative Radiology*, 43(6):395–402, June 2008.
- [181] J. R. Parker. *Algorithms for Image Processing and Computer Vision*. John Wiley & Sons, Inc., New York, NY, USA, December 1996.
- [182] M. Pechenizkiy, A. Tsymbal, and S. Puuronen. PCA-based feature transformation for classification: Issues in medical diagnostics. In *CBMS '04: Proceedings of the 17th IEEE Symposium on Computer-Based Medical Systems*, pages 535–540, Washington, DC, USA, June 2004. IEEE Computer Society.
- [183] P. Perner, T. P. Belikova, and N. I. Yashunskaya. 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*, volume 1121, pages 208–219, London, UK, August 1996. Springer-Verlag.
- [184] Z. Pizlo. *3D shape : its unique place in visual perception*. Cambridge, Massachussets. : The MIT Press, April 2008.
- [185] D. D. Y. Po and M. N. Do. Directional multiscale modeling of images using the contourlet transform. *IEEE Transactions on Image Processing*, 15(6):1610–1620, June 2006.
- [186] J. Portilla, V. Strela, M. J. Wainwright, and E. P. Simoncelli. Image denoising using scale mixtures of Gaussians in the Wavelet domain. *IEEE Transactions on Image Processing*, 12(11):1338–1351, November 2003.
- [187] M. Prasad and A. Sowmya. Multi-class unsupervised classification with label correction of HRCT lung images. In *Intelligent Sensing and Information Processing, 2004. Proceedings of International Conference on*, pages 51–56. IEEE Computer Society, January 2004.
- [188] M. Prasad and A. Sowmya. Multilevel emphysema diagnosis of HRCT lung images in an incremental framework. In J. M. Fitzpatrick and M. Sonka, editors, *Medical Imaging 2004: Image Processing*, volume 5370, pages 42–50. SPIE, June 2004.
- [189] M. Prasad and A. Sowmya. Multi-level classification of emphysema in HRCT lung images using delegated classifiers. In *Medical Image Computing and Computer-Assisted Intervention - MICCAI 2008*, volume 5241, pages 59–66. Springer, September 2008.
- [190] M. Prasad, A. Sowmya, and I. Koch. Feature subset selection using ICA for classifying emphysema in HRCT images. In *Pattern Recognition, 2004. ICPR 2004. Proceedings of the 17th International Conference on*, volume 4, pages 515–518. IEEE Computer Society, August 2004.
- [191] M. Prasad, A. Sowmya, and P. Wilson. Multi-level classification of emphysema in HRCT lung images. *Pattern Analysis and Applications*, 12(1):9–20, February 2009.

- [192] K. Price. A comparison of human and computer vision systems: a tutorial. *SIGART Bulletin*, 0(50):5–10, February 1975.
- [193] P. Pudil, J. Novovičová, and J. Kittler. Floating search methods in feature selection. *Pattern Recognition Letters*, 15(11):1119–1125, November 1994.
- [194] H. Qi and W. E. Snyder. Content-based image retrieval in PACS. *Journal of Digital Imaging*, 12(2):81–83, 1999.
- [195] R. J. Quinlan. Induction of decision trees. *Machine Learning*, 1(1):81–106, March 1986.
- [196] R. J. Quinlan. *C4.5: Programs for Machine Learning*. Morgan Kaufmann Publishers Inc., San Francisco, CA, USA, January 1993.
- [197] H. Rajasekaran, L. Lo Iacono, P. Hasselmeyer, J. Fingberg, P. Summers, S. Benkner, G. Engelbrecht, A. Arbona, A. Chiarini, C. Friedrich, M. Hofmann-Apitius, K. Kumpf, B. Moore, P. Bijlenga, J. Iavindrasana, H. Müller, R. Hose, R. Dunlop, and A. Frangi. Aneurist — towards a system architecture for advanced disease management through integration of heterogeneous data, computing, and complex processing services. In *21th IEEE Symposium on Computer-Based Medical Systems (CBMS)*, pages 361–366, Jyväskylä, Finland, June 2008.
- [198] T. Randen and J. H. Husøy. Filtering for texture classification: a comparative study. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 21(4):291–310, April 1999.
- [199] T. R. Reed and J. M. H. du Buf. A review of recent texture segmentation and feature extraction techniques. *Computer Vision, Graphics and Image Processing*, 57(3):359–372, May 1993.
- [200] A. P. Reeves, A. M. Biancardi, D. F. Yankelevitz, S. Fotin, B. M. Keller, A. C. Jirapatnakul, and J. Lee. A public image database to support research in computer aided diagnosis. In *31st Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, pages 3715–3718, September 2009.
- [201] C. C. Reyes-Aldasoro and A. Bhalerao. Volumetric texture description and discriminant feature selection for MRI. In C. J. Taylor and J. A. Noble, editors, *IPMI*, volume 2732 of *Lecture Notes in Computer Science*, pages 282–293. Springer, July 2003.
- [202] P. Robinson. Radiology’s Achilles’ heel: error and variation in the interpretation of the Rontgen image. *British Journal of Radiology*, 70(839):1085–1098, November 1997.
- [203] R. W. Rodieck. Quantitative analysis of cat retinal ganglion cell response to visual stimuli. *Vision Research*, 5(12):583–601, December 1965.
- [204] N. Roduit, R. Meyer, P. Bijlenga, and A. Geissbuhler. Image depersonalization: a new efficient approach. In *RSNA 2009: 95th Scientific Assembly and Annual Meeting of the Radiological Society of North America*. RSNA, December 2009.
- [205] J. Romberg. Sensing by random convolution. In *Computational Advances in Multi-Sensor Adaptive Processing, 2007. CAMPSP 2007. 2nd IEEE International Workshop on*, pages 137–140, December 2007.
- [206] A. Rosset, O. Ratib, A. Geissbuhler, and J.-P. Vallée. Integration of a multimedia teaching and reference database in a PACS environment. *RadioGraphics*, 22(6):1567–1577, December 2002.
- [207] A. Rosset, L. Spadola, and O. Ratib. OsiriX: An open-source software for navigating in multidimensional DICOM images. *Journal of Digital Imaging*, 17(3):205–216, September 2004.

- [208] Y. Rubner, C. Tomasi, and L. J. Guibas. The earth mover's distance as a metric for image retrieval. *International Journal of Computer Vision*, 40(2):99–121, November 2000.
- [209] M. Rudrapatna, A. Sowmya, T. Zrimec, P. Wilson, G. Kossoff, P. Lucas, J. Wong, A. Misra, and S. Busayarat. LMIK – learning medical image knowledge: an internet-based medical image knowledge acquisition framework. In S. Santini and R. Schettini, editors, *Internet Imaging V*, volume 5304, pages 307–318. SPIE, January 2003.
- [210] J. H. Ryu, E. J. Olson, D. E. Midthun, and S. J. Swensen. Diagnostic approach to the patient with diffuse lung disease. *Mayo Clinic Proceedings*, 77(11):1221–1227, November 2002.
- [211] S. Saita, M. Kubo, Y. Kawata, N. Niki, Y. Nakano, H. Ohmatsu, K. Tominaga, K. Eguchi, and N. Moriyama. Algorithm of pulmonary emphysema extraction using thoracic 3D CT images. In M. L. Giger and N. Karssemeijer, editors, *Medical Imaging 2007: Computer-Aided Diagnosis*, volume 6514, page 65143L. SPIE, February 2007.
- [212] R. Salgado, T. Mulkens, P. Bellinck, and J. L. Termote. Volume rendering in clinical practice. a pictorial review. *Journal Belge de Radiologie*, 86(4):215–220, July 2003.
- [213] G. Sasso, H. R. Marsiglia, F. Pigatto, A. Basilicata, M. Gargiulo, A. F. Abate, M. Nappi, J. Pulley, and F. S. Sasso. A visual query-by-example image database from chest CT images: Potential role as a decision and educational support tool for radiologists. *Journal of Digital Imaging*, 18(1):78–84, March 2005.
- [214] O. H. Schade. Optical and photoelectric analog of the eye. *Journal of the Optical Society of America*, 46(9):721–739, September 1956.
- [215] A. Schödl, R. Szeliski, D. H. Salesin, and I. Essa. Video textures. In *SIGGRAPH '00: Proceedings of the 27th annual conference on Computer graphics and interactive techniques*, pages 489–498, New York, NY, USA, July 2000. ACM Press/Addison-Wesley Publishing Co.
- [216] W. Schroeder, K. Martin, and B. Lorensen. *The Visualization Toolkit — An Object Oriented Approach to 3D Graphics*. Kitware, Inc., 3rd edition edition, December 2006.
- [217] A. Shamsheyeva and A. Sowmya. The anisotropic gaussian kernel for SVM classification of HRCT images of the lung. In *Proceedings of the 2004 Intelligent Sensors, Sensor Networks and Information Processing Conference*, pages 439–444, December 2004.
- [218] A. Shamsheyeva and A. Sowmya. Tuning kernel function parameters of support vector machines for segmentation of lung disease patterns in high-resolution computed tomography images. In *SPIE Medical Imaging*, volume 5370, pages 1548–1557, May 2004.
- [219] A. Shamsheyeva, A. Sowmya, and P. Wilson. Segmentation of lung patterns in high-resolution computed tomography images of the lung. *International Journal of Computational Intelligence and Applications*, 7(3):265–280, September 2008.
- [220] C.-R. Shyu, C. E. Brodley, A. C. Kak, A. Kosaka, A. M. Aisen, and L. S. Broderick. AS-SERT: A physician-in-the-loop content-based retrieval system for HRCT image databases. *Computer Vision and Image Understanding (special issue on content-based access for image and video libraries)*, 75(1/2):111–132, July/August 1999.
- [221] D. Simel and R. Drummond. *The rational clinical examination: evidence-based clinical diagnosis*. McGraw-Hill, August 2008.
- [222] U. Sinha, A. Bui, R. Taira, J. Dionisio, C. Morioka, D. Johnson, and H. Kangarloo. A review of medical imaging informatics. *Annals of the New York Academy of Sciences*, 980:168–197, December 2002.

- [223] M. Skurichina and R. P. Duin. Combining feature subsets in feature selection. In *Multiple Classifier Systems*, volume 3541 of *Lecture Notes in Computer Science*, pages 165–175. Springer, June 2005.
- [224] I. C. Sluimer, A. Schilham, M. Prokop, and B. van Ginneken. Computer analysis of computed tomography scans of the lung: a survey. *IEEE Transactions on Medical Imaging*, 25(4):385–405, April 2006.
- [225] I. C. Sluimer, P. F. van Waes, M. A. Viergever, and B. van Ginneken. Computer-aided diagnosis in high resolution CT of the lungs. *Medical Physics*, 30(12):3081–3090, December 2003.
- [226] A. W. M. Smeulders, M. Worring, S. Santini, A. Gupta, and R. Jain. Content-based image retrieval at the end of the early years. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 22(12):1349–1380, December 2000.
- [227] J. R. Smith, C.-Y. Lin, and M. Naphade. Video texture indexing using spatio-temporal wavelets. In *International Conference on image processing (ICIP 2002)*, volume 2, pages 437–440, September 2002.
- [228] P. B. A. Smits, J. H. A. M. Verbeek, and C. D. de Buissonje. Problem-based learning in continuing medical education: a review of controlled evaluation studies. *British Medical Journal*, 324(7330):153–156, January 2002.
- [229] A. J. Smola and B. Schölkopf. A tutorial on support vector regression. *Statistics and Computing*, 14(3):199–222, September 2004.
- [230] C. G. M. Snoek, M. Worring, and A. W. M. Smeulders. Early versus late fusion in semantic video analysis. In *MULTIMEDIA '05: Proceedings of the 13th annual ACM international conference on Multimedia*, pages 399–402, New York, NY, USA, November 2005. ACM.
- [231] M. Sonka, J. Tschirren, S. Ukil, X. Zhang, Y. Xu, J. MReinhardt, E. J. van Beek, G. McLennan, and E. A. Hoffman. Pulmonary CT image analysis and computer aided detection. In *4th IEEE International Symposium on Biomedical Imaging: From Nano to Macro, 2007. ISBI 2007.*, pages 500–503, April 2007.
- [232] L. Sørensen and M. De Bruijne. Dissimilarity representations in lung parenchyma classification. In N. Karssemeijer and M. L. Giger, editors, *Medical Imaging 2009: Computer-Aided Diagnosis*, volume 7260, page 72602Z. SPIE, February 2009.
- [233] L. Sørensen, P. Lo, H. Ashraf, J. Sporring, M. Nielsen, and M. De Bruijne. Learning COPD sensitive filters in pulmonary CT. In *Proc. Medical Image Computing and Computer Assisted Intervention (MICCAI'09)*, volume 5762 of *Lecture Notes in Computer Science*, pages 699–706, London, UK, September 2009. Springer.
- [234] L. Sørensen, S. Shaker, and M. De Bruijne. Texture classification in lung CT using local binary patterns. In *Proc. Medical Image Computing and Computer Assisted Intervention (MICCAI'08)*, volume 5241, pages 934–941. Springer, September 2008.
- [235] C. A. Souza, N. L. Müller, J. Flint, J. L. Wright, and A. Churg. Idiopathic pulmonary fibrosis: Spectrum of high-resolution CT findings. *American Journal of Roentgenology*, 185(6):1531–1539, December 2005.
- [236] P. Stark. *High resolution computed tomography of the lungs*. UpToDate, Waltham, MA, Denise S. Basow edition, August 2008.
- [237] H. Sumikawa, T. Johkoh, S. Yamamoto, K. Takahei, T. Ueguchi, Y. Ogata, M. Matsumoto, Y. Fujita, J. Natsag, A. Inoue, M. Tsubamoto, N. Mihara, O. Honda, N. Tomiyama, S. Hamada, and H. Nakamura. Quantitative analysis for computed tomography findings of various diffuse lung diseases using volume histogram analysis. *Journal of Computer Assisted Tomography*, 30(2):244–249, March 2006.

- [238] J. W. H. Tangelder and R. C. Veltkamp. A survey of content based 3D shape retrieval methods. *Multimedia Tools and Applications*, 39(3):441–471, September 2008.
- [239] R. L. Tannen, M. G. Weiner, and D. Xie. Use of primary care electronic medical record database in drug efficacy research on cardiovascular outcomes: comparison of database and randomised controlled trial findings. *British Medical Journal*, 338:b81, January 2009.
- [240] B. M. ter Haar Romeny. *Front-End Vision and Multi-Scale Image Analysis: Multi-scale Computer Vision Theory and Applications, written in Mathematica*. Springer Publishing Company, Incorporated, September 2009.
- [241] M. J. Thomeer, U. Costabel, G. Rizzato, V. Poletti, and M. Demedts. Comparison of registries of interstitial lung diseases in three European countries. *European Respiratory Journal*, 18(32):114–118, September 2001.
- [242] A. Tolouee, H. Abrishami-Moghaddam, R. Garnavi, M. Forouzanfar, and M. Giti. Texture analysis in lung HRCT images. In *Computing: Techniques and Applications, 2008. DICTA '08.Digital Image*, pages 305–311, December 2008.
- [243] M. F. Toubin, F. Truchetet, E. P. Verrecchia, C. Dumont, and M. A. Abidi. Multiresolution description of range images through 2D quincunx wavelet analysis. In H. H. Szu, editor, *Wavelet Applications VI*, volume 3723, pages 350–360. SPIE, April 1999.
- [244] G. D. Tourassi. Journey toward computer-aided diagnosis: Role of image texture analysis. *Radiology*, 213(2):317–320, July 1999.
- [245] G. T. Toussaint. The use of context in pattern recognition. *Pattern Recognition*, 10(3):189–204, January 1978.
- [246] Y. Uchiyama, S. Katsuragawa, H. Abe, J. Shiraishi, F. Li, Q. Li, C.-T. Zhang, K. Suzuki, and K. Doi. Quantitative computerized analysis of diffuse lung disease in high-resolution computed tomography. *Medical Physics*, 30(9):2440–2454, September 2003.
- [247] M. Unser. Sum and difference histograms for texture classification. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 8(1):118–125, January 1986.
- [248] M. Unser. Texture classification and segmentation using wavelet frames. *IEEE Transactions on Image Processing*, 4(11):1549–1560, November 1995.
- [249] M. Unser. Splines: a perfect fit for medical imaging. In M. Sonka and M. Fitzpatrick, editors, *Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series*, volume 4684, pages 225–236, May 2002.
- [250] M. Unser and A. Aldroubi. A review of wavelets in biomedical applications. *Proceedings of the IEEE*, 84(4):626–638, April 1996.
- [251] M. Unser, D. Sage, and D. Van De Ville. Multiresolution monogenic signal analysis using the Riesz–Laplace wavelet transform. *IEEE Transactions on Image Processing*, 18(11):2402–2418, November 2009.
- [252] M. Unser and D. Van De Ville. The pairing of a wavelet basis with a mildly redundant analysis via subband regression. *IEEE Transactions on Image Processing*, 17(11):2040–2052, November 2008.
- [253] R. Uppaluri, E. A. Hoffman, M. Sonka, P. G. Hartley, G. W. Hunninghake, and G. McLennan. Computer recognition of regional lung disease patterns. *American Journal of Respiratory and Critical Care Medicine*, 160(2):648–654, August 1999.
- [254] R. Uppaluri, E. A. Hoffman, M. Sonka, G. W. Hunninghake, and G. McLennan. Interstitial lung disease: A quantitative study using the adaptive multiple feature method. *American Journal of Respiratory and Critical Care Medicine*, 159(2):519–525, February 1999.

- [255] D. Van De Ville, T. Blu, and M. Unser. Isotropic polyharmonic B-Splines: Scaling functions and wavelets. *IEEE Transactions on Image Processing*, 14(11):1798–1813, November 2005.
- [256] D. Van De Ville, T. Blu, and M. Unser. On the multidimensional extension of the quincunx subsampling matrix. *Signal Processing Letters, IEEE*, 12(2):112–115, February 2005.
- [257] D. Van De Ville, D. Sage, K. Balać, and M. Unser. The Marr wavelet pyramid and multi-scale directional image analysis. In *Proceedings of the Sixteenth European Signal Processing Conference (EUSIPCO'08)*, Lausanne, Switzerland, August 2008.
- [258] D. Van De Ville and M. Unser. Complex wavelet bases, steerability, and the Marr-like pyramid. *IEEE Transactions on Image Processing*, 17(11):2063–2080, November 2008.
- [259] C. van der Walt and E. Barnard. Data characteristics that determine classifier performance. In *Proceedings of the Sixteenth Annual Symposium of the Pattern Recognition Association of South Africa*, pages 166–171, Parys, South Africa, November 2006.
- [260] B. van Ginneken and B. M. ter Haar Romeny. Multi-scale texture classification from generalized locally orderless images. *Pattern Recognition*, 36(4):899–911, November 2003.
- [261] A. Van Nevel. Texture classification using wavelet frame decompositions. In *Signals, Systems & Computers, 1997. Conference Record of the Thirty-First Asilomar Conference on*, volume 1, pages 311–314, November 1997.
- [262] V. N. Vapnik. *The Nature of Statistical Learning Theory*. Springer, New York, November 1995.
- [263] A. Vargas, A. Depeursinge, A. Platon, A. Geissbuhler, P.-A. Poletti, and H. Müller. A multimedia library of interstitial lung diseases at the University Hospitals of Geneva. In *Swiss Society of Radiology (SSR 2009)*, Geneva, Switzerland, June 2009.
- [264] K. Vo and A. Sowmya. Directional multi-scale modeling of high-resolution computed tomography (HRCT) lung images for diffuse lung disease classification. In *Proceedings of the 13th International Conference, CAIP '2009*, volume 5702, pages 663–671. Springer, September 2009.
- [265] J. F. Wade and T. E. King. Infiltrative and interstitial lung disease in the elderly patient. *Clinics in Chest Medicine*, 14(3):501–521, September 1993.
- [266] G. Wald. The receptors of human color vision. *Science*, 145:1007–1016, September 1964.
- [267] B. A. Wandell. *Foundations of Vision*. Sinauer Associates, Sunderland, Massachusetts, June 1995.
- [268] W. R. Webb, N. L. Müller, and D. P. Naidich, editors. *High-Resolution CT of the Lung*. Lippincott Williams & Wilkins, Philadelphia, PA, USA, 2001.
- [269] P. Welter, T. M. Deserno, B. Fischer, B. B. Wein, B. Ott, and R. W. Günther. Integration of CBIR in radiological routine in accordance with IHE. In K. M. Siddiqui and B. J. Liu, editors, *Medical Imaging 2009: Advanced PACS-based Imaging Informatics and Therapeutic Applications*, volume 7264, page 726404. SPIE, March 2009.
- [270] T. Westerveld. Image retrieval: Content versus context. In *Recherche d'Informations Assistée par Ordinateur (RIA'2000) Computer-Assisted Information Retrieval*, volume 1, pages 276–284, Paris, France, April 2000. CID.
- [271] P. Wighton, M. Sadeghi, T. K. Lee, and M. S. Atkins. A fully automatic random walker segmentation for skin lesions in a supervised setting. In G.-Z. Yang, D. J. Hawkes, D. Rueckert, J. A. Noble, and C. J. Taylor, editors, *MICCAI (1)*, volume 5762 of *Lecture Notes in Computer Science*, pages 1108–1115. Springer, September 2009.

- [272] P. M. Willy and K.-H. Kufer. Content-based medical image retrieval (CBMIR): An intelligent retrieval system for handling multiple organs of interest. In *CBMS '04: Proceedings of the 17th IEEE Symposium on Computer-Based Medical Systems*, pages 103–108, Washington, DC, USA, June 2004. IEEE Computer Society.
- [273] I. H. Witten and E. Frank. *Data Mining: Practical Machine Learning Tools and Techniques*. Morgan Kaufmann Series in Data Management Sys. Morgan Kaufmann, second edition, June 2005.
- [274] J. S. J. Wong and T. Zrimec. Automatic honeycombing detection using texture and structure analysis. In *Computational Intelligence Methods and Applications, 2005 ICSC Congress on*, pages 1–4. IEEE Computer Society, December 2005.
- [275] J. S. J. Wong and T. Zrimec. Classification of lung disease pattern using seeded region growing. In *Australian Conference on Artificial Intelligence*, pages 233–242, December 2006.
- [276] T.-F. Wu, C.-J. Lin, and R. C. Weng. Probability estimates for multi-class classification by pairwise coupling. *Journal of Machine Learning Research*, 5:975–1005, August 2004.
- [277] Y. Wu, E. Y. Chang, K. C.-C. Chang, and J. R. Smith. Optimal multimodal fusion for multimedia data analysis. In *MULTIMEDIA '04: Proceedings of the 12th annual ACM international conference on Multimedia*, pages 572–579, New York, NY, USA, October 2004. ACM.
- [278] D.-H. Xu, A. S. Kurani, J. Furst, and D. S. Raicu. Run-length encoding for volumetric texture. In *The 4th IASTED International Conference on Visualization, Imaging, and Image Processing – VIIP 2004*, Marbella, Spain, September 2004.
- [279] Y. Xu, M. Sonka, G. McLennan, J. Guo, and E. A. Hoffman. Sensitivity and specificity of 3-D texture analysis of lung parenchyma is better than 2-D for discrimination of lung pathology in stage 0 COPD. In A. A. Amini and A. Manduca, editors, *SPIE Medical Imaging*, volume 5746, pages 474–485. SPIE, February 2005.
- [280] Y. Xu, M. Sonka, G. McLennan, J. Guo, and E. A. Hoffman. MDCT-based 3-D texture classification of emphysema and early smoking related lung pathologies. *IEEE Transactions on Medical Imaging*, 25(4):464–475, April 2006.
- [281] Y. Xu, E. J. R. van Beek, Y. Hwanjo, J. Guo, G. McLennan, and E. A. Hoffman. Computer-aided classification of interstitial lung diseases via MDCT: 3D adaptive multiple feature method (3D AMFM). *Academic Radiology*, 13(8):969–978, August 2006.
- [282] J. I. Yellott. Spectral analysis of spatial sampling by photoreceptors: topological disorder prevents aliasing. *Vision Research*, 22(9):1205–1210, 1982.
- [283] O. Yu, Y. Mauss, I. J. Namer, and J. Chambron. Existence of contralateral abnormalities revealed by texture analysis in unilateral intractable hippocampal epilepsy. *Magnetic Resonance Imaging*, 19(10):1305–1310, December 2001.
- [284] S.-N. Yu and C.-T. Chiang. Similarity searching for chest CT images based on object features and spatial relation maps. In *Engineering in Medicine and Biology Society, 2004. EMBS '04. 26th Annual International Conference of the IEEE*, volume 1, pages 1298–1301, September 2004.
- [285] D. Zang and G. Sommer. Signal modeling for two-dimensional image structures. *Journal of Visual Communication and Image Representation*, 18(1):81–99, February 2007.
- [286] V. A. Zavaletta, B. J. Bartholmai, and R. A. Robb. High resolution multidetector CT-aided tissue analysis and quantification of lung fibrosis. *Academic Radiology*, 14(7):772–787, July 2007.

- [287] V. A. Zavaletta, B. J. Bartholmai, and R. A. Robb. Nonlinear histogram binning for quantitative analysis of lung tissue fibrosis in high-resolution CT data. In A. Manduca and X. P. Hu, editors, *Medical Imaging 2007: Physiology, Function, and Structure from Medical Images*, volume 6511, page 6511Q. SPIE, February 2007.
- [288] J. Zhang and T. Tan. Brief review of invariant texture analysis methods. *Pattern Recognition*, 35(3):735–747, March 2002.
- [289] A. Zien and C. S. Ong. Multiclass multiple kernel learning. In *ICML '07: Proceedings of the 24th international conference on Machine learning*, pages 1191–1198, New York, NY, USA, June 2007. ACM.
- [290] T. Zrimec and S. Busayarat. A system for computer aided detection of diseases patterns in high resolution CT images of the lungs. In *Twentieth IEEE International Symposium on Computer-Based Medical Systems, 2007. CBMS '07.*, volume 0, pages 41–46, Los Alamitos, CA, USA, June 2007. IEEE Computer Society.
- [291] T. Zrimec, S. Busayarat, and P. Wilson. A 3D model of the human lung with lung regions characterization. In *International Conference on Image Processing, 2004. ICIP '04*, volume 2, pages 1149–1152, October 2004.
- [292] T. Zrimec and J. Wong. Methods for automatic honeycombing detection in HRCT images of the lung. In *11th Mediterranean Conference on Medical and Biomedical Engineering and Computing 2007*, volume 16, pages 830–833, June 2007.
- [293] T. Zrimec and J. S. J. Wong. Improving computer aided disease detection using knowledge of disease appearance. In *MEDINFO 2007. Proceedings of the 12th World Congress on Health (Medical) Informatics*, volume 129, pages 1324–1328. IOS Press, August 2007.

# Index

- 4D imaging, 2  
Abstract, vii  
Accuracy, 95  
Achievements, 99  
Acknowledgements, ix  
Acute interstitial pneumonia, 19  
Affine invariance, 7  
Affine transformation, 7  
Annotation, 33  
Applications, 83  
B-spline wavelets, 50  
Bayes, 14  
Bias, 97  
Bibliography, 130  
Block size, 53  
Blockwise classification, 52, 91  
C4.5, 14  
Case selection, 33  
Case-based retrieval, 80  
CBIR-based CAD, 3  
Chapter five, 83  
Chapter four, 67  
Chapter one, 1  
Chapter six, 99  
Chapter three, 41  
Chapter two, 29  
Classifier comparison, 67  
Classifier ranking, 70  
Clinical features, 60  
Clinical parameter selection, 32  
Closing operation, 53  
Co-occurrence matrix, 10  
Color histograms, 12  
Computed tomography, 2  
Computer vision, 5  
Computer-aided diagnosis, 2  
Conclusions, 99  
Conditional entropy, 61  
Confusion matrix, 94  
Consistency of the multimodal feature space, 65  
Consolidation, 21  
Content-based image retrieval, 3  
Contents, iii  
Contextual medical image analysis, 73  
Contributions, 27  
Correlation matrix, 62  
Cross-validation, 93  
Cryptogenic organizing pneumonia, 19  
Data collection, 32  
Data entry, 36  
Database, 29  
Database browsing, 84  
Database contents, 36  
Decision trees, 14  
Desquamative interstitial pneumonia, 19  
Detection-based CAD, 2  
Difference of two Gaussians, 9  
Discrete wavelet frames, 50  
Distance measure, 81  
Dyadic lattice, 46  
Early fusion, 76  
Electrocardiogram, 2  
Electroencephalogram, 2  
Electromyogram, 2  
Electronic health record, 1  
Emphysema, 21  
Empirical risk, 13, 92  
Entropy, 61  
Eosinophilic pneumonia, 19  
Euclidean distance, 82  
Evaluation, 91  
Evaluation metrics, 94  
Evidence-based medicine, 1  
Expectation–maximization, 48  
Eye, 7  
F-measure, 96  
False negative, 94  
False positive, 94  
Feature extraction, 7  
Feature ranking, 60  
Feature vector, 53  
Features, 41  
Fibrotic tissue, 21  
Fovea, 9  
Functional MRI, 2

- Gaussian kernel, 16  
 Gaussian mixture model, 48  
 Generalization performance, 13, 92  
 Glossary, 123  
 Grey-level histograms, 42, 48  
 Grid computing, 102  
 Grid search, 68  
 Ground glass, 21  
 Healthy tissue, 21  
 High-resolution computed tomography, 17  
 Hounsfield Units, 17  
 HRCT protocol, 18  
 HTML form for clinical data, 115  
 Human cognition, 12  
 Human vision, 5  
 Hybrid CAD system, 85  
 Hypersensitivity pneumonitis, 19  
 Imbalanced class distributions, 102  
 Improvements, 101  
 Index, 150  
 Information gain, 61  
 Information gain ratio, 61  
 Information retrieval, 80  
 Interstitial lung diseases, 16  
 Introduction, 1  
 Iris, 30  
 Isotropic analysis, 43  
 Isotropic polyharmonic B-spline wavelets, 44  
 k-nearest neighbor, 14  
 Kernel modality fusion, 102  
 Langerhans cell histiocytosis, 19  
 Laplacian of a Gaussian, 10  
 Late fusion, 76  
 Learning medical image knowledge, 32  
 Learning to rank, 102  
 Leave-one-out, 93  
 Leave-one-patient-out, 94  
 Leave-one-ROI-out, 93  
 Lena, 30  
 Limitations, 99  
 Linear transformation, 7  
 List of Figures, 126  
 List of Tables, 128  
 Lung imaging database consortium, 31  
 Lung tissue pattern, 18  
 Lung tissue research consortium, 31  
 Lymphocytic interstitial pneumonia, 19  
 Magnetic resonance imaging, 2  
 Magnetoencephalogram, 2  
 McNemar's test, 70  
 Metrics, 94  
 Micronodules, 21  
 Monogenic directional texture analysis, 101  
 Multi-layer perceptron, 15  
 Multimedia database, 29  
 Multimodal distance measure, 81  
 Multimodal fusion, 76  
 Multiple classifier systems, 76  
 Multiresolution, 10  
 Naive Bayes, 14  
 Non-specific interstitial pneumonia, 19  
 Notation, 119  
 Optimization, 98  
 Overfitting, 72  
 Pattern locations, 34  
 Pearson's coefficient of regression, 62  
 Perspectives, 101  
 Picture archiving and communication system, 1  
 Pneumocystis pneumonia, 19  
 Positron emission tomography, 2  
 Precision, 95  
 Principal component analysis, 77  
 Product rule, 76  
 Pulmonary fibrosis, 19  
 Pyramidal image decomposition, 42  
 Qualitative evaluation, 98  
 Quality control, 36  
 Quincunx lattice, 46  
 Quincunx wavelet frames, 47  
 Résumé, v  
 Random drawing, 93  
 Raw primal sketch, 10  
 Recall, 95  
 References, 130  
 Region growing, 52  
 Respiratory bronchiolitis associated ILD, 19  
 Retina, 5  
 Rich Internet application, 83  
 ROC curves, 96  
 Rotation invariance, 43  
 Rotation-invariant, 10  
 Sampling, 7  
 Sampling factor, 50  
 Sarcoidosis, 19  
 Scale invariance, 43  
 Scale-invariant, 10  
 Scale-progression, 47  
 Scientific contributions, 27  
 Second opinion, 2  
 Segmentation of the lung volumes, 52, 91

- Semantic gap, 4
- Shift invariance, 9
- Shift-invariant WT, 42
- Single photon emission CT, 2
- Single testing accuracy, 61
- Spectral analysis, 11
- Stability, 72
- Supervised machine learning, 12
- Support vector machines, 15
  
- Table de matières, iii
- Table of contents, iii
- Texture, 10
- Texture features, 10
- Thanks, ix
- Thesis overview, 27
- Title, i
- True negative, 94
- True positive, 94
- Tuberculosis, 19
  
- University Hospitals of Geneva, 1
- Use cases, 83
  
- Validation, 92
- Visual cortex, 7
- Visual features, 41
- Visualization, 83
  
- Wavelet bases, 42
- Wavelet frames, 42
- Wavelet transform, 11
- Weighted Euclidean distance, 82