Università degli studi di Pisa

FACOLTÀ DI DI SCIENZE MATEMATICHE, FISICHE E NATURALI
Corso di Dottorato in Fisica Applicata
Settore Scientifico-Disciplinare FIS/07

Pattern recognition methods applied to medical imaging: lung nodule detection in computed tomography images

Candidato:
Niccolò Camarlinghi

Relatore:
Dott.ssa
Maria Evelina Fantacci

Anno Accademico 2011–2012,XXIII ciclo di dottorato

A Francesca e alla mia famiglia

Contents

List of Figures

List of acronyms

AL Agreement Level

ANN Artificial Neural Network

AUC Area Under the ROC curve

CAD Computer Aided Detection

CAD_I CAD for internal nodules

CAD_{JP} CAD for juxta-pleural nodules

CT Computed Tomography

DICOM Digital Imaging and Communications in Medicine format

FLD Fisher Linear Discriminator

FP False Positive

FN False Negative

FROCSV Free Receiver Operating Characteristic Score Value

FROC Free Receiver Operating Characteristic

GGO Ground Glass Opacity

HU Hounsfield Unit

IN Internal Nodule

JPN Juxta-Pleural Nodule

LDCT Low Dose Computed Tomography

LIDC Lung Image Database Consortium

LOO Leave One Out

LOPO Leave One Patient Out

MIP Maximum Intensity Projection

MRI Magnetic Resonance Imaging

MSDE Multi Scale Dot Enhancer

NLST National Lung Screening Trial

PET Positron Emission Tomography

PSN Pleura Surface Normal

ROC Receiver Operating Characteristic

ROI Region of Interest

SPECT Single Photon Emission Tomography

SSDE Single Scale Dot Enhancer

SVM Support Vector Machine

TN True Negative

TP True Positive

VBNA Voxel Based Neural Approach

Introduction

Lung cancer is one of the main public health issues in developed countries. The overall 5-year survival rate is only $10-16\%$ [1–4], although the mortality rate among men in the United States has started to decrease by about 1.5% per year since 1991 and a similar trend for the male population has been observed in most European countries.

By contrast, in the case of the female population, the survival rate is still decreasing, despite a decline in the mortality of young women has been observed over the last decade [5, 6].

Approximately 70% of lung cancers are diagnosed at too advanced stages for the treatments to be effective [7]. The five-year survival rate for early-stage lung cancers (stage I, see appendix A), which can reach 70% [8], is sensibly higher than for cancers diagnosed at more advanced stages.

Lung cancer most commonly manifests itself as non-calcified pulmonary nodules. The CT has been shown as the most sensitive imaging modality for the detection of small pulmonary nodules, particularly since the introduction of the multi-detector-row and helical CT technologies [9]. Screening programs based on Low Dose Computed Tomography (LDCT) may be regarded as a promising technique for detecting small, early-stage lung cancers [10,11]. The efficacy of screening programs based on CT in reducing the mortality rate for lung cancer has not been fully demonstrated yet, and different and opposing opinions are being pointed out on this topic by many experts [12, 13].

However, the recent results obtained by the National Lung Screening Trial (NLST), involving 53454 high risk patients, show a 20% reduction of mortality when the screening program was carried out with the helical CT, rather than with a conventional chest X-ray [14].

LDCT settings are currently recommended by the screening trial protocols. However, it is not trivial in this case to identify small pulmonary nodules,

due to the noisier appearance of the images in low-dose CT with respect to the standard-dose CT. Moreover, thin slices are generally used in screening programs, thus originating datasets of about $300 - 400$ slices per study. Depending on the screening trial protocol they joined, radiologists can be asked to identify even very small lung nodules, which is a very difficult and time-consuming task. Lung nodules are rather spherical objects, characterized by very low CT values and/or low contrast. Nodules may have CT values in the same range of those of blood vessels, airway walls, pleura and may be strongly connected to them. It has been demonstrated, that a large percentage of nodules ($20 - 35\%$) is actually missed in screening diagnoses [15].

To support radiologists in the identification of early-stage pathological objects, about one decade ago, researchers started to develop CAD methods to be applied to CT examinations [16–23].

Within this framework, two CAD sub-systems are proposed: CAD for internal nodules ($\mathrm{CAD_I}$), devoted to the identification of small nodules embedded in the lung parenchyma, i.e. Internal Nodules (INs) and $\mathrm{CAD_{JP}}$, devoted the identification of nodules originating on the pleura surface, i.e. Juxta-Pleural Nodules (JPNs) respectively.

As the training and validation sets may drastically influence the performance of a CAD system, the presented approaches have been trained, developed and tested on different datasets of CT scans (Lung Image Database Consortium (LIDC), ITALUNG − CT) and finally blindly validated on the ANODE09 dataset.

The two CAD sub-systems are implemented in the ITK framework [24], an open source C++ framework for segmentation and registration of medical images, and the rendering of the obtained results are achieved using VTK [25], a freely available software system for 3D computer graphics, image processing and visualization. The Support Vector Machines (SVMs) are implemented in $\mathrm{SVM_{Light}}$ [26]. The two proposed approaches have been developed to detect solid nodules, since the number of Ground Glass Opacity (GGO) contained in the available datasets has been considered too low.

This thesis is structured as follows: in the first chapter the basic concepts about CT and lung anatomy are explained. The second chapter deals with CAD systems and their evaluation methods. In the third chapter the datasets used for this work are described. In chapter 4 the lung segmentation algorithm is explained in details, and in chapter 5 and 6 the algorithms to detect internal and juxta-pleural candidates are discussed. In chapter 7 the reduc-

tion of false positives findings is explained. In chapter 8 results of the train and validation sessions are shown. Finally in the last chapter the conclusions are drawn.

The original contributions in this work are the following:

- The efficient implementation of the algorithm described in 5.5.2.
- The implementation of the CAD_{JP}, since it differs from [27, 28] in the way the normals to the pleura surface are computed and the nodule candidates are classified.
- New features are added to the classification procedure with respect to the implementation described in [17, 19, 22], and a new classifiers SVM is used.
- The lung segmentation described in chap.4 may be regarded as partially original, since it puts together different pre-existing procedures described in [29–31].
- Finally, the training and validation results are original parts.

Part I

Preliminaries

Chapter 1

Computed tomography

1.1 Computed Tomography

Computed tomography is a technique that allows to reconstruct a volume starting from a finite number of projections, each obtained using X-ray interaction with the target [32]. Modern Computed Tomography (CT) scanners typically employ a helical technology to scan the patient body, i.e. the table is translating while the X-ray source is rotating rapidly around the patient. The effect of this operation is that, in the frame of the table, the X-ray tube describes an helical path (see fig. 1.1.2).

Once reconstructed, a computed tomography chest exam is typically composed of $300-400$ two dimensional images (according to the patient's height and to the exam slice thickness). Each image is a slice of the reconstructed chest volume (see fig. 1.1.2).

The two dimensional images are gray level images containing, for each volume element, the reconstructed attenuation coefficient in Hounsfield Unit (HU).

The value in HU of a material X, is defined as

$$HU_X = \frac{\mu_X - \mu_{H_2O}}{\mu_{H_2O}} \times 1000,$$

where μ_{X} is the X-ray linear attenuation coefficient of the tissue contained in the voxel and $\mu_{\mathrm{H_2O}}$ is the attenuation coefficient of water.

The attenuation coefficients depend on the energy of the X-ray passing through the material. Since the spectra of a X-ray tube (see fig. 1.1.1) is not monochromatic, the HU value of each tissue is weighted at each energy

of the spectrum. Table (1.1) shows the HUs of some typical constituents of human body tissue.

The CT images are typically anisotropic, i.e. the voxels have different physical dimensions v_x, v_y, v_z with $v_x = v_y \sim 0.6$mm (referred as 2D pixel spacing) and $v_z > v_x, v_y$ (referred to as slice thickness) and v_z varied, in the analyzed cases, between 0.6 and 5 mm.

The images are commonly stored in Digital Imaging and Communications in Medicine format (DICOM), a standard for handling, storing, printing, and transmitting information in medical imaging [33]. The DICOM files include a full dictionary of metadata (i.e. slice thickness, patient 's orientation etc.) available together with the images.

Once the volume is reconstructed, it is possible to slice it in arbitrary planes. The most common planes for the visualization are those associated with the principal axes of the patient (see fig. 1.1.3).

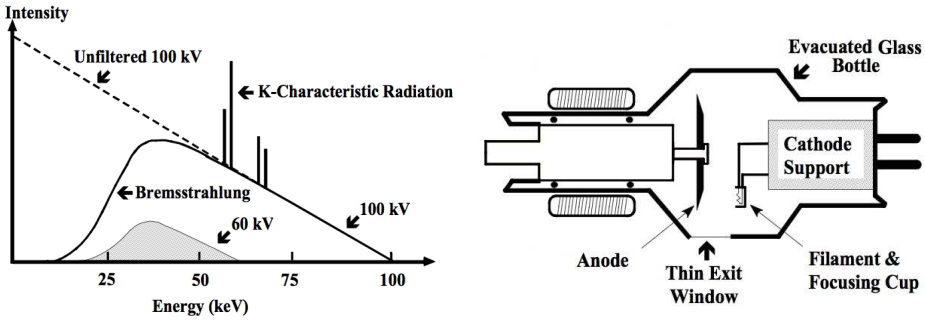
In the X-ray tube (see fig. 1.1.1), a beam of electrons is first accelerated through a potential gap and then decelerated by hitting a solid piece of metal with a high atomic number. The electron beam is generated using a filament heated by Joule effect, which then emits electrons through thermoionic emission.

The electrons are then accelerated across a potential gap, spanning between 40 kV to 150 kV, then strike the metallic target and produce X-ray radiation by bremsstrahlung and characteristic radiation.

The lower energy X-rays are absorbed within the X-ray tube, thus reducing the number of lower energy particles in the resultant spectrum. The theoretical and "real like" spectra of an X-ray source for a thick target are shown in fig. 1.1.1.

The main advantages of CT over the ordinary radiography consists in the resolution obtained in the exam and in a fully 3D reconstruction of the volume of interest. However, a higher resolution has to be paid in terms of a higher radiation dose delivered to the patient.

There are plenty of parameters to roughly estimate the dose delivered by a CT exam, the most important are those related to the X-ray tube: the kV and the mAs. The kV is the value of potential gap that accelerates the electrons in the X-ray tube, corresponding also to the maximum energy of the photon emitted by bremsstrahlung, while the mAs are an abbreviation for milli-ampere per second and are the product between the filament current



(a) (b)

Figure 1.1.1: a) Example of bremsstrahlung and characteristic radiation spectra for different kV and mAs. b) Scheme of an X-ray vacuum sealed tube, with a rotating anode.

Table 1.1: Hounsfield Units of some constituents of human body.

Substance	HU
Air	-1000
Fat	-120
Water	0
Muscle	+40
Bone	+400 or more
Pure lung tissue	50
Lung Parenchyma	-600

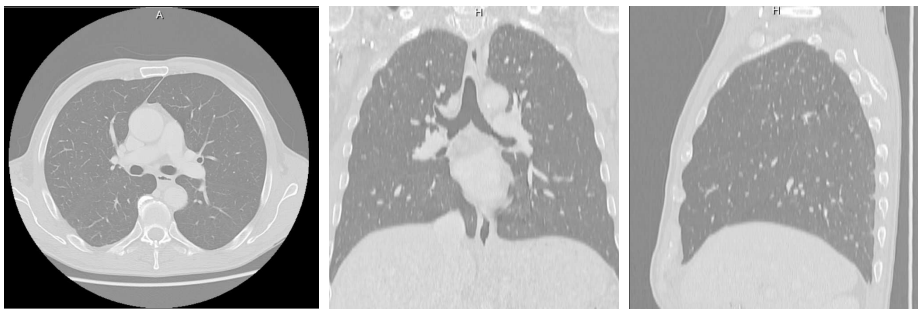
and exposure time. There is no direct relation between dose and image noise, however the noise generally decreases while increasing the dose.

1.2 Equivalent and effective dose

Ionizing radiations cause an energy transfer to the target that is described by the quantity dose. In the international System of Units (SI) the dose is expressed in gray (Gy) where the gray is defined as the absorption of

(a) (b)

Figure 1.1.2: a) Path described by the X-ray tube in helical CT b) The slices generated by a computed tomography exam.



(a) Axial Slice (b) Coronal Slice (c) Sagittal Slice

Figure 1.1.3: Three orthogonal projections of a CT scan.

Table 1.2: Table showing the value of w_r for alpha, beta and gamma radiations.

Radiation type	w_r
alpha	20
beta	1
gamma	1
protons (>2MeV)	5

one joule of ionizing radiations by one kilogram of matter. For a given dose radiations of different quality will induce different biological effects, e.g. one gray deposited by an alpha particle has a very different effect compared to one gray deposited by X-rays. For this reason a physical quantity called equivalent dose has been introduced:

$$H = w_r D,$$

where D is the dose in gray and w_r is a weighting factor that measures the particles delivering the dose. The unit of H is the sievert (Sv). Table 1.2 shows w_r for alpha, beta and gamma radiations.

Since biological damage depends on the targeted organ, it is possible to define a physical quantity that accounts for different types of tissue

$$E = \sum_{t=0}^{N} w_t H_t$$

where w_t are the weighting coefficients for different tissues and H_t the equivalent dose. E is referred to as effective dose and it is measured in sievert. Table 1.3 shows w_t values for different tissues and effective doses delivered in different examinations.

1.3 Lung anatomy as seen in a CT chest exam

The lungs are a pair of spongy, air-filled organs located on either side of the chest and are wrapped in a thin tissue layer called pleura. As shown in

Table 1.3: a) Weighting factor w_t for different tissue [32] b) Effective dose delivered by different examinations [14, 34].

(a)

Tissue	w_t
Lungs	0.12
Stomach	0.12
Skin	0.01
Gonads	0.2

(b)

Examination	Effective dose (mSv)
Chest X-ray	0.1
Head CT	1.5
Chest CT	8
Low dose chest CT	1.5
Screening Mammography	3

fig. (1.3.1), the right lung is divided into three lobes whereas the left one is divided into two. The physical boundaries between lobes are called fissures.

The trachea conducts inhaled air into the lungs through its tubular branches, called airways. The trachea is clearly visible in a CT chest exam as a tubular air-filled structure, near the head of the patient (see fig. 1.3.2).

Two branches of the airways, one for each lung, start from the trachea, and divide into smaller branches, finally becoming microscopic.

As reported in [35, 36], there are typically 23 generations of airways, and the relation between the trachea diameter d_0 and the diameter d_n of n^{th} branch follows the scaling law:

$$d_n = d_0 2^{-\frac{n}{3}}. \tag{1.3.1}$$

The eq. 1.3.1 holds up to the 16^{th} generation, after this generation the diameter of the airways remains approximatively constant until the 23^{th} generation.

According to the data reported in [36], the diameter of the trachea for adult subjects is (18 ± 2) mm, this number may slightly differ according to patient's sex, age and height. Using eq. (1.3.1) for the first generation of airways, a diameter of (14 ± 1.6) mm is found.

In a CT scan, a lot of structures are clearly visible; the most important are the lungs, ribs, trachea, esophagus, heart, fissures, blood vessels and the airway walls. In the CT slice shown in fig. (1.3.2) some of the mentioned organs are labeled.

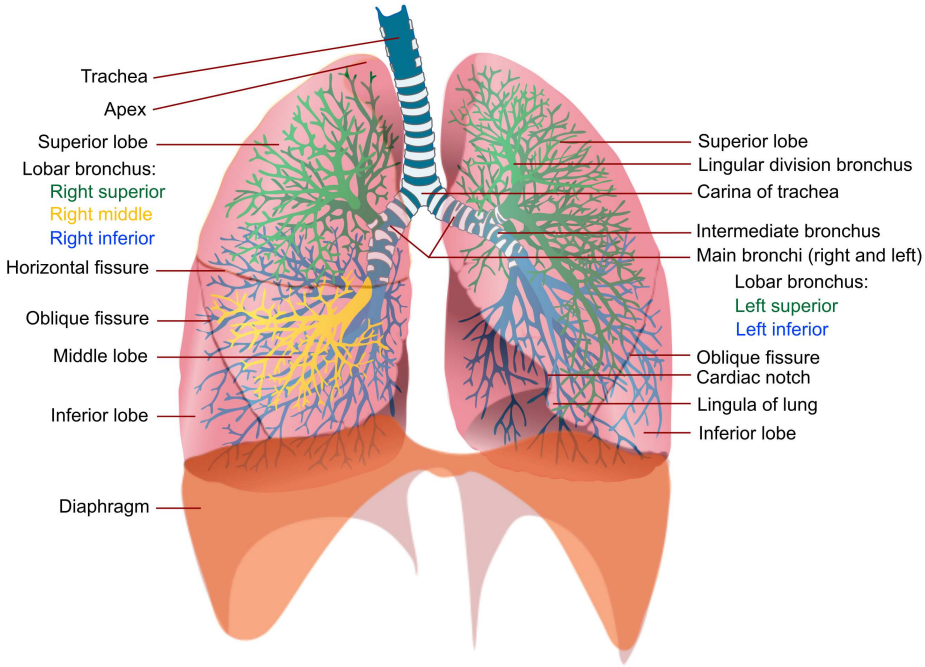


Figure 1.3.1: Figure representing the lung anatomy.

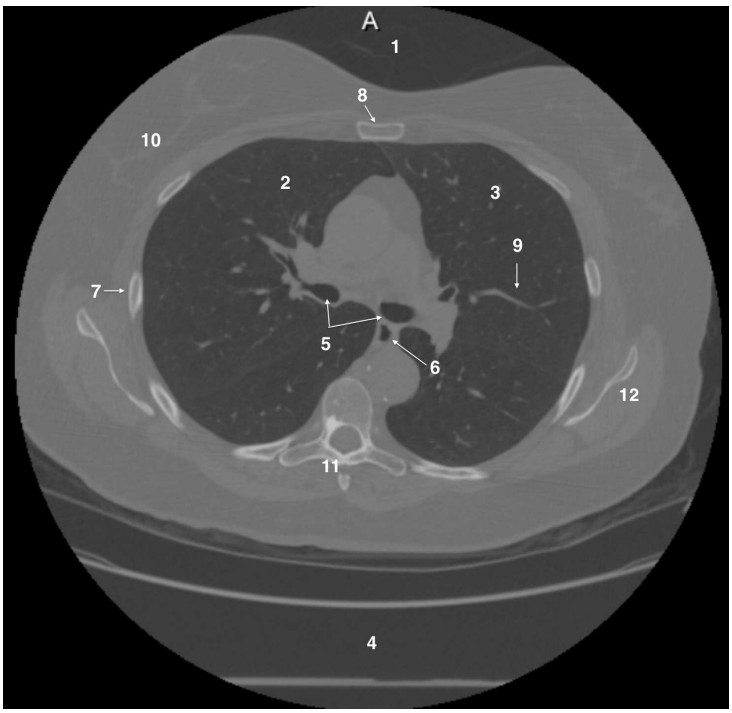


Figure 1.3.2: Slice of a chest CT exam. The principal organs and structures are labeled: 1) Air outside the patient's body, 2) Right lung, 3) Left lung, 4) Table of the scanner, 5) Trachea and large airways, 6) Esophagus, 7) Rib, 8) Sternum, 9) Blood vessels, 10) Fat tissue (patient's body), 11) Spine, 12) Scapula.

Chapter 2

Computer Aided Detection systems

2.1 What is a Computer Aided Detection (CAD) system?

CAD is a procedure that assists radiologists in the interpretation of medical images. Imaging techniques such as X-ray, MRI, and Ultrasound diagnostics yield plenty of information, which the radiologist has to analyze and evaluate comprehensively in a short time. CAD systems can help reading digital images, e.g. highlighting zones of the images which are suspected to contain a radiological sign of pathology (see fig. 2.6.1(b)).

CAD is a relatively young interdisciplinary technology combining elements of artificial intelligence and digital image processing with radiological image processing. A typical application is the detection of a tumor. Some hospitals use CAD to support radiologists in screening programs, e.g. as in mammography (for the diagnosis of breast cancer), and virtual colonoscopy (for the detection of polyps in the colon).

2.2 The role of CAD in a screening program

Screening is a strategy used in a high risk and asymptomatic population to detect a disease in individuals in absence of signs or symptoms of that disease.

The purpose of a screening program is to identify a disease in early stage, thus enabling earlier intervention and management in the hope to reduce mortality and suffering from that disease. Although screening may lead to an earlier diagnosis, not all screening tests have been shown to benefit the population being screened; over-diagnosis, misdiagnosis, and creating a false sense of security are some possible negative effects of screening. For these reasons, a test used in a screening program, especially for a disease with low incidence, must have good specificity in addition to good sensitivity [37].

In a lung cancer screening context, it might be possible to use a CAD system in different modes to improve radiologists's work, in particular the most common modalities are known as "first reader" and "second reader".

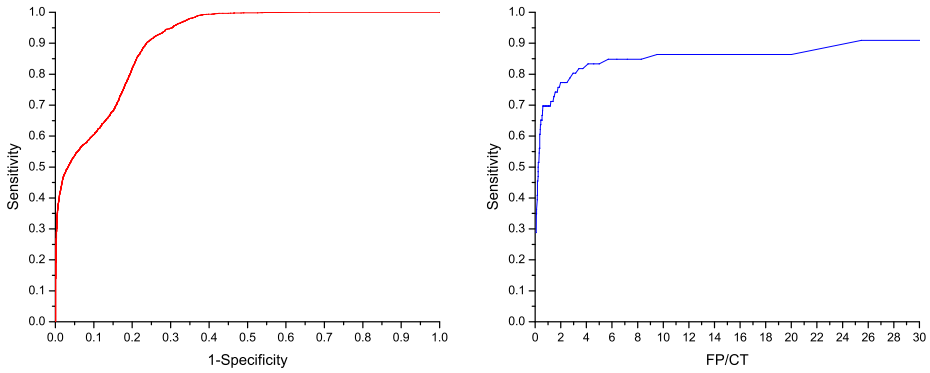
In first reader mode, the radiologists are allowed only to modify CAD findings. This mode is convenient when radiologists are asked to review a lot of images, most of those non pathological. Used in first reader mode the CAD is able to reduce reading time, and to prevent fatigue and inattentions. The main limitation of the first reader mode is that the radiologists' upper bound on the sensitivity is given by the CAD performance. For this reason the use of a CAD as first reader is advised against in general, and limited to the use as an "image sorter": the cases are selected by the CAD in order to be presented to radiologists, at first, those which are more probably pathological.

The most common use of a CAD system is the second reader mode. In the second reader mode, first radiologists read the exam "traditionally" and then with the help of the CAD. In principle this mode is slower than the first reader and the "traditional" modes, so this modality is acceptable only if it implies an enhancing of sensitivity at an acceptable rate of false positives.

2.3 How to evaluate a CAD performance

During the development and the validation of a CAD system, it is necessary to have a tool to compare its performance with those of other CADs or human readers. For this purpose, it is common to use two different figures of merit: Receiver Operating Characteristic (ROC) and Free Receiver Operating Characteristic (FROC) curves (see fig. 2.3.1), according to the task to be evaluated [38].

In this section, the word "observer", may refer both to CADs and to human readers, since the same considerations apply to both cases.



(a) ROC (b) FROC

Figure 2.3.1: Examples of ROC and FROC curves.

The ROC curve is used when binary classification is required, e.g. when an observer is asked to separate highly suspicious images for cancer from non pathological images. For such a kind of task, the observer is not asked to give only a binary answer, but to provide also a degree of suspicion p, e.g. a number in the $[0, 1]$ range, representing the confidence of its choice.

Once a set of images is processed by the observer, it is possible to evaluate its performance, varying a threshold t between 0 and 1, and evaluating for each t the number of True Positive (TP), False Positive (FP), True Negative (TN) and False Negative (FN) findings, with $p \geq t$ (see tab. (2.1)).

The ROC curve is obtained, evaluating the sensitivity $= TP/(TP + FN)$ and the specificity $= TN/(TN + FP)$ for each t, and plotting the sensitivity vs $(1 - \text{specificity})$.

The Area Under the ROC curve (AUC) is a good estimate of the observer performances, since its value is equal to the probability that a randomly chosen positive example is ranked higher than a randomly chosen negative example [39].

However, the ROC curve is not an optimal tool to evaluate tasks where the spatial identification of lesions is not trivial. For example, identifying all the lung nodules in a CT scan is not trivial, since there are lots of structures that mimic nodules and real nodules may be very subtle to detect.

To evaluate such a kind of task, it is necessary to include the identification

Table 2.1: Table showing all the possible outcomes of a test.

Test result	Gold standard	
	Positive	Negative
Positive	True Positive	False Positive
Negative	False Negative	True Negative

in the evaluation of the observer performances. This operation is typically carried out with the FROC analysis. In FROC the observer is asked to provide both a location and a degree of suspicion for each identified lesion.

If an annotated mark is not close enough to "real lesions", the mark is regarded as a FP. The main difference between the FROC and the ROC analysis, is that in the FROC the reader is free to annotate, in principle, an unlimited number of lesions. For this reason, it is not straightforward to compare two FROC curves as in the case of ROC curves. The FROC curve is obtained plotting the sensitivity versus the number of FP, similarly as for the ROC curve.

In this thesis, the FROC curves are compared using the criterion proposed in [17]: the values of sensitivity at seven predefined FP/CT rate (1/8, 1/4, 1/2, 1, 2, 4, 8) are averaged. The FROC analysis requires a localization criterion to be defined. In this thesis, a mark is considered corresponding to a nodule if its Euclidean distance from the annotated center is less than 1.5 times the radius of the annotated nodule.

2.4 The problem of the "gold standard"

In order to carry out FROC and ROC analysis, it is necessary to have a quality reference standard to compare with the method to be tested. In particular, to assess the performance of CAD systems for lung nodule detection, the list of nodule locations annotated by radiologists is needed.

The creation of the reference standard is carried out using the so called *"gold standard",* i.e. the most accurate exam that is able to detect a certain pathology.

Opposite to other cases, e.g. breast cancer, the "gold standard" for detection of lung cancer cannot be based on histological exams, since lung biopsy is a dangerous procedure for the patient, and it is carried out only for big nodules and nodules strongly suspected to be cancer. Therefore, in principle, the biopsy would not be applicable to detect early cancers in asymptomatic population.

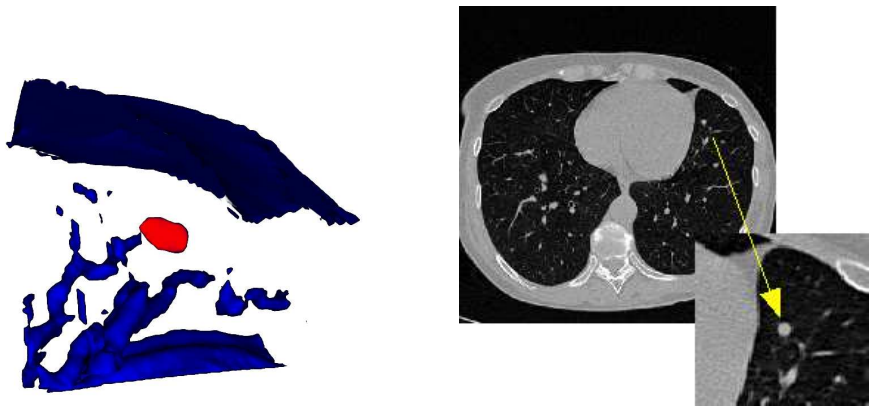
For this reason, the most accurate and less invasive exam for early lung cancer detection is a chest CT red by expert radiologists. However, since the inter-reader variability is very high for this kind of task, the problem of finding a "gold standard" for a certain number of cases is not as easy as it might seem in principle. this problem many solutions were proposed, but the discussion on this topic is nowadays still open.

The most common choice to create a reference standard is to form a panel of experts to review each scan, first independently and then, in case of disagreement, to jointly come together to arrive at a consensus decision. This solution is very difficult to apply in many cases and consensus panels are frequently believed to reflect the opinion of the "strongest" member of the panel [40]. Moreover, in forced consensus studies the inter-reader variability is completely lost.

To overcome these problems, the Lung Image Database Consortium (LIDC) [41] adopted a new reading modality. The LIDC designed a two-phase data collection process that allows multiple expert readers to review each CT scan, first independently (reading phase) and then, with their own and all the blinded annotations of the other readers (unblinded reading phase). Using this method, readers are free to adjust their own annotations without being forced to change them. At the end of this procedure there is no forced consensus, and the final reports contain, separately, for each case, all the unblinded marks of all the readers.

2.5 CAD system architecture

It is convenient, for detection purpose, to divide lung nodules in two different categories, according to their location and morphology: internal nodules and juxta-pleural nodules. Internal nodules (see fig. 2.5.1) originate in the lung parenchyma and can be regarded as spherical objects fully embedded into it, whereas juxta-pleural nodules (see fig. 2.5.2) originate on the pleural



(a) Rendering of the internal part of lung, in red is shown an internal nodule

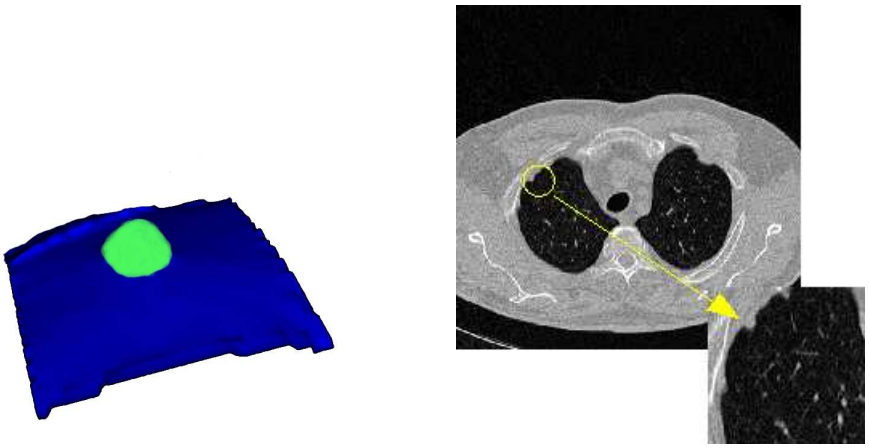
(b) A CT image containing an internal nodule

Figure 2.5.1: Pictures representing an internal nodule.

surface and can be modeled as hemispherical objects. Since these two main categories are very different, the proposed approach implements two different procedures, each devoted to the identification of one typology of nodules:

- CAD_I for the identification of internal nodules;
- CAD_{JP} for the identification of juxta-pleural nodules.

The two CADs share the lung segmentation procedure described in chap.4. After the lung segmentation two procedures devoted to the identification of internal and juxta-pleural nodule candidates are implemented. The procedure to detect internal nodules candidates is described in chap.5, and it is based on a Hessian eigenvalues algorithm. The procedure to detect juxta-pleural candidates is described in chap.6 and it is based on the detection of candidates trough pleura surface normals. After the candidate detection, the classification procedures described in chap.7 are implemented. Finally the two CADs system are combined using the procedure described in 8.4. In fig. 2.5.3 a flowchart of the two CAD system procedures is shown.



(a) Rendering of the pleura surface, in green is shown a juxta-pleural nodule

(b) A CT image containing a juxta-pleural nodule

Figure 2.5.2: Pictures representing a juxta-pleural nodule.

2.6 OsiriX plugin

In order to allow the annotation of the images and the visualization of the CAD findings, a dedicated plugin for the OsiriX DICOM viewer [42] has been implemented within the Cocoa environment of MacOSX. OsiriX is an open source DICOM image processing and visualization software for different imaging types, e.g. Magnetic Resonance Imaging (MRI), CT, Positron Emission Tomography (PET), PET-CT, Single Photon Emission Tomography (SPECT)-CT, Ultrasounds, etc., which allows a simple mechanism for writing plugins that expand its functionalities. The plugin allows annotating nodules and visualizing in different colors the separate CAD sub-system results, their combination and the radiologists' annotations. A screenshot of the plugin is shown in fig. 2.6.1.

Using the plugin, radiologists can visualize and process the CT data using all the tools included in the OsiriX software (zooming, changing the intensity windowing, browsing slices, Maximum Intensity Projection (MIP) projections, etc.), identify nodules and annotate them. The annotations consist of simple circular Region of Interests (ROIs), identified by their center and radius. Using the sliders it is also possible to select a working point, to visualize only findings above the selected thresholds.

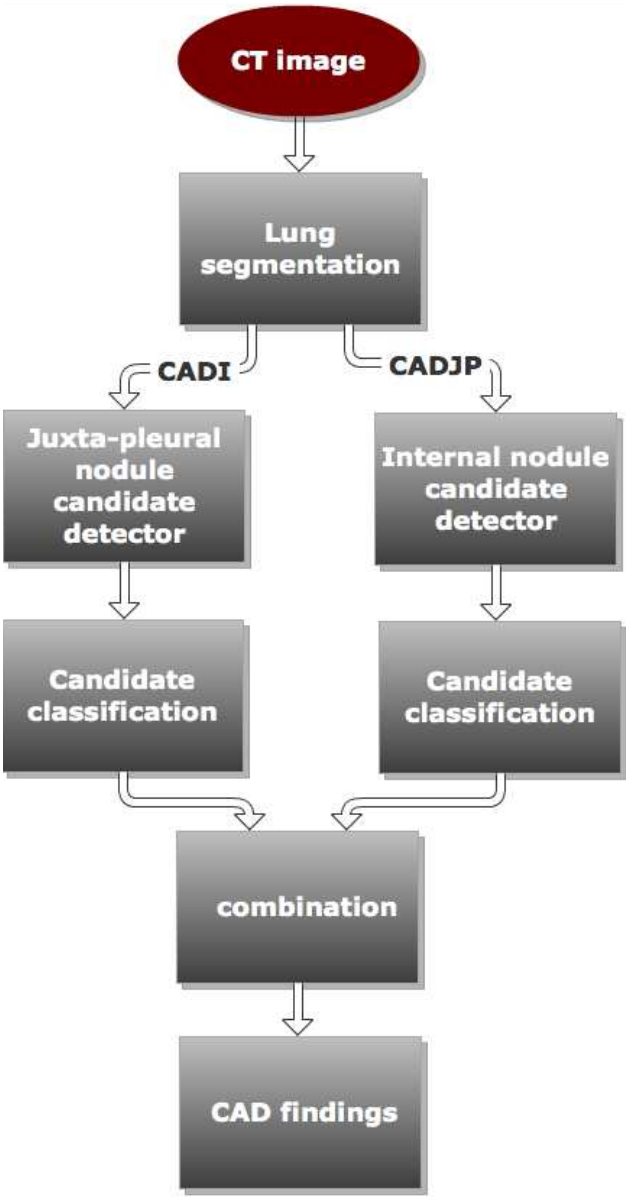
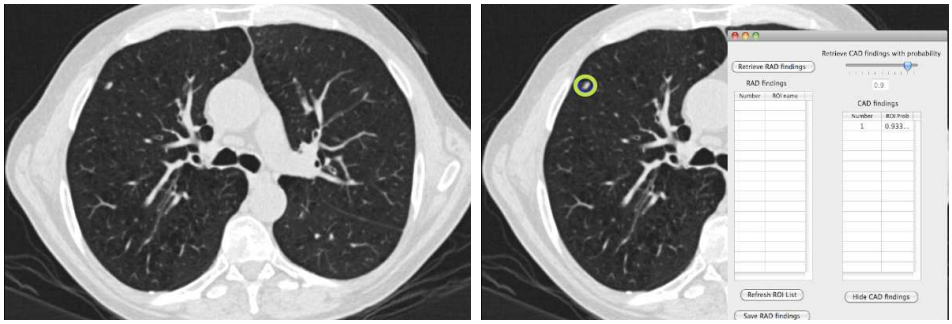


Figure 2.5.3: Flowchart of the CAD sub-systems for internal nodule and juxta-pleural nodule detection.



(a) (b)

Figure 2.6.1: a) Screenshot of an image as seen in the OsiriX software. b) A nodule annotated by the CAD system. The plugin window is visible in the right part of the image.

Chapter 3

Datasets

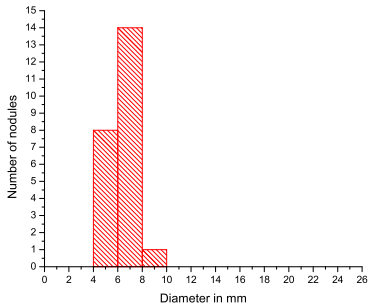
3.1 Introduction

As discussed in sec.2.4, to obtain a reliable annotated dataset of CT may require a lot of effort. Therefore, it is natural to consider the data acquisition and annotation procedure a very important part of a CAD system development. Data coming from clinical environment are often not suitable to train or validate a CAD, since annotations provide qualitative explanation of the lesion position and type. For this reason, and many others, the typical choice to develop a CAD system is to rely on public datasets.

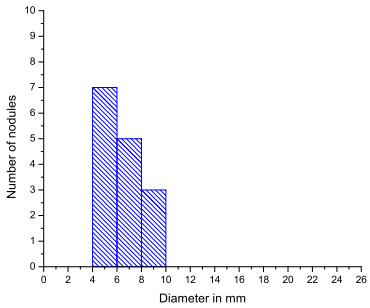
Results obtained on public datasets have the advantage to be directly comparable with those present in literature, therefore, for scientific works, public datasets should be preferred. In this thesis three datasets, two of which are public, are used to train and to validate the proposed CAD scheme: the ITALUNG − CT, the LIDC and the ANODE09 database.

3.2 The ITALUNG − CT database

The presented CAD system has been originally developed and trained on a subset of the ITALUNG − CT database, collected by the Pisa center of the ITALUNG−CT trial, which is the first Italian randomized controlled trial for the screening of lung cancer [43]. The CT scans were acquired with a 4-slice spiral CT scanner (Siemens Volume Zoom) according to a low-dose protocol (tube voltage: 140 kV, tube current: 20 mA, mean equivalent dose 0.6 mSv),



(a) Histogram of diameters of the internal nodules.



(b) Histogram of the diameters of the juxta-pleural nodules.

Figure 3.2.1: Diameter of the nodules annotated in the 20 ITALUNG − CT CT scans.

with 1.25 mm slice collimation. Each scan is stored in DICOM format. Slices were reconstructed at 1 mm thickness, using a medium sharp reconstruction kernel (Siemens B50f). The number of slices per scan is approximately 300, each slice being a 512 by 512 pixel matrix, with pixel sizes ranging from 0.53 to 0.74 mm and 12 bit gray levels in Hounsfield Units. The annotations were marked by experienced radiologists participating to the Italian MAGIC-5 research project [44]. In this thesis, a dataset of 20 CT scans, containing 23 internal nodules and 15 juxta-pleural nodules with the size distribution shown in fig. 3.2.1, is used.

3.3 LIDC database

The LIDC database is the largest collection of annotated and publicly available CTs. LIDC is a multi center and multi manufacturer database, with cases of different collimation, kV and tube current and reconstructed with different slice thickness.

On the contrary to the previous release of the LIDC database, where the annotations were obtained by a forced consensus among different readers, in the current release, to capture the inter-reader variability, four different

annotations made by four expert radiologists for each case in a two phase reading modality are provided, as explained in sec.2.4.

The LIDC annotations contain three kinds of objects [45]: nodules with diameters ≥ 3 mm, nodules with diameters < 3 mm and "non nodules" with diameters > 3 mm. The contours of the objects marked as nodules with diameters ≥ 3 mm are provided for every reader together with eight subjective characteristics in a $1-5$ scale: subtlety, internal structure, calcification, sphericity, margin, spiculation, texture, malignancy. For nodules with diameters < 3 mm and non nodules with diameters > 3 mm only a centroid is provided and no information on the size is available.

However, for a detection purpose LIDC data are redundant, since only a centroid and an estimate of the lesion size are needed.

Starting from the annotated contours, it is possible to obtain the information needed with many different algorithms. However, in order for the results to be directly comparable with those of literature, the annotations provided by the Cornell University [46] are used. These annotations are built with the purpose to be a reference standard to compare results of different data groups, and are generated according to the prescriptions given in [47].

First, the correspondence among contours drawn by different readers are evaluated: a mask containing all the lesion voxels is generated starting from each contour, then the masks with at least one voxel in common are considered referring to the same object.

To evaluate the center and the dimension of a lesion, first each contour referring to a lesion is assigned a center of mass, then the median value of the center in each direction provides the center of the lesion.

To estimate the size of the lesion, the volume of voxels contained in each contour is evaluated, and the median volume is considered the volume of the lesion. To obtain an estimation of the nodule size, the diameter of the sphere having the same volume as the nodule estimated volume is evaluated, i.e.

$$diameter = 2 \cdot \left(\frac{3 \cdot V}{4\pi} \right)^{\frac{1}{3}}$$

where V is the estimated volume of the lesion.

It is common to report the results on this database using four Agreement Level (AL) among the single reader annotations, i.e. considering the nodules

Table 3.1: Number of nodules in the datasets $LIDC_1$ and $LIDC_2$ divided by category.

Dataset	Internal Nodule (IN) (relevant/irrelevant)	Juxta-Pleural Nodule (JPN) (relevant/irrelevant)
$LIDC_1$	137 (96/41)	72 (42/30)
$LIDC_2$	137 (95/42)	36 (19/17)

Table 3.2: Number of nodules in the datasets $LIDC_1$ and $LIDC_2$ divided by AL.

Dataset	AL 1	AL2	AL3	AL4	number of CTs
$LIDC_1$	209	138	98	67	69
$LIDC_2$	173	114	86	42	69

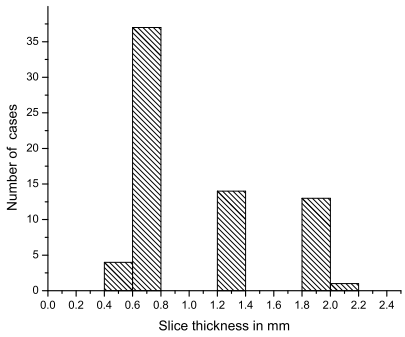
annotated by at least one (AL1), two (AL2), three (AL3) and four radiologists (AL4).

At present, the LIDC database contains 1000 CTs. However, in this thesis, only 138 CT scans are used. The 138 CTs result from selecting cases with slice thickness $\leqq$ 2 mm, without contrast media and annotation included in the Cornell University report. The 138 CTs were divided in two subsets named $LIDC_1$ and $LIDC_2$, consisting of 69 CTs each.

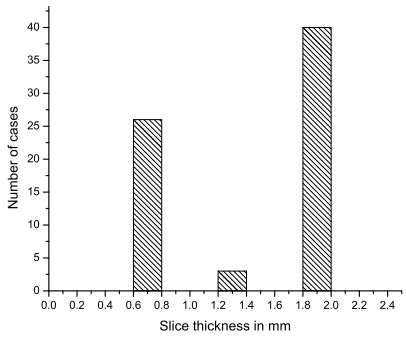
Nodules with AL1 are referred to as irrelevant in this thesis, since all the training procedure is carried out using nodules with at least AL2.

Table 3.1 shows the number of nodules contained in the two subsets, divided into two categories. Table 3.2 shows the agreement level for the datasets $LIDC_1$ and $LIDC_2$. In fig. 3.3.1 the slice thickness of the two subsets are shown.

Figures 3.3.2 and 3.3.3 show the nodule diameter distributions for relevant and irrelevant nodules. Since the LIDC annotations provide no information on the nodule typology, the nodules are divided in two categories, namely internal or juxta-pleural, on the basis of visual assessment.



(a) Slice thickness distribution of the $LIDC_1$ dataset.



(b) Slice thickness distribution of the $LIDC_2$ dataset.

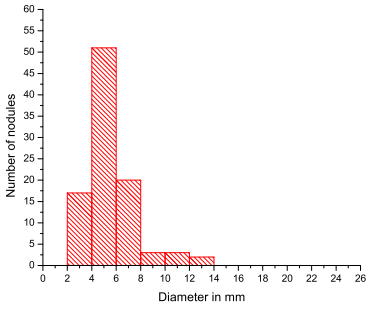
Figure 3.3.1: Slice thickness of the CT scans of $LIDC_1$ and $LIDC_2$ datasets.

3.4 ANODE09 database

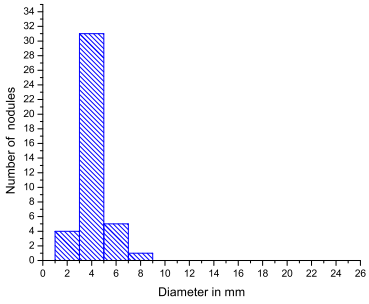
The ANODE09 [48] is an international initiative devoted to compare objectively different CAD systems, able to perform automatic detection of pulmonary nodules in chest CT scans on a single common database, with a single evaluation protocol. Data is provided by the Nelson study, the largest CT lung cancer screening trial in Europe. The images, acquired according to a low dose protocol, have slice thickness between 0.7 and 1 mm and an average number of slices equal to 430. Any team, whether from academia or industry, can join this study. The database of this study consists in five example CTs with publicly available annotations and 50 low dose thin-slice CT scans without public availability of the annotations. The 50 CTs are intended as a validation dataset, so it is forbidden to use these data to train any CAD system. The only information known about the 50 CTs, is the distribution of the nodule diameters: 40% of the nodules are below 4 mm in diameter, 40% have a diameter between 4 and 6 mm and 20% are larger.

Since annotations for the 50 CT scans are not publicly available, all the findings must be inserted in a file together with their coordinates and degree of suspicion, and uploaded to the ANODE09 website [48], to receive in return the FROC curves.

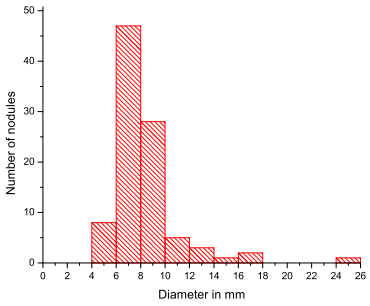
The FROC is evaluated using the same criterion used in this thesis, which



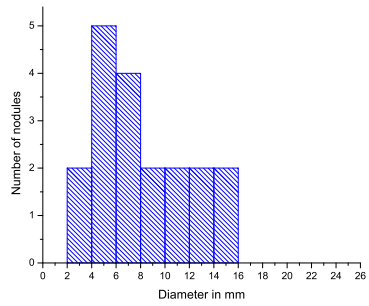
(a) LIDC_1 relevant nodule distribution.



(b) LIDC_1 irrelevant nodule distribution.

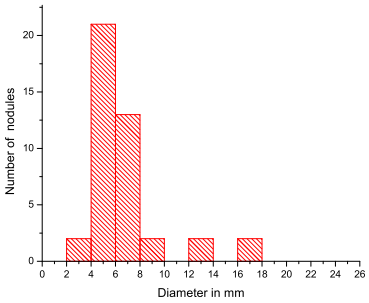


(c) LIDC_2 relevant nodule distribution.

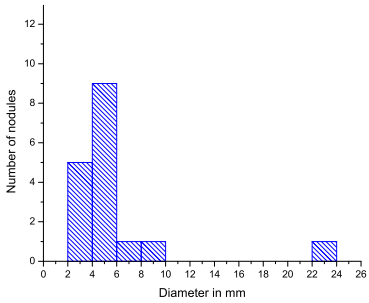


(d) LIDC_2 irrelevant nodule distribution.

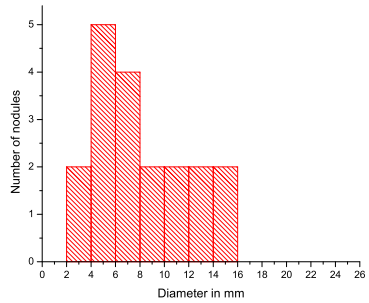
Figure 3.3.2: LIDC_1 and LIDC_2 internal nodule size distributions.



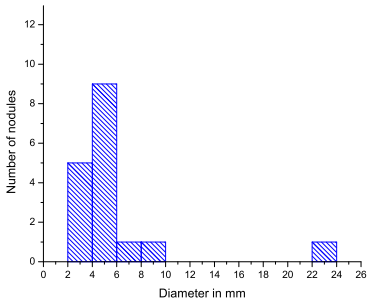
(a) $LIDC_1$ relevant nodule distribution.



(b) $LIDC_1$ irrelevant nodule distribution.

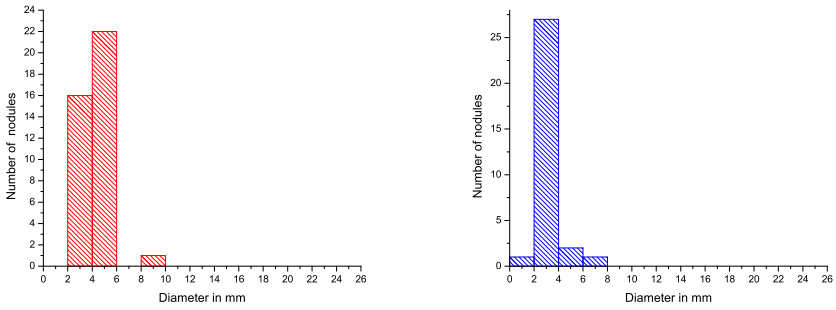


(c) $LIDC_2$ relevant nodule distribution.



(d) $LIDC_2$ irrelevant nodule distribution.

Figure 3.3.3: $LIDC_1$ and $LIDC_2$ juxta-pleural nodule size distributions.



(a) Relevant nodule distribution. (b) Irrelevant nodule distribution.

Figure 3.4.1: Diameters of the nodules contained in the five example CTs of the ANODE09 dataset.

allows for a direct comparison.

As discussed in sec. 2.3, to extract a single score value from the FROC curve, the average sensitivity at seven predefined false positive rates is computed: $1/8, 1/4, 1/2, 1, 2, 4$, and 8 FP/CT. This value is referred to as Free Receiver Operating Characteristic Score Value (FROCSV) in the following sections.

The nodules of the ANODE09 dataset are divided into six categories according to their location and shape, thus allowing a more accurate scoring of each category independently: nodules bigger than 5 mm, nodules smaller than 5 mm, isolated nodules, juxta-pleural nodules, juxta-vascular nodules and peri-fissural nodules.

The dataset contains also irrelevant findings, i.e. findings that are counted neither as FP nor as TP in the evaluation of the FROC curve. There are three types of irrelevant findings: findings that mimic a nodule but that an expert observer believes not to be a nodule, nodules with benign characteristics and nodules that are too small to be relevant.

The nodules contained in the five example CTs were divided, according to visual assessment, into internal nodules and juxta-pleural nodules, thus resulting in 36 relevant internal nodules and only three relevant juxta-pleural nodules.

Part II

Image analysis algorithms

Chapter 4

Lung segmentation

4.1 Introduction

In computer vision, segmentation refers to the process of partitioning a digital image into multiple segments, i.e. sets of pixels. The goal of segmentation is to simplify and/or change the representation of an image into more meaningful and easier information to analyze. In this case, the lung segmentation module of the CAD is essential to avoid marking findings outside lungs, and to provide an accurate representation of the pleura surface, necessary to detect juxta-pleural nodules. Since segmentation methods always rely on anatomical models, the information provided in sec. 1.3 is often used in this chapter.

The chapter is organized as follows: first, the image preprocessing and the algorithm to find low intensity voxels inside the patient body are shown, then the algorithm for the trachea and large airways segmentation and then the algorithm to separate the lungs and to fill vessels and airways are described, finally, the results obtained on a subset of the LIDC dataset are shown.

4.2 Image preprocessing

Shape identifier and morphological algorithms are designed to work better with isotropic images, i.e. images with cubic voxels. However, as it is discussed in sec. 1.1, CT images are typically anisotropic. To improve the

performances of morphological and shape identifier filters, an isotropization procedure is implemented.

Among many possible and arbitrary choices to resample the CT volume, it was chosen to resample the image in order to obtain a spacing of 1 mm and to smooth the image with a Gaussian kernel proportional to the original spacing of the images, as suggested in [24, 49].

4.3 Identification of low intensity voxels inside the patient's body

To identify the low intensity voxels inside the patient body, a procedure similar to [29] is implemented. This procedure consists in a simple thresholding [50] between $S_{up} = -600\,HU$ and $S_{low} = -1000\,HU$ followed by a tridimensional connected component labeling. Then, searching the biggest connected component not lying on the boundary of the volume (see fig. 4.3.1) provides the lungs together with air voxels contained in the trachea and the large airways. If two connected components, with a volume ratio of at least 0.5 are found, the algorithm goes to the "vessel and airway walls removal" and the lungs are considered to be separated.

The output of this procedure is a mask M where the low intensity voxels are labeled as foreground, i.e. are set to 1, while the rest of the image is labeled as background, i.e. is set to 0.

As shown in fig. 4.3.2, the air voxels inside the trachea act as a bridge between the two lungs. To obtain the segmentation for each lung separately, it is necessary to segment out from M the trachea and the large airways.

4.4 Trachea segmentation

Trachea segmentation is a relatively common task in lung CT segmentation. There are plenty of algorithms available in literature to segment trachea, the presented approach is a mixture of different elements, with the purpose of roughly identifying the trachea and the large airways connected to it.

(a) (b) (c)

Figure 4.3.1: Steps of the identification of the low intensity voxels inside the patient's body a) Thresholding between S_{up} and S_{low}: the low intensity voxels are labeled as foreground b) Connect component and labeling operation. The labels are shown in different gray levels c) Selection operation: the biggest connected component not lying on the boundary of the image is selected, whereas the others are discarded.

The algorithm is divided into four steps:

- optimal threshold identification;
- threshold and connect component analysis;
- search of connected component representing the trachea;
- trachea removal.

Optimal threshold identification

To segment out from M the trachea and the large airways, a two materials decomposition approach is implemented [30]. This approach relies on the hypothesis that the voxels contained in the mask M can be divided into two categories: lung voxels and air voxels. Since trachea and large airways are mostly filled with air, the first step to identify them is to find all the voxels of air contained in M.

In particular, assuming that the lungs are composed by air and lung tissue, it is possible to write the two material decomposition formula

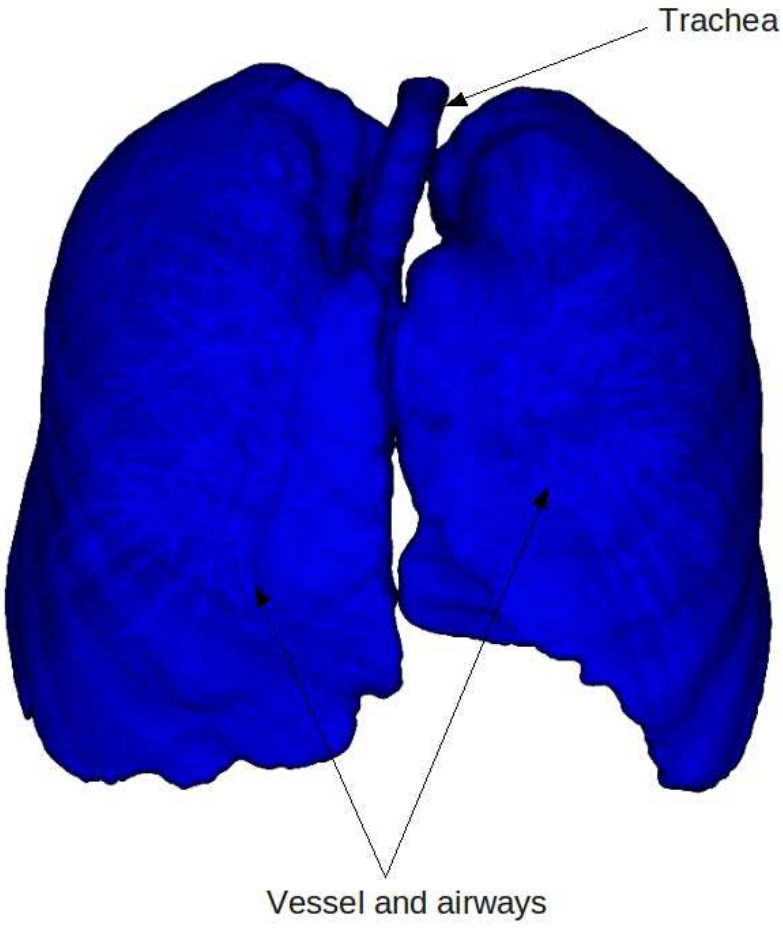


Figure 4.3.2: Rendering of the low intensity voxels obtained in first step of the segmentation procedure.

$$n_M \bar{H}_M = n_{Air} \bar{H}_{Air} + n_{Lung} \bar{H}_{Lung} \tag{4.4.1}$$

where $\bar{\mathrm{H}}_{\mathrm{M}}, \bar{\mathrm{H}}_{\mathrm{Air}}, \bar{\mathrm{H}}_{\mathrm{Lung}}$ are respectively the average density of the voxels on the mask M, the average intensity of the air and the average intensity of the lung tissue, and $\mathrm{n}_{\mathrm{Air}}, \mathrm{n}_{\mathrm{Lung}}, \mathrm{n}_{\mathrm{M}}$ are respectively the number of air voxels, the number lung tissue voxels and the total number of voxels in M, with

$$n_M = n_{Air} + n_{Lung}. \tag{4.4.2}$$

Given $\bar{\mathrm{H}}_{\mathrm{Air}}, \bar{\mathrm{H}}_{\mathrm{Lung}}$, and using eq. 4.4.2, it is possible to obtain $\mathrm{n}_{\mathrm{Air}}$ inverting eq. 4.4.1

$$n_{Air} = n_M \frac{\bar{H}_M - \bar{H}_{Lung}}{\bar{H}_{Air} - \bar{H}_{Lung}}.$$

By evaluating the histogram of HU values included in the mask M, it is possible to find the threshold $\mathrm{T}_{\mathrm{Air}}$ so that

$$T_{Air} = \max_{T \in [T_{min}, T_{max}]} \{n_T \leq n_{Air}\}$$

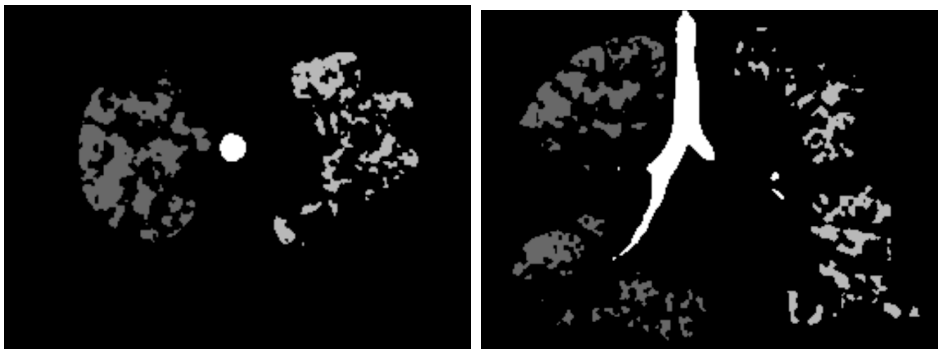
where n_{T} is the number of voxels, inside the mask M with intensity values $\leq \mathrm{T}$ and $\mathrm{T}_{\mathrm{min}}$ and $\mathrm{T}_{\mathrm{max}}$ are the maximum and minimum values of the voxels inside M. The values of $\bar{\mathrm{H}}_{\mathrm{Air}} = -999\,\mathrm{HU}$ and $\bar{\mathrm{H}}_{\mathrm{Lung}} = +48\,\mathrm{HU}$ are selected according to the prescription given in [30].

Threshold and connected component analysis

Once $\mathrm{T}_{\mathrm{Air}}$ is found, to identify all the air voxels inside the mask M, a thresholding operation between $\mathrm{T}_{\mathrm{min}}$ and $\mathrm{T}_{\mathrm{Air}}$ is applied. Then a connect component analysis is able to discard all the isolated air voxels. The results of this operation is shown in fig. 4.4.1, where a connected component representing the trachea is clearly visible.

Search of connected component representing the trachea

To identify the trachea among all the connected components shown in fig. 4.4.1, a circular 2D shape, corresponding to the trachea axial section, is searched near the head of the patient, i.e. in the upper central region of the CT. The identification of the circular section is carried out using the Hough transform for circle [50].



(a) Axial view (b) Coronal view

Figure 4.4.1: Air segmentation: the connected component found selecting those voxels with intensity $\leq T_{Air}$. a) In axial view connected components are shown in different gray levels b) the same connected components of a) viewed in coronal view.

The Hough transform is an operation that allows the identification of different shape classes (e.g. line, circles, ellipses, etc etc...) in binary images, using a voting procedure. The voting procedure is carried out in a parameter space, and the object candidates are obtained as local maxima in a so-called accumulator space, constructed by the algorithm for computing the Hough transform. A simple explanation of the Hough transform follows.

Hough transform for circles of known radius

Suppose that a circle C of known radius r has to be found in a bidimensional image.

The parametric form of the C is written

$$\begin{aligned} x &= r \cdot \cos(\theta) + a \\ y &= r \cdot \sin(\theta) + b \end{aligned} \tag{4.4.3}$$

where $\theta \in [0, 2\pi]$ is a polar angle and a,b are the coordinates of the circle center.

Inverting the eq. 4.4.3 with respect to a and b, yields to

$$\begin{aligned} a &= x - r \cdot \cos(\theta) \\ b &= y - r \cdot \sin(\theta) \end{aligned} \tag{4.4.4}$$

which is a circle of radius r centered in x, y.

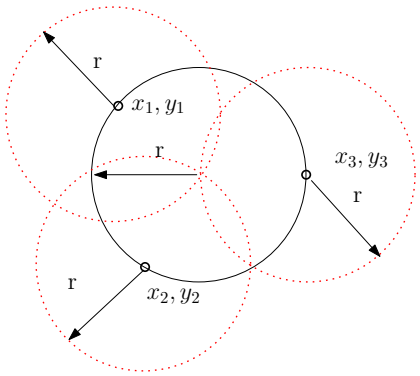


Figure 4.4.2: Sketch representing the idea of the Hough transform for circles: the black circle C is the one to be detected, while the three dotted circles of radius r and centers (x_1, y_1), (x_2, y_2), (x_3, y_3) are circles drawn in the parameter space. As shown in the figure, the point where the three red circles intersect is the center of C.

Since the radius of the circle to be found is known, the parameter space is two dimensional, i.e. the two dimensions are the coordinated (a, b). According to the transform eq. 4.4.4, each point (x, y), lying on C, becomes a circle of radius r itself in the parameter space.

Each pixel in the parameter space accumulates a score proportional to the number of circles passing through it (see fig. 4.4.2). Once this procedure is applied, the center of C is given by the pixel that accumulates the higher score.

Hough transform for circles of unknown radius

If the radius of the circle to be found is not known, the procedure is similar, except that the range $[r_{min}, r_{max}]$ of allowed radii is spanned. Then for each $r \in [r_{min}, r_{max}]$, the Hough procedure with a known radius is repeated.

To apply the Hough algorithm to the trachea identification, first a morphological operation to evaluate the border of the mask is implemented, and then the Hough transform is applied on the upper central region of the CT image. The pixel with highest score in the accumulator is then used as the seed to identify the trachea, while the other connected components are discarded.

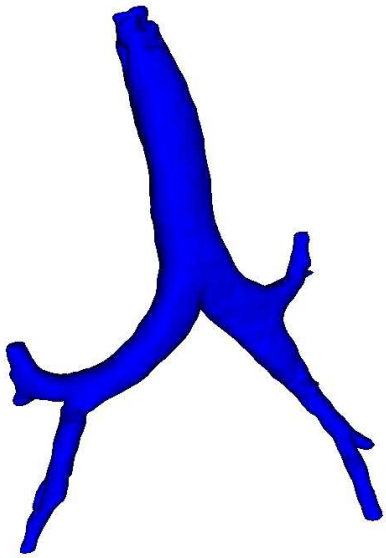


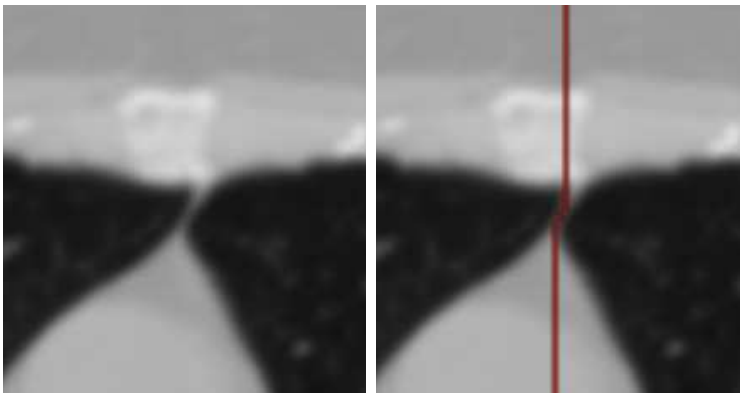
Figure 4.4.3: Rendering of the mask T, representing the trachea and the large airways.

Following the prescriptions given in sec. 1.3, values of $r_{min} = 7$ mm and $r_{max} = 11$ mm for the trachea radius are selected, and it is chosen, on the basis of experiments, to search the trachea only in the upper 10% of the slices and in a band of 20 mm around the image center. The output of this procedure is the mask T shown in fig. 4.4.3.

Trachea removal

Once the trachea and the large airways are correctly identified, a dilation operation with a spherical kernel of radius 2 mm is applied to the mask T to include the high intensity walls of trachea and the airways. Then each voxel belonging to the dilated T mask is deleted from the mask M, thus generating a new mask M_1.

At the end of this procedure lungs may be still fused, due to the low intensity of the anterior and posterior junction. In case of failure, a dedicated procedure is implemented to identify lung junctions.



(a) (b)

Figure 4.5.1: a) The anterior lung junction b) The anterior lung junction where the maximum cost path, found in the lung separation step, is superimposed.

4.5 Lung separation

If the lungs are still fused, a dedicated procedure similar to that reported in [31,51] is implemented. The procedure search, for each slice, the maximum path weighted on HU between the anterior and posterior part of the patient.

This method relies on the assumption that the junction tissue has a larger HU value compared to the tissue generally contained in the mask M_1. The junction lines are searched only in band of size s = 30 mm near the center of mass of the segmented region contained in M_1. Before starting the path search, the voxel values outside the mask M_1 are set to +2000, in order to prevent the path to enter in the lungs when it's not necessary.

The maximum cost of the path is evaluated using a dynamic programming approach [52]. The results of this procedure are shown in fig. 4.5.1. Once the optimal path is found, the voxels belonging to it are subtracted from the M_1 mask and the connected components are searched with a 6-pixel connectivity rule.

(a) (b)

Figure 4.6.1: a) The segmented lung before vessels and airway walls removal b) The segmented lung after vessels and airway walls removal

4.6 Vessels and airway walls removal

As shown in fig. 4.6.1, vessels and airways walls are not included in the segmented lung at this stage. To include them without modifying the pleura surface morphology, i.e. without modifying the shape of the juxta-pleural nodules, a combination of morphological operators is applied [50]. In particular, a sequence of the dilation and the erosion operators with spherical kernels r_d and r_e, with $r_e > r_d$ is applied. Finally, the logical OR operation between the obtained mask and the original lung mask provides the final mask, where the vessels and the airway walls are filled in, while maintaining the original shape of the lung mask border [22]. The shape of the juxta-pleural nodules is not modified by this procedure (see fig. 4.6.1). Once removed vessels and airways, the pleura surface is defined as the surface separating parenchyma from the rest of the image (see fig. 4.7.1).

The optimal parameter $r_d = 8$ mm is chosen to fill holes whose dimension is approximatively equal to that of the first generation of airways (see sec.1.3). Whereas the parameter $r_d = 16$ mm is chosen to erode the pleura surface of 16 mm more than the dilation operation, thus avoiding to undersegment juxta-pleural nodules with a radius up to 16 mm.

4.7 Automatic error check

The segmentation of anatomical structures in CT scans is a challenging task in medical imaging. Therefore, it is reasonable to expect any algorithm to fail in some cases, due to small anatomical differences among patients and different parameters in image acquisition and reconstruction.

For this reason, it is important to implement a tool able to automatically check the organ segmentations and to assess, on the basis of observables, if the segmentations are likely to be correct or not.

The procedures to be checked against failures are the trachea segmentation and the lung separation.

The trachea segmentation is considered successful, when a connected component with a volume in the range of $V_{min} = 10000\,mm^3$ to the $V_{max} = 70000\,mm^3$ is found, otherwise the segmented region is discarded. The values V_{min} and V_{max} are chosen on the basis of the range of values found on the $LIDC_1$ set described in sec. 3.3.

The trachea segmentation failure is mainly due to the selection of a threshold T_{Air} unable to separate the trachea air from the lungs. The lung separation is checked against failure and the identified connected component are filtered according to their dimensions. The lungs are finally considered to be separated when two connected components with at least a volume ratio of 0.5 are found.

Failures in separation are mainly due to the fact that, separating lungs slice by slice, doesn't imply a tridimensional separation. In case of failures, the default behavior of the procedure consists in applying the fill vessels procedure explained in sec. 4.6 to the whole block including the two lungs.

4.8 Results

The segmentation algorithm is tested and optimized on a subset of LIDC consisting in 138 CTs from different centers and acquired with scanners from different manufacturers. The $LIDC_1$ is used to set segmentation parameters, while $LIDC_2$ is used to validate the results. In tab. 4.1, the results of the segmentation procedure are shown, the failure of the algorithm is assessed on the basis of the automatic error check and visual assessment.

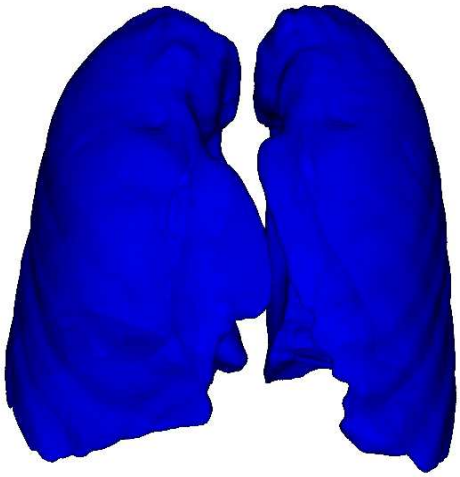


Figure 4.7.1: Rendering of the lung segmentation mask.

Table 4.1: Table showing the results of the segmentation procedure obtained on the LIDC subsets.

Procedure	Dataset	
	$LIDC_1$	$LIDC_2$
Trachea segmentation failure	1/69	1/69
Lung separation failure	6/69	7/69

The results of this procedure show that the trachea segmentation fails once on the $LIDC_1$ dataset, once on the $LIDC_2$; whereas the lung separation procedure fails on 6 cases of the $LIDC_1$ and 7 of the $LIDC_2$. In all the other cases, a good segmentation was achieved. The workflow of the segmentation algorithm is resumed in fig. 4.8.1.

These tests show that the segmentation is stable even though it can fail in some cases.

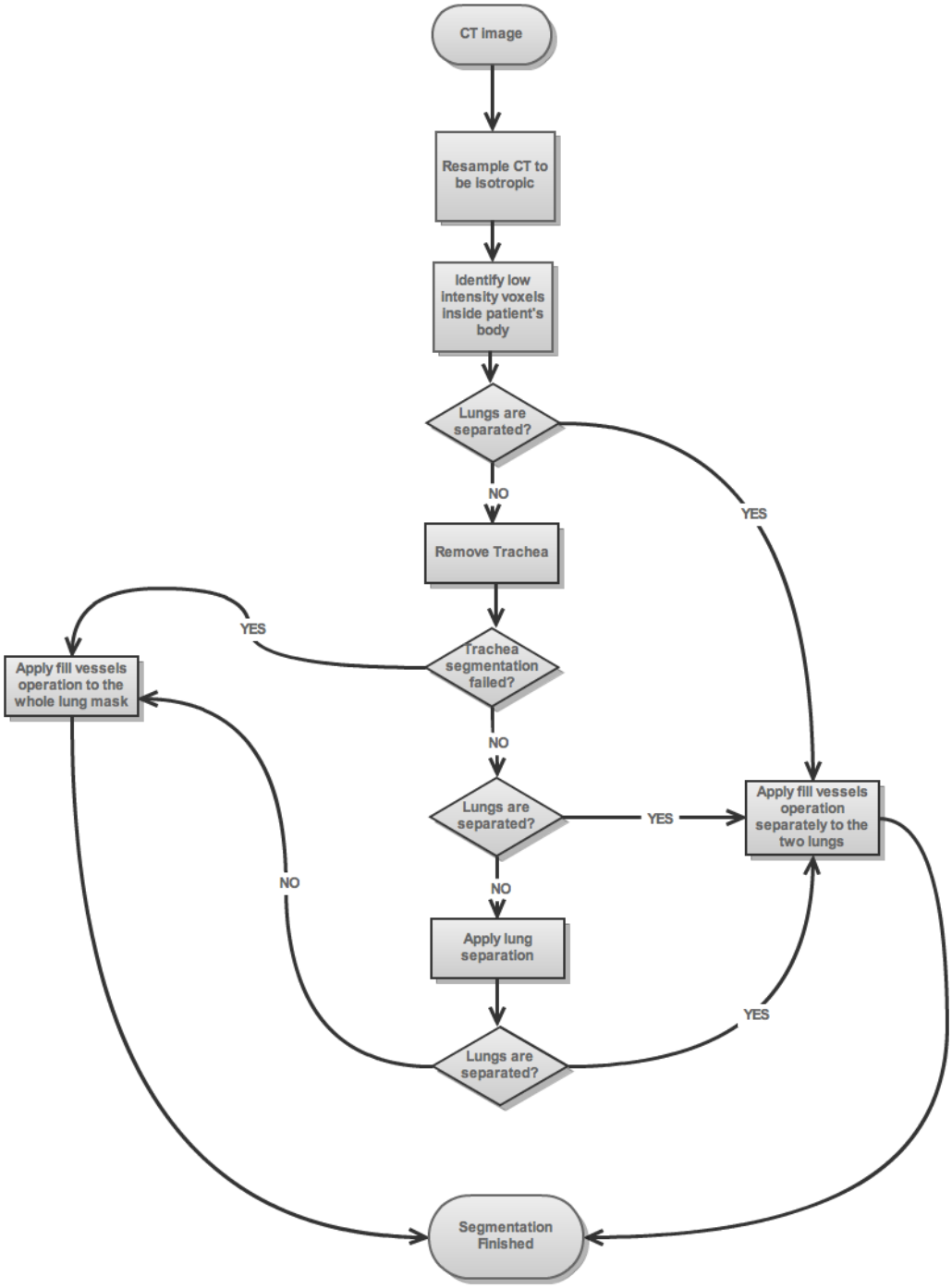


Figure 4.8.1: Flow Chart representing the lung segmentation process.

Chapter 5

Detection of internal nodules

5.1 Introduction

Hessian algorithms are very common image analysis, since they provide a robust tool to find shapes like cylinders or dots in bidimensional and tridimensional images. In this chapter the algorithm used to identify solid internal nodules is shown. The chapter is organized as follows: first the Single Scale Dot Enhancer (SSDE) and Multi Scale Dot Enhancer (MSDE) algorithms [20] are explained, then an efficient implementation of the algorithm is described and the numerical results of the implemented MSDE together with the theoretical algorithm are shown.

5.2 Single Scale Dot Enhancer (SSDE) filter

As discussed in [20], Gaussian blobs are an effective model to represent internal nodules. To enhance such a kind of objects inside the lung volume, a SSDE filter is implemented.

For example, in a tridimensional space the expression of a Gaussian is

$$I(x, y, z) = H_0 \cdot \exp(-\frac{x^2}{2\sigma_x^2} - \frac{y^2}{2\sigma_y^2} - \frac{z^2}{2\sigma_z^2}). \tag{5.2.1}$$

where H_0 is a constant, σ_x^2, σ_y^2, σ_z^2 are the variances along the three axis of the Gaussian, i.e. the dimensions of the object.

According to the values assumed by $\sigma_x, \sigma_y, \sigma_z$, the function $I(x, y, z)$ represents different structures.
It is possible to distinguish three main cases:

1. Blob shape: The variances have the same order of magnitude, e.g. $\sigma_x \approx \sigma_y \approx \sigma_z \approx \sigma$, $I_{blob}(x, y, z) = H_0 \cdot \exp(-\frac{x^2+y^2+z^2}{2\sigma^2})$.

2. Cylinder shape: One of the variances much bigger than the others, e.g. $\sigma_x \to \infty$ represents a tubular structure with the axis parallel to the x axis $I_{cylinder}(x, y, z) \approx H_0 \cdot \exp(-\frac{y^2}{2\sigma_y^2} - \frac{z^2}{2\sigma_z^2})$.

3. Plane shape: two variances have $\sigma_x \to \infty$ and $\sigma_y \to \infty$, $I_{plane} \approx H_0 \cdot \exp(-\frac{z^2}{2\sigma_z^2})$.

By evaluating the Hessian matrix of eq. 5.2.1, the eq. 5.2.2 is obtained

$$H(x,y,z) = \begin{pmatrix} (\frac{x^2}{\sigma_x^4} - \frac{1}{\sigma_x^2}) & \frac{xy}{\sigma_x^2\sigma_y^2} & \frac{xz}{\sigma_x^2\sigma_z^2} \\ \frac{xy}{\sigma_x^2\sigma_y^2} & (\frac{y^2}{\sigma_y^4} - \frac{1}{\sigma_y^2}) & \frac{yz}{\sigma_y^2\sigma_z^2} \\ \frac{xz}{\sigma_x^2\sigma_z^2} & \frac{yz}{\sigma_y^2\sigma_z^2} & (\frac{z^2}{\sigma_z^4} - \frac{1}{\sigma_z^2}) \end{pmatrix} I(x,y,z) \qquad (5.2.2)$$

where $H(x, y, z)$ is the matrix form of the eq. 5.2.3

$$H_{i,j}(x,y,z) = \partial_i \partial_j I(x,y,z) \qquad (5.2.3)$$

for $i, j = 1, .., 3$.
By computing $H(0, 0, 0)$, i.e. evaluating $H(x, y, z)$ in the Gaussian center, a diagonal matrix is obtained

$$H(0,0,0) = H_0 \begin{pmatrix} -\frac{1}{\sigma_x^2} & 0 & 0 \\ 0 & -\frac{1}{\sigma_y^2} & 0 \\ 0 & 0 & -\frac{1}{\sigma_z^2} \end{pmatrix}. \qquad (5.2.4)$$

This means that the eigenvalues of the Hessian matrix, in the center of the Gaussian, are function of the sizes of the object to be enhanced. Thus, using the Hessian eigenvalues, it is possible to evaluate a score proportional to the probability to have a Gaussian shape centered in that point.

In practice, since the analytic form of $H(x, y, z)$ is not known, the Hessian needs to be evaluated numerically, for example using a finite difference approach or a recursive approach [53]. Computing Hessian involves computing derivatives. However, since derivatives without any prior smoothing are known to enhance noise [54], it's necessary to define a scale of interest.

The definition of a scale of interest in computer vision is a fundamental concept, in order to give the algorithms an idea of the dimension of the objects to be searched. Once the eigenvalues $\lambda_1, \lambda_2, \lambda_3$ of the Hessian matrix are evaluated for each voxel, it is possible to calculate the score

$$z_{dot}(\lambda_1, \lambda_2, \lambda_3) = \begin{cases} \frac{|\lambda_3|^2}{|\lambda_1|} & \text{if } \lambda_1, \lambda_2, \lambda_3 < 0 \\ 0 & \text{otherwise} \end{cases}; \tag{5.2.5}$$

where $\mid \lambda_1 \mid \geq \mid \lambda_2 \mid \geq \mid \lambda_3 \mid$.

Evaluating the z_{dot} value, for a blob, a cylinder and plane shapes, it is possible to show that the z_{dot} vanishes in the center of the planar and cylindrical structures, while for blobs of variance σ^2, $z_{dot}(I_{blob}) \sim \frac{1}{\sigma^2}$. This means that the SSDE is able to discriminate blobs from planar and cylindrical structures and that its score is inversely proportional to the magnitude of the blob. However since, in general, it is preferable to assign a comparable score to all the nodules in a certain range (e.g. from 3 mm to 10 mm in diameter), a multiscale approach has to be followed.

5.3 Multi Scale Dot Enhancer (MSDE) filter

The MSDE algorithm consists in combining, according to the prescription given in [20], the z_{dot} functions evaluated at several scales. This procedure is based on an *a priori* knowledge of the sizes of the objects to be enhanced.

For what concerns nodules, assuming that a nodule can be approximated by a 3D Gaussian with scale parameter σ, the nodule diameter can be denoted with 4σ, thus taking into account for more than 95% of the nodule volume. If the nodule diameters to be enhanced are in the range $[d_{min}, d_{max}]$, the scales to be considered for the Gaussian filter are in the range $[\sigma_{min}, \sigma_{max}]$, where $\sigma = d/4$. Within that range, the N intermediate scales are computed as $\sigma_i = r^{i-1}\sigma_{min}$ where $i = 1, .., N$ and $r = (\frac{d_{max}}{d_{min}})^{1/(N-1)}$.

The resulting filter value is then:

$$z_{norm}(\sigma_i) = \sigma_i^2 z_{dot}(\sigma_i) \tag{5.3.1}$$

$$z_{max} = max(z_{norm}(\sigma_i)) \tag{5.3.2}$$

with $i = 1, .., N$. The final output of the MSDE is a matrix, referred in the following section as $Z(x, y, z)$, where each voxel contains the value of the obtained z_{max}. The range of diameters for the objects to be enhanced, and

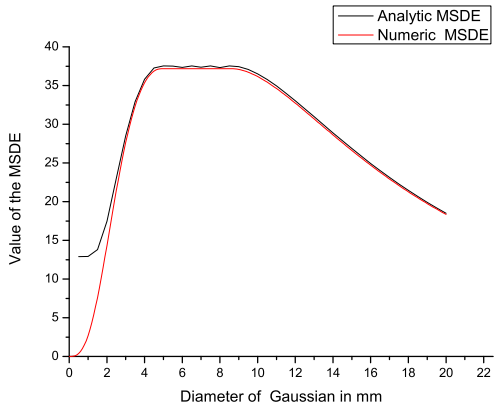


Figure 5.3.1: Analytic and numeric response curve of the MSDE filter. The filter settings are $\mathrm{d}_0 = 4$, $\mathrm{d}_1 = 7$ with $\mathrm{N} = 5$ (number of multiscale steps). The filter is run on Gaussian blobs with diameter in the range $[0, 10\,\mathrm{mm}]$. From the picture is visible that the filter response is almost flat in the interval $[5, 9\,\mathrm{mm}]$.

the number of intermediate smoothing scales have to be determined on the basis of the target dataset of nodules.

Moreover, using the MSDE, it is possible to have an estimation of the candidate size, using the scale σ_i corresponding to the best response of the filter. This information is collected in an image, storing, for each voxel, the value σ_i of the best scale.

A peak detection algorithm is then applied to $\mathrm{Z}(\mathrm{x}, \mathrm{y}, \mathrm{z})$ result to detect local maxima in the filter output matrix. The final output is a list of locations of nodule candidates ordered by decreasing value of the corresponding $\mathrm{z}_{\max}$. Of course, in addition to the true nodules, this list is expected to contain also a quite large number of false positive findings, that needs to be reduced in further steps.

5.4 Analytic evaluation of the N-dimensional MSDE

It is possible to evaluate theoretically the response of the MSDE on n-dimensional Gaussian objects of scale A

$$G_A = H_0 \exp(-\frac{|\vec{x}|^2}{2A^2}),$$

where $\tilde{x} = (x_1, .., x_n)$ and H_0 is a positive constant.

To apply the MSDE filter, the object G_A has first to be convolved with a Gaussian of standard deviation σ representing the scale of interest

$$G_\sigma = \left(\frac{1}{\sqrt{2\pi}\sigma}\right)^n \exp(-\frac{|\vec{x}|^2}{2\sigma^2}).$$

This operation becomes a simple product if evaluated using the Fourier transform. Once returned to ordinary space the result of the convolution product is

$$G_A * G_\sigma = H_0 \left(\frac{\sqrt{2\pi}A}{\sqrt{2\pi(\sigma^2 + A^2)}}\right)^n \exp(-\frac{|\vec{x}|^2}{2(\sigma^2 + A^2)}), \tag{5.4.1}$$

that is a Gaussian of scale $\sqrt{\sigma^2 + A^2}$.

In an N-dimensional context the z_{dot} define in eq. 5.2.5 becomes

$$z_{dot}(\lambda_1, .., \lambda_n) = \begin{cases} \frac{|\lambda_n|^2}{|\lambda_1|} & \text{if } \lambda_1, .., \lambda_n < 0 \\ 0 & \text{otherwise} \end{cases} \tag{5.4.2}$$

with $|\lambda_1| \geq .. \geq |\lambda_n|$.

Using the N-dimensional version of the eq. (5.2.4) and eq. (5.4.1), is obtained

$$z_{dot} = H_0 \frac{A^n}{(\sigma^2 + A^2)^{\frac{n}{2}+1}}. \tag{5.4.3}$$

Studying the eq. (5.4.3) it is possible to show that it has a maximum for $A = \sigma\sqrt{n/2}$. This means that the SSDE filter has the maximum value for $A = \sigma$ only for $n = 2$ and that for $n \neq 2$ the scales of the filter have to be calibrated. For example, for $n = 3$ the upper and lower bound of the multiscale filter should be multiplied by $\sqrt{3/2}$. As shown in fig. 5.3.1, this behavior is found also in the numeric implementation of the MSDE algorithm.

Investigating the behavior of the MSDE, it is possible to show that

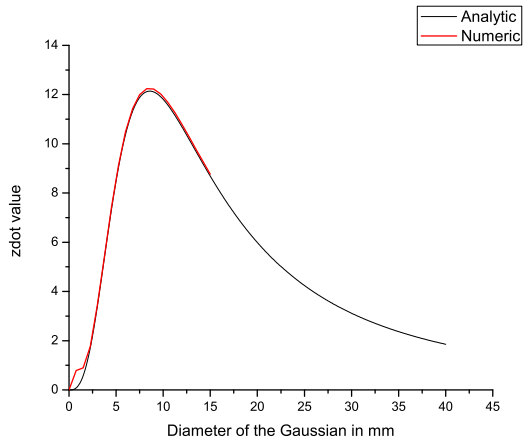


Figure 5.4.1: Result of the SSDE filter run with a fixed scale $\mathrm{d} = 7$ mm on Gaussian blobs with diameters in the range $[0, 40]$ mm. The $\mathrm{z_{dot}}$ function from eq. (5.4.3) and the $\mathrm{z_{dot}}$ obtained from tests on artificial objects are shown.

$$z_{max} = max(z_{norm}) = max(\sigma^2 z_{dot}) = max(\sigma^2 H_0 \frac{A^n}{(\sigma^2 + A^2)^{\frac{n}{2}+1}}). \qquad (5.4.4)$$

and that the eq. 5.4.4 has a maximum for $\sigma_{\mathrm{best}} = \mathrm{A}\sqrt{2/\mathrm{n}}$.

Substituting σ_{best} in eq.(5.4.4) yields to

$$z_{max} = z_{norm}(\sigma_{best}) = H_0 \frac{\frac{2}{n}}{(1 + \frac{2}{n})^{\frac{n}{2}+1}},$$

which doesn't depend on A.

In particular if the Gaussian blob diameter is in the range of enhancing $[\sigma_{\mathrm{min}}, \sigma_{\mathrm{max}}]$, the score is $\mathrm{z_{max}} = \mathrm{z_{norm}}(\sigma_{\mathrm{best}})$; otherwise its score is $\mathrm{z_{norm}}(\sigma_{\mathrm{max}})$ if $\sigma_{\mathrm{max}} < \sigma_{\mathrm{best}}$, or $\mathrm{z_{norm}}(\sigma_{\mathrm{min}})$ if $\sigma_{\mathrm{min}} > \sigma_{\mathrm{best}}$.

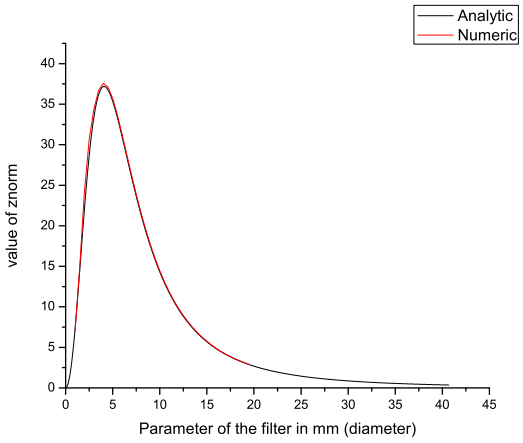


Figure 5.4.2: Result of SSDE filter applied to a blob of diameter 5 mm. The filter is run with scale parameter varying in the range $[0, 40]$ mm. The z_{norm} function from eq. (5.3.1) and the z_{norm} obtained from tests on artificial objects are shown.

5.5 Efficient numerical implementation of the MSDE

The algorithm consists of two main building blocks: the Hessian evaluation and the eigenvalues computation. The Hessian computation is carried out using a recursive algorithm [53]. The recursive algorithm is able to perform in a single step both the Gaussian convolution and the second derivatives. For the eigenvalues computation, the Lienard Chipart criterion is implemented to enhance the computational efficiency.

5.5.1 Evaluation of the Hessian

Evaluating the Hessian of an image is one of the most common task in computer vision. The computation of the Hessian can be very intensive since it requires first to smooth the image with a low pass filter, i.e. a Gaussian, to attenuate the noise that may be enhanced by the derivatives and then to apply the derivative operator.

The direct implementation of a convolution is straightforward but inefficient.

In particular, if a signal of dimension N has to be convolved with a kernel of dimension K, the number of operations needed to compute the convolution, are $\mathcal{O}(\mathrm{K} \cdot \mathrm{N})$.

For a tridimensional matrix $\mathrm{N} \times \mathrm{N} \times \mathrm{N}$ this leads to an algorithm that scales as $\mathcal{O}(3 \cdot \mathrm{K} \cdot \mathrm{N}^3)$ for a separable kernel of dimension $\mathrm{K} \times \mathrm{K} \times \mathrm{K}$, i.e. a kernel that can be evaluated separately in the three directions, or $\mathcal{O}(\mathrm{K}^3 \cdot \mathrm{N}^3)$ for a non separable kernel.

For Gaussian kernels, K is typically proportional to the Gaussian's variance, thus for large variance the value of K can make the convolution algorithm very inefficient. To compute the Hessian, after the convolution operation is computed, the derivatives have to be applied and this adds another algorithm with complexity $\mathcal{O}(3 \cdot \mathrm{D} \cdot \mathrm{N}^3)$, where D is the dimension of the derivative kernel.

To improve the computational efficiency of the algorithm, it is possible to implement a recursive solution that scales as $\mathcal{O}(\mathrm{Z} \cdot \mathrm{N}^3)$, where Z is a constant that doesn't depend on the dimension of the Gaussian [53].

Moreover, since the recursive filter is built to commute with the smoothing operation, it is possible to convolve directly the second derivative of the Gaussian with image. The recursive algorithm provides a fast and accurate tool to evaluate Gaussian convolutions and their derivatives, that is faster than the naive approach for big variance.

5.5.2 Eigenvalues computation

There are many standard algorithms to find the eigenvalues of matrices as the Jacobi algorithm, the QR and QL algorithms [55], however most of the algorithms are suitable for matrices of arbitrary size. Since the MSDE algorithm is typically used for images of low dimensionality, i.e. tridimensional or bidimensional, good results, in terms of computational cost, can be obtained using a dedicate algorithm to solve low dimensionality problems.

The eigenvalues of a $\mathrm{n} \times \mathrm{n}$ matrix H can be computed directly evaluating the roots of the matrix characteristic polynomial, which are found solving with respect to λ

$$det(H - \lambda \cdot I) = 0 \tag{5.5.1}$$

where I is the identity matrix, det is the determinant operator.

Therefore the direct implementation formula, is suitable up to n = 4, since, in general, no solution is available for polynomial of the degree > 4.
For n = 3, the characteristic polynomial can be written as

$$P(H) = -\lambda^3 + trace(H)\lambda^2 - Z(H)\lambda + det(H)$$

where

$$Z(H) = \sum_i \sum_j \frac{1}{2}(H_{ij}H_{ji} - H_{ii}H_{jj})$$

and trace(H) = $\sum_{i=1}^{n} H_{ii}$.
Finding the roots of a polynomial up to the fourth order can be implemented using a direct formula [56]. However, as pointed out in [55], the direct formula is known to be prone to numerical instabilities. The precision of eigenvalues computation is not a key factor for this approach and an accuracy of 1% is enough to obtain satisfactory results. The accuracy of the direct algorithm has been shown [55] to be at least 10^{-3}, for this reason this approach, leading to a fast computation, is implemented.
However since, for MSDE the eigenvalues need to be computed only if the $\lambda_1, \lambda_2, \lambda_3 < 0$, it is possible to reduce the computational cost of the algorithm by computing only those eigenvalues that have negative real parts.
Finding the signs of the eigenvalues, without actually evaluating them is possible using the Lienard Chipart criterion [57].

Theorem 1. *Lienard Chipart Criterion*
Given a polynomial P with real coefficients

$$\lambda^n + a_1\lambda^{n-1} + ... + a_n$$

then a necessary and sufficient condition for all the roots of P to have negative real parts is that

$$a_n > 0\,, a_1 > 0\,, a_3 > 0,\ a_5 > 0$$

and the Hurwitz determinants Δ_i

$$\Delta_{n-1} > 0,\ \Delta_{n-3} > 0\,, \Delta_{n-5} > 0$$

where

$$\Delta_i = \begin{vmatrix} a_1 & a_3 & a_5 & . & . & a_{2i-1} \\ 1 & a_2 & a_4 & . & . & a_{2i-2} \\ 0 & a_1 & a_3 & . & . & a_{2i-3} \\ 0 & 1 & a_2 & . & . & a_{2i-4} \\ . & . & . & . & . & . \\ . & . & . & . & . & . \\ 0 & 0 & . & . & . & a_i \end{vmatrix}.$$

Using the theorem 1 for a third order equation results in three relations:

$$\begin{gathered} \mathrm{trace(H)} < 0 \\ \det(\mathrm{H}) < 0 \\ \det(\mathrm{H}) - \mathrm{trace(H)} \cdot \mathrm{Z(H)} > 0 \end{gathered} \tag{5.5.2}$$

if one of this check fails the eigenvalues are not computed and the z_{dot} function is set to 0. If all the three checks in eq. 5.5.2 are satisfied, the eigenvalues are evaluated using the direct formula.

The eigenvalues of a symmetric matrix are real, so since the direct formula may lead to complex solutions, the imaginary parts of the eigenvalues are considered as numerical errors and are neglected.

5.6 Results

The analytic MSDE and SSDE, shown in sec. 5.4 and sec. 5.3, can be directly compared with the numerical results of the implemented filter. As shown in fig. 5.4.1 and fig. 5.4.2, the results of the numerical implemented filter are in good agreement with the analytic formulation. Moreover, the test performed on real CT datasets, composed by $LIDC_1$ and $LIDC_2$, shows that the direct implementation of the eigenvalues together with the Lienard Chipart criterion improve the algorithm performance, on average, of 2.5 times. The average time of eigenvalues computation passes from 69 seconds per CT, using the iterative version, to 27.8 seconds on an i7 950 3.07 GHz processor with 12 GB of RAM.

Chapter 6

Detection of juxta-pleural nodules

6.1 Introduction

Juxta-pleural nodules can be regarded as concave hemi-spherical objects connected to the pleura surface (see fig. 2.5.2). To detect such a kind of objects, a method based on the Pleura Surface Normal (PSN) overlap is implemented. This approach consists in searching the points where many surface normals intersect (see fig. 6.4.1). This approach can be implemented in many different ways, since many choices are available in the literature to represent the surface and to evaluate the normals and their intersections. In the next sections the implementation choices are described.

6.2 Surface representation

The pleura surface may be defined as the surface associated with the mask obtained in the segmentation step (see chap. 4). However, representing a surface using cubes (i.e. voxels) is an ill-posed problem. To overcome this limit, it is common to use an algorithm to transform a binary mask, e.g. to transform a segmented volume, in a polygon mesh [58]. A polygon mesh is a collection of vertices, edges and faces that defines the shape of a polyhedral object in 3D computer graphics and solid modeling. The faces usually consist of triangles, quadrilaterals or other simple polygons. One of the most used algorithm to obtain a mesh from a binary mask is the "marching cube algorithm" [59].

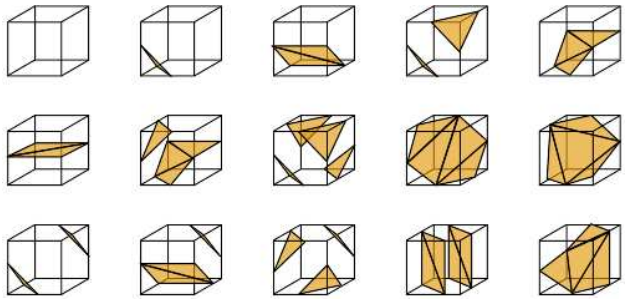


Figure 6.2.1: The 15 unique combinations of the marching cube algorithm.

The algorithm proceeds through the scalar field, taking eight neighbor locations at a time (thus forming an imaginary cube), then determining the polygon(s) needed to represent the part of the iso-surface that passes through this cube. The individual polygons are then fused into the desired surface.

This is done by creating an index to a precalculated array of 256 possible polygon configurations ($2^8 = 256$) within the cube. By means of reflections and symmetrical rotations it is possible to reduce the configurations to 15 unique cases (see fig. 6.2.1). The application of the marching cube leads to a mesh as the one shown in fig. 4.7.1.

6.3 Normals computation

The implemented idea was inspired by [22,27]. The main difference between these approaches and the one proposed is the way the normals are computed. To evaluate the normals a procedure based on a representation of the pleura surface as a polygonal mesh is implemented, rather than a voxel based one as described in [22,27].

Normals are evaluated using the triangular mesh representing the pleura surface. In particular, the normal to each triangle is calculated by using the vector product between the triangle edges; then the normals are evaluated for each vertex by averaging all the normals to the neighbor triangles. This procedure is used in computer graphics to evaluate light reflection on surfaces and it is called "Gouraud shading" [60].

6.4 Normals intersection

Since the evaluation of the intersections in the real 3D space is a complex and computationally intensive operation, it is implemented in the voxel space. This means that each voxel accumulates a score proportional to the number of normals passing through it. This information is collected in the score matrix S(x, y, z) (see fig. 6.4.1). To implement this operation a line rasterization algorithm is needed [61], and the "Bresenham's line algorithm" [62] is implemented.

The "Bresenham line" is an algorithm which determines which points in an n-dimensional raster should be plotted in order to form a close approximation to a straight line between two given points. It is commonly used to draw lines on a computer screen, as it uses only integer addition, subtraction and bit shifting, all of which are very cheap operations in standard computer architectures.

Since the intersections in the voxels space are prone to numerical errors and to errors in the computation of the normals, it is preferable to consider Gaussians cylinders instead of line segments. This operation is equivalent to a Gaussian smoothing of scale σ_{cylinder} of the matrix S(x, y, z) [27]

$$A(x,y,z) = \int\int\int S(x',y',z') \frac{e^{-\frac{(x-x')^2+(y-y')^2+(z-z')^2}{2\sigma^2_{cylinder}}}}{(\sigma_{cylinder}2\pi)^3} dx'dy'dz' \qquad (6.4.1)$$

At the end of this procedure, it is possible to collect the list of locations of nodule candidates looking for the local maxima in the matrix A(x, y, z). Of course, in addition to the true nodules, this list is expected to contain also a quite large number of false positive findings.

6.5 Tests on artificial objects

A theoretical validation of the response of the PSN filter is difficult, however it is possible to qualitatively assess the correct work of the filter using artificial shapes of different size as those shown in fig. 6.5.1.

For an hemispherical object of radius r, the expected PSN score obtained using normals of length l = r is expected to be

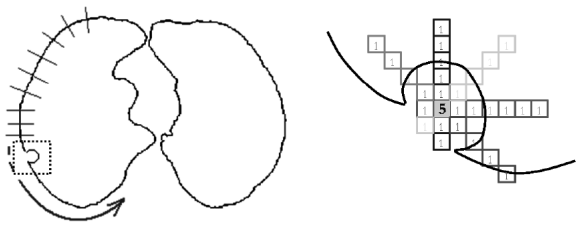


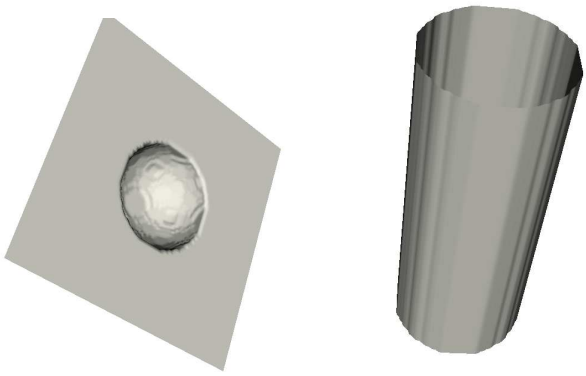
Figure 6.4.1: Pictures representing the matrix $\mathrm{S(x, y, z)}$: each voxel accumulates a score proportional to the number of normals passing through it.

$$PSN_{hemisphere} \propto r^2$$

whereas for a cylinder objects the expected score is

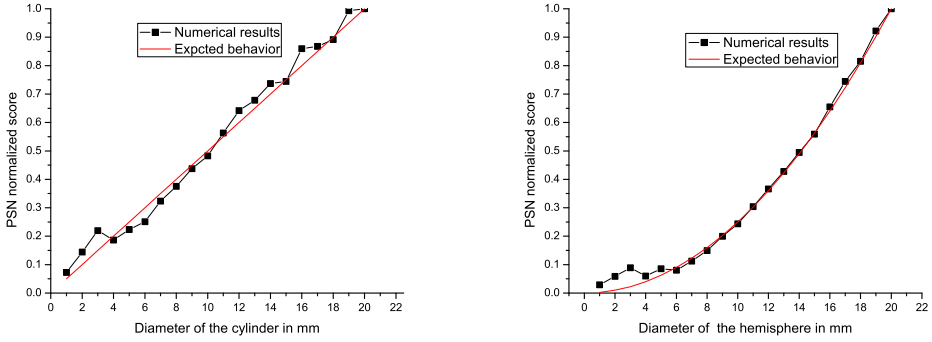
$$PSN_{cylinder} \propto r.$$

The results of the tests are shown in fig. 6.5.2, the PSN score is normalized to the max output obtained in the range, in order to factor out the constant. The results of these tests are obtained with a $\sigma_{\mathrm{cylinder}} = 3$ mm.



(a) Hemisphere shape. (b) Cylinder shape.

Figure 6.5.1: Artificial objects used for testing the PSN filter.



(a) Cylinder (b) Hemisphere

Figure 6.5.2: PSN filter results on artificial objects of radius in the range$[0, 20]$ mm, the parameter of the filter l in this test is equal to the radius of the object.

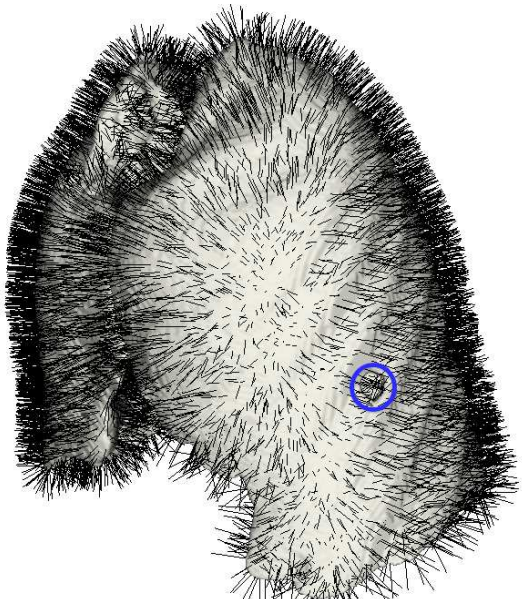


Figure 6.5.3: Picture showing the normals computed on the pleura surface. The blue circle underlines a juxta-pleural nodule.

Chapter 7

Reduction of false positive findings

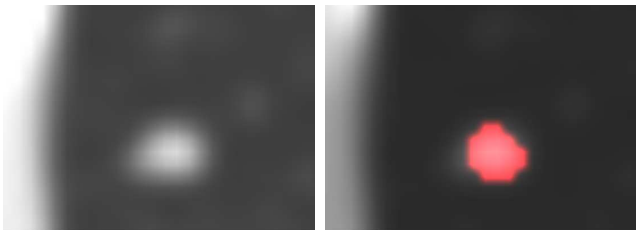
7.1 Introduction

As discussed in chap. 5 and in chap. 6, candidate nodule detectors provide a lot of false positives that must be reduced, in order to obtain a reasonable number of findings. Since radiologists' experience is the leading factor to decide between nodules and non nodules, it is natural to think of a classifier [63] as an approach to reduce the amount of false positive generated by the algorithms.

The challenging task to build a successful classifier is how to encode the information contained in the images in few relevant numbers, called features. The approach followed to reduce the amount of false positive findings is an upgrade of the Voxel Based Neural Approach (VBNA), presented in [16, 17, 19, 28, 64, 65]. In the following sections the features to be extracted together with the candidate segmentation algorithm are shown.

7.2 Candidate segmentation

As discussed in chap. 5 and chap. 6, internal and juxta-pleural nodule candidate locations are obtained computing local maxima in the $Z(x, y, z)$ and $A(x, y, z)$ matrices. In order to obtain a reasonable number of findings, the



(a) (b)

Figure 7.2.1: Example of a candidate nodule segmentation: a) a nodule, b) the corresponding segmentation.

lists of local maxima are sorted according to their values and only those candidates above the two values $T_{internal}$ and $T_{juxta-pleural}$ are kept, thus obtaining a first selection of candidates with low specificity and an high sensitivity.

However, it is possible to further use the information contained in $Z(x, y, z)$ and $A(x, y, z)$ to obtain also a rough segmentation of the nodule candidate. The segmentation of candidates is implemented with the same algorithm both for internal and juxta-pleural candidates.

First, a thresholding operation, with a fixed and empirical threshold $T_{candidate}$, followed by a connect component analysis is applied to the $Z(x, y, z)$ and $A(x, y, z)$ images. Then, a dimensional analysis is applied to the components found in the previous step, and those with the major axis outside an interval $[d_{min}, d_{max}]$ are discarded.

This operation provides the segmentations of a large number of candidates, however the features are extracted only for those candidates that are contained in the lists of local maxima. Those candidates with an empty segmentation are discarded, while for the others the features described in the following sections are evaluated.

7.3 VBNA features

The VBNA method consists in assigning to each voxel of a candidate a vector of features. The features are grey level intensity values, morphological features of the voxel neighborhood, candidate detector features and ROI level features.

The idea behind the method is to give the classifier raw data in order to favor the learning of hidden correlations.

However, this method implies that the classifier must be trained with a lot of patterns to infer the relation. In the following sections the used features are described in details.

7.3.1 Gray level intensity features

Each voxel of a ROI is characterized by the gray level intensity values of its neighborhood. The CT values of the voxel 3D neighborhood are rolled down into a feature vector to be analyzed by a Support Vector Machine (SVM) classifier (see fig. 7.3.1). The radius of extraction of the features is 2 voxels; this means that the collected intensity values for each voxel are 125 ($5\times5\times5$).

Other experiments employing different radiuses of extraction were carried out using the values of 1 and 3 voxels. The former resulted in worst performance during the training phase while the latter resulted in 343 gray level features making the training unfeasible.

7.3.2 Morphological features

The eigenvalues of the gradient and the Hessian matrices are computed for each voxel. However, as discussed in [54], before computing the derivatives a smoothing procedure has to be applied.

For the internal candidates, the variance of the smoothing is provided by the scale which led to the best z_{max} value; whereas for juxta-pleural candidates the variance is fixed and it is equal to σ_{cylinder}.

The gradient matrix is computed as

$$G_{i,j} = \sum \partial_{x_i} I(x_1, x_2, x_3) \partial_{x_j} I(x_1, x_2, x_3) \quad for \quad i, j = 1, .., 3 \qquad (7.3.1)$$

where the sums are over the neighborhood area, and $I(x_1, x_2, x_3)$ are the values of the image (see fig. 7.3.1).

The Hessian matrix is then computed using the expression

$$H_{i,j} = \partial_{x_i} \partial_{x_j} I(x_1, x_2, x_3) \quad for \quad i, j = 1, .., 3.$$

The eigenvalues of $G_{i,j}$ and $H_{i,j}$ are then computed for each voxel of the candidate and added to the list of features.

These six features, exploiting the morphology of the voxel neighborhood, are added to the textural features given by the intensity values, in order to improve the method discriminating power.

7.3.3 Candidate detector features

To improve the classic VBNA method, each voxel is assigned the local value extracted by the matrices generated by the candidate detectors $Z(x, y, z)$ and $A(x, y, z)$.

7.3.4 ROI based features

The main improvement of this analysis is the introduction of two ROI level features: the size in mm^3 of the candidate and the integrated ROI candidate detector score, i.e. the sum over the ROI of the local candidate value.

Therefore, the total number of features for each voxel is 134: 125 gray level intensity features, 6 features exploiting the morphology of the voxel neighborhood, 2 features extracted from candidate detectors and the volume of the candidate in mm^3.

7.3.5 Classifier selection

There is no theoretical principle to build an optimal classifier for solving a specific problem. The only general purpose approach available at present is to choose the best among different models. More specifically, it is possible to view a classifier selection problem as choosing, within a set of candidate model structures, the "best" one according to a certain criterion.

In this context, a standard tool in statistics known as K-fold cross-validation [63, 66] may be used.

The K-fold cross validation consists in dividing the dataset in K different subsets, training the classifier on $K - 1$ subsets and validating it on the remaining one. This operation can be repeated in K different ways, using each time different $K - 1$ subsets for training and one for performance assessment. At the end of the procedure each subset is unbiasedly evaluated by a classifier, i.e. by a classifier that was not trained with it.

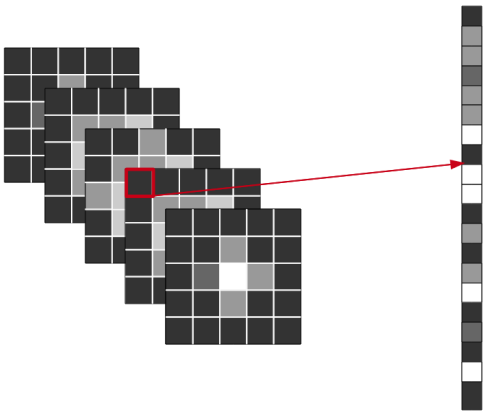


Figure 7.3.1: Basic idea of the VBNA false positive reduction: each voxel is characterized by a feature vector constituted by the intensity values of its 3D neighbors ($5 \times 5 \times 5$).

The K classifiers provide an estimate of the performance of the classifier trained on the whole datasets and validated on a dataset of homogeneous characteristics. This method is very useful when the availability of labeled data is reduced, since it allows to use the same set of data both for training and assessing the performance. A special case of the K-fold cross validation is the Leave One Out (LOO) cross validation, where K is equal to the number of pattern N. Therefore, for each fold, there is only one pattern that is not used for the training.

In developing CAD systems it often happens to train classifiers with patterns that belong to the same exam. These patterns may be highly correlated, therefore a K-fold cross validation at pattern level may be not the optimal approach. However, a LOO at patient/exam level may be implemented [67], i.e. patterns are divided in folds according to the patient/exam they belong to. This special LOO validation is often referred as Leave One Patient Out (LOPO) cross validation.

An SVM classifier (see appendix C) is trained to assign each voxel either to the nodule or to the normal tissue target class. At the end of this procedure, each ROI is assigned a score, averaging all the values attributed by the SVM to the voxels. With this score, it is possible to evaluate the FROC curve as discussed in sec. 2.3.

7.4 Discussion

The segmentation of nodule candidates is a necessary step in order to include in the analysis only those voxels belonging to the structures of interest and in order to exclude from the evaluation voxels belonging to anatomical structures surrounding the candidates, e.g. lung parenchyma, pleura surface, etc etc..

The main advantage of the VBNA method is that, in principle, it doesn't require a very accurate segmentation, since no shape based features are extracted from each ROI. This means that, if the difference between the "optimal" segmentation and the obtained one is sufficiently small, the overall score of the ROI is almost unaltered.

On the contrary, the approaches based on shape features extraction may be very sensitive to the output of the segmentation, since small differences in the segmentation may completely change the values of the features, thus leading to very different results.

Part III

Results

Chapter 8

CAD training and validation

8.1 Introduction

As different training sets may lead to significantly different results, to assess the stability of the developed CAD it is necessary to carry out several training and validation sessions. The datasets used in this chapter are those described in chap. 3. The datasets $LIDC_1$ and $LIDC_2$ are used as training sets, for both the CAD for internal nodules (CAD_I) and the CAD for juxta-pleural nodules (CAD_{JP}). Only nodules with at least AL2, i.e. only those nodules annotated by at least two radiologists, are considered relevant for the training.

In the following sections the results of the experiments made with the CAD_I and the CAD_{JP} are shown. At the end of the training session, the CAD_I and the CAD_{JP} are combined using the approach explained in sec. 8.4 and discussed in [17]. The best CAD, among all those trained, is then applied to the ANODE09 50 CT dataset. The result of the ANODE09 validation must be intended as a real life unbiased measurement of the CAD performance.

8.2 Training on $LIDC_1$

The $LIDC_1$ data set consists of 69 CTs containing 96 relevant internal nodules and 42 relevant juxta-pleural nodules (see sec. 3.3). Here follows a description of the training sessions.

8.2.1 CAD_I training

The CAD_I is applied to the 69 $LIDC_1$ CTs, using the segmentation settings described in chap. 4 and the MSDE with parameters $d_0 = 4$ mm, $d_1 = 7$ mm and $N = 5$. An example of the MSDE filter application to a real CT is shown in fig. 8.2.3. After the computation of the MSDE, local maxima, with a basis of 5 mm are searched in the $Z(x, y, z)$ matrix.

Analyzing the distribution of FPs and TPs, it is found that discarding findings below the $T_{internal} = 15$ score a sensitivity of $\frac{91}{96} \sim 95\%$ is achieved at 155 FP/CT on average.

The candidate feature vectors are then classified with a linear SVM (see appendix C), using the features and the methods described in chap. 7, an empirical threshold $T_{candidate} = 2$ and discarding those candidates with major axis outside the $[d_{min} = 2, d_{max} = 20]$ mm range.

The candidate feature vectors are sampled with a 2 to 1 FP to TP ratio for a total of 22413 patterns.

To search for the best model, a LOPO cross validation (see appendix B.6) is carried out, varying the penalty parameter C of the linear SVM in the range $[2^{-20}, 2^{20}]$. The best model is then selected according to the AUC.

The results of the SVM training are shown in fig. 8.2.1. The results of the LOPO training show that the performance of the classifiers is stable in a wide range of values, with a peak value of AUC=0.98 for $C = 0.0625$.

It is possible to evaluate the performance of the CAD_I on the $LIDC_1$ in LOPO mode, by using the 69 models originating from the best C training, in oder to assess the quality of the model trained. Results of the LOPO are shown in fig. 8.2.2 and in tab. 8.1, the score is evaluated only on internal nodules.

Table 8.1: FROCSV of the CAD_I trained on $LIDC_1$ and validated on the $LIDC_1$ in LOPO mode at the four LIDC ALs.

Agreement level	FROCSV
AL1	0.472
AL2	0.605
AL3	0.676
AL4	0.743

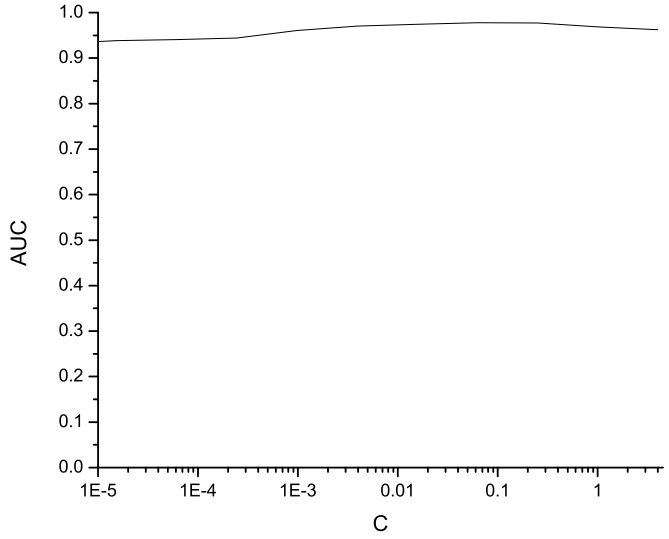


Figure 8.2.1: Results in terms of the AUC of the SVM training on the $\mathrm{LIDC_1}$ with the features extracted by the $\mathrm{CAD_I}$ as a function of the penalty parameter C.

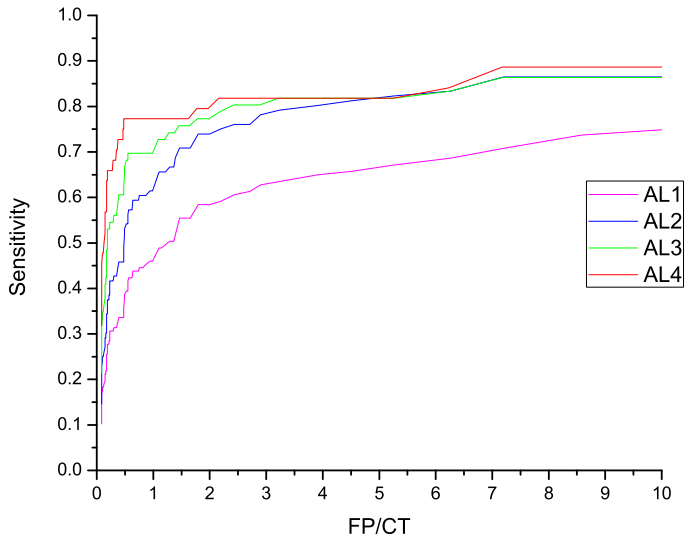
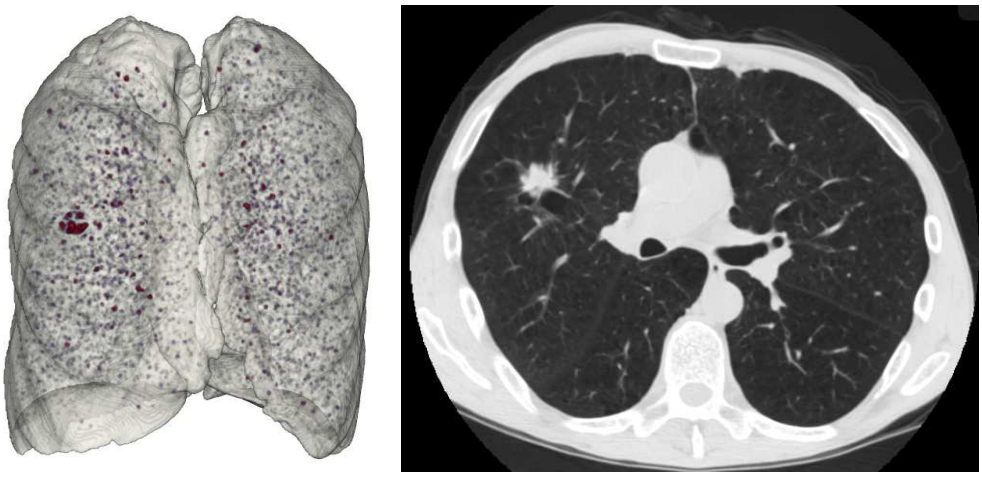


Figure 8.2.2: FROC curves obtained by the $\mathrm{CAD_I}$ at the four ALs in LOPO mode on the $\mathrm{LIDC_1}$.



(a) (b)

Figure 8.2.3: a) Rendering of the MSDE filter results applied to a CT scan. The voxels with higher score are colored in red. b) One slice of the CT scan shown in a).

8.2.2 CAD_I validation

The first validation of the CAD_I is carried out using the $LIDC_2$ dataset, the $ITALUNG-CT$ dataset and the five ANODE09 examples. The best model found and the settings described in sec. 8.2.1 are used. As shown in tab. 8.2, the system performs better on the $LIDC_1$ than on the $LIDC_2$, whereas it obtains a very good result on the $ITALUNG-CT$ dataset. However, the performance on the five ANODE09 example CTs is instead comparable to those obtained on the $LIDC_1$ and the $LIDC_2$ with AL_1.

8.2.3 CAD_{JP} training

CAD_{JP} is applied to the 69 CTs of the $LIDC_1$, using a 5 mm long normal and $\sigma_{cylinder} = 0.75$ mm. After the computation of the PSN, local maxima with a basis of 5 mm are searched in the $A(x, y, z)$ matrix. An example of the PSN filter application to a real CT is shown in fig. 8.2.6.

Analyzing the distribution of FPs and TPs, it is possible to show that when discarding findings below the score $T_{juxta-pleural} = 6$, a sensitivity of $\frac{40}{42} \sim 95\%$ is obtained at 80 FP/CT on average.

The candidate feature vectors are then classified with a linear SVM, using the features and methods described in chap. 7, an empirical threshold $T_{candidate} = 2$ and discarding those candidates with major axis outside the $[d_{min} = 0, d_{max} = 40]$ mm range. The features of the candidates are sampled with a 3 to 1 FP to TP ratio for a total of 39168 of training patterns.

To search for the best model a LOPO cross validation is carried out, varying the penalty parameter C of the linear SVM in the range $[2^{-20}, 2^{20}]$. The best

Table 8.2: FROCSV obtained by the CAD_I trained on the $LIDC_1$ and validated on the ANODE09 five example CTs, the $LIDC_2$ and the $ITALUNG-CT$ datasets.

Dataset	FROCSV
ANODE09 5 examples	0.462
$LIDC_1$(LOPO)	0.472/0.605/0.676/0.743
$LIDC_2$	0.471/0.548/0.592/0.615
$ITALUNG-CT$	0.685

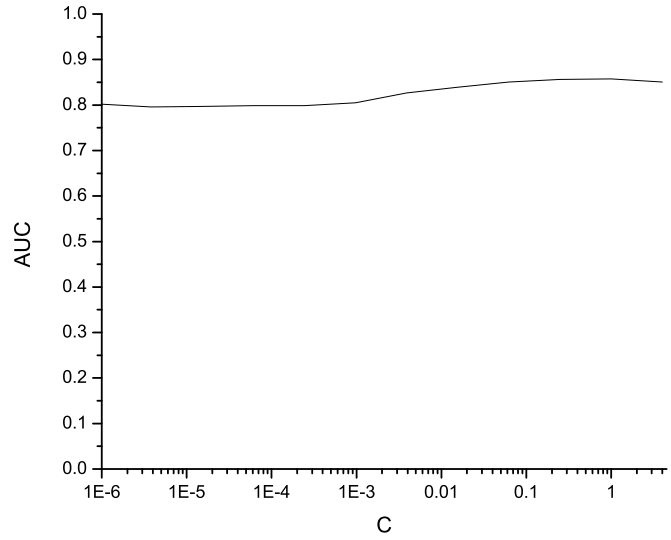


Figure 8.2.4: Results in terms of AUC of the SVM training on the LIDC_1 with the features extracted by the $\mathrm{CAD}_{\mathrm{JP}}$ as a function of the penalty parameter C.

model is then selected according to the AUC.

The results of the SVM training are shown in fig. 8.2.4. The results of the LOPO training show that the performances of the classifier are stable in a wide range of values, with a peak value of AUC = 0.86 for C = 1.0. The evaluation of the $\mathrm{CAD}_{\mathrm{JP}}$ performance in LOPO mode on the LIDC_1 leads to the results shown in fig. 8.2.5 and in tab. 8.3, the score is evaluated only on juxta-pleural nodules.

8.2.4 $\mathrm{CAD}_{\mathrm{JP}}$ validation

The validation of the $\mathrm{CAD}_{\mathrm{JP}}$ is carried out using the LIDC_2 dataset, the $\mathrm{ITALUNG} - \mathrm{CT}$ dataset and the five ANODE09 examples. The best model found and the settings described in sec. 8.2.3 are used. The results in terms of FROCSV are shown in tab. 8.4.

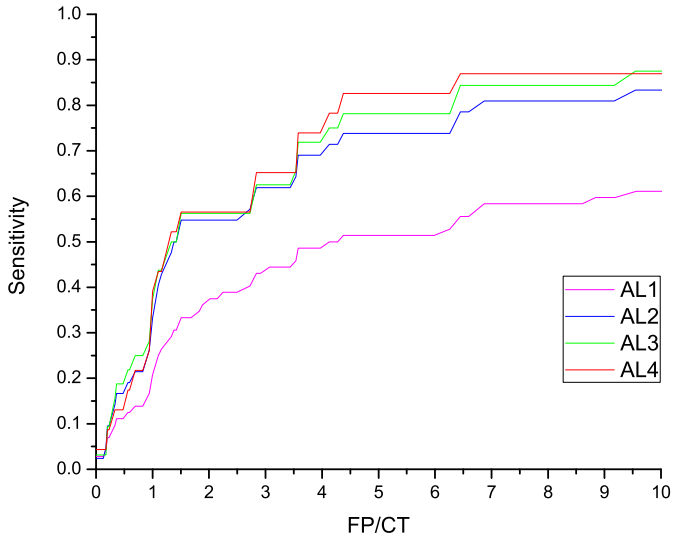
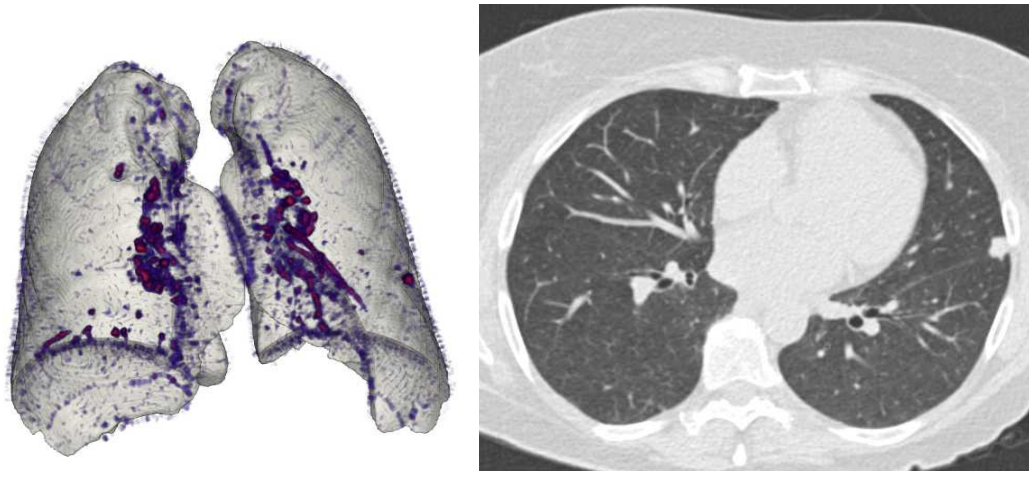


Figure 8.2.5: FROC curves obtained by the $\mathrm{CAD}_{\mathrm{JP}}$ at the four ALs in LOPO mode on the LIDC_1 (as described in sec. 3.3).

Table 8.3: FROCSV obtained by the $\mathrm{CAD}_{\mathrm{JP}}$ trained on the LIDC_1 and validated on the LIDC_1 in LOPO mode, at the four LIDC ALs.

Agreement level	FROCSV
AL1	0.276
AL2	0.384
AL3	0.405
AL4	0.407



(a) (b)

Figure 8.2.6: a) Rendering of the results of the PSN filter applied to a CT scan. The voxels with higher score are colored in red. b) One slice of the CT scan shown in a).

Table 8.4: FROCSV obtained for the $\mathrm{CAD_{JP}}$ trained on the $\mathrm{LIDC_1}$ and validated on ANODE09 five examples, the $\mathrm{LIDC_2}$ and the$\mathrm{ITALUNG-CT}$ datasets.

Dataset	FROCSV
ANODE09 5 examples	0.381
$\mathrm{LIDC_1(LOPO)}$	0.276/0.384/0.405/0.405
$\mathrm{LIDC_2}$	0.251/0.330/0.314/0.286
$\mathrm{ITALUNG-CT}$	0.238

8.3 Training on $LIDC_2$

The $LIDC_2$ consists of 69 CTs, containing 95 relevant internal nodules and 19 relevant juxta-pleural nodules (see sec. 3.3). Here follows a description of the training sessions.

8.3.1 CAD_I training

The CAD_I is applied to the 69 CTs of the $LIDC_2$, using the segmentation settings described in chap. 4 and MSDE with parameters $d_0 = 4$ mm, $d_1 = 7$ mm and $N = 5$. After the computation of the MSDE, local maxima,with a basis of 5mm are searched in the $Z(x, y, z)$ matrix.

Analyzing the distribution of FPs and TPs, it is found that discarding findings below the score $T_{internal} = 15$, a sensitivity of $\frac{84}{95} \sim 88\%$ is achieved at 148 FP/CT on average.

The candidate feature vectors are then classified with a linear SVM (see appendix C), using the methods described in chap. 7, an empirical threshold $T_{candidate} = 2$ and discarding those candidates with major axis outside $[d_{min} = 2, d_{max} = 20]$ mm range.

The features of the candidates are sampled with a 2 to 1 FP to TP ratio, for a total of 31509 train patterns. To search for the best model a LOPO cross validation is carried out, varying the penalty parameter C of the linear SVM in the range $[2^{-20}, 2^{20}]$. The best model is then selected according to the AUC.

The results of the SVM training are shown in fig. 8.3.1. The LOPO results show that the performances of the classifier are stable in a wide range of values, with a peak value of AUC = 0.97 for C = 1.0.

It is possible to evaluate the performance of the CAD_I on the $LIDC_2$ in LOPO mode, using the 69 models originated from the training session, thus assessing the quality of the model trained. Results are shown in fig. 8.3.2 and in tab. 8.5, the score is evaluated only on internal nodules.

8.3.2 CAD_I validation

The validation of the CAD_I is carried out using the $LIDC_2$ dataset, the ITALUNG − CT dataset and the five ANODE09 examples. The best model

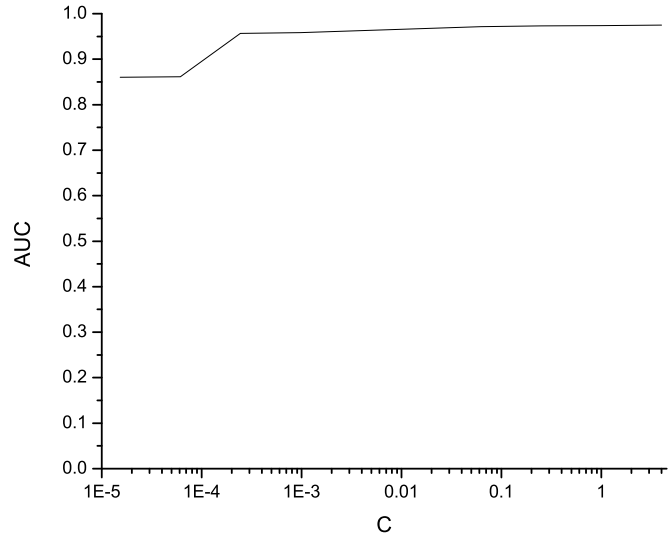


Figure 8.3.1: Results in terms of AUC of the SVM training on the LIDC_2 with the features extracted by the CAD_I as a function of the penalty parameter C.

Table 8.5: FROCSV of the CAD_I trained on the LIDC_2 for the four different LIDC ALs.

Agreement level	FROCSV
AL1	0.465
AL2	0.538
AL3	0.577
AL4	0.597

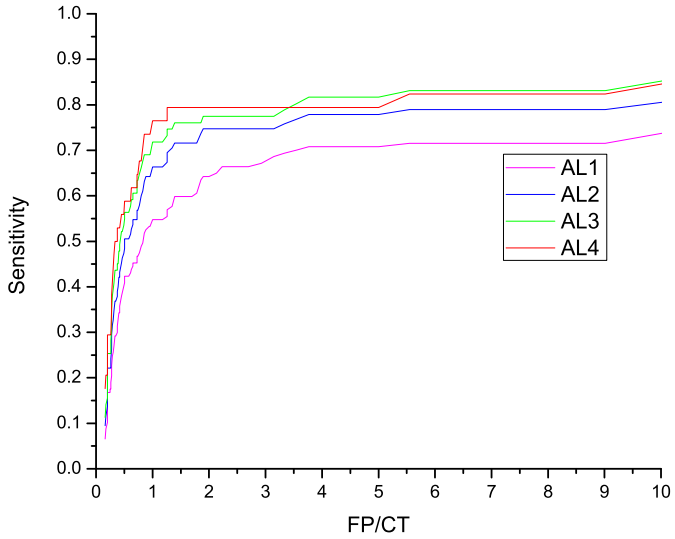


Figure 8.3.2: FROC curves obtained by the CAD_I on the LIDC_2 at the four ALs in LOPO mode.

Table 8.6: Average score obtained by the CAD_I trained on the $LIDC_2$ and validated on ANODE09 five examples, the $LIDC_1$ and the $ITALUNG-CT$ datasets.

Dataset	FROCSV
ANODE09 5 examples	0.428
$LIDC_1$	0.466/0.591/0.669/0.740
$LIDC_2$(LOPO)	0.465/0.538/0.577/0.597
$ITALUNG-CT$	0.661

found and all the settings described in sec. 8.3.1 are used. The obtained results are shown in tab. 8.6.

8.3.3 CAD_{JP} training

The CAD_{JP} is applied to the 69 CTs of the $LIDC_2$, using a 5 mm long normal and $\sigma_{cylinder} = 0.75$ mm. After the computation of the PSN local maxima, with a basis of 5 mm are searched in the $A(x, y, z)$ matrix.

Analyzing the distribution of FPs and TPs, it is possible to show that discarding findings below the score $T_{juxta-pleural} = 5$, a sensitivity of $\frac{18}{19} \sim 95\%$ is achieved 107 FP/CT on average.

The candidate feature vectors are then classified with a linear SVM using the methods described in chap. 7, an empirical threshold $T_{candidate} = 2$ and discarding those candidates with major axis outside the $[d_{min} = 0, d_{max} = 40]$ mm range. The features of the candidates are sampled, with a 6 to 1 FPs to TPs ratio for a total 33453 of train patterns.

To search for the best model a LOPO cross validation is carried out, varying the penalty parameter C of the linear SVM in the range $[2^{-20}, 2^{20}]$. The best model is then selected according to the AUC.

The results of the SVM training are shown in fig. 8.3.3. The results of the LOPO training show that the performance of the classifiers are stable in a wide range of values, with a peak value of AUC = 0.91 for C = 1.0.

It is possible to evaluate the performance of the CAD_{JP} on the $LIDC_2$ in LOPO mode, using the 69 models originating from the train session, thus assessing the quality of the model trained. Results are shown in fig. 8.3.4 and in tab. 8.7, the score is evaluated only on juxta-pleural nodules.

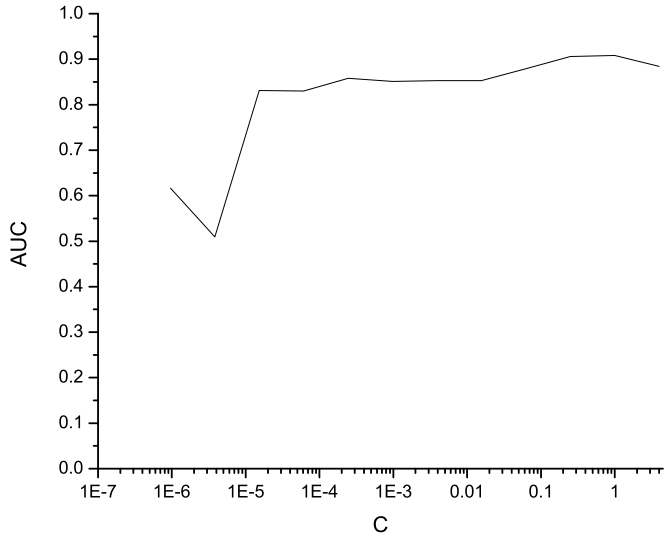


Figure 8.3.3: Results in terms of AUC of the SVM training on the LIDC_2 with the features extracted by the $\mathrm{CAD}_{\mathrm{JP}}$ as a function of the penalty parameter C.

Table 8.7: FROCSV of the $\mathrm{CAD}_{\mathrm{JP}}$ trained on the LIDC_2 for the four different LIDC ALs.

Agreement level	FROCSV
AL1	0.267
AL2	0.409
AL3	0.476
AL4	0.482

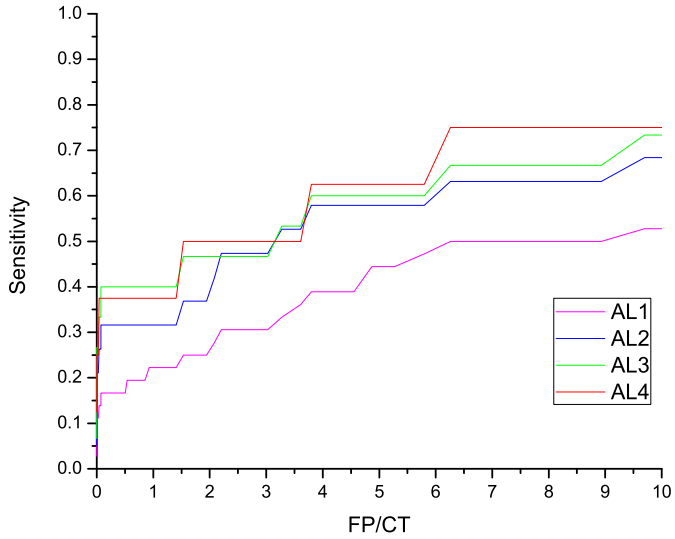


Figure 8.3.4: FROC curves obtained by the $\mathrm{CAD_{JP}}$ on the $\mathrm{LIDC_2}$ at the four ALs in LOPO mode.

Table 8.8: Average score obtained for the CAD_I trained on the $LIDC_2$ and validated on the ANODE09 5 example, the $LIDC_1$ and the $ITALUNG - CT$ datasets.

Agreement level	FROCSV
ANODE09 5	0.458
$LIDC_1$	0.344/0.495/0.553/0.575
$LIDC_2$(LOPO)	0.267/0.409/0.476/0.482
$ITALUNG - CT$	0.340

8.3.4 CAD_{JP} validation

The validation of the CAD_{JP} is carried out using the $LIDC_1$ dataset, the $ITALUNG - CT$ dataset and the ANODE09 five CTs examples. The best model found and all the settings described in sec. 8.3.1 are used. The obtained results are shown in tab. 8.8.

8.4 CAD_I and CAD_{JP} combination

The CAD_I and the CAD_{JP} are intended as complementary procedures to detect the two main nodule categories identified in this study. However, in many cases it is preferable to obtain a list of all the lung candidates; the procedure that merges the list of candidates belonging to two different CADs is referred to as combination.

To combine efficiently two CAD systems may be regarded as a non trivial task. The side effect of simply merging findings is to obtain a CAD of unpredictable performance. Therefore a procedure to combine the two approaches should be implemented.

To combine the CAD_I and the CAD_{JP}, the algorithm discussed in [16,17,68] is implemented. This algorithm may be divided in two steps: normalization and clustering.

The normalization procedure is applied in order to make findings that come from different CAD schemes directly comparable.

For each CAD, a function assigning to each degree of suspicion p a new value f(p) is evaluated according to the rule

Table 8.9: Summary of the FROCSV values obtained by CAD_I and CAD_{JP}. The results obtained on the training dataset are evaluated in LOPO mode.

Training dataset	CAD	$LIDC_1$(AL1/../AL4)	$LIDC_2$(AL1/../AL4)	Anode5	ITALUNG – CT
$LIDC_1$	CAD_I	0.472/0.605/0.676/0.743	0.471/0.548/0.592/0.615	0.462	0.685
$LIDC_1$	CAD_{JP}	0.276/0.384/0.405/0.405	0.251/0.330/0.314/0.286	0.381	0.238
$LIDC_2$	CAD_I	0.466/0.591/0.669/0.740	0.465/0.538/0.577/0.597	0.428	0.661
$LIDC_2$	CAD_{JP}	0.344/0.495/0.553/0.575	0.267/0.409/0.476/0.482	0.458	0.340

Table 8.10: FROCSV values obtained combining the CAD_I and CAD_{JP}, using the normalization function obtained on the training dataset using AL2 nodules.

Train dataset	Validation dataset				Average performance
	$LIDC_1$(AL2)	$LIDC_2$(AL2)	Anode5	ITALUNG – CT	
$LIDC_1$	0.467	0.507	0.418	0.405	0.449
$LIDC_2$	0.382	0.520	0.397	0.316	0.404

$$p \rightarrow f(p) = \frac{TP(p)}{TP(p) + FP(p) + 1} \tag{8.4.1}$$

where TP(p) and FP(p) are the number of TPs and FPs with degree of suspicion $\geq$ p. Eq. 8.4.1 is, except for the term +1, the positive predictive value of the CAD at the threshold p.

The normalization procedure is implemented on a set for which the annotations are known, and then the obtained function (p, f(p)) is used to normalize findings belonging to different datasets.

Once all the findings are normalized, it is possible to combine the information provided by different CAD schemes, by clustering those findings lying within a predefined distance and by summing their f(p).

This procedure is applied both to the CAD_I and CAD_{JP}, and the normalization functions are those evaluated on the training dataset using AL2 nodules. The results of the combination are summarized in tab. 8.10.

8.5 Validation on the ANODE09 dataset

As shown in tab. 8.10, the CAD trained on the $LIDC_1$ performs, on average, considerably better than the CAD trained on the $LIDC_2$ and moreover it shows more stable performances across the $LIDC_1$ and the $LIDC_2$ datasets.

The CAD system trained on the $LIDC_1$ has been therefore applied to the ANODE09 50 CT dataset. The CAD_I is applied using the settings described in sec. 8.2.1 on the 50 ANODE09 CT dataset producing 144 candidates per CT on average. The CAD_{JP} has been run with the settings described in sec. 8.2.3 on the 50 ANODE09 CT dataset producing 38 candidates per CT on average.

Since in the ANODE09 evaluation at most 2000 findings are processed, the list of findings of the 50 CTs was sorted according to their degree of suspicion and the 2000 most significative findings have been retained.

The FROCSV obtained on the 50 ANODE09 CTs is 0.393 with a peak sensitivity of 81% obtained at 32.76 FP/CT.

The FROCSVs of each category are summarized in tab. 8.11. The overall performance of the CAD is slightly worse than the estimation provided by the five example CTs (0.418).

This is probably due to the underrepresentation of pleural nodules in the five ANODE09 example CTs (only three relevant juxtapleural nodules are annotated), thus leading to an inaccurate estimation of the CAD_{JP} performance.

Indeed, as it is clear from tab. 8.11, the CAD_{JP} is lowering the average FROCSV of the overall CAD. This low performance may be explained by taking into account the combination procedure. As shown in tab. 8.9 the CAD_I always outperforms the CAD_{JP} in terms of FROCSV thus resulting in normalization factors that penalize the CAD_{JP}.

However, the maximum sensitivity obtained by the CAD_{JP} is 74% at 32.76 FP/CT. This shows that, even though the performance in the low FP/CT range is not good, the CAD_{JP} may be improved modifying the classification step.

8.6 Analysis of the FPs of CAD_I and CAD_{JP}

The list of FPs generated by the CAD_I and the CAD_{JP} on the $LIDC_2$ dataset was analyzed: the 50 FPs with the highest score findings of each CAD were

Table 8.11: Score obtained on the ANODE09 50 CTs for the six categories of nodules.

Nodule category	FP/CT							FROCSV
	1/8	1/4	1/2	1	2	4	8	
small nodules	0.095	0.154	0.287	0.350	0.453	0.516	0.616	0.353
large nodules	0.279	0.317	0.371	0.455	0.504	0.551	0.634	0.444
isolated nodules	0.143	0.266	0.367	0.453	0.54	0.600	0.62	0.421
vascular nodules	0.259	0.314	0.442	0.523	0.605	0.616	0.663	0.489
juxtapleural nodules	0	0	0	0	0.062	0.193	0.407	0.09
peri-fissural nodules	0.314	0.371	0.474	0.6	0.657	0.669	0.715	0.543
all nodules	0.175	0.225	0.324	0.396	0.475	0.531	0.624	0.393

reviewed and classified according to the visual assessment.

The vast majority of CAD_I FP findings (38/50) are related to vessels, i.e. they are vessels crossing, branching or large airways. The remaining 12 findings are small nodular objects that were not contained in the annotations (see fig. 8.6.1). The FPs have probabilities in the range of 0.41 to 1 corresponding to a FROC point with 65% sensitivity.

The 50 most important CAD_{JP} FPs are in correspondence of ribs and clavicles (8/50), the azygos vein (13/50), apical scars (4/50) and pleura thickenings (17/50) (see fig. 8.6.2).

The FPs have probabilities in the range of 0.91 to 1 corresponding to a FROC point with 16% sensitivity.

8.7 Comparison with literature

In this section, some of the results found in literature are reported. Most of the mentioned approaches were selected among those systems participating to the ANODE09 initiative.

ISI CAD

The ISI CAD is described in [17] and [69]. This method implements a region growing and morphological smoothing of the lung boundaries to segment the lungs. To extract nodule candidates, the shape index and curvedness are

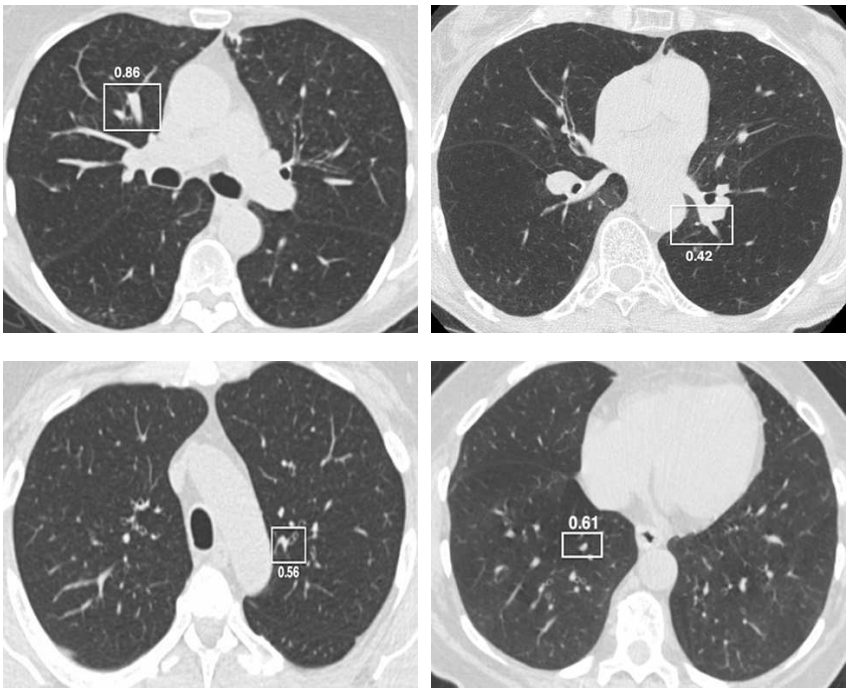
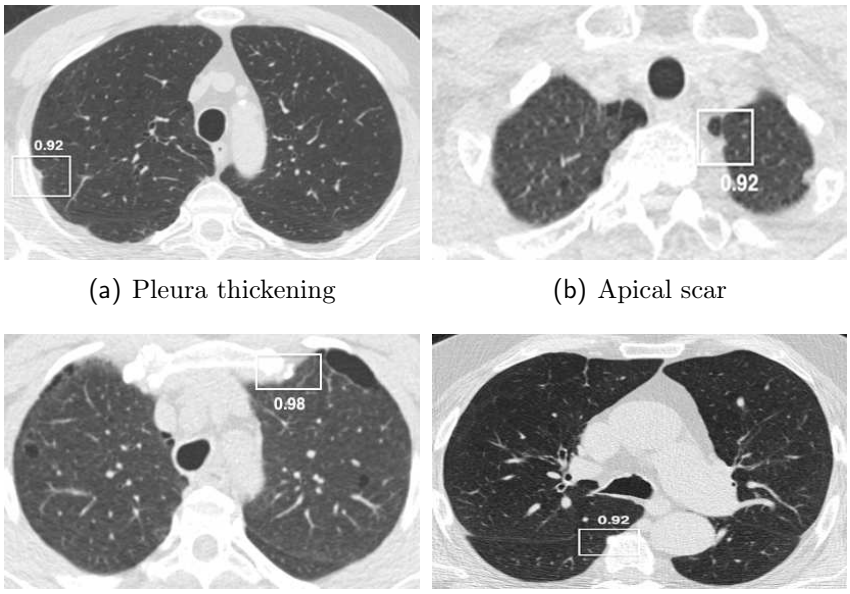


Figure 8.6.1: Four example FP findings generated by the $\mathrm{CAD_I}$.



(a) Pleura thickening (b) Apical scar

(c) Clavicle (d) Azygous vein

Figure 8.6.2: Four FP findings generated by the $\mathrm{CAD}_{\mathrm{JP}}$.

computed at a fixed scale of 1 voxel. Those voxels having these values in a predefined range are clustered into candidates.

FP candidates are then removed by means of a two-step approach using k-Nearest Neighbor classification (kNN). First a kNN classifier is trained using simple features such as intensity value and image gradients to discard the most obvious candidates, then a second KNN classifier employing 19 more complex features, is trained to finally classify the candidates.

In [17] the ISI CAD is reported to have 80% sensitivity with an average of 4.2 FP/CT on 813 CTs scan belonging to the NELSON study. This method applied to the ANODE09 datasets led to a FROCSV of 0.632.

VBNA CAD

The VBNA method, which may be regarded as an older version of the here proposed CAD approach, is described in [17, 22, 65]. This method uses a thresholding and morphological approach to segment the lungs. A MSDE and a PSN filter are then implemented to identify the nodule candidates. A

total of 131 features are extracted for each voxel candidate and classified by a neural network (see appendix B) to reduce FPs.

This method is claimed to have a sensitivity of 78% at 8 FP/CT and 70% at 4 FP/CT on a 30 CT dataset of the ITALUNG − CT database, containing 35 internal nodules and 32 juxta-pleural nodules. The same approach applied to the ANODE09 dataset resulted in a FROCSV of 0.293.

The results of an improved version of the VBNA CAD, trained on the $LIDC_1$ are reported in [16], where a sensitivity of $\sim 60\%$ at 4 FP/CT and and a FROCSV of 0.43 were obtained on the $LIDC_2$ dataset using nodules with AL2.

CAM CAD

The Channeler Ant Model (CAM) method is described in [17]. The method employs a 3D region growing method to segment the lung parenchyma and a wave-front algorithm for the definition of the lung surface on the inner side.

Two virtual ant colonies are then started from the root of the vessel tree to segment the bronchial tree. Then the same method is employed to segment the nodule candidates. When all the ants have died, candidates are obtained thresholding the pheromone map and classifying them using five intensity features by means of a neural network classifier.

This method claimed a sensitivity of 46% and 64% obtained at an average of 2 and 6 FP/CT respectively, on a subset of the ITALUNG-CT database.The same method applied on the ANODE09 dataset resulted in a FROCSV score of 0.254.

The results of an improved version of the CAM CAD are presented in [16] where a FROCSV of 0.57 is obtained by training the system on $LIDC_1$ and validating it on $LIDC_2$.

RGVP CAD

The Region Growing Volume Plateau (RGVP) CAD is described in [17]. This method uses the same lung segmentation method of the CAM approach and identifies nodule candidates evaluating an adaptive threshold for each and then filtering the candidates according to their roundness and volume. The list of obtained candidates is then classified using seven intensity based

features. The system obtained a FROCSV value of 0.291 on the ANODE09 database.

The results of an improved version of the RGVP CAD are presented in [16] where a FROCSV of 0.39 is obtained training the system on $LIDC_1$ and validating on $LIDC_2$.

Flyerscan CAD

The "Flyerscan" CAD is described in [21]. This approach first segment the lung by using a thresholding operator, then big airways are segmented and the lungs are separated with a dedicated procedure. After preprocessing, a nodule candidate detection and segmentation algorithm is employed to locate potential lung nodules. First the candidate segmentations are obtained using 15 predefined thresholds, then an expert system is trained to discards those segmentations not belonging to nodules and then the obtained segmentations at the different thresholds are merged by means of an OR operator. Finally each candidate is classified using a Fisher Linear Discrimination (FLD) classifier and 40 2D and 3D intensity and geometric features.

In [21] this approach is claimed with a sensitivity of 82.66% with an average of 3 FP/CT on a set of 84 LIDC CTs containing 143 nodules. The same algorithm applied to the ANODE09 50 CTs dataset led to a FROCSV score of 0.552.

MIG CAD

In [70] the MIG CAD is described. This method first segments the lungs using histogram thresholding, seeded region growing and morphological operators. Then a fast radial filter is applied to the lung to detect candidates. Then a Volume Of Interest (VOI) is extracted from each candidate and is used to evaluate the Maximum Intensity Projections (MIP) along the three principal axis.

The three MIP images are used to extract the Zernike moments and the resulting three feature vectors are classified using an SVM. A region is classified as a true nodule if at least two over three vectors are classified as positive.

The system is reported to have a sensitivity of 71% at 6.5 FP/CT and 60% at 2.5 FP/CT on a subset of the LIDC database containing 154 CTs using

nodules with AL4 only. The same CAD applied to the ANODE09 dataset led to a score of 0.310.

Li et al. CAD

In [23]the Li et al. CAD is presented. First the lung is segmented using a threshold operation followed by a procedure devoted to the inclusion of juxta-pleural nodules in the volume of interest. Then an MSDE filter is applied to the whole lung and the candidates are detected thresholding the MSDE matrix and searching connected components within a range of dimensions. To each candidates is then assigned a vector of 18 intensity and morphological based features. The candidates are then classified with a rule based classifier.

Li et al. [23] reported a 86% sensitivity at 6.6 FP/CT, 81% at 3.3 and 75% at 1.6 FP/CT obtained on a set of 117 not public scans. Unfortunately the CAD never joined the ANODE09 initiative.

Discussion

The performances of the proposed approach are comparable with those presented in this section (see tab. 8.12). However, there are few elements that may be improved in order to solve the issues highlighted in sec. 8.6.

First, as the vast majority of the CAD_I FPs are related to vessel branchings and crossings, the CAD_I may be improved using a more refined vessel and airway segmentation as the one implemented in the CAM CAD or implementing a dedicated procedure to remove the findings corresponding to these structures.

However, from sec. 8.5 it is also clear that the weak point of the proposed approach is the CAD_{JP}. This approach may be improved implementing other features to classify the juxta-pleural candidates, in particular positional features, i.e. features that describe the position of the candidate in the lung with respect to the organs and shape based features.

Moreover as the number of candidates generated by the MSDE and the PSN is very high, i.e. order of 100 candidates per filter, it should be useful to implement a pre-filtering to reduce the number of candidates before the classification step. This filtering should be based on features other than VBNA and should be able to discard the most obvious FPs.

Table 8.12: Comparison of results found in literature. Note that since the FROC points are not directly available on the ANODE09 website [48] the results in terms of sensitivity are only approximatively estimated from the provided FROC curve images.

CAD	Declared performance	Dataset	ANODE09	
			FROCSV	max sens.
ISI [17]	80% sens. at 4.2 FP/CT	813 CT Nelson study	0.632	$\sim 80\%$
VBNA [17]	78% sens. at 8 FP/CT	30 CT ITALUNG-CT	0.293	$\sim 70\%$
CAM [17]	64% sens. at 6 FP/CT	ITALUNG-CT	0.254	$\sim 65\%$
RGVP [17]	-	-	0.291	$\sim 60\%$
Flyerscan [21]	82.66% sens. at 3 FP/CT	84 CTs LIDC	0.552	$\sim 85\%$
MIG [70]	71% sens. at 6.5 FP/CT	154 CTs LIDC	0.310	$\sim 50\%$
Li et al. [23]	86% sens. at 6.6 FP/CT	117 not public CTs	-	-
Proposed CAD	75% sens. at 8.2 FP/CT	69 CTs LIDC2	0.393	$\sim 81\%$

As a final remark, it should be noticed that the proposed approach provides the unique characteristic of implementing two dedicate procedures to detect internal and juxta-pleural nodules. Therefore the combination of this two approaches affects the overall performance and thus may be regarded as an algorithm to be optimized.

Conclusions

Computer aided detection systems are becoming a widespread technology to support radiologists in many hospitals around the world. However, for problems as lung cancer early detection, CAD systems are far from being "perfect" and even though many improvements occurred in the last decade the research on this topic is still active.

Within this framework a CAD system for the identification of lung nodules in CT images has been proposed. The CAD system may be divided in two subprocedures the CAD_I and the CAD_{JP} each dedicated to detect a specific lung nodule category. The two systems consist of three modules: lung segmentation, nodule candidate detection and nodule candidate classification. The lung segmentation is built using standard elements of image analysis and dynamic programming, whereas the candidate detectors are based on hessian eigenvalues filters (MSDE) and on pleura surface normals intersections (PSN). Finally the candidate classification is carried out using the VBNA method, which was improved by adding new features.

These procedures were completely optimized and tested on CTs belonging to the largest collection of annotated chest CTs publicly available at present, the LIDC [41].

The reason why the LIDC database has been extensively used in this work is that it provides a variety of CTs acquired and reconstructed with different protocols and annotated with the same multi-reader paradigm. Therefore a CAD trained on such a dataset is expected to be more robust and less database dependent than a CAD trained on a homogeneous set.

However, as discussed in sec. 3.3, the LIDC provides a lot of information that has to be processed in order to be used for detection purpose. In particular, it was chosen to use the annotations provided by [46] and to train the algorithms only with nodules annotated by at least two radiologists (AL2).

The AL2 level of agreement is a tradeoff between including in the annotations objects with an uncertain nature, e.g. nodules annotated by only one radiologist, and having enough nodules to train the CAD algorithms.

Since the performance of a CAD may radically change according to the training set, different training and validation sessions were carried on the LIDC and ITALUNG-CT datasets.

The most stable obtained CAD was then blindly applied to the 50 CT ANODE09 dataset, resulting in a FROCSV of 0.393. This means that the system is able to detect, on average, 39% of the nodules in the range of 1/8 to 8 FP/CT.

The application to the ANODE09, ITALUNG-CT and LIDC dataset has shown that there is still room for improvement for both the CAD_I and the CAD_{JP}, in particular the CAD_{JP} may be improved including new features to classify candidates in order to improve the separation between FPs and true juxtapleural nodules.

However, the proposed CAD performed considerably better than the CAD scheme presented in [22] that participated to the original ANODE09 competition and that may be regarded as a previous version of this approach.

In literature a wealth of publications about CAD for lung nodules detection are available. However, as training, validation sets and evaluation procedures play a prominent role in the evaluation of a CAD performance, it is always difficult to compare the results of the described approaches. Therefore it was chosen to compare mostly the performance of the CAD that participated to the ANODE09, as this competition provides an unbiased measurement of the CAD performances obtained with the same evaluation protocol.

Another important topic described in this work is the combination of the CAD systems. In the context of this work, the combination algorithm has been used only to combine the CAD_I and the CAD_{JP}, but in general it may be considered a tool to improve the single CAD performance. Indeed, as shown in [16, 17, 68] the combination of CADs may be regarded as a promising technique to improve the results.

Finally, it should be noticed that lung nodules are common to many pathologies, thus it might be possible to apply the proposed CAD scheme to structures related to pathologies other than lung cancer, e.g. detection of silicosis nodules in dust exposed patients and detection of lung secondary metastasis in patients that underwent chemotherapy cure.

As discussed in sec. 2.2, the CAD systems are typically intended to be used by radiologists as a "second reader", therefore the real value of a CAD has to be assessed in a clinical trial context, where it will be possible to verify if the system is able to help the radiologist's work. For this purpose, a dedicated plugin for the OsiriX DICOM viewer [42] was developed to allow easy chest CT annotations and visualization of the CAD findings.

Acknowledgements

There are plenty of persons I would like to express my sincere gratitude to and probably they would not fit in a single encyclopedia, therefore I'm sorry if I forgot someone.

First of all, I would like to thank Prof.ssa Maria Evelina, Alessandra and Ilaria for always helping me and for their support: I owe you a lot since you have shown me how to organize and set up a scientific work. I would like also to thank Prof. Bram Van Ginneken for hosting me in his group in Nijmegen: it was a very beautiful experience and I learnt a lot of new things during my stay and I was very happy to spend my time with the DIAG members, in particular with Colin, Tao and Medhat.

Then I'd like to thank all my friends: Monsieur Crescio, Madame Giulia ("La zia"), Miss Ietta ("La boss"), Dr. Biscio, Michele ("Maestro Pagliacci aka Sciabarra aka Maicor"), Vjola, Goran and last but not least Vedran. Special thanks are addressed to the "dudes": Gallorini ("I' Regio"), Galli" aka Gallaccio Riglionis" (even tough I'll never ask for your car :D), Gallucci (in particular for RPG discussions), Nicolino (for linux and IT discussions) and Simeone for providing me funny moments during lunch break and during "calcetto".

Thanks are due also to Prof. Alberto Del Guerra, Prof.ssa Valeria Rosso, Prof.ssa Cristiana Peroni for giving me the possibility to complete this work and to my "new" colleagues: Katrin, Giancarlo, Nicola, Melanie, Nahema, Sascha, Stefano and Matteo for standing me in this uneasy period. I also would like to thank again Francesca, Ietta and Katrin for helping me in writing the thesis and for your suggestions. Thanks are due also to all the members of the MAGIC5 collaboration and in particular to Piergiorgio Cerello.

Finally, I would like to thank again Maddalena, Franco, Ietta, Francesca and my family for always supporting me, even in the hardest moments.

Appendix A

Lung Cancer

The staging of the lung cancer is an important factor to assess the patient's prognosis. There are mainly two categories of lung cancer: Small Cell Lung Cancers (SCLC) and Non-Small Cell Lung cancers (NSCLC). These two types of cancers grow and spread in different ways and may have different treatment options, so a distinction between these two types is important.

For a SCLC, the TNM classification is associated with the stages of the tumor [71]: T describes the size of tumor, N the involvement of the regional lymph node and M the presence of distant metastasis.

The variable T ranges in a 1-4 scale proportional to the size of the primary tumor.

The N variable ranges in 0-3 scale, where:

- N0: tumor cells absent from regional lymph nodes
- N1: regional lymph node metastasis present; (at some sites: tumor spread to closest or small number of regional lymph nodes)
- N2: tumor spread to an extent between N1 and N3 (N2 is not used at all sites)
- N3: tumor spread to more distant or numerous regional lymph nodes (N3 is not used at all sites)

The variable M ranges

- M0: no distant metastasis

Stage	TNM subset	Stage	TNM subset
0	carcinoma in situ	IIIB	T4N0M0
IA	T1 N0 M0		T4 N1 M0
IB	T2 N0 M0		T4 N2 M0
IIA	T1 N1 M0		T1 N3 M0
IIB	T2 N1 M0		T2 N3 M0
	T3 N0 M0		T3N3M0
IIIA	T3 N1 M0		T4 N3 M0
	T1 N2 M0	IV	Any T Any N M1
	T2 N2 M0		
	T3 N2 M0		

Table A.1: Cancer stages according to TNM classification.

- M1: metastasis to distant organs

There are four main stages for lung cancer according to the spread of the pathology that may be resume in tab. A.1.

Appendix B

Artificial Neural Networks

Introduction

Classifiers are often used in modern science to infer from data some relations, that may be hidden at first sight. In particular, their usage is common in fields like pattern recognition, where there is no strong theoretical basis. There are a lot of classifiers available in literature, and among all it is possible to identify two main families: classifiers that need to have labeled data to be trained (supervised classifiers) and classifiers which don't (unsupervised classifiers). Typical examples of supervised classifiers are support vector machines and neural networks, while examples of unsupervised classifiers are Self Organizing Map (SOM) and K-means algorithms [63].

In this thesis only supervised classifiers are used, and in the following appendices two different kinds of classifiers are described: Artificial Neural Network (ANN) and Support Vector Machine (SVM).

B.1 What is a neural network?

The term Artificial Neural Network [63] is traditionally referred to mathematical models inspired by the structure and the functional aspects of biological neural networks. An ANN consists of an interconnected group of artificial neurons, and it processes information using a connectionist approach to computation. Modern ANNs are non-linear statistical data modeling tools and they are used to model complex relationships between inputs and outputs,

to find patterns in data and, in general, to solve problems hard to model theoretically.

There are several ANN architectures available in literatures, the most common are the Single Layer Perceptron (SLP) and the Multi-Layer Perceptron (MLP). The MLP consists of neurons distributed in layers of different sizes. The SLP is an MLP without hidden layers, i.e. where the input layer is directly connected to the output layer. As it has been shown [63, 72], the SLP is able to solve only linear separable problems, thus it cannot be regarded as an optimal tool for solving general problems. Instead, it was shown that a MLP with one hidden layer is able to approximate an arbitrary continuous function (see universal approximation theorem).

B.2 Neurons

The basic element of an ANN is the neuron. There are several neuron models available in literature, the most common is the McCulloch–Pitts [63]. The neurons are connected themselves trough synapses w_{ij} and have a special synapse, called bias b_i, whose input is always 1. The value of each neuron depends on the sum of values of all the neurons connected to it, each weighted on the synaptic connection and filtered with the neuron activation function $\varphi(\cdot)$

$$V_j = \varphi\Big(\sum_{i \in neurons\, connected\, to\, j} w_{ij} V_i + b_i\Big),$$

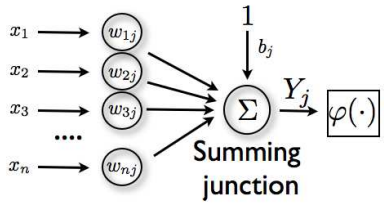
where w_{ij} is the value of the synapse between the neurons i and j and V_i and V_j are the values of the neurons and b_i is the bias of the $i-th$ neuron. The single neuron can be interpreted as the union of a summing junction directly connected to an activation function $\varphi(\cdot)$ (see fig. B.3.1(a)).

To ensure the universality of the MLP (see universal approximation theorem), the function $\varphi(\cdot)$, also referred as neuron activation function, must be a non constant bounded monotone continuous function. Typical choices for $\varphi(\cdot)$ are the sigmoid function

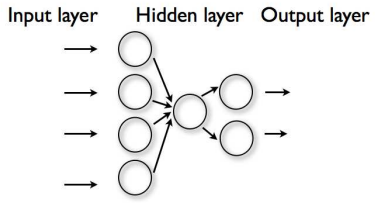
$$\varphi(V) = \frac{1}{1 + exp(-\alpha V)} \tag{B.2.1}$$

or the hyperbolic tangent

$$\varphi(V) = a * \arctan(bV).$$



(a) Scheme of a neuron represented as a summing junction connected to the activating function



(b) Picture representing a three-layer MLP with 4 input neurons, 1 neuron in the hidden layer and 2 output neurons

Figure B.3.1: Pictures representing a neuron and MLP.

B.3 Multi-layer perceptron architecture

The MLP consists of three or more layers (an input and an output layer with one or more hidden layers) of nonlinearly-activating nodes. Each node in one layer is connected with a certain weight w_{ij} to every node in the following layer. The first layer is commonly referred as "input layer" while the last one is referred as "output layer". All the layers between the last and the first are typically called "hidden" layers (see fig. B.3.1(b)). The input vector of an ANN is often referred to as "features".

For the MLP it is possible to prove the following theorem:

Theorem. *Universal approximation theorem*

Let be $\varphi(\cdot)$ be a non constant, bounded and monotone-increasing continuous function. Let I_{m_0} denote the m_0-dimensional unit hypercube $[0,1]^{m_0}$. The space of continuous functions on I_{m_0} is denoted by $C(I_{m_0})$. Then, given any function $f \in C(I_{m_0})$ and $\epsilon > 0$, there exist an integer m_1 and sets of real constants α_i, b_i, and w_{ij}, where $i = 1, .., m_1$ and $j = 1, .., m_0$ such that we may define

$$F(x_1, ..., x_{m0}) = \sum_{i=1}^{m_1} \alpha_i \varphi(\sum_{j=1}^{m_0} w_{ij} x_j + b_i)$$

as an approximate realization of the function $f(\cdot)$; that is

$$|F(x_1, ..., x_{m_0}) - f(x_1, ..., x_{m_0})| < \epsilon$$

for all $x_1, x_2, ..., x_{m_0}$ that lie in the input space.

The universal approximation theorem is directly applicable to the MLP, where :

- $\varphi(\cdot)$ is the activation function
- m_0 and m_1 are the input nodes and number of hidden neuron
- $x_1, .., x_{m_0}$ are the input features of the ANN
- $w_{i_1}, ..., w_{m_0}$ are the hidden neurons synaptic weights and b_i the bias
- $\alpha_1, ..., \alpha_{m_1}$ are synaptic weights of the output layer

The theorem shows that a single hidden layer network is enough for a MLP to compute an approximation to a given training set. However, this theorem does not say that a single layer is optimum for generalization or how many neurons should be used in the hidden layers. The problem of finding the optimal architecture for an MLP to solve a problem is addressed in the following sections.

B.4 Learning process

B.4.1 Back-propagation algorithm

An ANN learns about its environment through an interactive process of adjustment applied to its synaptic weight and bias levels. This process is also called training phase and for the MLP can be carried out using the Back-Propagation (BP) algorithm described in [63]. The BP is an iterative algorithm that modifies the MLP weights w_{ij} in order to reproduce the relation between the train patterns $i = 1, .., N$ and their desired value d_i.

Each step of the iteration n produces a new set of w_{ij}, and for each set it is possible to evaluate the distance between the desired output d_i and the obtained output y_i.

For a single pattern, the distance can be evaluated as

$$\mathcal{E}(n) = \frac{1}{2} \sum_{j \in output\, layer} e_j^2(n)$$

with $e_j = d_j - y_j$, and obtain the average distance on a set averaging all the $\mathcal{E}(n)$

$$\mathcal{E}_{average}(n) = \frac{1}{N} \sum_{n=1}^{N} \mathcal{E}(n) \tag{B.4.1}$$

The number of times the entire train dataset has been presented to the ANN is called epoch. The BP determines the weights using an iterative steepest descend algorithm to find the minimum of $\mathcal{E}$. The requirement for the BP algorithm to work, is that $\varphi(\cdot)$ must be a derivable function, for a complete description of the algorithm see [63].

The $\mathcal{E}$ is function of the w_{ij}. To find the minimum of such a function, each weight is adjusted in the opposite direction to the gradient of $\mathcal{E}$

$$\Delta w_{ij} = -\eta \frac{\partial \mathcal{E}(w_{ij})}{\partial w_{ij}} \tag{B.4.2}$$

where η is the learning rate parameter that controls the size of the step against the direction of the gradient.

There are two different ways of implementing the BP algorithm on a dataset: the batch mode and the sequential mode. The batch mode consists in updating the weights of the ANN only at the end of an epoch, whereas the sequential mode consists in randomizing the order of the patterns each epoch and update the weights after each pattern is presented to the ANN.

If avoiding to be stuck in a local minima is the most important factor, the sequential mode should be preferred to the batch mode, because presenting the patterns in a random order each epoch make it less likely for the BP to be trapped in a local minima. Instead the batch algorithm, which is very easy to parallelize, can be used to solve large problems. In this thesis the sequential mode is always used to train the ANN classifiers. The ANN classifiers can be used to solve two different kinds of problems: classification and regression. Classification consists in dividing the input data in two or more classes, while regression consists in fitting a continue value function.

B.5 Normalization

Each feature vector must be scaled in order to avoid numerical problems in the algorithms and to avoid that attributes in greater numeric ranges dominate those in smaller numeric ranges. Note that the statistics needed to normalize the input, such as the mean and standard deviation, are computed from the training data, not from the validation or test data. The validation and test data must be standardized using the statistics computed from the training data.

Linear scaling

$$x_i \rightarrow \frac{2x_i - (max_i + min_i)}{max_i - min_i} \tag{B.5.1}$$

where $\max_i$ and $\min_i$ are the max and min value of the $i-th$ feature. Using this normalization it is possible to obtain all the features in the training set in the range $[-1, 1]$.

Standardization

The standardization consists in replacing each feature x_i with

$$x_i \rightarrow \frac{x_i - \mu}{\sigma}$$

where μ_i and σ_i are the mean and std-deviation of the $i-th$ feature on training set with

$$\mu_i = \frac{\sum x_i}{N}$$

and

$$\sigma_i = \sqrt{\frac{\sum (x_i - \mu)^2}{N-1}}$$

B.6 Cross validation techniques

There is no theoretical principle to build an optimal ANN to solve a problem, i.e. it is not possible to establish how many layers and hidden neuron to use.

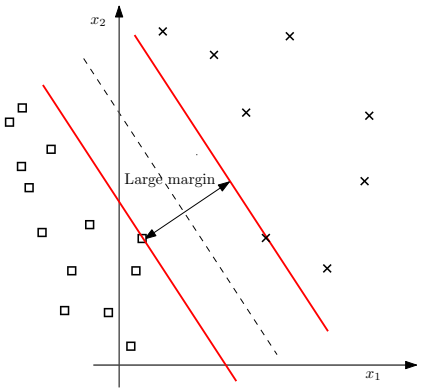
The only available approach at present is to test empirically which is the best model to fit the problem. More specifically, it is possible to view a classifier selection problem as choosing, within a set of candidate model structures, the "best" one according to a certain criterion. In this context, a standard tool in statistics known as k-fold cross-validation [63, 66] may be used. The k-fold cross validation consists in dividing the dataset in K different subsets, to train the classifier on $K-1$ subsets and to "validate" it on the remaining one. This can be done in K different ways, using each time different $K-1$ subsets for training and one for performance assessment. At the end of the procedure each subset was unbiasedly evaluated by a classifier, i.e. it was evaluated by a classifier that was not trained with it. The K classifiers provide an estimate of the performance of the classifier trained on the whole datasets and validated on a dataset of homogeneous characteristics. This method is very useful when there is not big availability of labeled data, because it allows to use the same set of data for training and assessing the performance. A special case of the k-fold cross validation is the Leave One Out (LOO) cross validation, where K is equal to the number of pattern N. Therefore, for each fold there is only one pattern that is not used for the training. In the context of CAD systems it often happens to train classifiers with patterns that come from the same exam. These patterns may be highly correlated, so doing a k-fold division at pattern level may be not the best approach. Instead a LOO at patient/exam level may be implemented [67], i.e. patterns are divided in folds according to the patient/exam they belong to. This special LOO validation is often referred as Leave One Patient Out (LOPO) cross validation.

Appendix C

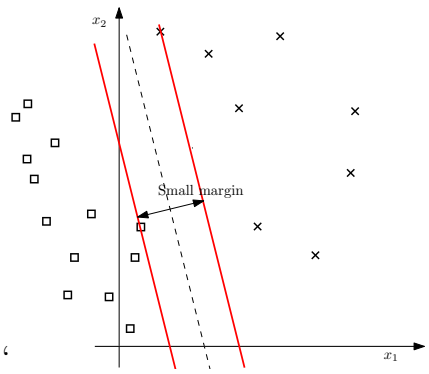
Linear Support Vector Machine

C.1 Introduction

Support Vector Machines (SVMs) [63, 72] are a set of supervised learning methods that analyze data and recognize patterns, typically used for classification and regression analysis. The SVM paradigm is not based on a biological analogy, like it is for ANN, but it is based on finding an hyperplane separating data with the largest possible margin (see fig. C.1.1). The SVM classifiers can perform both multi-class classification and regression; for the sake of simplicity, only the two class SVM, which is the only one used in this thesis, is described. In the following sections the SVM for linearly and non linearly separable patterns are described together with their algorithms. In general it is possible to implement SVM with kernel other than linear, i.e. the input features are non linearly mapped in a higher dimensional space before applying the standard SVM algorithm. In this thesis, such a kind of SVM are not discussed, since the number of features used to classify patterns is always very high (order of 100). In fact when the number of features is large, it is not necessary to map data to a higher dimensional space and the nonlinear mapping does not improve the performance [73]. Using the linear kernel is good enough, and it is possible to search only for the optimal parameter C.



(a) Hyperplane separating the 2-dimensional data with a "large" margin



(b) Hyperplane separating the 2-dimensional data with a "small" margin

Figure C.1.1: Two hyperplanes separating the same set of data.

C.2 Optimal hyperplane for linearly separable patterns

Suppose, two classes of data have to be separated, and denote the optimal hyperplane with

$$\mathbf{w_0^T x} + b_0 = 0$$

where $\mathbf{w_0}, \mathrm{b} \in \mathbb{R}^{\mathrm{n}}$ are the parameters of the hyperplane. It is possible to show that the distance of a point $\tilde{\mathbf{x}}$ from the hyperplane is

$$r = \frac{g(\tilde{\mathbf{x}})}{\|\mathbf{w_0}\|} \tag{C.2.1}$$

where $\mathrm{g}(\mathbf{x})$ is

$$g(\mathbf{x}) = \mathbf{w_0^T x} + b_0.$$

The issue is now to find a criterion to determine $\mathbf{w_0}$ and b_0 and to define the concept of the optimal separating hyperplane given the set of training $\mathbf{x_i}, \mathrm{d_i}$. In this context, it is assumed that the data are divided in two classes $\mathrm{d_{i=1,..,N}} = \pm 1$.

The $\mathbf{w_0}$ and b_0 of the optimal hyperplane, must satisfy the constraints

$$\begin{aligned} \mathbf{w_0^T x_i} + \mathrm{b}_0 \geq 1 \quad \text{for} \quad \mathrm{d_i} = 1 \\ \mathbf{w_0^T x_i} + \mathrm{b}_0 \leq -1 \quad \text{for} \quad \mathrm{d_i} = -1 \end{aligned}$$

which may be rewritten in a single equation

$$d_i(\mathbf{w_0^T x_i} + b) \geq 1. \tag{C.2.2}$$

Of course it is implicit the assumption that such a $\mathbf{w_0}$ and b exist, and this is true only for linearly separable patterns.

The particular data points $\mathbf{x_i}, \mathrm{d_i}$ for which the equality eq. C.2.2 holds, are called Support Vectors (SV). This vectors play a prominent role in the operation of this class of learning machine. In conceptual terms the SV are those data points that lie closest to the decision surface and are therefore the most difficult to classify. As such, they have a direct bearing on the optimum location of the decision surface. Using the eq. C.2.1, it is possible to evaluate the distance from the optimal hyperplane for patterns with $\mathrm{d_i} = 1$

$$r = \frac{g(x)}{\|w_0\|} = \frac{1}{\|w_0\|}$$

and $\mathrm{r} = -\frac{1}{\|\mathrm{w_0}\|}$ for a pattern for $\mathrm{d_i} = -1$.

The margin between these hyperplanes, is defined as $\rho = 2\mathrm{r} = \frac{2}{\|\mathbf{w_0}\|}$, and to maximize ρ is equal to find the minimum of $\|\mathbf{w_0}\|$ subject to the constraints eq. C.2.2.

The problem of finding the optimal hyperplane for a set of N patterns $\mathbf{x_i}, \mathrm{d_i}$ can be expressed in terms of finding the minimum of

$$\mathcal{L}(\mathbf{w}, b) = \frac{1}{2}\mathbf{w^T w} \tag{C.2.3}$$

subject to the constraints eq. C.2.2, where the factor $\frac{1}{2}$ is for convenience.

The constrained problem eq. C.2.3 can be solved finding the "saddle point" [63, 72] of the Lagrange function

$$\overline{\mathcal{L}}(\mathbf{w}, b, \alpha) = \frac{1}{2}\mathbf{w^T w} - \sum_{i=1}^{N} \alpha_i [d_i(\mathbf{w^T x_i} + b) - 1] \tag{C.2.4}$$

where α_i are the Lagrange multipliers and $\mathbf{w}$ and b are the parameters of the hyperplane. The saddle point can be found looking for the minimum of $\overline{\mathcal{L}}(\mathrm{w}, \mathrm{b}, \alpha)$ with respect to the variables $\mathbf{w}$ and b and looking for the maximum with respect to the α_i variables. This problem can be solved either in the primal space (the space of the parameter $\mathbf{w}$ and b) or in the dual space

(the space of the Lagrange multipliers α_i). It is convenient to solve the dual problem, because this problem can be expressed only in terms of the $\mathbf{x_i}, \mathrm{d_i}$.

To find local extrema of a function subject to inequality constraints, the Kuhn-Tucker (KT) may be used

$$\begin{aligned} \alpha_\mathrm{i} &\geq 0 \\ \alpha_\mathrm{i}[\mathrm{d_i}(\mathbf{w^T x_i} + \mathrm{b}) - 1] &= 0 \end{aligned} \tag{C.2.5}$$

in this case both the eq. C.2.3 and the constraints are convex, and the KT conditions are necessary and sufficient for a maximum with respect to α_i, of the function eq. C.2.4.

In particular, the KT conditions imply in this case that $\alpha_\mathrm{i} \neq 0$ only for SV.

C.3 Solution of the dual problem

It is possible to re-factor the "primal problem", using the relations obtained imposing the partial derivatives, with respect to $\mathbf{w}$ and b, to be equal to 0

$$\frac{\partial \overline{\mathcal{L}}(\mathbf{w}, b, \alpha)}{\partial \mathbf{w}} = 0 \;\Rightarrow \mathbf{w} = \sum_{i=1}^{N} \alpha_i d_i \mathbf{x_i} \tag{C.3.1}$$

$$\frac{\partial \overline{\mathcal{L}}(\mathbf{w}, b, \alpha)}{\partial b} = 0 \;\Rightarrow \sum_{i=1}^{N} \alpha_i d_i = 0 \tag{C.3.2}$$

in particular substituting the eq. C.3.1 and the eq. C.3.2 in the eq. C.2.4.

After this substitution the new problem becomes

$$\mathcal{Q}(\alpha) = \sum_{i=1}^{N} \alpha_i - \frac{1}{2} \sum_{i=1}^{N} \sum_{j=1}^{N} \alpha_i \alpha_j d_i d_j x_i^T x_j \tag{C.3.3}$$

subject to the constraints

$$\sum_{i=1}^{N} \alpha_i d_i = 0$$

and

$$\alpha_i \geq 0$$

The equation eq. C.3.3 is also called "dual problem" and it is expressed as a function of $\mathbf{x_i}, \mathrm{d_i}$.
Finding the maximum of the "dual problem" $\mathcal{Q}(\alpha_\mathbf{0})$ is equivalent to find the minimum of $\mathcal{L}(\mathbf{w_0}, \mathrm{b}_0)$ under the constraints eq. C.2.2, and the solution of the "primal problem" can be expressed as

$$\mathbf{w_0} = \sum_{i=1}^{N} \alpha_{\mathbf{0,i}} d_i \mathbf{x_i} \tag{C.3.4}$$

where $\alpha_{\mathbf{0,i}}$ is the solution of the dual problem. It is important to note that only the SVs contribute to the eq. C.3.4, since for other vectors $\alpha_\mathrm{i} = 0$.
Instead the bias can be computed using the relation

$$b_0 = \frac{1}{N_{SV}} \sum_{s=1}^{N_{SV}} (d_s - \mathbf{w_0^T x_s})$$

where the sum is over all the SVs and $\mathrm{N_{SV}}$ is the number of SVs.

C.4 Optimal hyperplane for non linearly separable patterns

If the input patterns are not linearly separable, the relation eq. C.2.2 cannot be satisfied. To allow for deviance from the linearly separable problem, it is possible to modify the eq. C.2.2 introducing the slack variables ξ_i.
The relation eq. C.2.2 becomes

$$d_i(\mathbf{w^T x_i} + b) \geq 1 - \xi_i \tag{C.4.1}$$

and the corresponding function to minimize is

$$\mathcal{L}(\mathbf{w}, \xi) = \frac{1}{2}\mathbf{w^T w} + C \sum_{i=1}^{N} \xi_i \tag{C.4.2}$$

with the constraints eq. C.4.1, where C is a parameter selected by the user, that controls the tradeoff between complexity of the machine and the number of non separable points. After using the Lagrange multipliers, the eq. C.4.2 becomes

$$\overline{\mathcal{L}}(\mathbf{w}, b, \xi_i, \alpha_i, \mu_i) = \frac{1}{2}\mathbf{w^T w} + C \sum_{i=1}^{N} \xi_i - \sum_{i=1}^{N} \alpha_i [d_i(\mathbf{w^T x_i} + b) - 1 + \xi_i] - \sum_{i=1}^{N} \mu_i \xi_i \tag{C.4.3}$$

where α_i and μ_i are the Lagrange multipliers.
The KT conditions become

$$\begin{aligned} \alpha_\mathrm{i} &\geq 0 \\ \alpha_\mathrm{i}[\mathrm{d_i}(\mathbf{w^T x_i} + \mathrm{b}) - 1 + \xi_\mathrm{i}] &= 0 \end{aligned}$$

and

$$\begin{aligned} \mu_\mathrm{i} &\geq 0 \\ \mu_\mathrm{i}\xi_\mathrm{i} &= 0 \end{aligned}$$

Using the saddle point criterion, for eq. C.4.3 and using the obtained relations, it is possible to obtain the equation

$$\mathcal{Q}(\alpha) = \sum_{i=1}^{N} \alpha_i - \frac{1}{2} \sum_{i=1}^{N} \sum_{j=1}^{N} \alpha_i \alpha_j d_i d_j x_i^T x_j \tag{C.4.4}$$

with the constraint

$$\sum_{i=1}^{N} \alpha_i d_i = 0.$$

Taking the derivative of

$$\frac{\partial \mathcal{L}(\mathbf{w}, \xi_i)}{\partial \xi_i} = 0 \;\Rightarrow\; C = \mu_i + \alpha_i$$

and since both α_i and μ_i are ≥ 0

$$\mu_i = C - \alpha_i \geq 0 \Rightarrow \; \alpha_i \leq C$$

which, together with the constraint $\alpha_\mathrm{i} \geq 0$

$$0 \leq \alpha_i \leq C. \tag{C.4.5}$$

The eq. C.4.4 is the same obtained for linearly separable patterns eq. C.3.3 except for the constraint eq. C.4.5. It is important to note that neither the slack variables ξ_i, nor their Lagrange multipliers μ_i, appear in the dual problem.
It is possible to rewrite the equation C.4.4 in matrix form

$$\frac{1}{2}\alpha^T H \alpha - f^T \alpha \tag{C.4.6}$$

where $\mathrm{f} = \{1, .., 1\}$, $\alpha = \{\alpha_1, .., \alpha_\mathrm{N}\}$ and $\mathrm{H_{ij}} = \mathrm{d_i d_j x_i^T x_j}$. The eq. C.4.6 is a convex optimization problem with linear constraints, and can be solved using standard programming techniques [63, 72].

C.5 Normalization

As discussed in sec. B.5, all the input features to a classifier need to be normalized in order to avoid that attributes in greater numeric ranges dominating those in smaller numeric ranges. The scaling applied in this thesis for the SVM is a linear scaling eq. B.5.1.

C.6 Cross validation techniques

As shown in B.6 one of the most common way to choose among models, i.e. to choose between SVM with different C parameters, is the cross validation. In particular in this thesis the LOPO is often used to assess the best model. Since there is no a-priori indication on what is the best value of C for different datasets, it is common to span a range value of C going from 2^{-20} to 2^{20}. The best model is then selected according to AUC obtained by the LOPO validation.

Bibliography

[1] M. Quinn E. Mugno R. Capocaccia A. Micheli, P. Baili and P. Grosclaude (The EUROCARE Working Group). Life expectancy and cancer survival in the eurocare-3 cancer registry areas. *Ann. Oncol*, 2003.

[2] Ahmedin Jemal, Taylor Murray, Elizabeth Ward, Alicia Samuels, Ram C. Tiwari, Asma Ghafoor, Eric J. Feuer, and Michael J. Thun. Cancer statistics, 2005. *CA Cancer J Clin*, 55(1):10–30, 2005.

[3] Cancer facts & figures main. `http://www.cancer.org/Research/CancerFactsFigures/index`.

[4] Ahmedin Jemal, Rebecca Siegel, Jiaquan Xu, and Elizabeth Ward. Cancer statistics, 2010. *CA: A Cancer Journal for Clinicians*, 60(5):277–300, September 2010.

[5] Joanna Didkowska, Marta Manczuk, Ann McNeill, John Powles, and Witold Zatonski. Lung cancer mortality at ages 35-54 in the european union: ecological study of evolving tobacco epidemics. *BMJ*, 331(7510):189–191, July 2005.

[6] C Bosetti, F Levi, F Lucchini, E Negri, and C La Vecchia. Lung cancer mortality in european women: recent trends and perspectives. *Annals of Oncology: Official Journal of the European Society for Medical Oncology / ESMO*, 16(10):1597–1604, October 2005. PMID: 16014639.

[7] D C Ihde and J D Minna. Non-small cell lung cancer. part i: Biology, diagnosis, and staging. *Current Problems in Cancer*, 15(2):61–104, April 1991. PMID: 1649734.

[8] J C Nesbitt, J B Putnam, G L Walsh, J A Roth, and C F Mountain. Survival in early-stage non-small cell lung cancer. *The Annals of Thoracic Surgery*, 60(2):466–472, August 1995. PMID: 7646126.

[9] Diederich, S. Diederich, Lentschig, M. Lentschig, Overbeck, T. Overbeck, Wormanns, D. Wormanns, Heindel, and W. Heindel. Detection of pulmonary nodules at spiral CT: comparison of maximum intensity projection sliding slabs and single-image reporting. *European Radiology*, 11(8):1345–1350, 2001.

[10] S. Arahata T. Kodaira T. Isomura T. Kato K. Yamakawa K. Maruyama S. Itoh, M. Ikeda and T. Ishigaki. Lung cancer screening: Minimum tube current required for helical ct. *Radiology*, pages 175–183, 2000.

[11] D.F. Yankelevitz D.P. Naidich G. McGuinness O.S. Miettinen D.M. Libby M.W. Pasmantier J. Koizumi N.K. Altorki C.I. Henschke, D.I. McCauley and J.P. Smith. Early lung cancer action project: overall design and findings from baseline screening. *The Lancet*, pages 99–105, 1999.

[12] S.J. Swensen. Screening for lung cancer with computed tomography. *BMJ*, pages 894–895, 2003.

[13] Edward F. Patz, Stephen J. Swensen, and James E. Herndon. Estimate of lung cancer mortality from Low-Dose spiral computed tomography screening trials: Implications for current mass screening recommendations. *J Clin Oncol*, 22(11):2202–2206, June 2004.

[14] Reduced Lung-Cancer mortality with Low-Dose computed tomographic screening. *The New England Journal of Medicine*, June 2011. PMID: 21714641.

[15] Heidi C. Roberts, Demetris Patsios, Michael Kucharczyk, Narinder Paul, and Timothy P. Roberts. The utility of computer-aided detection (CAD) for lung cancer screening using low-dose CT. *International Congress Series*, 1281:1137–1142, May 2005.

[16] Niccoló Camarlinghi, Ilaria Gori, Alessandra Retico, Roberto Bellotti, Paolo Bosco, Piergiorgio Cerello, Gianfranco Gargano, Ernesto

Lopez Torres, Rosario Megna, Marco Peccarisi, and Maria Evelina Fantacci. Combination of computer-aided detection algorithms for automatic lung nodule identification. *International Journal of Computer Assisted Radiology and Surgery*, July 2011. PMID: 21739112.

[17] Bram van Ginneken, Samuel G. Armato III, Bartjan de Hoop, Saskia van Amelsvoort-van de Vorst, Thomas Duindam, Meindert Niemeijer, Keelin Murphy, Arnold Schilham, Alessandra Retico, Maria Evelina Fantacci, Niccolò Camarlinghi, Francesco Bagagli, Ilaria Gori, Takeshi Hara, Hiroshi Fujita, Gianfranco Gargano, Roberto Bellotti, Sabina Tangaro, Lourdes Bolanos, Francesco De Carlo, Piergiorgio Cerello, Sorin Cristian Cheran, Ernesto Lopez Torres, and Mathias Prokop. Comparing and combining algorithms for computer-aided detection of pulmonary nodules in computed tomography scans: The anode09 study. *Medical Image Analysis*, 14(6):707 – 722, 2010.

[18] Bruno Golosio, Giovanni Luca Masala, Alessio Piccioli, Piernicola Oliva, Massimo Carpinelli, Rosella Cataldo, Piergiorgio Cerello, Francesco De Carlo, Fabio Falaschi, Maria Evelina Fantacci, Gianfranco Gargano, Parnian Kasae, and Massimo Torsello. A novel multithreshold method for nodule detection in lung CT. *Medical Physics*, 36(8):3607–3618, August 2009. PMID: 19746795.

[19] I. Gori, M. E Fantacci, A. Preite Martinez, and A. Retico. An automated system for lung nodule detection in low-dose computed tomography. http://adsabs.harvard.edu/abs/2007arXiv0704.2728G, March 2007.

[20] Qiang Li, Shusuke Sone, and Kunio Doi. Selective enhancement filters for nodules, vessels, and airway walls in two- and three-dimensional CT scans. *Medical Physics*, 30(8):2040–2051, August 2003. PMID: 12945970.

[21] Temesguen Messay, Russell C. Hardie, and Steven K. Rogers. A new computationally efficient cad system for pulmonary nodule detection in ct imagery. *Medical Image Analysis*, 14(3):390 – 406, 2010.

[22] Alessandra Retico, Francesco Bagagli, Niccolò Camarlinghi, Carmela Carpentieri, Maria Evelina Fantacci, and Ilaria Gori. A voxel-based neural approach (VBNA) to identify lung nodules in the ANODE09 study. In *Medical Imaging 2009: Computer-Aided Diagnosis*, volume

7260, pages 72601S–8, Lake Buena Vista, FL, USA, February 2009. SPIE.

[23] Qiang Li, Feng Li, and Kunio Doi. Computerized detection of lung nodules in thin-section ct images by use of selective enhancement filters and an automated rule-based classifier. *Academic Radiology*, 15(2):165 – 175, 2008.

[24] See the information available at the web site "`http://www.itk.org/`".

[25] See the information available at the web site "`http://www.vtk.org/`".

[26] See the information available at the website "`http://svmlight.joachims.org/`". accessed 5 september 2011.

[27] David S Paik, Christopher F Beaulieu, Geoffrey D Rubin, Burak Acar, R Brooke Jeffrey, Judy Yee, Joyoni Dey, and Sandy Napel. Surface normal overlap: a computer-aided detection algorithm with application to colonic polyps and lung nodules in helical CT. *IEEE Transactions on Medical Imaging*, 23(6):661–675, June 2004. PMID: 15191141.

[28] I. Gori P. Kasae B. Golosio A. Piccioli P. Cerello G. De Nunzio S. Tangaro A. Retico, M.E. Fantacci. Pleural nodule identification in low-dose and thin-slice lung computed tomography. *Comput Biol Med, in press.*, 2009.

[29] A. Geissbuhler J. Heuberger and H. Muller. Lung ct segmentation for image retrieval using the insight toolkit (itk). *Medical Imaging and Telemedicine (MIT)*, August 2005.

[30] Carlos Vinhais and Aurelio Campilho. Lung parenchyma segmentation from ct images based on material decompositiong. In Aurelio Campilho and Mohamed Kamel, editors, *Image Analysis and Recognition*, volume 4142 of *Lecture Notes in Computer Science*, pages 624–635. Springer Berlin Heidelberg, 2006. 10.1007/11867661_56.

[31] Eva M van Rikxoort, Bartjan de Hoop, Max A Viergever, Mathias Prokop, and Bram van Ginneken. Automatic lung segmentation from thoracic computed tomography scans using a hybrid approach with error detection. *Medical Physics*, 36(7):2934–2947, July 2009. PMID: 19673192.

[32] J. Beutel. *Handbook of medical imaging: Physics and psychophysics*. Number v. 1. SPIE Press, 2000.

[33] DICOM. see information available at "http://medical.nema.org/".

[34] See the information available at "http://en.wikipedia.org/wiki/X-ray_computed_tomography".

[35] E R Weibel and D M Gomez. Architecture of the human lung. use of quantitative methods establishes fundamental relations between size and number of lung structures. *Science (New York, N.Y.)*, 137:577–585, August 1962. PMID: 14005590.

[36] E Breatnach, G C Abbott, and R G Fraser. Dimensions of the normal human trachea. *AJR. American Journal of Roentgenology*, 142(5):903–906, May 1984. PMID: 6609569.

[37] See the information available at the web site "http://en.wikipedia.org/wiki/Screening_(medicine)".

[38] See the information available at the web site "http://www.devchakraborty.com/FROC_vs_free_response.html".

[39] D P Chakraborty and L H Winter. Free-response methodology: alternate analysis and a new observer-performance experiment. *Radiology*, 174(3 Pt 1):873–881, March 1990. PMID: 2305073.

[40] Michael F. McNitt-Gray, Samuel G. Armato III, Charles R. Meyer, Anthony P. Reeves, Geoffrey McLennan, Richie C. Pais, John Freymann, Matthew S. Brown, Roger M. Engelmann, Peyton H. Bland, Gary E. Laderach, Chris Piker, Junfeng Guo, Zaid Towfic, David P.-Y. Qing, David F. Yankelevitz, Denise R. Aberle, Edwin J.R. van Beek, Heber MacMahon, Ella A. Kazerooni, Barbara Y. Croft, and Laurence P. Clarke. The lung image database consortium (LIDC) data collection process for nodule detection and annotation. *Academic Radiology*, 14(12):1464–1474, December 2007.

[41] See the information available at the web site "http://imaging.cancer.gov/programsandresources/InformationSystems/LIDC".

[42] "http://www.osirix-viewer.com/" accessed 8 october 2011.

[43] Andrea Lopes Pegna, Giulia Picozzi, Mario Mascalchi, Francesca Maria Carozzi, Laura Carrozzi, Camilla Comin, Cheti Spinelli, Fabio Falaschi, Michela Grazzini, Florio Innocenti, Cristina Ronchi, and Eugenio Paci. Design, recruitment and baseline results of the ITALUNG trial for lung cancer screening with low-dose CT. *Lung Cancer (Amsterdam, Netherlands)*, 64(1):34–40, April 2009. PMID: 18723240.

[44] R. Bellotti, P. Cerello, S. Tangaro, V. Bevilacqua, M. Castellano, G. Mastronardi, F. De Carlo, S. Bagnasco, U. Bottigli, R. Cataldo, E. Catanzariti, S. C. Cheran, P. Delogu, I. De Mitri, G. De Nunzio, M. E. Fantacci, F. Fauci, G. Gargano, B. Golosio, P. L. Indovina, A. Lauria, E. Lopez Torres, R. Magro, G. L. Masala, R. Massafra, P. Oliva, A. Preite Martinez, M. Quarta, G. Raso, A. Retico, M. Sitta, S. Stumbo, A. Tata, S. Squarcia, A. Schenone, E. Molinari, and B. Canesi. Distributed medical images analysis on a grid infrastructure. *Future Gener. Comput. Syst.*, 23(3):475–484, 2007.

[45] See the information available at the web site "`https://wiki.nci.nih.gov/display/Imaging/LIDC`".

[46] See the information available at "`http://www.via.cornell.edu/lidc`".

[47] A Reeves, A Biancardi, T Apanasovich, C Meyer, H Macmahon, E Vanbeek, E Kazerooni, D Yankelevitz, M Mcnittgray, and G Mclennan. The lung image database consortium (LIDC)1A comparison of different size metrics for pulmonary nodule measurements. *Academic Radiology*, 14(12):1475–1485, December 2007.

[48] See the information available at the web site "`http://anode09.isi.uu.nl/`".

[49] See the itk software guide available at the web site "`www.itk.org/ItkSoftwareGuide.pdf`". pages 257–252.

[50] Rafael C. Gonzalez and Richard Eugene Woods. *Digital image processing*. Prentice Hall, 2008.

[51] Shiying Hu, Eric A. Hoffman, and Joseph M. Reinhardt. Automatic lung segmentation for accurate quantitation of volumetric x-ray ct images. *IEEE Transactions on Medical Imaging*, 20:490–498, 2001.

[52] See the information available at the website "http://en.wikipedia.org/wiki/Dynamic_programming". Accessed 5 september 2011.

[53] R. Deriche. Fast algorithms for Low-Level vision. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 12:78–87, January 1990. ACM ID: 81086.

[54] See the information available at the web site "http://www.insight-journal.org/browse/publication/179".

[55] J. Kopp. Efficient Numerical Diagonalization of Hermitian 3 × 3 Matrices. *International Journal of Modern Physics C*, 19:523–548, 2008.

[56] Milton Abramowitz and Irene A. Stegun. *Handbook of Mathematical Functions: with Formulas, Graphs, and Mathematical Tables*. Dover Publications, June 1965.

[57] S. Barnett. A new formulation of the Lienard-Chipart stability criterion. *Proc. Cambridge Phil. Soc*, 70:269, 1971.

[58] See the information available at the web site "http://en.wikipedia.org/wiki/Polygon_mesh".

[59] William Lorensen and Harvey Cline. Marching cubes: A high resolution 3D surface construction algorithm. In *SIGGRAPH '87: Proceedings of the 14th annual conference on Computer graphics and interactive techniques*, volume 21, pages 169, 163. ACM Press, July 1987.

[60] See the information available at the web site "http://it.wikipedia.org/wiki/Gouraud_shading".

[61] See the information available at the web site "http://en.wikipedia.org/wiki/Rasterisation".

[62] See the information available at the web site "http://en.wikipedia.org/wiki/Bresenham's_line_algorithm".

[63] Simon Haykin. *Neural Networks: A Comprehensive Foundation*. Prentice Hall, 2 edition, July 1998.

[64] I. Gori, F. Bagagli, M. E. Fantacci, A. Preite Martinez, A. Retico, I. De Mitri, S. Donadio, C. Fulcheri, G. Gargano, R. Magro, M. Santoro, and S. Stumbo. Multi-scale analysis of lung computed tomography images. *Journal of Instrumentation*, 2(09):P09007, 2007.

[65] I. Gori, F. Bagagli, N. Camarlinghi, M. E. Fantacci, A. Retico, Barattini, L. Bolanos, F. Falaschi, G. Gargano, A. Massafra, and C. Spinelli. Methodology for automated detection of internal and juxtapleural lung nodules in computed tomography images. *International Journal of Computer Assisted Radiology and Surgery*, 4(Suppl 1):S360, June 2009.

[66] M. Stone. Cross-Validatory Choice and Assessment of Statistical Predictions. *Journal of the Royal Statistical Society. Series B (Methodological)*, 36(2):111–147, 1974.

[67] 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, 2004. CARS 2004 - Computer Assisted Radiology and Surgery. Proceedings of the 18th International Congress and Exhibition.

[68] M. Niemeijer, M. Loog, M.D. Abramoff, M.A. Viergever, M. Prokop, and B. van Ginneken. On combining computer-aided detection systems. *Medical Imaging, IEEE Transactions on*, 30(2):215 –223, feb. 2011.

[69] K. Murphy, B. van Ginneken, A.M.R. Schilham, B.J. de Hoop, H.A. Gietema, and M. Prokop. A large-scale evaluation of automatic pulmonary nodule detection in chest ct using local image features and k-nearest-neighbour classification. *Medical Image Analysis*, 13(5):757 – 770, 2009.

[70] Alessandro Riccardi, Todor Sergueev Petkov, Gianluca Ferri, Matteo Masotti, and Renato Campanini. Computer-aided detection of lung nodules via 3D fast radial transform, scale space representation, and zernike MIP classification. *Medical Physics*, 38(4):1962–1971, April 2011. PMID: 21626929.

[71] Clifton F. Mountain. Staging classification of lung cancer: A critical evaluation. *Clinics in Chest Medicine*, 23(1):103 – 121, 2002. Lung Cancer.

[72] V Kecman. *Learning and Soft Computing, Support Vector Machines, Neural Networks and Fuzzy Logic Models*. MIT Press, 2001.

[73] See the informations available at "`http://www.csie.ntu.edu.tw/~cjlin/papers/guide/guide.pdf`".