

**A COMPUTER AIDED DIAGNOSIS SYSTEM (CAD)
FOR THE DETECTION OF PULMONARY
NODULES ON CT SCANS**

By

Eng. Michael Samir Labib Habib

Systems and Biomedical Engineering Department
Faculty of Engineering, Cairo University

**A Thesis submitted to the
Faculty of Engineering at Cairo University
in Partial Fulfillment of the
Requirements for the Degree of**

**(MASTER OF SCIENCE)
IN
(SYSTEMS AND BIOMEDICAL ENGINEERING)**

FACULTY OF ENGINEERING, CAIRO UNIVERSITY
GIZA, EGYPT
2009

**A COMPUTER AIDED DIAGNOSIS SYSTEM (CAD)
FOR THE DETECTION OF PULMONARY
NODULES ON CT SCANS**

By

Eng. Michael Samir Labib Habib

Systems and Biomedical Engineering Department
Faculty of Engineering, Cairo University

A Thesis submitted to the
Faculty of Engineering at Cairo University
in Partial Fulfillment of the
Requirements for the Degree of

(MASTER OF SCIENCE)
IN
(SYSTEMS AND BIOMEDICAL ENGINEERING)

Under the Supervision of

Prof. Dr. Yasser M. Kadah

Dr. Ahmed Salah El-Din Mohamed

Systems and Biomedical Engineering Department
Faculty of Engineering, Cairo University

FACULTY OF ENGINEERING, CAIRO UNIVERSITY
GIZA, EGYPT
2009

A COMPUTER AIDED DIAGNOSIS SYSTEM (CAD) FOR THE DETECTION OF PULMONARY NODULES ON CT SCANS

By

Eng. Michael Samir Labib Habib

Systems and Biomedical Engineering Department
Faculty of Engineering, Cairo University

A Thesis submitted to
The Faculty of Engineering at Cairo University
in Partial Fulfillment of the Requirements
for the Degree of

(MASTER OF SCIENCE)
IN
(SYSTEMS AND BIOMEDICAL ENGINEERING)

Approved by the
Examining Committee:

Prof. Dr. Yasser M. Kadah, Thesis Main Advisor

Prof. Dr. Abdalla S. A. Mohamed, Member

Prof. Dr. Samia Abd-Elrazek Mashali, Member

FACULTY OF ENGINEERING, CAIRO UNIVERSITY
GIZA, EGYPT
2009

نظام للتشخيص بمساعدة الحاسب الألي لأورام الرئة في صور الأشعة المقطعية

إعداد

مهندس / مايكل سمير نبيب حبيب

قسم الهندسة الطبية و المنظومات
كلية الهندسة – جامعة القاهرة

رسالة مقدمة إلي كلية الهندسة – جامعة القاهرة
كجزء من متطلبات الحصول على درجة الماجستير
في الهندسة الحيوية الطبية و المنظومات

كلية الهندسة – جامعة القاهرة
الجيزة – جمهورية مصر العربية
2009

نظام للتشخيص بمساعدة الحاسب الآلي لأورام الرئة في صور الأشعة المقطعية

إعداد

مهندس / مايكل سمير نبيب حبيب

قسم الهندسة الطبية و المنظومات
كلية الهندسة – جامعة القاهرة

رسالة مقدمة إلي كلية الهندسة – جامعة القاهرة
كجزء من متطلبات الحصول على درجة الماجستير
في الهندسة الحيوية الطبية و المنظومات

تحت إشراف

د / أحمد صلاح الدين محمد

أ.د / ياسر مصطفى قدح

قسم الهندسة الطبية و المنظومات
كلية الهندسة – جامعة القاهرة

كلية الهندسة – جامعة القاهرة
الجيزة – جمهورية مصر العربية
2009

نظام للتشخيص بمساعدة الحاسب الآلي لأورام الرئة في صور الأشعة المقطعية

إعداد

مهندس / مايكل سمير نبيب حبيب

قسم الهندسة الطبية و المنظومات
كلية الهندسة – جامعة القاهرة

رسالة مقدمة إلي كلية الهندسة – جامعة القاهرة
كجزء من متطلبات الحصول على درجة الماجستير
في الهندسة الحيوية الطبية و المنظومات

يعتمد من لجنة الممتحنين:

أ.د / ياسر مصطفى قدح – المشرف الرئيسي

أ.د / عبد الله سيد أحمد محمد – عضو

أ.د / سامية عبد الرازق مشالي - عضو

كلية الهندسة – جامعة القاهرة
الجيزة – جمهورية مصر العربية
2009

ملخص الرسالة

سرطان الرئة يحتل الترتيب الثاني للسرطانات الأكثر تشخيصاً في الولايات المتحدة ويتصدر أكثر أنواع السرطانات المؤدية للوفاة في العالم. بل ان عدد الوفيات الناتجة عنه يفوق عدد الوفيات الناتجة عن سرطانات الثدي والبروستاتا والمستقيم والقولون مجموعة معاً. و مما يبدو ان هذا المعدل في ازدياد مطرد , ففي الولايات المتحدة الأمريكية يتم اكتشاف أكثر من 170000 حالة سنوياً و 45000 حالة في إنجلترا, و تقدر عدد الوفيات السنوية في الولايات المتحدة الأمريكية ب 165000 . أيضا تعتبر الرئة من الأماكن الأكثر عرضه للإصابة بسرطان ثانوي (مثل سرطان الثدي والأمعاء).

كثير من الأبحاث كشفت إمكانية الاكتشاف المبكر لأورام الرئة في مراحله الأولى عن طريق إستعمال الأشعة المقطعية الحلزونية ذات الجرعات المنخفضة . و يعتبر الاكتشاف المبكر لسرطان الرئة أكثر الطرق فاعلية لزيادة فرص النجاة من هذا السرطان . حيث يساعد الاكتشاف المبكر لعمل متابعه دورية لمراقبة تطور درجة السرطان من حيث زيادة عدد الأورام أو زيادة حجمها و من ثم يمكن عمل تقييم دورى لمدى فاعلية العلاج المتبع وتعديله اذا لزم الأمر

ايضاً ساعد تطوير الأشعة المقطعية و ظهور الأشعة المقطعية المتعددة المجسات إلى توفير برامج ودراسات لعمل مسح دورى بخاصة للحالات الأكثر عرضه للإصابة بالسرطان باستعمال صور اشعه اقل سمكاً التي تكشف عن أورام صغيرة جداً, و لكن ادى ذلك الى زيادة فى عدد الصور الناتجة لتصوير الحالة الواحدة, مما ادى في النهاية إلى الاحتياج إلى تقييم كم هائل من الصور من قبل أخصائيو الأشعة. الأمر الذى جعل عملهم أكثر ارهاقاً و فى نفس الوقت عرضة لأن يتغافلوا عن ملاحظة بعض الأورام . و من ثم اصبح الاحتياج الى توافر نظام مميكن يساعد أخصائيو الاشعه امر ملح

. يقدم هذا البحث محاولة لتطوير نظام مميكن لاكتشاف امراض الرئة من خلال الأشعة المقطعية تم فى هذا البحث إتباع ثلاثة محاور مختلفة و مقارنة النتائج التي تم الحصول عليها من كل محور على حدة . فى المحور الأول تم النظر للأورام على اعتبارها أجسام ثنائية البعد يكون كل تواجد للورم فى كل صورة مقطعية (ثنائية البعد) هو تواجد مستقل . فتم اولاً عمل إستخراج للرئة من منطقته الصدر فى الصورة المقطعية و تم بعد ذلك عمل إستخراج للأجسام التى تتشابه فى خصائصها الهندسية مع الأورام من حيث الاستدارة و المساحة والكثافة . ثم تم فى النهاية إستخراج خصائص اخرى لملمس الأجسام و فى النهاية . تم عمل فصل للأجسام بناء على كل الخصائص التى تم إستخراجها

فى المحور الثانى تم بناء مجسمات ثلاثية البعد من الأجسام ثنائية البعد وذلك للاستفادة من الفرق الجوهري الذى يميز الأورام عن بقية الأجسام الاخرى و بخاصة الاوعية الدموية إذ تأخذ الأورام فى الغالب شكلاً كروياً بينما الاوعية الدموية تأخذ شكلاً أسطوانياً فتم إستخراج الأجسام التى تتشابه مع الأورام و أستبعاد الأخرى

فى المحور الثالث تم عمل نوع من الدمج بين ما تم أستخراجه من كلا المحورين السابقين و جائت نتائج البحث لتكشف عن الدقة العالية للمحور الذى أعتمد على البعد الثنائى للأجسام و خاصه تلك لها التى قطر صغير جدا يقترب من 1.2 مم ولكن عدد الأجسام الكاذبة التى تم إستخراجها ايضاً كان كماً هائلاً مما يجعل الأعتدال على الخصائص ثنائية البعد للأجسام فقط امر غير مجدى

اما نتائج استعمال المحور الثانى الذى يعتمد على البناء الثلاثى البعد للأجسام و تقييمهم بناء على ذلك فكان الأكثر قبولاً . إذ جائت الدقة اقل من النظام ثنائى البعد من حيث عدد الأجسام التى أكتشفت و من حيث اقل قطر تم اكتشافه وهو 3 مم . ولكن عدد الأجسام الكاذبة كان أقل بكثير

اما نتائج المحور الثالث فكانت تحمل قيم متوسطة من عدد الأجسام الكاذبة التى تم استخراجها وعدد الاجسام الصحيحة

Acknowledgment

First I thank God for helping me accomplishing this research. His hands were seen more than clear in all the arrangements He made to me.

I would like to deeply thank Prof. Dr. Yasser Kadah for his continuous and great support on both the technical and the psychological levels, his enthusiastic encouragement and guidance during my research period kept motivating me to delve deeply in the research and to endure going through many previous experimental attempts until I finally managed to improve my algorithm to the stage it is in now.

I would also like to thank Dr. Ahmed Salah who taught me the various techniques of image processing during my undergraduate studies, which were of great aid to me in my research, and one of things that impulsed me to go for that research that depends deeply on solid understanding to the image processing principles.

I am really indebted to many of my colleagues who never let me down in any quick technical or administrative advice I sought from them.

I am also indebted to the American University in Cairo (AUC) for providing me with access to its library.

Many thanks goes to my former supervisors at work, Ms Perveen Ali, Mr. Tarek Badawy, Prof. Michael Kagan, and Mr. Maged Tadros, who provided me with all the support I needed to be able to allocate time for my studies.

Lastly, I would like to thank my family (my father, my sister, and my brothers) who endured my long absences from being with them to work on my research. I really apologize for missing being with you.

Abstract

Lung cancer is the second most commonly diagnosed cancer in the United States, and it is the leading cancer related death in the world, with the current fatality rate exceeding that of the next three most common cancers (breast, prostate, and colorectal) combined. At the same time, it appears that the rate has been steadily increasing. Around 170,000 is discovered every year in USA, around 45,000 in UK. Estimated 165,000 deaths every year in US. Also, the lung is a frequent site of metastasis from other cancers (such as breast and bowel) that manifest as pulmonary nodules.

Multiple studies have shown that low-dose screening helical CT scans can detect peripheral lung cancers at an early stage. Detection of suspicious lesions in the early stages of cancer is considered the most effective way to improve survival. Because of the prevalence of benign, stable lung nodules in current and former smokers, once a nodule is detected, serial follow-up scans must be performed to detect growth. The introduction of multi-detector CT (MDCT) scanners has led to lung cancer screening studies with a larger number of thinner slices, resulting in the detection of more nodules. This increase in the number of images per CT examination makes the process of CT interpretation more time consuming and tedious for the radiologist. This can lead to decreased detection sensitivity for nodules caused by reviewer fatigue, apart from the fact that the majority of the screening cases are normal, and hence diagnostic reading errors may be hard to avoid. Therefore, computerized methods for nodule detection to assist the radiologist became important.

In this research, we considered the problem of developing an automated system for detecting the presence of pulmonary nodules in the lung CT, that can be considered as a second reader to the radiologists. The essence of developing a system like that needed to focus on detecting nodules in their early stages, which are the very small nodules that are likely to be overlooked by the radiologists, and hence the focus in this research was in lung CT images that have slice thickness of 2mm.

In developing this system; First, The CT images in DICOM format were read by the system, and then some DICOM information, such the slice vertical position, the horizontal and the vertical resolution, were extracted. The slices then were sorted according the their vertical position to build a stack of images. Second, the lungs were extracted from the thorax to minimize the Region Of Interest (ROI), several segmentation related issues needed to be tackled to overcome the loss of the nodules that are attached to the lung wall (pleural membrane) during the segmentation of the lungs. Third, multiple gray level thresholding, labeling, and

region growing techniques were applied to extract objects that have shape features resemble those of the nodules. Finally, several features (geometric and texture) were extracted, to be used in the classification stage for false positives reduction.

In the false positive (FP) reduction stage, three different approaches were used; First, by building the 3D structure of the objects from the corresponding 2D objects, and then extracting the objects that have nodule like 3D structure (which is almost spherical) were extracted; A further reduction in the FPs was tried by using the Support Vector Machine (SVM) classifier. Second, by doing a cross consultation between the objects extracted from the 2D and the 3D approaches, so that objects that were filtered by the 3D approach, will also be filtered from the 2D list of extracted objects. That finally led to significant reduction in the number of FPs. Third, just by relying on the SVM classifier, treating each 2D object as a separate stand alone object, the 2D features extracted were fed into the classifier for final judgment.

The results showed that, the 2D approach resulted in a very high accuracy in the extraction of the nodules (82.75%) but with huge number of FPs (49/Slice). The combined 2D&3D approaches led to significant reduction in the FPs (17/Slice) but on the expense of decreasing the accuracy of the number of nodules extracted (64.5%). The 3D approach was the best in getting acceptable extraction accuracy (74.49%) while leading to considerable reduction in the number of FPs (1.42/Slice).

The use of a classifier with all the stages led to very high specificity, but to a very poor sensitivity value. The main cause of that is mainly because of the very few number of nodules used in the training compared to the huge number of candidate objects extracted.

Contents

Acknowledgment	ii
Abstract	iii
Contents	v
List of Tables	viii
List of Figures	ix
List of Abbreviations	x
Chapter 1: Introduction	1
1.1 Thesis Objectives.....	2
1.2 Thesis Organization.....	3
Chapter 2: Anatomy of the lung and lung cancer overview	4
2.1 Introduction.....	4
2.2 Anatomy of the lungs and the respiratory system.....	4
2.3 The gas exchange system.....	4
2.4 The breathing process.....	5
2.5 Common lung diseases.....	7
2.6 Lung Cancer.....	8
2.7 Causes of lung cancer.....	9
2.8 Types of lung cancer.....	12
2.9 The signs and symptoms of lung cancer.....	13
2.10 Diagnosing lung cancer.....	14
2.11 Staging of lung cancer.....	18
2.12 Treatment of lung cancer.....	18
2.13 The prognosis (outcome) of lung cancer.....	23
Chapter 3: CT and CAD Systems for Lung Cancer	25
3.1 CT versus Chest X-Ray.....	25
3.2 Lung CT.....	26
3.3 What is Computer Aided Diagnosis (CAD).....	27
3.4 CAD motivation.....	27
3.5 Nodules.....	29
3.5.1 What are nodules.....	29
3.5.2 Shape of the nodule.....	30
3.5.3 Characteristics of nodules in CT.....	30
3.5.4 Types of nodules (Location).....	30
3.5.5 Types of nodules (Texture).....	32
3.6 Role of CAD with pulmonary nodules.....	32
3.7 Summary.....	33

Chapter 4: Literature review.....	35
4.1 Overview of the most relevant systems and method in CAD literature.....	35
4.2 Summary.....	43
Chapter 5: The Developed CAD System.....	44
5.1 Introduction.....	44
5.1.1 The dataset.....	44
5.1.2 The annotation of the radiologists.....	45
5.1.3 Tools and H/W used.....	56
5.2 The algorithm of the 2D approach used in developing a CAD system.....	47
5.2.1 Reading the scan set system.....	48
5.2.2 Segmentation of the Lung.....	48
5.2.2.1 The Segmentation process.....	49
5.2.2.2 Challenges and Post -Segmentation Treatment.....	49
5.2.2.3 Challenges still present in the segmentation operation.....	52
5.2.3 Multiple Gray Level Thresholding.....	53
5.2.3.1 Challenges tackled in the thresholding stage.....	55
5.2.3.2 Challenges still present in the thresholding stage.....	56
5.2.4 Labeling.....	57
5.2.4.1 Challenges tackled in the labeling stage.....	58
5.2.4.2 Challenges still present in the labeling stage.....	58
5.2.5 Shape Features Extraction (Stage1).....	60
5.2.5.1 Challenges tackled in this stage.....	61
5.2.6 Shape Features Filter (Stage1).....	62
5.2.7 Region Growing.....	62
5.2.7.1 Implementation of Region Growing.....	63
5.2.7.2 Challenges in the Region Growing Stage.....	64
5.2.8 Shape Features Extraction (Stage2).....	68
5.2.9 Shape Features Filter (Stage2).....	68
5.2.10 Texture feature extraction.....	68
5.3 Summary.....	68
Chapter 6: False Positives Reduction.....	72
6.1 Introduction	72
6.2 Developing a CAD system using 3D approach.....	72
6.2.1 The algorithm for the 3D approach.....	73
6.2.1.1 3D labeling.....	73
6.2.1.2 3D Region Growing.....	73
6.2.1.3 3D features extraction	75
6.2.1.4 3D filter.....	76
6.2.1.5 Classification.....	77

6.2.2 Major challenges in the 3D approach.....	77
6.2.3 Summary.....	78
6.3 Classification of the extracted 2D candidate objects using SVM Classifier.....	79
6.4 Combined 2D & 3D candidate objects.....	79
Chapter 7: Results.....	80
7.1 Results obtained before using the classifier.....	81
7.1.1 Result of the 2D approach.....	81
7.1.2 Result of combined 2D and 3D approach.....	81
7.1.3 Result of the 3D approach.....	81
7.1.4 Detected Nodules according to the diameter.....	81
7.1.5 Detected Pleural Nodules.....	82
7.1.6 Missed 2D Nodules.....	82
7.1.7 Missed 3D Nodules.....	82
7.2 Results obtained after using the classifier.....	83
7.3 Discussion.....	84
Chapter 8: Conclusions and Future Work.....	85
8.1 Conclusions.....	85
8.2 Future work.....	86
References.....	87

List Of Tables

Table 5.1	The attenuation values of the different tissues in HU.....	45
Table 7.1	All results obtained before using the classifier.....	81
Table 7.2	Detected nodules according to the diameter.....	81
Table 7.3	Detected Pleural Nodules.....	82
Table 7.4	Analysis on the missed 2D nodules.....	82
Table 7.5	Result after applying the SVM.....	83

List Of Figures

Figure 2.1	A:Anatomy of the lungs B:The internal structure of the lung...	5
Figure 2.2	A: The internal air duct system of the lung, B:The alveoli surrounded by blood capillaries, C: Breathing process.....	6
Figure 2.3	Lung Cancer.....	9
Figure 3.1	Several abnormalities in X-Ray images.....	26
Figure 3.2	Lung CT.....	27
Figure 3.3	The change over time (123 days) in the size of a nodule.....	29
Figure 3.4	Several shapes of nodules.....	31
Figure 3.5	Nodules tend to have Gaussian distribution.....	32
Figure 5.1	A CT image of the lung.....	45
Figure 5.2	The annotation of the radiologist.....	45
Figure 5.3	The detailed block diagram of the 2D approach followed in developing a CAD system.....	47
Figure 5.4	The histogram of the thorax image. A: In HU units, B: In gray level values with the best threshold marked.....	48
Figure 5.5	Different stages of lung segmentation.....	50
Figure 5.6	Tackling the challenges in the lung segmentation.....	51
Figure 5.7	Challenges still present in lung segmentation.....	53
Figure 5.8	Thresholding at different gray level values.....	54
Figure 5.9	The core part of nodules extracted at each GLV	55
Figure 5.10	The Labeling technique.....	57
Figure 5.11	Remaining parts of the lung wall were assigned the same label as the pleural nodule.....	58
Figure 5.12	Nodules attached to vessels failed to be extracted during the thresholding operation.....	59
Figure 5.13	Some Shape Features Extracted.....	61
Figure 5.14	Creating an outer frame to be used to label pleural objects.....	62
Figure 5.15	Concept of Region Growing.....	63
Figure 5.16	Inhomogeneous objects after being thresholded.....	66
Figure 5.17	Region growing to irregular shape objects.....	67
Figure 5.18	Inhomogeneous and Irregular shape nodules.....	67
Figure 5.19	Inhomogeneous but have regular shape nodules.....	67
Figure 6.1	The whole algorithm used for developing a CAD system using 2D, 3D, and combined 2D& 3D approaches.....	74
Figure 6.2	A:10-Neighborhood connectivity labeling 3D objects, B:The vascular structure built using 3D labeling.....	75

List Of Abbreviations

ACTH	Adrenocorticotrophic hormone
CAD	Computer Aided Diagnosis
CFTR	Cystic fibrosis transmembrane conductance regulator
COPD	Chronic Obstructive Pulmonary Disease
CT	Computed Tomography
EGFR	Epidermal growth factor receptor
FN	False Negative
FNA	Fine needle aspiration
FP	False Positive
GLV	Gray Level Value
MDCT	Multi-detector CT
NSCLCs	Non-small cell lung cancers
PDT	Photodynamic therapy
RFA	Radiofrequency ablation
SCLCs	Small cell lung cancer
SVM	Support Vector Machine
TN	True Negative
TP	True Positive

Chapter 1

Introduction

Lung cancer is one of the leading causes of death in USA [15] and Europe. Surgery, radiation therapy, and chemotherapy are used in the treatment of lung carcinoma. In spite of that, the five-year survival rate for all stages combined is only 14%.

CT is considered to be the most accurate imaging modality available for early detection and diagnosis of lung cancer. It allows detecting pathological deposits as small as 1mm in diameter. These deposits are called lung nodules. However, the large amount of data per examination makes the interpretation tedious and difficult, leading to a high false-negative rate for detecting small nodules.

Retrospective analysis of CT scans often shows undetected nodules on the initial scans of oncological patients [16].

Image processing and visualization techniques for volumetric CT data sets may improve the radiologist's ability to detect small lung nodules. For example, reconstruction of CT images with narrow interscan spacing and interpretation of images using cine rather than film-based viewing technique [1], have been reported to improve small nodule detection.

Computer-aided diagnosis (CAD) provides a computer output as a “second opinion” in order to assist radiologists in the diagnosis of various diseases on medical images. Currently, a significant research effort is being devoted to the detection and characterization of lung nodules in thin-section computed tomography (CT) images, which represents one of the newest direction of CAD development in thoracic imaging.

One of the most important applications of CAD is the detection and characterization of lung cancer. Some evidence suggests that early detection of lung cancer may allow for timely therapeutic intervention and thus a favorable prognosis for the patients [17,18]. It is reported [19] that the survival rate for early-stage localized cancer (stage I) is 49%.

For nodule detection in chest radiography, CAD schemes have been developed by many investigators. The typical performance of current detection schemes in chest radiography is a 70–75% sensitivity with 1.5–3 false positives per image. For nodule characterization in chest radiography, semi-automated and automated CAD schemes have also been developed by a number of investigators.

1.1 Thesis Objectives

The main two objectives of this thesis are the development of a CAD system to be used in the detection of the pulmonary nodules in the lung CT, and in reaching this objective, the second objective comes to be in assessing the performance of developing such a system relying solely on the 2D features of the objects extracted from the lung CT, and developing a CAD system that makes use of the 3D features of the object, and developing a CAD system that combines both 2D and 3D.

In this research, three approaches were implemented. In the 2D approach, the CT images were read, lungs were segmented, 2D candidate objects were extracted based on the fulfillment of some shape requirements. In the 3D approach, the extracted 2D objects were aligned to build volumetric 3D objects. Again, the features of these objects were extracted, and those objects that have features resemble those of nodules were promoted to be considered candidate objects. In the combined 2D and 3D approach, objects were filtered twice, according to their 2D features by 2D filters, and according to their 3D features by 3D filters.

As a primary step in all the approaches, lung segmentation was done using thresholding technique followed by some morphological operations to smooth the outer frame of the extracted lungs preventing loss of nodules attached to the pleural membrane.

In all approaches, the objects were extracted based on the intensity difference nodules are known to have compared to that of the surrounding tissues and vessels. Thresholding technique was used to extract the objects. Because of the Gaussian texture structure nodules are known to have, region growing technique was implemented.

Finally, the results obtained from all the approaches were assessed in terms of the number of nodules extracted, the number of False Positives (FPs) extracted, the diameter of the nodules extracted.

An attempt to use Support Vector Machine (SVM) classifier was also tried, to see the effectiveness of getting further reduction in the number of FPs.

1.2 Thesis Organization

The thesis consists of eight chapters organized as following:

- **Chapter 1** gives an introduction on the thesis and its objectives.
- **Chapter 2** gives a medical overview on the lungs, its function, its common diseases, the lung cancer, it causes, its treatment.
- **Chapter 3** discusses the CAD systems, and their motivations, particularly for lung cancer.
- **Chapter 4** gives an insight on the literature review of CAD system developed for lung cancer in CT
- **Chapter 5** covers the algorithm and the techniques used in this research for developing a CAD system
- **Chapter 6** mentions the techniques used to make reduction in the number of False Positives (FPs)
- **Chapter 7** discusses the final results obtained from applying the three different approaches, the accuracy of each technique, analyze the nodules detected and nodule missed in terms of their location (pleural, parenchymal, or juxta-vascular), and in terms of their diameter size.
- **Chapter 8** concludes the thesis by giving some ideas about proposed future work to extend the work done in this research and to tackle the challenges that are still present.

Chapter 2:

Anatomy of the lung and lung cancer overview

2.1 Introduction

In this chapter we will have a quick overview on the anatomy of the lungs, which is very important to understand many of the structures we will see while interpreting the lung CT. We will also have a quick look on the respiratory system, which will help in understanding the internal structure of the lung. Then, we will cover some of the most common lung diseases. Lastly, we will talk about lung cancer, the core disease we are working in developing a system to diagnose.

2.2 Anatomy of the lungs and the respiratory system

As we can see in Figure 2.1-A, the right lung consists of three lobes, while the left has only two lobes. Each lung has two layers of pleura, which is filled in with a fluid called pleural fluid produced by the pleurae to lubricate the surfaces of the pleura.

The main entrance of the trachea into the lung is called the primary bronchus (Figure 2.1-B), which is subdivided into secondary bronchi, then to tertiary bronchi, and so on. The last end of the wind track are the alveoli, which are the final branchings of the respiratory tree and act as the primary gas exchange units of the lung.

2.3 The gas exchange system

The alveolar sacs (Fig 2.2-A) consist of many alveoli and are composed of a single layer of epithelial tissue. There are about 300 million alveoli in the adult lung. The alveoli are considered the functional unit of the lung. In the spaces between the alveoli of the lungs is elastic connective tissue which is important for exhalation. The alveoli are surrounded by a network of pulmonary capillaries (Fig. 2.2-B). These capillaries are made of simple squamous epithelium, therefore there are only two cells between the air in the alveoli and the blood in the pulmonary capillaries which permits efficient diffusion of gases. It is through the moist walls of both the alveoli and the capillaries that rapid exchange of CO₂ and O₂ occurs. Carbon dioxide diffuses from the red blood cells through the capillary walls, into the alveoli. CO₂ leaves the alveoli, exhaled through the nose and mouth. The opposite process occurs with O₂, which diffuses from the alveoli into the capillaries, and from there into the red blood cells.

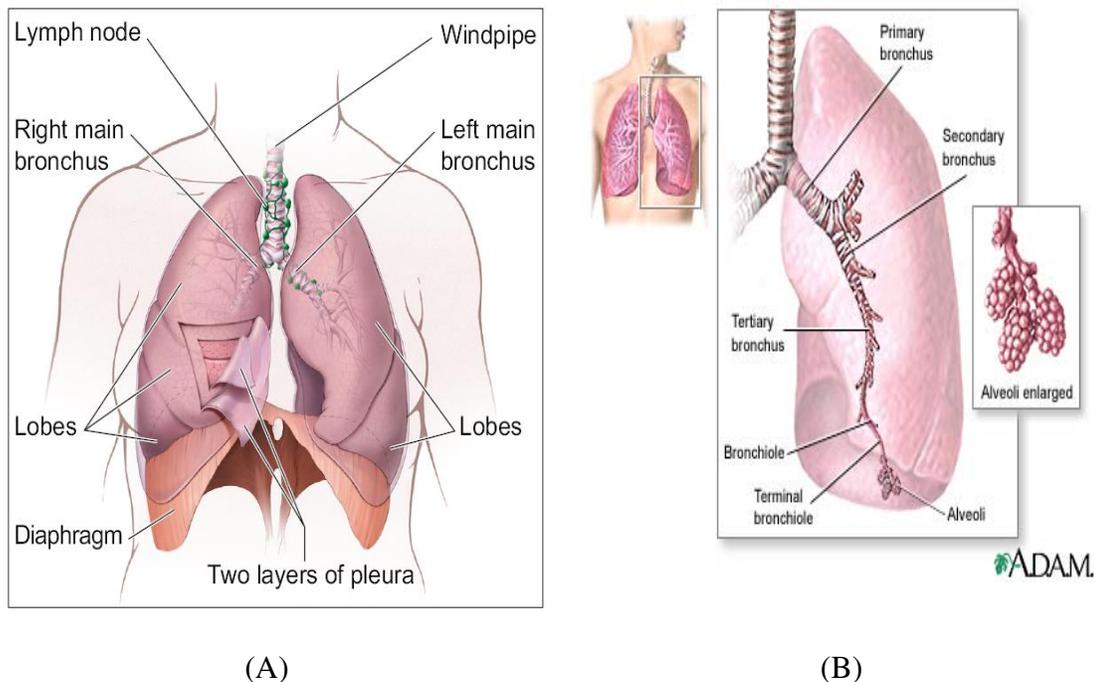


Fig. 2.1 A:Anatomy of the lungs B:The internal structure of the lung

2.4 The breathing process

The diaphragm divides the body cavity into the

- **abdominal cavity**, which contains the viscera (e.g., stomach and intestines) and the
- **thoracic cavity**, which contains the heart and lungs.

The inner surface of the thoracic cavity and the outer surface of the lungs are lined with **pleural membranes** which adhere to each other. If air is introduced between them, the adhesion is broken and the natural elasticity of the lung causes it to collapse. This can occur from trauma. And it is sometimes induced deliberately to allow the lung to rest. In either case, reinflation occurs as the air is gradually absorbed by the tissues.

Because of this adhesion, any action that increases the volume of the thoracic cavity causes the lungs to expand, drawing air into them.

Inhalation is also called inspiration and is the movement of air into the lungs (Fig. 2.3-C). Changes in the shape and size of the thoracic cavity result in changes in the air pressure within that cavity and in the lungs. The difference in air pressure causes the movement of air into and out of the lungs. Air moves from an area where pressure is high to area where pressure is lower. Respiratory muscles are responsible for changes in the shape of the thoracic cavity that cause the air

movements involved in breathing.

- During inspiration (inhaling),
 - The external intercostal muscles contract, lifting the ribs up and out.
 - The diaphragm contracts, drawing it down .
- During expiration (exhaling), these processes are reversed and the natural elasticity of the lungs returns them to their normal volume. At rest, we breath 15-18 times a minute exchanging about 500 ml of air.
- In more vigorous expiration,
 - The internal intercostal muscles draw the ribs down and inward
 - The wall of the abdomen contracts pushing the stomach and liver upward.

Under these conditions, an average adult male can flush his lungs with about 4 liters of air at each breath. This is called the **vital capacity**. Even with maximum expiration, about 1200 ml of **residual air** remain.

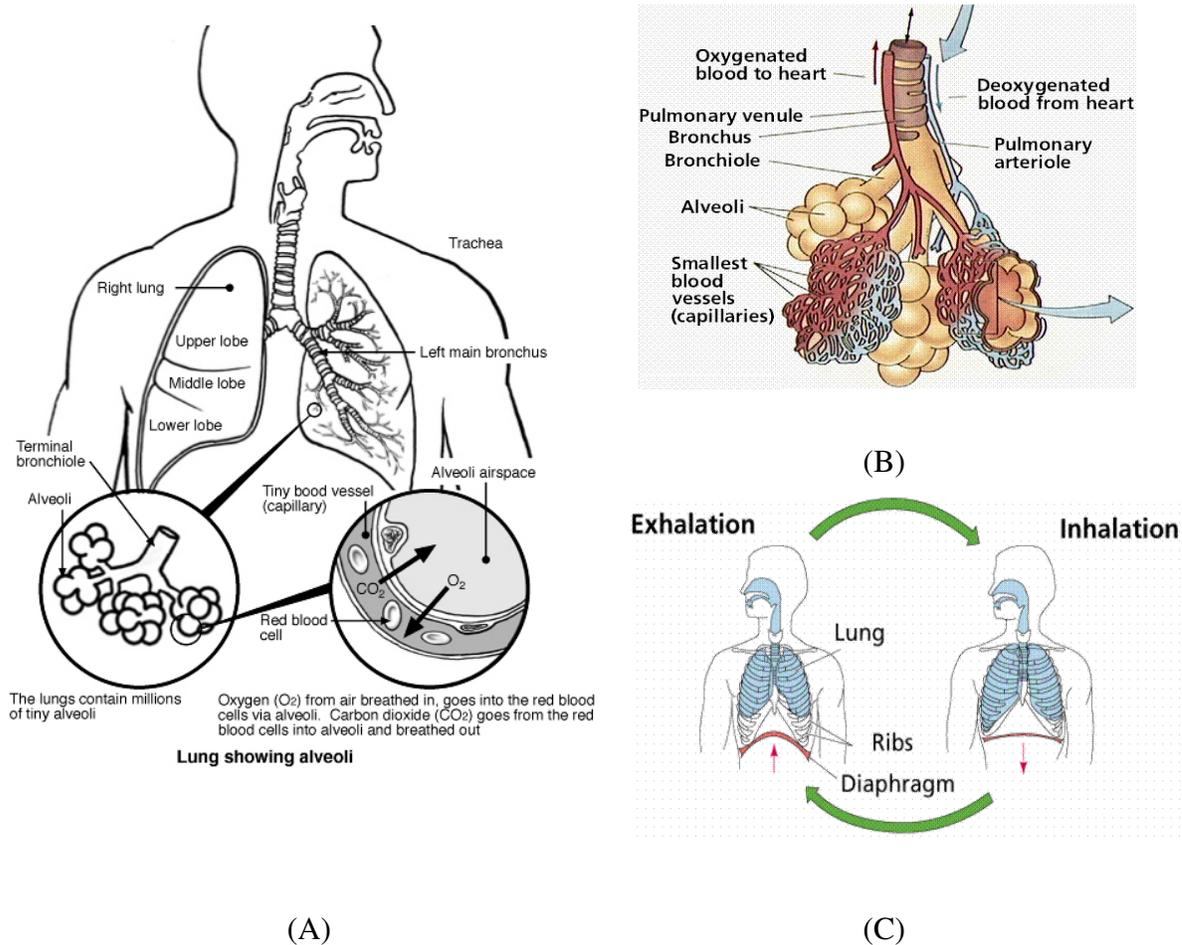


Fig. 2.2 A: The internal air duct system of the lung, B: The alveoli surrounded by blood capillaries, C: Breathing process

2.5 Common lung diseases

1. Pneumonia

Pneumonia is an infection of the alveoli. It can be caused by many kinds of both bacteria (e.g., *Streptococcus pneumoniae*) and viruses. Tissue fluids accumulate in the alveoli reducing the surface area exposed to air. If enough alveoli are affected, the patient may need supplemental oxygen.

2. Asthma

In asthma, periodic constriction of the bronchi and bronchioles makes it more difficult to breathe in and, especially, out. Attacks of asthma can be:

- triggered by airborne irritants such as chemical fumes and cigarette smoke
- airborne particles to which the patient is **allergic**.

3. Emphysema

In this disorder, the delicate walls of the alveoli break down, reducing the gas exchange area of the lungs. The condition develops slowly and is seldom a direct cause of death. However, the gradual loss of gas exchange area forces the heart to pump ever-larger volumes of blood to the lungs in order to satisfy the body's needs. The added strain can lead to heart failure.

The immediate cause of emphysema seems to be the release of proteolytic enzymes as part of the inflammatory process that follows irritation of the lungs. Most people avoid this kind of damage during infections, etc. by producing an enzyme inhibitor (a serpin) called **alpha-1 antitrypsin**. Those rare people who inherit two defective genes for alpha-1 antitrypsin are particularly susceptible to developing emphysema.

4. Chronic Bronchitis

Any irritant reaching the bronchi and bronchioles will stimulate an increased secretion of mucus. In chronic bronchitis the air passages become clogged with mucus, and this leads to a persistent cough. Chronic bronchitis is usually associated with cigarette smoking.

5. Chronic Obstructive Pulmonary Disease (COPD)

Irritation of the lungs can lead to asthma, emphysema, and chronic bronchitis. And, in fact, many people develop two or three of these together. This constellation is known as **chronic obstructive pulmonary disease (COPD)**.

Among the causes of **COPD** are

- cigarette smoke (often)
- cystic fibrosis (rare)

6. Cystic fibrosis

It is a genetic disorder caused by inheriting two defective genes for the **cystic fibrosis transmembrane conductance regulator (CFTR)**, a transmembrane protein needed for the transport of Cl^- ions out of the epithelial cells of the lung thus enabling water to follow by osmosis. Diminished CFTR function reduces the water content of the fluid in the lungs making it more viscous and difficult for the ciliated cells to move it up out of the lungs. The accumulation of mucus plugs the airways interfering with breathing and causing a persistent cough. Cystic fibrosis is the most common inherited disease in the U.S. white population.

2.6 Lung Cancer [12]

Tumors can be benign or malignant; when we speak of "cancer," we refer to those tumors that are considered malignant. Benign tumors can usually be removed and do not spread to other parts of the body. Malignant tumors, on the other hand, grow aggressively and invade other tissues of the body, allowing entry of tumor cells into the bloodstream or lymphatic system and then to other sites in the body. This process of spread is termed **metastasis**; the areas of tumor growth at these distant sites are called metastases. Since lung cancer tends to spread or metastasize very early in its course, it is a very life-threatening cancer and one of the most difficult cancers to treat. While lung cancer can spread to any organ in the body, certain organs -- particularly the adrenal glands, liver, brain, and bone -- are the most common sites for lung-cancer metastasis.

The lung is also a very common site for metastasis from tumors in other parts of the body. Tumor metastases are made up of the same type of cells as the original, or primary, tumor. For example, if prostate cancer spreads via the bloodstream to the lungs, it is metastatic prostate cancer in the lung and is not lung cancer.

Lung cancer is the most common cancer and the most common cause of cancer deaths in U.S. males. Although more women develop breast cancer than lung cancer, since 1987 U.S. women have been dying in larger numbers from lung cancer than from breast cancer.

Lung cancer, like all cancer, is an uncontrolled proliferation of cells (Fig. 2.3). There are several forms of lung cancer, but the most common (and most rapidly increasing) types are those involving the epithelial cells lining the bronchi and bronchioles.

Ordinarily, the lining of these airways consists of two layers of cells.

Chronic exposure to irritants

- causes the number of layers to increase. This is especially apt to happen at forks where the bronchioles branch.
- The ciliated and mucus-secreting cells disappear and are replaced by a disorganized mass of cells with abnormal nuclei.
- If the process continues, the growing mass penetrates the underlying basement membrane.
- At this point, malignant cells can break away and be carried in lymph and blood to other parts of the body where they may lodge and continue to proliferate.
- It is this metastasis of the primary tumor that eventually kills the patient.

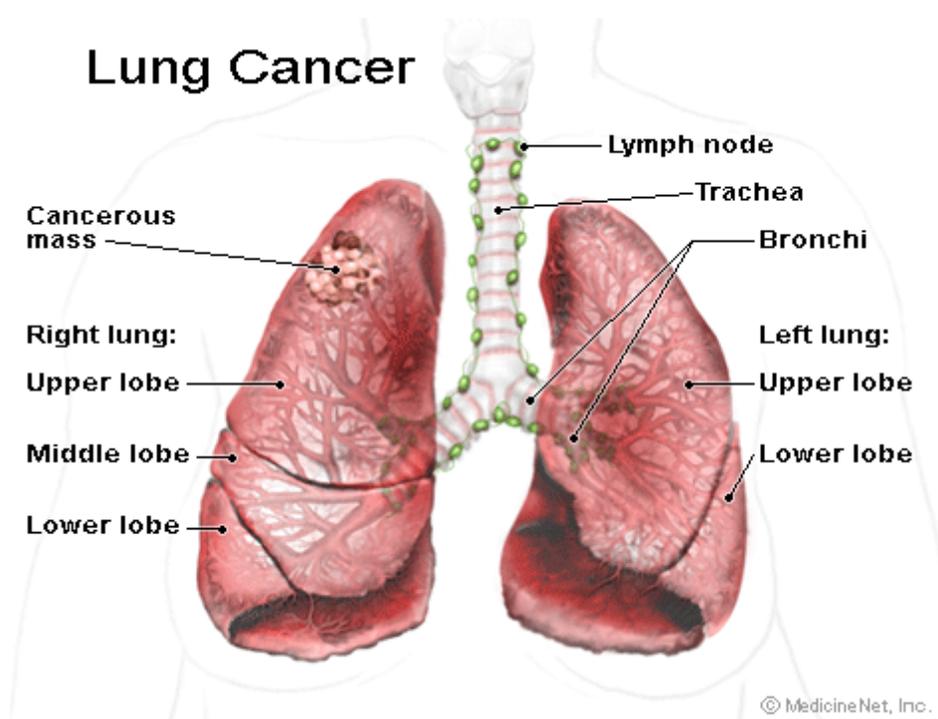


Fig. 2.3 Lung Cancer

2.7 Causes of lung cancer [12]

1-Smoking

The incidence of lung cancer is strongly correlated with cigarette smoking, with about 90% of lung cancers arising as a result of tobacco use. The risk of lung cancer increases with the number of cigarettes smoked over time; doctors refer to this risk in terms of pack-years of smoking history (the number of packs of

cigarettes smoked per day multiplied by the number of years smoked). For example, a person who has smoked two packs of cigarettes per day for 10 years has a 20 pack-year smoking history. While the risk of lung cancer is increased with even a 10-pack-year smoking history, those with 30-pack-year histories or more are considered to have the greatest risk for the development of lung cancer. Among those who smoke two or more packs of cigarettes per day, one in seven will die of lung cancer.

Pipe and cigar smoking can also cause lung cancer, although the risk is not as high as with cigarette smoking. While someone who smokes one pack of cigarettes per day has a risk for the development of lung cancer that is 25 times higher than a nonsmoker, pipe and cigar smokers have a risk of lung cancer that is about five times that of a nonsmoker.

Tobacco smoke contains over 4,000 chemical compounds, many of which have been shown to be cancer-causing, or carcinogenic. The two primary carcinogens in tobacco smoke are chemicals known as nitrosamines and polycyclic aromatic hydrocarbons. The risk of developing lung cancer decreases each year following smoking cessation as normal cells grow and replace damaged cells in the lung. In former smokers, the risk of developing lung cancer begins to approach that of a nonsmoker about 15 years after cessation of smoking.

2-Passive smoking

Passive smoking, or the inhalation of tobacco smoke from other smokers sharing living or working quarters, is also an established risk factor for the development of lung cancer. Research has shown that nonsmokers who reside with a smoker have a 24% increase in risk for developing lung cancer when compared with other nonsmokers. An estimated 3,000 lung cancer deaths occur each year in the U.S. that are attributable to passive smoking.

3-Asbestos fibers

Asbestos fibers are silicate fibers that can persist for a lifetime in lung tissue following exposure to asbestos. The workplace is a common source of exposure to asbestos fibers, as asbestos was widely used in the past as both thermal and acoustic insulation. Today, asbestos use is limited or banned in many countries, including the U.S. Both lung cancer and mesothelioma (cancer of the pleura of the lung as well as of the lining of the abdominal cavity called the peritoneum) are associated with exposure to asbestos. Cigarette smoking drastically increases the chance of developing an asbestos-related lung cancer in exposed workers. Asbestos workers who do not smoke have a fivefold greater risk of developing lung cancer than nonsmokers, and those asbestos workers who smoke have a risk that is 50 to 90 times greater than nonsmokers.

4-Radon gas

Radon gas is a natural, chemically inert gas that is a natural decay product of uranium. Uranium decays to form products, including radon, that emit a type of ionizing radiation. Radon gas is a known cause of lung cancer, with an estimated 12% of lung-cancer deaths attributable to radon gas, or 15,000-22,000 lung-cancer-related deaths annually in the U.S., making radon the second leading cause of lung cancer in the U.S. As with asbestos exposure, concomitant smoking greatly increases the risk of lung cancer with radon exposure. Radon gas can travel up through soil and enter homes through gaps in the foundation, pipes, drains, or other openings. The U.S. Environmental Protection Agency estimates that one out of every 15 homes in the U.S. contains dangerous levels of radon gas. Radon gas is invisible and odorless, but it can be detected with simple test kits.

5-Familial predisposition

While the majority of lung cancers are associated with tobacco smoking, the fact that not all smokers eventually develop lung cancer suggests that other factors, such as individual genetic susceptibility, may play a role in the causation of lung cancer. Numerous studies have shown that lung cancer is more likely to occur in both smoking and nonsmoking relatives of those who have had lung cancer than in the general population. Recent research has localized a region on the long (q) arm of human chromosome number 6 that is likely to contain a gene that confers an increased susceptibility to the development of lung cancer in smokers.

6-Lung diseases

The presence of certain diseases of the lung, notably chronic obstructive pulmonary disease (COPD), is associated with an increased risk (four to six times the risk of a nonsmoker) for the development of lung cancer even after the effects of concomitant cigarette smoking are excluded.

7-Prior history of lung cancer

Survivors of lung cancer have a greater risk than the general population of developing a second lung cancer. Survivors of non-small cell lung cancers (NSCLCs) have an additive risk of 1%-2% per year for developing a second lung cancer. In survivors of small cell lung cancers (SCLCs), the risk for development of second cancers approaches 6% per year.

8-Air pollution

Air pollution from vehicles, industry, and power plants can raise the likelihood of developing lung cancer in exposed individuals. Up to 1% of lung cancer deaths are attributable to breathing polluted air, and experts believe that prolonged exposure to highly polluted air can carry a risk for the development of lung cancer similar to

that of passive smoking.

2.8 Types of lung cancer [12]

Lung cancers, also known as bronchogenic carcinomas (carcinoma is another term for cancer), are broadly classified into two types: small cell lung cancers (SCLC) and non-small cell lung cancers (NSCLC). This classification is based upon the microscopic appearance of the tumor cells themselves. 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.

1-SCLC comprise about 20% of lung cancers and are the most aggressive and rapidly growing of all lung cancers. SCLC are strongly related to cigarette smoking, with only 1% of these tumors occurring in nonsmokers. SCLC metastasize rapidly to many sites within the body and are most often discovered after they have spread extensively. Referring to a specific cell appearance often seen when examining samples of SCLC under the microscope, these cancers are sometimes called oat cell carcinomas.

2-NSCLC are the most common lung cancers, accounting for about 80% of all lung cancers. NSCLC can be divided into three main types that are named based upon the type of cells found in the tumor:

a-Adenocarcinomas are the most commonly seen type of NSCLC in the U.S. and comprise up to 50% of NSCLC. While adenocarcinomas are associated with smoking like other lung cancers, this type is observed as well in nonsmokers who develop lung cancer. Most adenocarcinomas arise in the outer, or peripheral, areas of the lungs. **Bronchioloalveolar carcinoma** is a subtype of adenocarcinoma that frequently develops at multiple sites in the lungs and spreads along the preexisting alveolar walls.

b-Squamous cell carcinomas were formerly more common than adenocarcinomas; at present, they account for about 30% of NSCLC. Also known as epidermoid carcinomas, squamous cell cancers arise most frequently in the central chest area in the bronchi.

c-Large cell carcinomas, sometimes referred to as undifferentiated carcinomas, are the least common type of NSCLC.

Mixtures of different types of NSCLC are also seen.

Other types of cancers can arise in the lung; these types are much less common than NSCLC and SCLC and together comprise only 5%-10% of lung cancers:

3-Bronchial carcinoids account for up to 5% of lung cancers. These tumors are

generally small (3-4 cm or less) when diagnosed and occur most commonly in people under 40 years of age. Unrelated to cigarette smoking, carcinoid tumors can metastasize, and a small proportion of these tumors secrete hormone-like substances that may cause specific symptoms related to the hormone being produced. Carcinoids generally grow and spread more slowly than bronchogenic cancers, and many are detected early enough to be amenable to surgical resection.

4-Cancers of supporting lung tissue such as smooth muscle, blood vessels, or cells involved in the immune response can rarely occur in the lung.

As discussed previously, metastatic cancers from other primary tumors in the body are often found in the lung. Tumors from anywhere in the body may spread to the lungs either through the bloodstream, through the lymphatic system, or directly from nearby organs. Metastatic tumors are most often multiple, scattered throughout the lung, and concentrated in the peripheral rather than central areas of the lung.

2.9 The signs and symptoms of lung cancer [12]

Symptoms of lung cancer are varied depending upon where and how widespread the tumor is. Warning signs of lung cancer are not always present or easy to identify. A person with lung cancer may have the following kinds of symptoms:

1-No symptoms:

In up to 25% of people who get lung cancer, the cancer is first discovered on a routine chest X-ray or CT scan as a solitary small mass sometimes called a coin lesion, since on a two-dimensional X-ray or CT scan, the round tumor looks like a coin. These patients with small, single masses often report no symptoms at the time the cancer is discovered.

2-Symptoms related to the cancer:

The growth of the cancer and invasion of lung tissues and surrounding tissue may interfere with breathing, leading to symptoms such as cough, shortness of breath, wheezing, chest pain, and coughing up blood (hemoptysis). If the cancer has invaded nerves, for example, it may cause shoulder pain that travels down the outside of the arm (called Pancoast's syndrome) or paralysis of the vocal cords leading to hoarseness. Invasion of the esophagus may lead to difficulty swallowing (dysphagia). If a large airway is obstructed, collapse of a portion of the lung may occur and cause infections (abscesses, pneumonia) in the obstructed area.

3-Symptoms related to metastasis:

Lung cancer that has spread to the bones may produce excruciating pain at the sites of bone involvement. Cancer that has spread to the brain may cause a number of neurologic symptoms that may include blurred vision, headaches,

seizures, or symptoms of stroke such as weakness or loss of sensation in parts of the body.

4-Paraneoplastic symptoms:

Lung cancers frequently are accompanied by symptoms that result from production of hormone-like substances by the tumor cells. These paraneoplastic syndromes occur most commonly with SCLC but may be seen with any tumor type. A common paraneoplastic syndrome associated with SCLC is the production of a hormone called adrenocorticotrophic hormone (ACTH) by the cancer cells, leading to oversecretion of the hormone cortisol by the adrenal glands (Cushing's syndrome). The most frequent paraneoplastic syndrome seen with NSCLC is the production of a substance similar to parathyroid hormone, resulting in elevated levels of calcium in the bloodstream.

5-Nonspecific symptoms:

Nonspecific symptoms seen with many cancers, including lung cancers, include weight loss, weakness, and fatigue. Psychological symptoms such as depression and mood changes are also common.

2.10 Diagnosing lung cancer [12]

Doctors use a wide range of diagnostic procedures and tests to diagnose lung cancer. These include...

- The **history and physical examination** may reveal the presence of symptoms or signs that are suspicious for lung cancer. In addition to asking about symptoms and risk factors for cancer development such as smoking, doctors may detect signs of breathing difficulties, airway obstruction, or infections in the lungs. Cyanosis, a bluish color of the skin and the mucous membranes due to insufficient oxygen in the blood, that suggests compromised function of the lung. Likewise, changes in the tissue of the nail beds, known as clubbing, may also indicate lung disease.
- The **chest X-ray** is the most common first diagnostic step when any new symptoms of lung cancer are present. The chest X-ray procedure often involves a view from the back to the front of the chest as well as a view from the side. Like any X-ray procedure, chest X-rays expose the patient briefly to a minimum amount of radiation. Chest X-rays may reveal suspicious areas in the lungs but are unable to determine if these areas are cancerous. In particular, calcified nodules in the lungs or benign tumors called hamartomas may be identified on a chest X-ray and mimic lung cancer.

- **CT (computerized axial tomography scan, or CAT scan) scans** may be performed on the chest, abdomen, and/or brain to examine for both metastatic and primary tumor. A CT scan of the chest may be ordered when X-rays do not show an abnormality or do not yield sufficient information about the extent or location of a tumor. One advantage of CT scans is that they are more sensitive than standard chest X-rays in the detection of lung nodules. Sometimes intravenous contrast material is given prior to the procedure to help delineate the organs and their positions. A CT scan exposes the patient to a minimal amount of radiation. The most common side effect is an adverse reaction to intravenous contrast material that may have been given prior to the procedure. There may be resulting itching, a rash, or hives that generally disappear rather quickly. Severe anaphylactic reactions (life-threatening allergic reactions with breathing difficulties) to contrast material are rare. CT scans of the abdomen may identify metastatic cancer in the liver or adrenal glands, and CT scans of the head may be ordered to reveal the presence and extent of metastatic cancer in the brain.
- A technique called a **low-dose helical CT scan** (or spiral CT scan) is sometimes used in screening for lung cancers. This procedure requires a special type of CT scanner and has been shown to be an effective tool for the identification of small lung cancers in smokers and former smokers.
- Magnetic resonance imaging (**MRI**) scans may be appropriate when precise detail about a tumor's location is required. The MRI technique uses magnetism, radio waves, and a computer to produce images of body structures. As with CT scanning, the patient is placed on a moveable bed which is inserted into the MRI scanner. There are no known side effects of MRI scanning, and there is no exposure to radiation. The image and resolution produced by MRI is quite detailed and can detect tiny changes of structures within the body. People with heart pacemakers, metal implants, artificial heart valves, and other surgically implanted structures cannot be scanned with an MRI because of the risk that the magnet may move the metal parts of these structures.
- Positron emission tomography (**PET**) scanning is a specialized imaging technique that uses short-lived radioactive drugs to produce three-dimensional colored images of those substances in the tissues within the body. While CT scans and MRI scans look at anatomical structures, PET scans measure metabolic activity and functioning of tissue. PET scans can determine whether a tumor tissue is actively growing and can aid in determining the type of cells within a particular tumor. In PET scanning, the patient receives a short half-lived radioactive drug and receives approximately the amount of radiation exposure as two chest X-rays. The

drug discharges particles known as positrons from wherever they are taken up and used in the body. As the positrons encounter electrons within the body, a reaction producing gamma rays occurs. A scanner records these gamma rays and maps the area where the radioactive drug is located. For example, combining glucose (a common energy source in the body) with a radioactive substance will show where glucose is rapidly being used, for example, in a growing tumor.

- **Bone scans** are used to create images of bones on a computer screen or on film. Doctors may order a bone scan to determine whether a lung cancer has metastasized to the bones. In a bone scan, a small amount of radioactive material is injected into the bloodstream and collects in the bones, especially in abnormal areas such as those involved by metastatic tumors. The radioactive material is detected by a scanner, and the image of the bones is recorded on a special film for permanent viewing.
- **Sputum cytology:** The diagnosis of lung cancer always requires confirmation of malignant cells by a pathologist, even when symptoms and X-ray studies are suspicious for lung cancer. The simplest method to establish the diagnosis is the examination of sputum under a microscope. If a tumor is centrally located and has invaded the airways, this procedure, known as a sputum cytology examination, may allow visualization of tumor cells for diagnosis. This is the most risk-free and inexpensive tissue diagnostic procedure, but its value is limited since tumor cells will not always be present in sputum even if a cancer is present. Also, noncancerous cells may occasionally undergo changes in reaction to inflammation or injury that makes them look like cancer cells.
- **Bronchoscopy:** Examination of the airways by bronchoscopy (visualizing the airways through a thin, fiberoptic probe inserted through the nose or mouth) may reveal areas of tumor that can be sampled (biopsied) for diagnosis by a pathologist. A tumor in the central areas of the lung or arising from the larger airways is accessible to sampling using this technique. Bronchoscopy may be performed using a rigid or a flexible, fiberoptic bronchoscope and can be performed in a same-day outpatient bronchoscopy suite, an operating room, or on a hospital ward. The procedure can be uncomfortable, and it requires sedation or anesthesia. While bronchoscopy is relatively safe, it must be carried out by a lung specialist (pulmonologist or surgeon) experienced in the procedure. When a tumor is visualized and adequately sampled, an accurate cancer diagnosis usually is possible. Some patients may cough up dark-brown blood for one to two days after the procedure. More serious but rare complications include a greater amount of bleeding, decreased levels of oxygen in the

blood, and heart arrhythmias as well as complications from sedative medications and anesthesia.

- **Needle biopsy:** Fine needle aspiration (FNA) through the skin, most commonly performed with radiological imaging for guidance, may be useful in retrieving cells for diagnosis from tumor nodules in the lungs. Needle biopsies are particularly useful when the lung tumor is peripherally located in the lung and not accessible to sampling by bronchoscopy. A small amount of local anesthetic is given prior to insertion of a thin needle through the chest wall into the abnormal area in the lung. Cells are suctioned into the syringe and are examined under the microscope for tumor cells. This procedure is generally accurate when the tissue from the affected area is adequately sampled, but in some cases, adjacent or uninvolved areas of the lung may be mistakenly sampled. A small risk (3%-5%) of an air leak from the lungs (called a pneumothorax, which can easily be treated) accompanies the procedure.
- **Thoracentesis:** Sometimes lung cancers involve the lining tissue of the lungs (pleura) and lead to an accumulation of fluid in the space between the lungs and chest wall (called a pleural effusion). Aspiration of a sample of this fluid with a thin needle (thoracentesis) may reveal the cancer cells and establish the diagnosis. As with the needle biopsy, a small risk of a pneumothorax is associated with this procedure.
- **Major surgical procedures:** If none of the aforementioned methods yields a diagnosis, surgical methods must be employed to obtain tumor tissue for diagnosis. These can include mediastinoscopy (examining the chest cavity between the lungs through a surgically inserted probe with biopsy of tumor masses or lymph nodes that may contain metastases) or thoracotomy (surgical opening of the chest wall for removal or biopsy of a tumor). With a thoracotomy, it is rare to be able to completely remove a lung cancer, and both mediastinoscopy and thoracotomy carry the risks of major surgical procedures (complications such as bleeding, infection, and risks from anesthesia and medications). These procedures are performed in an operating room, and the patient must be hospitalized.
- **Blood tests:** While routine blood tests alone cannot diagnose lung cancer, they may reveal biochemical or metabolic abnormalities in the body that accompany cancer. For example, elevated levels of calcium or of the enzyme alkaline phosphatase may accompany cancer that is metastatic to the bones. Likewise, elevated levels of certain enzymes normally present within liver cells, including aspartate aminotransferase (AST or SGOT) and alanine aminotransferase (ALT or SGPT), signal liver damage, possibly

through the presence of metastatic tumor

2.11 Staging of lung cancer [12]

The stage of a cancer refers to the extent to which a cancer has spread in the body. Staging involves both evaluation of a cancer's size as well as the presence or absence of metastases in the lymph nodes or in other organs. Staging is important for determining how a particular cancer should be treated, since lung-cancer therapies are geared toward specific stages. Staging of a cancer is also critical in estimating the prognosis of a given patient, with higher-stage cancers generally having a worse prognosis than lower-stage cancers.

Doctors may use several tests to accurately stage a lung cancer, including laboratory (blood chemistry) tests, X-rays, CT scans, bone scans, and MRI scans. Abnormal blood chemistry tests may signal the presence of metastases in bone or liver, and radiological procedures can document the size of a cancer as well as possible spread to other organs.

NSCLC are assigned a stage from I to IV in order of severity:

- In stage I, the cancer is confined to the lung.
- In stages II and III, the cancer is confined to the chest (with larger and more invasive tumors classified as stage III).
- Stage IV cancer has spread from the chest to other parts of the body.

SCLC are staged using a two-tiered system:

- Limited-stage SCLC refers to cancer that is confined to its area of origin in the chest.
- In extensive-stage SCLC, the cancer has spread beyond the chest to other parts of the body.

2.12 Treatment of lung cancer [12]

Treatment for lung cancer can involve surgical removal of the cancer, chemotherapy, or radiation therapy, as well as combinations of these treatments. The decision about which treatments will be appropriate for a given individual must take into account the localization and extent of the tumor as well as the overall health status of the patient.

As with other cancers, therapy may be prescribed that is intended to be curative (removal or eradication of a cancer) or palliative (measures that are unable to cure a cancer but can reduce pain and suffering). More than one type of therapy may be prescribed. In such cases, the therapy that is added to enhance the effects of the

primary therapy is referred to as adjuvant therapy. An example of adjuvant therapy is chemotherapy or radiotherapy administered after surgical removal of a tumor in order to be certain that all tumor cells are killed.

1-Surgery

Surgical removal of the tumor is generally performed for limited-stage (stage I or sometimes stage II) NSCLC and is the treatment of choice for cancer that has not spread beyond the lung. About 10%-35% of lung cancers can be removed surgically, but removal does not always result in a cure, since the tumors may already have spread and can recur at a later time. Among people who have an isolated, slow-growing lung cancer removed, 25%-40% are still alive five years after diagnosis. Surgery may not be possible if the cancer is too close to the trachea or if the person has other serious conditions (such as severe heart or lung disease) that would limit their ability to tolerate an operation. Surgery is less often performed with SCLC because these tumors are less likely to be localized to one area that can be removed.

The surgical procedure chosen depends upon the size and location of the tumor. Surgeons must open the chest wall and may perform a wedge resection of the lung (removal of a portion of one lobe), a lobectomy (removal of one lobe), or a pneumonectomy (removal of an entire lung). Sometimes lymph nodes in the region of the lungs are also removed (lymphadenectomy). Surgery for lung cancer is a major surgical procedure that requires general anesthesia, hospitalization, and follow-up care for weeks to months. Following the surgical procedure, patients may experience difficulty breathing, shortness of breath, pain, and weakness. The risks of surgery include complications due to bleeding, infection, and complications of general anesthesia.

2-Radiation

Radiation therapy may be employed as a treatment for both NSCLC and SCLC. Radiation therapy uses high-energy X-rays or other types of radiation to kill dividing cancer cells. Radiation therapy may be given as curative therapy, palliative therapy (using lower doses of radiation than with curative regimens), or as adjuvant therapy in combination with surgery or chemotherapy. The radiation is either delivered externally, by using a machine that directs radiation toward the cancer, or internally through placement of radioactive substances in sealed containers within the area of the body where the tumor is localized. Brachytherapy is a term used to describe the use of a small pellet of radioactive material placed directly into the cancer or into the airway next to the cancer. This is usually done

through a bronchoscope.

Radiation therapy can be given if a person refuses surgery, if a tumor has spread to areas such as the lymph nodes or trachea making surgical removal impossible, or if a person has other conditions that make them too ill to undergo major surgery. Radiation therapy generally only shrinks a tumor or limits its growth when given as a sole therapy, yet in 10%-15% of people it leads to long-term remission and palliation of the cancer. Combining radiation therapy with chemotherapy can further increase the chances of survival when chemotherapy is administered. External radiation therapy can generally be carried out on an outpatient basis, while internal radiation therapy requires a brief hospitalization. A person who has severe lung disease in addition to a lung cancer may not be able to receive radiotherapy to the lung. A type of external radiation therapy called the "gamma knife" is sometimes used to treat single brain metastases. In this procedure, multiple beams of radiation are focused on the tumor over a few minutes to hours while the head is held in place by a rigid frame.

For external radiation therapy, a process called simulation is necessary prior to treatment. Using CT scans, computers, and precise measurements, simulation maps out the exact location where the radiation will be delivered, called the treatment field or port. This process usually takes 30 minutes to two hours. The external radiation treatment itself generally is done over four or five days a week for several weeks.

Radiation therapy does not carry the risks of major surgery, but it can have unpleasant side effects including fatigue and lack of energy. A reduced white blood cell count (rendering a person more susceptible to infection) and low blood platelet levels (making blood clotting more difficult) can also occur with radiation therapy. If the digestive organs are in the field exposed to radiation, patients may experience nausea, vomiting, or diarrhea. Radiation therapy can irritate the skin in the area that is treated, but this irritation generally improves with time after treatment has ended.

3-Chemotherapy

Both NSCLC and SCLC may be treated with chemotherapy. Chemotherapy refers to the administration of drugs that stop the growth of cancer cells by killing them or preventing them from dividing. Chemotherapy may be given alone, as an adjuvant to surgical therapy, or in combination with radiotherapy. While a number of chemotherapeutic drugs have been developed, the class of drugs known as the platinum-based drugs have been the most effective in treatment of lung cancers.

Chemotherapy is the treatment of choice for most SCLC, since these tumors are

generally widespread in the body when they are diagnosed. Only half of people who have SCLC survive for four months without chemotherapy. With chemotherapy, their survival time is increased up to four- to fivefold. Chemotherapy alone is not particularly effective in treating NSCLC, but when NSCLC have metastasized, it can prolong survival in many cases.

Chemotherapy may be given as pills, as an intravenous infusion, or as a combination of the two. Chemotherapy treatments are usually given in an outpatient setting. A combination of drugs is given in a series of treatments, called cycles, over a period of weeks to months, with breaks in between cycles. Unfortunately, the drugs used in chemotherapy also kill normally dividing cells in the body, resulting in unpleasant side effects. Damage to blood cells can result in increased susceptibility to infections and difficulties with blood clotting (bleeding or bruising easily). Other side effects include fatigue, weight loss, hair loss, nausea, vomiting, diarrhea, and mouth sores. The side effects of chemotherapy vary according to the dosage and combination of drugs used and may also vary from individual to individual. Medications have been developed that can treat or prevent many of the side effects of chemotherapy. The side effects generally disappear during the recovery phase of the treatment or after its completion.

4-Brain prophylactic radiation

SCLC often spreads to the brain. Sometimes people with SCLC that is responding well to treatment are treated with radiation therapy to the head to treat very early spread to the brain (called micrometastasis) that is not yet detectable with CT or MRI scans and has not yet produced symptoms. Brain radiation therapy can cause short-term memory problems, fatigue, nausea, and other side effects.

5-Treatment of recurrence

Lung cancer that has returned following treatment with surgery, chemotherapy, and/or radiation therapy is called recurrent or relapsed. If a recurrent cancer is confined to one site in the lung, it may be treated with surgery. Relapsed tumors generally do not respond to the chemotherapeutic drugs that were previously administered. Since platinum-based drugs are generally used in initial chemotherapy of lung cancers, these agents are not useful in most cases of recurrence. A type of chemotherapy referred to as second-line chemotherapy is used to treat recurrent cancers that have previously been treated with chemotherapy, and a number of second-line chemotherapeutic regimens have been proven effective at prolonging survival. People with recurrent lung cancer who are well enough to tolerate therapy are also good candidates for experimental

therapies, including clinical trials.

6-Targeted therapy

One alternative to standard chemotherapy is the drug erlotinib (Tarceva) which may be used in patients with NSCLC who are no longer responding to chemotherapy. It is a so-called targeted drug, a drug that more specifically targets cancer cells, resulting in less damage to normal cells. Erlotinib targets a protein called the epidermal growth factor receptor (EGFR) that helps cells to divide. This protein is found at abnormally high levels on the surface of some types of cancer cells, including many cases of non-small cell lung cancer. Erlotinib is taken by mouth in pill form.

Other attempts at targeted therapy include drugs known as antiangiogenesis drugs, which block the development of new blood vessels within a cancer. Without adequate blood vessels to supply oxygenated blood, the cancer cells will die. The antiangiogenic drug bevacizumab (Avastin) has recently been found to prolong survival in advanced lung cancer when it is added to the standard chemotherapy regimen. Bevacizumab is given intravenously every two to three weeks. However, since this drug may cause bleeding, it is not appropriate for use in patients who are coughing up blood, if the lung cancer has spread to the brain, or in people who are receiving anticoagulation therapy ("blood thinner" medications). Bevacizumab is also not used in cases of squamous cell cancer, because it leads to bleeding from this type of lung cancer.

7-Photodynamic therapy (PDT)

One newer therapy used for different types and stages of lung cancer (as well as some other cancers) is photodynamic therapy. In photodynamic treatment, a photosensitizing agent (such as a porphyrin, a naturally occurring substance in the body) is injected into the bloodstream a few hours prior to surgery. During this time, the agent deposits itself selectively in rapidly growing cells such as cancer cells. A procedure then follows in which the physician applies a certain wavelength of light through a handheld wand directly to the site of the cancer and surrounding tissues. The energy from the light activates the photosensitizing agent, causing the production of a toxin that destroys the tumor cells. PDT has the advantages that it can precisely target the location of the cancer, is less invasive than surgery, and can be repeated at the same site if necessary. The drawbacks of PDT are that it is only useful in treating cancers that can be reached with a light source and is not suitable for treatment of extensive cancers. Research is ongoing to further determine the effectiveness of PDT in lung cancer.

8-Radiofrequency ablation (RFA)

Radiofrequency ablation is being studied as an alternative to surgery, particularly in cases of early stage lung cancer. In this newer type of treatment, a needle is inserted through the skin into the cancer, usually under guidance by CT scanning. Radiofrequency (electrical) energy is then transmitted to the tip of the needle where it produces heat in the tissues, killing the cancerous tissue and closing small blood vessels that supply the cancer. RFA usually is not painful and has been approved by the U.S. Food and Drug Administration for the treatment of certain cancers including lung cancers. Studies have shown that this treatment can prolong survival similarly to surgery, when used to treat early stages of lung cancer, but without the risks of major surgery and the prolonged recovery time associated with major surgical procedures.

9-Experimental therapies

Since no therapy is currently available that is absolutely effective in treating lung cancer, patients may be offered a number of new therapies that are still in the experimental stage, meaning that doctors do not yet have enough information to decide whether these therapies should become accepted forms of treatment for lung cancer. New drugs or new combinations of drugs are tested in so-called clinical trials, which are studies that evaluate the effectiveness of new medications in comparison with those treatments already in widespread use. Experimental treatments known as immunotherapies are being studied that involve the use of vaccine-related therapies or other therapies that attempt to utilize the body's immune system to fight cancer cells.

2.13 The prognosis (outcome) of lung cancer [12]

The prognosis of lung cancer refers to the chance for cure or prolongation of life (survival), and is dependent upon where the cancer is localized, the size of the cancer, the presence of symptoms, the type of lung cancer, and the overall health status of the patient.

SCLC has the most aggressive growth of all lung cancers, with a median survival time of only two to four months after diagnosis when untreated. (That is, by two to four months, half of all patients have died.) However, SCLC is also the type of lung cancer most responsive to radiation therapy and chemotherapy. Because SCLC spreads rapidly and is usually disseminated at the time of diagnosis, methods such as surgical removal or localized radiation therapy are less effective in treating this tumor type. However, when chemotherapy is used alone or in combination with other methods, survival time can be prolonged four- to fivefold;

however, of all patients with SCLC, only 5%-10% are still alive five years after diagnosis. Most of those who survive have limited-stage SCLC.

In non-small cell lung cancer (NSCLC), results of standard treatment are generally poor in all but the most localized cancers that can be surgically removed. However, in stage I cancers that can be completely removed, five-year survival approaches 75%. Radiation therapy can produce a cure in a small minority of patients with NSCLC and leads to relief of symptoms in most patients. In advanced-stage disease, chemotherapy offers modest improvements in survival time, although overall survival rates are poor.

The overall prognosis for lung cancer is poor when compared with some other cancers. Survival rates for lung cancer are generally lower than those for most cancers, with an overall five-year survival rate for lung cancer of about 16% compared to 65% for colon cancer, 89% for breast cancer, and over 99% for prostate cancer.

Chapter 3

CT and CAD Systems for Lung Cancer

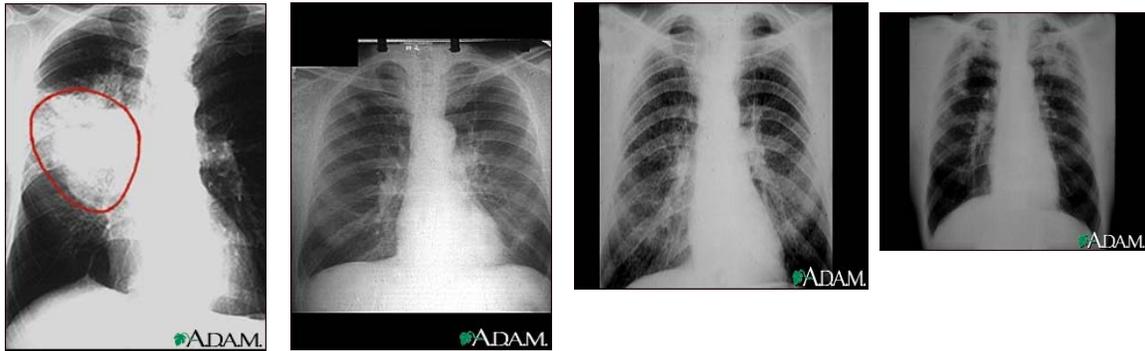
3.1 CT versus Chest X-Ray

X-Ray has been used for long in the diagnosis of lung cancer and still being used in many medical facilities. However, it is very difficult for radiologists to detect and to diagnose lung cancer on chest x-ray images because of following reasons:

- 1-There are many tissues that overlap each other on chest radiographs.
- 2-The presence of cancerous tumors is obscured by the overlying ribs, bronchia, blood vessels and other normal anatomic structures.
- 3-The shadows of cancerous tumors seen on chest radiograph are usually vague and subtle, and they tend to be missed.

Moreover, in X-Ray examination, the same abnormality might indicate more than disease and there is usually no decisive diagnosis. In (Figure 3.1) [13], we can see various diagnostic Chest X-Ray images. As we can see in the X-Ray in (Fig. 3.1-A), a patient with central cancer of the right lung, we can easily see a white big mass in the middle portion of the right lung (seen on the left side of the picture). This is a very clear but terminal cancer found. In (Fig. 3.1-B) shows adenocarcinoma of the lung. There is a rounded light spot in the right upper lung at the level of the second rib. The light spot has irregular and poorly defined borders and is not uniform in density; diseases that may cause this type of x-ray result would be tuberculous or fungal granuloma, and malignant or benign tumors. In (Fig. 3.1-C) we can see stage II coal worker's pneumoconiosis, there are diffuse, small light areas on both sides of the lungs. Other diseases that may explain these x-ray findings include simple silicosis, disseminated tuberculosis, metastatic lung cancer, and other diffuse, infiltrative pulmonary diseases. In (Fig. 3.1-D), we can see a complicated coal workers pneumoconiosis. In this image, there are diffuse, small, light areas (3 to 5 mm) in all areas on both sides of the lungs; Also there are large light areas which run together with poorly defined borders in the upper areas on both sides of the lungs. Diseases which may explain these X-ray findings include complicated coal workers pneumoconiosis, silico-tuberculosis, disseminated tuberculosis, metastatic lung cancer, and other diffuse infiltrative pulmonary diseases.

The most serious drawback of chest X-ray is, for the lung cancer to be seen in chest X-ray, it needs to be at least a centimeter in diameter to be detectable by an ordinary X-ray. However, by the time a tumor has reached this size, the original cell which became cancerous has divided (or doubled) 36 times. As death usually results after 40 such cell divisions, it is clear that the X-ray diagnoses lung cancer very late when the cancer is terminal.



(A)

(B)

(C)

(D)

Fig. 3.1 several abnormalities in X-Ray images

3.2 Lung CT

As we can in (Fig. 3.2), a typical lung CT consists of the following components:

1-The two lungs (left and right)

2-A cross section of a vertebra

3-Bone, only for the sections where a cross section of the ribs was present, for the slices taken at the intercostal spaces, bone part in the CT will not appear.

4-Sternum, similar to bone, sternum is present only in the top CT slices of the thorax.

5-Trachea, a big hole that vary in size along the CT slices of the thorax, and diminishes completely when the trachea branches to the two bronchi.

6-Mediastinum, which is the part in the thorax that contains the heart and is bound by the sternum from the front and spinal cord from the back.

7-Fat and muscle.

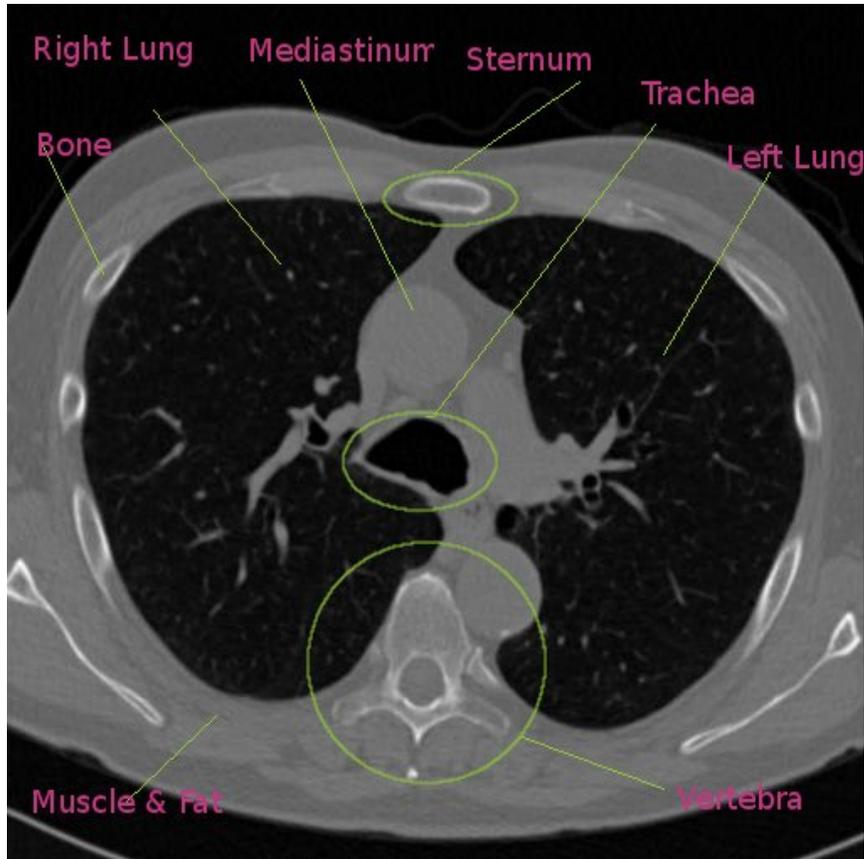


Fig. 3.2 Lung CT

3.3 What is Computer Aided Diagnosis (CAD)

CAD has been defined as a diagnosis made by a radiologist with the benefit of information generated by computerized image analysis [4]. Although some investigators distinguish between the concepts of computer-aided detection and CAD, others have interpreted CAD broadly as encompassing both the detection task and the classification task.

3.4 CAD motivation

The history of CAD for chest radiography has a long history [5], and was faced badly by the main challenge of extracting the nodules from a chest background, which are formed from a superimposed three-dimensional structure that is projected into a two-dimensional image.

1-The advent of CT imaging

The introduction of helical CT has created significant advances in the detection and characterization of disease throughout the body. Helical CT offers a series of advantageous features that include eliminating respiratory misregistration and hence, solving the problem of missing small lesions because of variations in the patient's depth of respiration on successive breath-holds. This phenomenon is most pertinent in the lower lungs. Several experimental and clinical studies have shown that the detection rate of pulmonary nodules is significantly higher on helical CT scans than on conventional CT scans. Another advantage of helical CT is that maximum lesion conspicuity can be achieved by reconstructing the acquired helical CT data set in overlapping sections. However, overlapping section reconstruction leads to a large number of images, which can create problems for film-based viewing, particularly when follow-up studies are being performed.

2-Screening program of high-risk populations

The early detection and treatment of lung cancer has proven to give very positive rate of prognosis. Hence, with high sensitivity achieved by the CT, screening program of high-risk populations (such as previous smokers or people who have been subject to exposure to some substances such as asbestos) started to emerge, using low-dose helical CT protocol. These programs led to increased number of slices that radiologists need to examine, taking into consideration the large number of slices of each scan set because of the capability to get smaller slice thicknesses. That all made the job of radiologists truly tedious, and very likely to miss nodules. More importantly, the fact that a failure to identify a nodule may crucially affect a patient prognosis.

Consequently, researchers were encouraged to think of devising a computer-aided diagnosis (CAD) system that could be considered as a second reader to radiologists in improving the nodule detection, with more hope to be developed to the level that it will be reliable enough for the radiologists to consider it as a first reader. In addition to that, CAD also is very useful in quantifying the volume and number of nodules.

Moreover, CAD systems have also been used in assisting the radiologist in the follow-up of CT scans (Fig. 3.3), by evaluating the progress of nodules over the time, in terms of their number and the change in their sizes. That help in giving better objective evaluation to the effectiveness of the treatment.

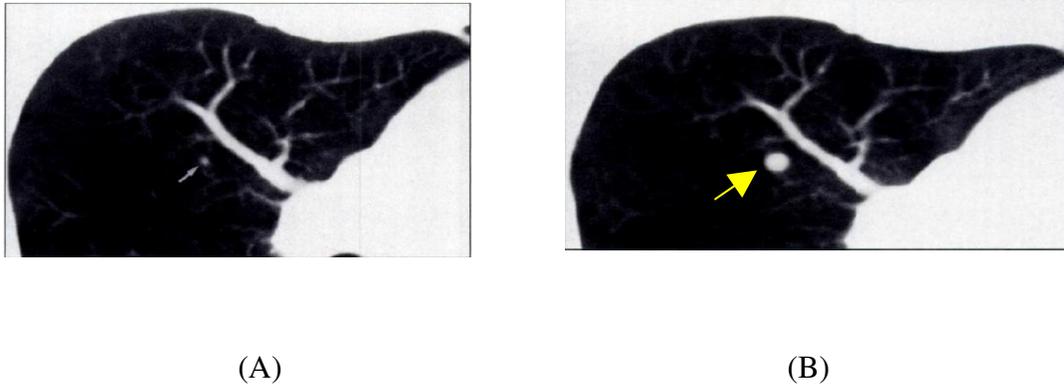


Fig 3.3 The change over time (123 days) in the size of a nodule. 48-year-old woman with histologically proven carcinoma of breast. A, Helical CT scan of chest reveals pulmonary nodule (arrow) that is 4 mm in diameter. B, Follow-up helical CT scan obtained 123 days after A confirms presence of nodule, which has enlarged.

3.5 Nodules

3.5.1 What are nodules

In 1984, the Fleischner Society published a glossary of terms for thoracic radiology [10], in which a lung nodule was defined as “any pulmonary or pleural lesion represented in a radiograph by a sharply defined, discrete, nearly circular opacity 2–30 mm in diameter.” Twelve years later, the Fleischner Society published a glossary of terms specifically for thoracic CT [8] in which a lung nodule was defined as a “round opacity, at least moderately well margined and no greater than 3 cm in maximum diameter.” The Fleischner Society’s pathologic definition of a nodule as a “small, approximately spherical, circumscribed focus of abnormal tissue” [8] reflects the three-dimensional nature of the physical lesion manifested radiologically as a nodule on CT scans.

From the above, it could be concluded that:

- The term “nodule” represents a spectrum of abnormalities (irrespective of presumed histology), which is itself a subset of a broader spectrum of abnormalities termed “focal abnormality;” a lesion should be considered a “nodule” if it satisfies the definition of “nodule” (the most essential component of which is its “nodular” morphology—the remaining components will be determined by the visual nodule library)
- Nodules may represent primary lung cancers, metastatic disease, or non-cancerous processes.

3.5.2 Shape of the nodule

Nodules have diameter between 3-30 mm, could get enlarged to be at maximum 6cm. Nodules vary in shape depending on many factors, whether they are isolated, attached to the plural, attached to a vessel, solid, non solid, air filled, homogeneous, or non homogeneous. (Fig. 3.4)

Isolated nodules are usually spherical in shape (in 3D view), and they are seen as circular objects in the 2D slices. The nodule appears in more than one slice if it has a relatively big size. Small size nodules are always found in one slice only. Nodules attached to the wall are semi-sphere in shape, and they might span more than one slices depending on their sizes.

3.5.3 Characteristics of nodules in CT

Nodules could be characterized by radiologists according to the following features:

1-Subtlety: in terms of its difficulty in detection.

2-Internal structure: or expected internal composition of the nodule (soft tissue, fluid, fat, air).

3-Calcification: pattern of calcification if present.

4-Sphericity: the three dimensional shape of the nodule in terms of its roundness:

5-Margin: description of how well defined the margins of the nodule is.

6-Speculation: amount of speculation present in nodule

7-Texture: internal texture or composition of nodule in terms of solid and ground glass components.

8-Malignancy: Radiologist subjective assessment of likelihood of malignancy of this nodule.

3.5.4 Types of nodules (Location)

1.Isolated Nodules

Nodules that are located with the lung area.

2.Juxta-Pleural

Nodules that are attached to the lung wall

3.Juxta-Vascular

Nodules that are attached to the vessels that are inside the lung area. Vessels are hardly seen close to the lung wall, and when run perpendicular to the slices, they tend to have circular shape.

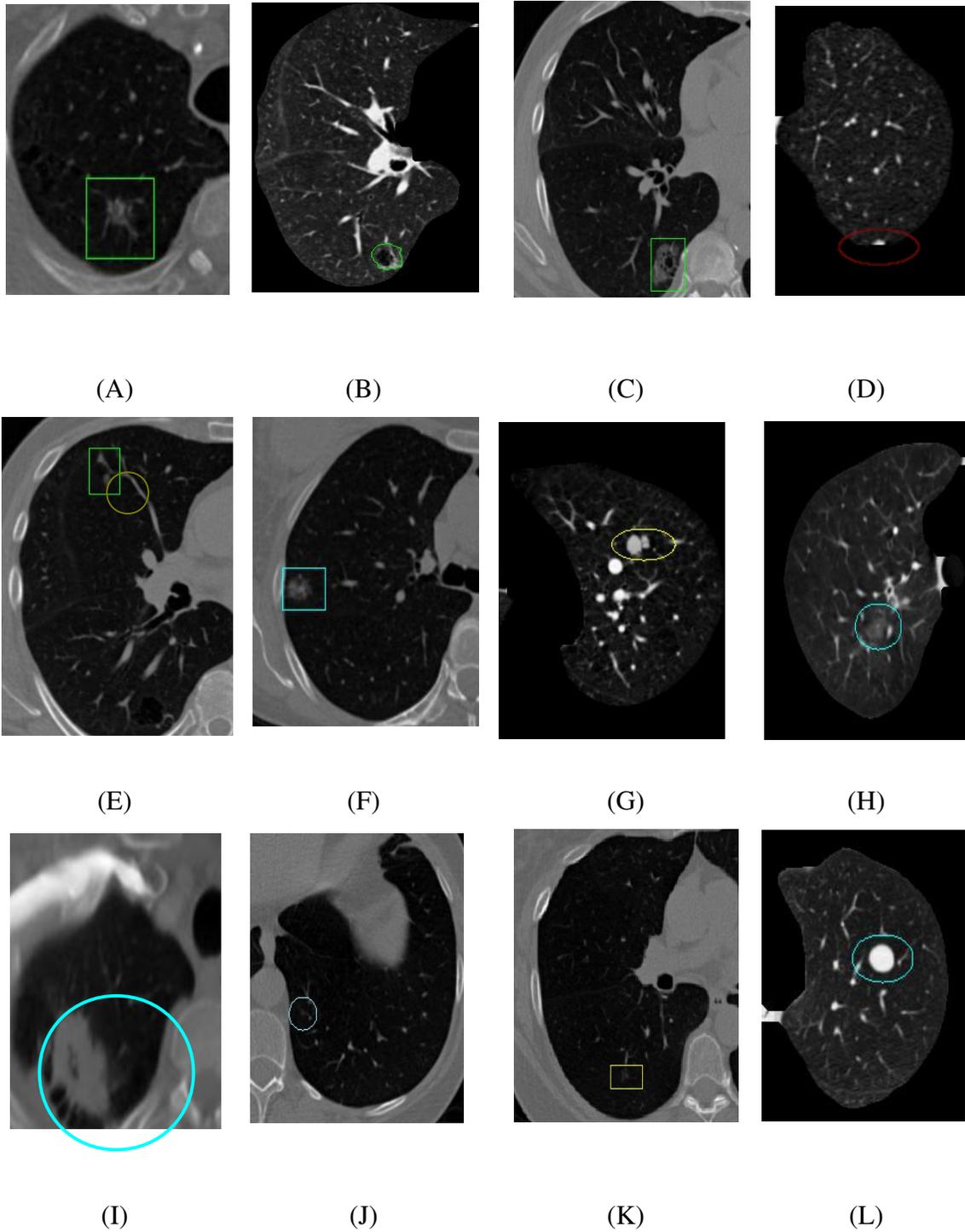


Fig. 3.4 several shapes of nodules

A:Irregular in shape and attached to vessels, B & C: Air filled nodules, D:Juxta-Pleural, E:Juxta-Vascular, F:Non homogeneous, G:Multi-Part nodule, H:Low Gray Level Value, I:Huge mass, J, K:Extremely small, L:Ideal Nodule

3.5.5 Types of nodules (Texture)

1.Subsolid Nodules:

Are nodules that don't completely obscure the lung parenchyma within it.

2.Ground glass Nodules:

Semi solid nodules that have solid component within them that obscure the lung parenchyma.

3.Cavitary nodule

The nodule that contains an air-filled necrotic region is termed cavitary.

Nodules are generally of relatively high contrast, circular shape, and uniform distribution of density. Although vessels running perpendicular to the slices might appear circular in shape, nodules are usually of higher gray level value than them. Compact nodules were found to have a Gaussian Gray Level Distribution [7] as shown in Fig. 3.5:

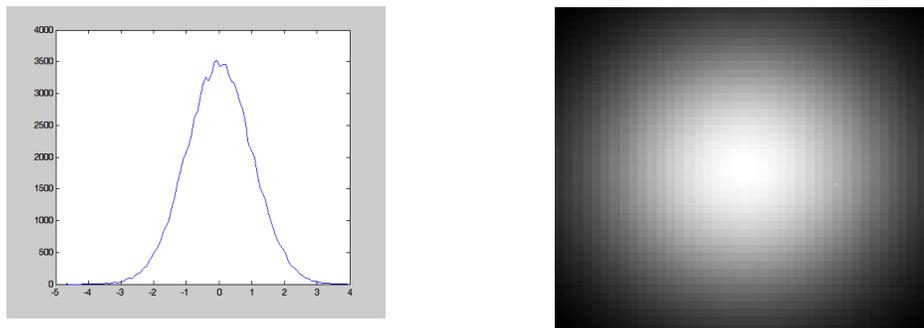


Fig. 3.5 Nodules tend to have Gaussian distribution

3.6 Role of CAD with pulmonary nodules

CAD systems can aid radiologists by providing a “second opinion” and may be used in the first stage of examination in the near future, the main role of using a CAD with pulmonary nodules lies in three main applications:

- 1.Improve nodule detection by the radiologists
- 2.Quantify nodule volume and number
- 3.Assess change over time in nodule number and diameter, which could be used in the assessment of the efficiency of the treatment.

3.7 Summary

Two main developments motivated and accelerated the development of CAD systems, the first is, the advent of multi-detector row CT scanners combined with gantry rotation times of less than 500 msec per rotation. These scanners generate an extensive amount of image data per examination. A single thoracic CT examination routinely generates more than 300 thin section images from a single breath hold with state-of-the-art 16-section scanners. Thus, it is now possible to acquire volumetric image data of the thorax with images composed of isotropic voxels with millimeter and submillimeter resolution. The second development is the growing awareness of lung cancer screening by using a low-dose helical CT protocol. These developments, combined with the fact that CT examinations are being used more often for a wide range of diagnostic tasks, are dramatically increasing the workload of radiologists, which, in turn, may result in more errors of omission. Consequently, CAD techniques may become a practical necessity in the interpretation of CT scans; for example, CAD may be used in the detection, classification, quantification of the number, volume, and growth of lung nodules.

Chapter 4

Literature Review

4.1 Overview of the most relevant systems and method in CAD literature

In 1997, Manfred Tillich et al. [1] examined whether Cine and Film-Based Viewing improves the detection of pulmonary nodules in patients with known extrathoracic malignancy. They reviewed the CT of 60 patients with known extrathoracic malignancy by four radiologists. They divided the nodules into four groups according to the maximum diameter: nodules smaller than or equal to 5 mm, nodules larger than 5 mm but smaller than 10 mm, nodules larger than 10 mm but smaller than 20 mm, and nodules larger than 20 mm. They also assigned a four points conspicuity score, three parameters were successively analyzed for each nodule: nodule diameter, density difference between nodule and adjacent blood vessels, and margin of each nodule (smooth or irregular, classified as well or ill defined). They discovered that, interpreters saw 266 nodules on cine viewing, whereas 237 nodules were seen with static film-based viewing. They reached the conclusions that Cine viewing of helical CT scans significantly increases the detection rate of pulmonary nodules that are smaller than or equal to 5 mm in diameter. However, they found no significant difference between cine and film-based viewing for the detection of nodules that were larger than 5 mm in diameter. They concluded that, the advantage of cine viewing may be attributed to both the larger image size and the ability to scroll through images for improved differentiation between vessels and nodules.

In 1999, Samuel G. Armato et al. [2] were ones of the leading researchers to incorporate the data assembled from consequent 2D slices to complement each other, giving at the end 3D features of the objects detected. They applied gray level thresholding on each slice of a CT scan set to segment the thorax from the background, and then the lungs from the thorax. They then applied rolling ball algorithm to the lung segmentation contours to avoid the loss of juxta-pleural nodules. Afterward, they applied multiple gray level thresholds to the volumetric lung regions to identify nodule candidates. For each nodule candidate 2D and 3D geometric and gray level features are computed. They computed nine features for each nodule candidate; six geometric (volume sphericity, radius of the equivalent sphere, maximum compactness, maximum circularity, and maximum eccentricity) and three gray-level features: mean gray level and standard deviation within

the structure, and gray level threshold at which the volume of the structure first decreases below the upper volume bound. Then, the values of these features were analyzed by a Linear Discriminant Analysis (LDA) classifier, which has substantially reduced the number of false positives. They applied their algorithm on 17 cases that comprised 493 sections with 187 pulmonary nodules. The effective diameter of the found nodules ranged from 3.1-27.8 mm. The algorithm led to a detection sensitivity of 82% with 796 FPs. After incorporating the LDA classifier, and using a Leave-One-Out to evaluate its performance, an area of 0.93 under the curve of the ROC was yielded, which was an operating point of 85% sensitivity and 89% specificity, and indicated an overall sensitivity for nodule detection of 70%, with an average of three false positives per section. This outcome corresponds to 89% reduction in the number of false-positive findings after the application of LDA.

In 2001, Jane P. Ko et al. [3] developed a computer system that automatically identifies nodules at chest CT, quantifies their diameters, and assessed for changes at follow up. They showed that the computer system they developed in their study has quantitative abilities that could affect many facets of oncology care, from diagnosis of metastases to assessment of their response to chemotherapy. To segment the lung parenchyma, they first applied a threshold to get a binary image for the slice, and then to detect (extract) the thorax and the lung borders, they made analysis to the vertical and horizontal profiles of the binary image. These profiles were used to identify an initial point on the lung border. Beginning at this point, the border was traced with a method based on a backtracking algorithm. They had to do lung border correction in order to avoid losing the pleural nodules. They did that by comparing the curvatures at points on the lung border. A rapid change in curvature indicated a nodule, large vessel, or bronchus that formed an acute or obtuse angle with the lung border, and the lung border was then corrected by means of insertion of a border segment. To extract the objects, They used multiple gray level thresholding and sequential labeling algorithm. Afterward, they analyzed the candidates according to their location and shape. For the location, they divided the lung into five regions, and computed the centroid of each lung and the whole image. They followed an approach that, any object found at a distance less than 5mm from the lung wall is likely to be a nodules, as vessels are rarely seen in these regions. They then did a shape determination to the objects extracted, by measuring how elongated an object is. They also formed the 3D structure of consecutive slices, to get 3D measurements of the objects by computing their volumes. They used the centroid of the trachea in each slice as reference point to register all the slices. To analyze the change over time for the nodules, they had to face a big challenge, which is the complication of registering the CT images of the two different studies, that are attributed to the differences in patient

positions and inspiration. To tackle this problem, for each image in the initial study, they used the computer to help finding a matching image in the next study. The matching was performed using the centroids of anatomic structures such as the sternum, lungs, vertebra, and trachea. Given the centroid of a nodule in one image, a projected centroid of the same nodule in an image from a subsequent study was calculated with the translational and rotational parameters generated from the global registration of thoracic structures. A fixed area of 10^2 pixels around this projected centroid was searched to locate the centroid of the corresponding nodule. The initial study was similarly searched to identify nodules that corresponded with nodules identified on the follow-up study. The sizes of corresponding nodules in the two studies were then compared. They applied rule-based classification to identify 'highly likely nodules', 'likely nodules', and 'normal structure'. Comparing the results they got for the two studies with that of the radiologists. They identified 318 (86%) of 370 nodules in 16 studies (eight initial and eight follow-up studies). Twenty-four (46%) of the 52 missed nodules were 3 mm or less in diameter. Thirty-eight (73%) of the missed nodules were small, 12 (23%) were medium, and two (4%) were large. Thirty (58%) of the missed nodules contacted the lung border, and 13 (25%) were adjacent to a vessel. The result they reached was that, the CAD assessment of the change in size of the nodules in the eight patients did not differ a lot than that reached by the radiologists. They also reached a very important conclusion about the difficulty of having an accurate measurement to the sizes of the nodules, because it is affected by the threshold values for pixel intensities that are used to distinguish brighter soft-tissue structures from the darker air-containing components of the lung. The CT scanner, kilovolt potential, reconstruction algorithm, section thickness, and nodule location in the field of view have been shown to affect pixel intensity [8] and, therefore, apparent nodule diameter.

Joseph M. Reinhardt et al. in 2001 devised an algorithm for automatic lung segmentation for accurate quantification of volumetric X-Ray CT images [8]. In doing lung segmentation, they also separated the left and the right lungs by identifying the anterior and posterior junctions by dynamic programming. Finally, a sequence of morphological operations is used to smooth the irregular boundary along the mediastinum in order to obtain results consistent with those obtained by manual analysis, in which only the most central pulmonary arteries are excluded from the lung region. The method has been tested by processing 3-D CT data sets from eight normal subjects, each imaged three times at biweekly intervals with lungs at 90% vital capacity. The main stages they had were could be summarized in:

1-Threshold selection

It was done using, what is called optimal thresholding, which is an automatic threshold selection method that allows us to accommodate the small variations in tissue density expected across a population of subjects. For this step, they assumed that the image volume contains only two types of voxels: 1) voxels within the very dense body and chest wall structures (the body voxels) and 2) low-density voxels in the lungs or in the air surrounding the body of the subject (the nonbody voxels). They then used optimal thresholding to select a segmentation threshold to separate the body from the nonbody voxels, and then identified the lungs as the low-density cavities inside of the body. The segmentation threshold is selected through an iterative procedure.

Let T^i be the segmentation threshold at step i . To choose a new segmentation threshold, you apply T^i to the image to separate the voxels into body and nonbody voxels.

Let μ_b and μ_n be the mean gray-level of the body voxels and nonbody voxels after segmentation with threshold T^i . Then the new threshold for step $i+1$ is:

$$T^{(i+1)} = (\mu^b + \mu^n) / 2 \quad (4.1)$$

This iterative threshold update procedure is repeated until there is no change in the threshold, i.e., $T^{i+1} = T^i$. The initial threshold T^0 is selected based on the CT number for pure air (-1000 HU) and the CT number for voxels within the chest wall/body (> 0 HU).

2-Segmentation of large Airways

They first identified the initial location of the trachea and did a slice-by-slice region growing tracking the whole trachea till it starts to branch to the two bronchi. They then removed the whole trachea and the left and right mainstem bronchi.

3-Left and Right Lung Separation

Using a repeated sequence of image erosion until the number of objects in the image became two, they then, using a conditional dilation can retain the same size again, just before they get merged again. Then they locate the junction between them.

Yongbum Lee et al. in 2001 [6] proposed a novel template matching technique based on Genetic Algorithm Template Matching (GATM) for detecting nodules existing within the lung area; the GA was used to determine the target position in the observed image and to select an adequate template image from several reference patterns for quick template matching. In addition, a conventional template matching was employed to detect nodules existing on the lung wall area, lung wall template matching (LWTM), where semicircular models were used as reference patterns; the semicircular models were rotated according to the angle of the target point on the contour of the lung wall. They formed the models of

the nodules to have Gaussian Gray Level Distribution, and they used the Genetic Algorithms Template Matching to search for the location of spherical nodules within the volumetric lung parenchyma, and to select an adequate template image from reference images. The search operation was regraded as an optimization problem, in which the input was the 3D CT images, and the 3D templates were constructed by three consecutive 2D template image, each has a Gaussian gray level distribution. Four different diameter sizes were employed, ranging in sizes from 10-40 pixels. After initial detecting candidates using the two template matching methods, they extracted a total of 13 feature values and used them to eliminate false-positive findings. Twenty clinical cases involving a total of 557 sectional images were used in this study. 71 nodules out of 98 were correctly detected by this scheme (i.e., a detection rate of about 72%), with the number of false positives at approximately 1.1/sectional image.

To stimulate the advancement of computer-aided diagnostic (CAD) research for lung nodules in thoracic computed tomography (CT), the National Cancer Institute launched in 2004 a cooperative effort known as the Lung Image Database Consortium (LIDC) [4]. The LIDC is composed of five academic institutions from across the United States that are working together to develop an image database that now serves as an international research resource for the development, training, and evaluation of CAD methods in the detection of lung nodules on CT scans.

The database has been since then available online to the public, and have been used by many researchers. One of the main motivations of building this database, was the absence of a common reference among researchers of CAD systems to assess the efficiency of their techniques in a robust and reliable fashion. Hence, a reliable comparison of CAD methods reported in the literature was impossible. Furthermore, researchers in LIDC realized the hurdles that impede the advance of CAD which lie in: First, the difficulty for researchers to get access to patient data and the barriers they will face because of the regulations that govern the transmission of this data. Second, because of the substantial differences in the judgments among experienced radiologists, there will be a necessity to the construction of a panel of experienced thoracic radiologists to provide, what could be defined as “the truth” for the detected nodule, the thing that is so expensive.

The LIDC summarized its mission in:

- 1-To develop an image database as a web accessible international research resource for the development, training, and evaluation of CAD methods for lung cancer detection and diagnosis using helical computed tomography (CT).
- 2-The database should enable the correlation of performance of CAD methods for

detection and classification of lung nodules with spatial, temporal and pathological ground truth.

3-To provide a true research resource, the database must contain more than images. Consequently, the database now consists of an image repository and an associated relational database in which nodule features (eg, radiologist outlines, subjective subtlety ratings, and lobar location); technical parameters of the scan available in the digital imaging and communications in medicine, or DICOM, header (eg, exposure rate, reconstruction algorithm, and scanner model); and patient information (eg, age, sex, smoking history, and any available diagnostic information, such as the results of follow-up studies or pathologic examinations) are recorded. The relational database component will give users of the database the ability to extract customized image subsets based on search results.

Process Model:

Protected patient information was removed from the header of the DICOM images. Information collected for each lesion includes lesion type (eg, scar or nodule), the radiologist's subjective level of confidence that the lesion represents a focal abnormality in general or a nodule more specifically, radiologic texture (eg, solid, part solid, or nonsolid) if considered a nodule, a five-point lesion subtlety score (ranging from "obvious" to "extremely subtle"), presence of calcifications, and lobar location.

Once the four radiologists (each from a different institution) have performed the blinded review, the results of the blinded review of each radiologist is made available to all of the other radiologists who reviewed the scan. Each radiologist then performs an unblinded review of the scan with the additional information provided by the other radiologists. During this unblinded review, the radiologists review all marked structures (eg, their own markings, as well as the markings of the other radiologists who reviewed the scan) and decide whether to include each marked structure as a nodule. It is important to note that a forced consensus is not be imposed; rather, all of the nodules indicated by the reviewing radiologists is tallied and recorded in the database. Information obtained from all radiologists during both blinded and unblinded reviews is included in the database to provide a rich source of data for investigators. For example, nodules recorded by only two of the radiologists during the blinded review will constitute a different detection target than nodules initially identified by all radiologists during the blinded review. Even more interesting might be nodules recorded by only two of the radiologists during the blinded review and then recorded by only the same two radiologists during the unblinded review, which implies that other radiologists observed this structure and declined to consider it a nodule. As another example, the spatial extent of a nodule may be described in probabilistic terms on the basis of the number of radiologists' outlines that encompass

each pixel. The challenges the LIDC faces were: First, coping with the rapid advance in CT technology; Second, the quality of the images included; Third, defining the nodule; Fourth, defining the truth for the nodules; Fifth, outlining detected nodules; Sixth, verifying diagnosis.

1-Coping with the rapid advance in CT technology

LIDC has included only scans that have section section less than 5mm, however, because of the fast pace of development in CT imaging, scans included now in LIDC that have slice thickness of 3mm for example, may be useless to use in the near future, particularly, when compared with scans that have slice thickness of 1.25 mm

2-Image Quality

Image artifacts that are caused by patient factors, such as respiratory motion, and scanner factors, such as beam hardening, are a reality in medical imaging, CAD developers will need to contend with their presence. Consequently, scans with high levels of noise or with streak, motion, or metal artifacts is included. However, they were assessed by LIDC radiologists and denoted by a “marginal” or “unacceptable” rating in the image quality field, so that investigators may explicitly exclude or include images with marginal or unacceptable image quality from an image dataset.

3-Defining the nodule

A utilitarian definition of a nodule may not be straightforward, since the notion of a nodule may not represent a single entity capable of verbal definition. The term nodule is more appropriately applied to a spectrum of abnormalities (irrespective of presumed histologic findings), which is itself a subset of a broader spectrum of abnormalities that we will term focal abnormalities. On the basis of this conceptualization, all nodules are focal abnormalities, but not all focal abnormalities are nodules.

LIDC is now creating a visual nodule library that seeks to capture image-based examples of lesions that span the spectrum of nodule and the superset spectrum of focal abnormality.

4-Truth assessment

In an attempt to articulate this issue, a panel of experienced radiologists were formed. However, reaching a reliable definition of truth might require extra data such as: follow-up CT scans (radiologic history of the patient), that enable radiologists to evaluate nodule growth in terms of number and volume, and also pathology reports, which will help in reaching a decisive adjudication, particularly in cases of substantial differences in radiologists' opinions. Nevertheless, since the variability among radiologists in the

detection and classification of lung nodules is a reality and has to be understood and appreciated, and since getting a pathologic report for every single detected nodules is not realistic, and since having a complete radiologic history of patients is often difficult in practice and subsequently is not available for all patients, truth in LIDC was defined based on the consensus reached after blinded and unblinded readings by radiologists was reached benefiting by the available pathologic reports and radiologic history whenever available. LIDC did not impose forced consensus, and hence all the blinded and unblinded readings of the radiologists were mentioned.

5-Outlining the nodules

The outlines for each qualifying nodule is obtained to record not only the position (eg, a centroid calculated from the radiologist outline) but also the spatial extent of the lesions that are the focus of the LIDC database. In addition, the approximate centers of nodules that are smaller than 3 mm and suspicious for cancer is indicated by the LIDC radiologists; outlines for these small nodules is not recorded. The anticipated degree of interradiologist variability in the outlining task was a concern for the LIDC. Specifically, they were concerned that inconsistencies among the outlines constructed by different radiologists might render the resulting spatial extent estimates impractical for use by CAD researchers. LIDC concluded that the nodule outlining process truly has no reference standard, and hence the variation in nodule outlines among the panel of radiologists is included with the database to provide a statistical map describing the combined opinion of experienced readers. Nodule size will be derived from these nodule outlines.

6-Verifying the diagnosis

Every effort is made to obtain a verified diagnosis for each qualifying nodule as either cancerous or noncancerous when it becomes available through chart review. In addition, the pathologic subtype of nodules (both cancerous and noncancerous), which is based on the basis of the World Health Organization classification is extracted from patient charts and included with the database. To confirm the patient records, the pathology reports for all patients and any available histopathologic slides or surgical specimens for a subset of the patients (5%) is independently reviewed by the LIDC panel of experienced pathologists. As LIDC recognizes that many scans may not have available pathologic information; furthermore, they recognize that not all nodules on any one scan will have such information. The database documents the scans and nodules for which a verified diagnosis has been obtained with fine-needle aspiration biopsy, surgical resection, or extended radiologic observation in which no growth is demonstrated over a 2-year period.

In 2007, Serhat et al. [11] used Genetic Cellular Neural Network (G-CNN) to segment lung regions from the CT images, and then for each lung region, ROIs were specified with using the density values of the pixels while doing 8-directional search. The 3D ROI image was obtained by combining all the 2D ROI images. Then, they used 3D template matching, by doing convolution with the 3D objects, and the similarity resulted for each object was used to strengthen those ROIs that have 3D structure similar to that of the nodules. Finally, they applied fuzzy rule based thresholding to extract the ROIs. To test the system's efficiency, they used 16 cases with 425 slices, and found that the system achieved 100% sensitivity with 13.37 FPs/case when the nodule thickness was greater than 5.625mm.

In 2007, Martin Dolejsi [20] presented a computer-aided diagnosis (CAD) system to detect small-size (from 2mm to around 10mm) pulmonary nodules from helical CT scans. Their system used two different schemes to locate juxtapleural nodules and parenchymal nodules. For juxtapleural nodules, morphological closing and thresholding was used to find nodule candidates. To locate non-pleural nodule candidates, 3D blob detector uses multiscale LoG filter was employed. Ellipsoid model was fitted on nodules. To define which of the nodule candidates are in fact nodules, an additional classification step was applied. Linear and multi-threshold classifiers were used. System was tested on 18 cases (4853 slices) with total sensitivity of 96%, with about 12 false positives/slice. The classification step reduced number of false positives to 9 per slice without significantly decreasing sensitivity (89,6%).

4.2 Summary

Many researches have been conducted in the last 6-10 years for developing a CAD system. The approaches followed could be grouped into two main categories, either model based or density based. In model based, nodules were regarded as spherical object that has varying diameters within the known range of 3mm-30mm, and the goal of the research was to find a technique to search for objects having similar shape. The main two strategies followed were by either doing template matching, or by applying a filter that can strengthen the objects that have similar geometric shape as the model the filter was build on. Recently, nodules are regarded more generally as ellipsoid objects[20] rather than spherical. The second approach is the density based, which relies on the fact that nodules tend to have higher density value than the surrounding environment they are in. In this approach, techniques such as multiple thresholding and region growing are usually employed. The most outstanding work done in the last few years was the development of the LIDC database which has led to waves of researches in developing CAD systems for detecting pulmonary nodules.

Chapter 5:

The Developed CAD System

5.1 Introduction

5.1.1 The dataset

The dataset used are from Lung Image Database Consortium, it is a publicly available, well characterized repository of Lung CT images, with annotations of more than one experienced radiologists done by consensus between them. 18 scan set were used, each one consists on an average of 180 slices. The images were acquired with a 512*512 matrix and quantized with 16 bits. They were transferred into the Digital Imaging and Communications in Medicine (DICOM) format at which, the Hounsfield units (**Table 5.1**) for attenuation were translated into brightness values, as shown in **Figure 5.1**

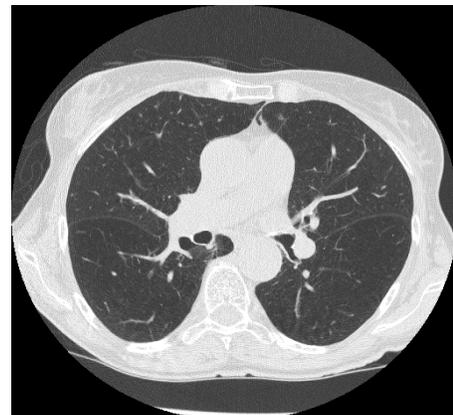
Substance	HU Range
AIR	-1000
LUNG	-400 -600
NODULE	-150
FAT	-60 -100
WATER	0
SOFT TISSUE	+40 +80
BONE	+400 +1000

Table 5.1

The attenuation values of the different tissues in HU

Fig. 5.1

A CT image of the lung, showing the two lungs, the sternum, the bone and fat around them.



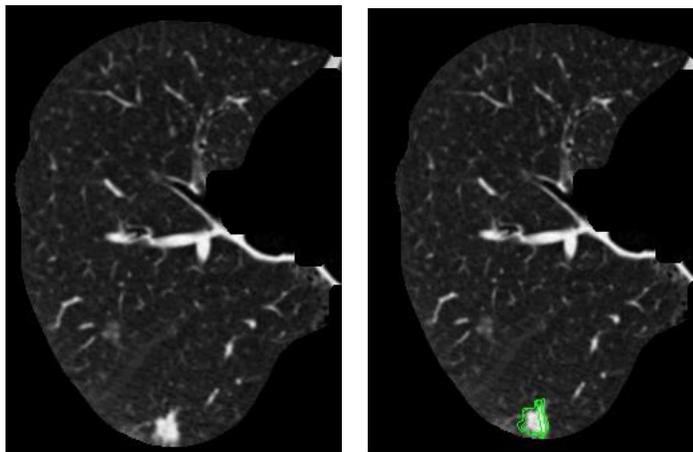
5.1.2 The annotation of the radiologists

The annotation of the radiologists are in XML format, taking style similar to the one shown below:

```
<unblindedReadNodule>
<noduleID>5069</noduleID>
<roi>
<imageZposition>-429.100006</imageZposition>
<imageSOP_UID>1.3.6.1.4.1.9328.50.3.2396</imageSOP_UID>
<inclusion>TRUE</inclusion>
<edgeMap>
<xCoord>308</xCoord>
<yCoord>252</yCoord>
<xCoord>309</xCoord>
<yCoord>254</yCoord>
<xCoord>310</xCoord>
<yCoord>255</yCoord>
<xCoord>311</xCoord>
<yCoord>255</yCoord>
.....
</edgeMap>
</roi>
</unblindedReadNodule>
```

In this format, each nodule is given a unique ID and is tagged by either “blindedRead” or “unblindedRead”. The slice number (labeled by the slice file name) is mentioned. An edge map is stated, in which the nodule is marked by a contour specified by an x-y set of coordinates. An example of a contour marking read nodules is seen below in **Fig. 5.2**

Fig. 5.2 The annotation of the radiologist. The image in the right shows a contour specified by the radiologist that marks his finding of a nodule.



The annotation of the radiologists also include the abnormalities that could be wrongly diagnosed as nodules, while they are not, these objects are tagged in the XML file with the tag “NotNodule”, and they also have their contour (set of x-y vertices) specified.

In addition to that, radiologists were asked to subjectively assess several characteristics of the nodule (each characteristic on a 1-5 scale):

1. ***Subtlety***: in terms of its difficulty in detection
2. ***Internal structure***: or expected internal composition of the nodule (soft tissue, fluid, fat, air)
3. ***Calcification***: pattern of calcification if present
4. ***Sphericity***: the three dimensional shape of the nodule in terms of its roundness:
5. ***Margin***: description of how well defined the margins of the nodule is.
6. ***Spiculation***: amount of speculation present in nodule
7. ***Texture***: internal texture or composition of nodule in terms of solid and ground glass components .
8. ***Malignancy***: Radiologist subjective assessment of likelihood of malignancy of this nodule

5.1.3 Tools and H/W used

- 1.Processor Core 2 Duo, 1.8 GHz, 3GB RAM
- 2.Linux 2.6.28-14
- 3.Matlab R2008b
- 4.Perl Scripting Language
- 5.LIBSVM 2.89 classification and regression tool

5.2 The algorithm of the 2D approach used in developing a CAD system:

The block diagram shown below (Fig. 5.3) gives an insight on the flow of the algorithm developed, and it shows the various techniques incorporated such as: Multiple Gray Level Thresholding, 2D labeling, region growing, and so on.

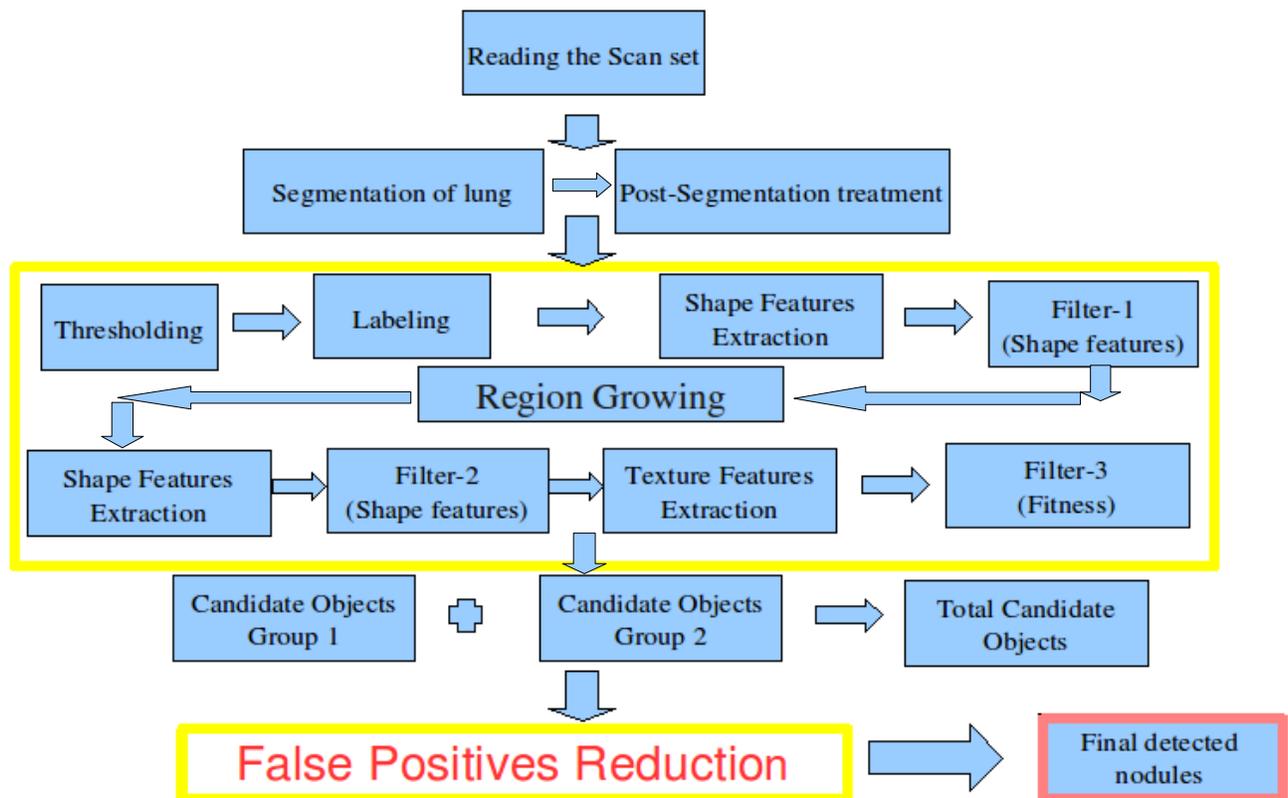


Fig. 5.3

The detailed block diagram of the 2D approach followed in developing a CAD system

5.2.1 Reading the scan set system

In this step the whole slices of the scan are read, then their DICOM information is extracted to get:

- 1-The horizontal and vertical resolution, which are the pixel sizes along the x-y dimension and z-dimension respectively.
- 2-Slice position, which will help in sorting the slices to be in order.

A stack matrix was constructed to contain all the slices in order, which will be the input of the segmentation process.

5.2.2 Segmentation of the Lung

The input images are usually in Hounsfield unit, having histogram similar to the one shown in (Fig. 5.4), where the attenuation values of the different tissues are shown in (Table 5.1).

The corresponding gray level values could to be obtained according to the equation:

$$\text{Gray Level Value} = HU + 1024 \quad (5.1)$$

The input histogram of the thorax image in gray level values is shown in (Fig. 5.4-A), from the histogram we can see that by selecting a threshold value between the two peaks of the lungs parenchyma and the fat (Fig. 5.4-B), we could easily extract the lungs from the thorax.

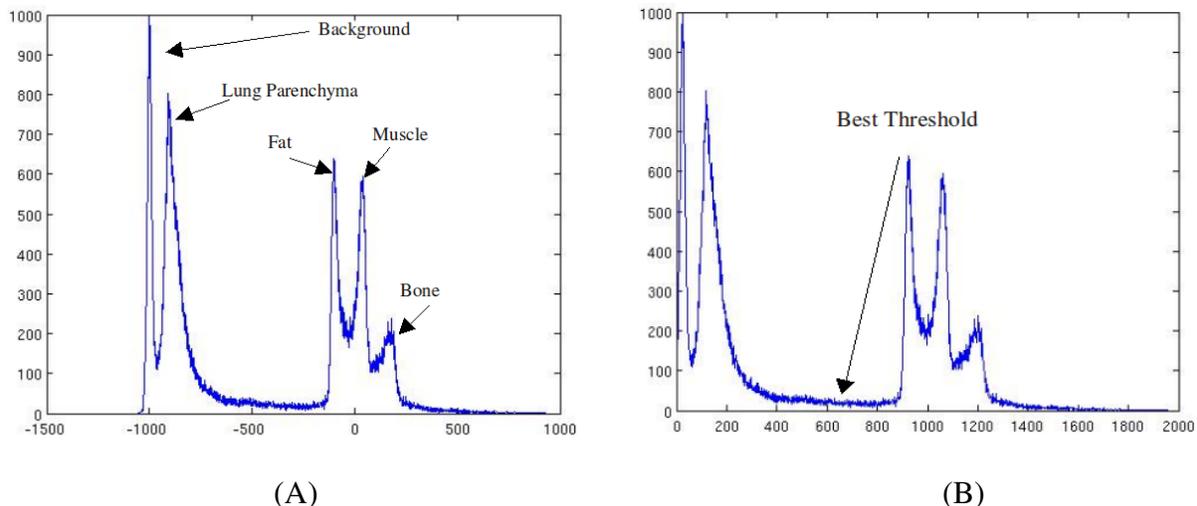


Fig. 5.4 The histogram of the thorax image. A: In HU units, B: In gray level values with the best threshold marked.

5.2.2.1 The Segmentation process

First, a threshold was applied to the input CT image of the thorax (Fig. 5.5-A), to get the image in (Fig. 5.5-B). The best threshold found experimentally was at HU=-420. A closure morphological operation is done using a disk kernel of radius 3 to clean the background (Fig. 5.5-C), then the borders was cleared, and hence lungs parenchyma (Fig. 5.5-D) were extracted

5.2.2.2 Challenges and Post -Segmentation Treatment

1.Big Plural nodules missed during the segmentation

As can be seen in Fig. 5.5-A, the pleural nodule marked by the arrow was missed during the segmentation. The closure operation using a bigger in size kernels will lead to the parenchyma got merged with the outer border as shown in (Fig. 5.5-E). Therefore, a light closure operation should be applied first, then after removing the borders, we would be able to apply a more harsh closure operation to be able to extract the lungs with all pleural nodules included, as shown in (Fig. 5.5-F)

The kernel applied in the first closure operation was radius in shape having a radius=7, the second more harsh one was also radius in shape and have a radius=12.

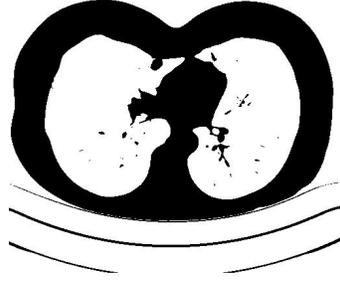
2-Pleural nodules are labeled with parts of the lung wall in low thresholding

To prepare the lungs parenchyma for the steps following it which are, multiple gray level thresholding and then labeling. As we can see in (Fig. 5.5-H) the outcome of doing high thresholding usually lead to full removal of the lung wall. However, at low level thresholding, parts of the lung wall remain (Fig. 5.5-I), the thing that makes many objects got extracted with parts of the lung wall attached to it (Fig. 5.5-J), which will eventually be removed by the filters as it will have geometric features that does not match those of the known geometric features of the nodules.

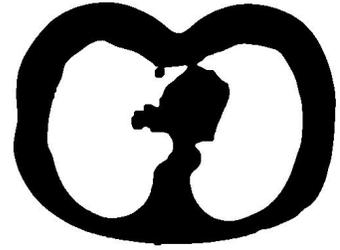
A solution to tackle this issue is to remove the lung wall (Fig. 5.5-K). The lung wall is removed by subtracting an eroded image from the original image (Fig. 5.5-L). Image erosion is done using the convolution of a kernel radius in shape with radius=3.



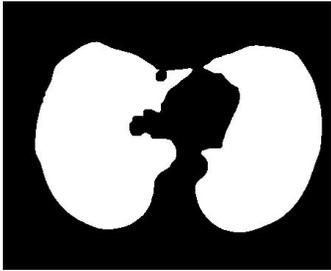
(A)



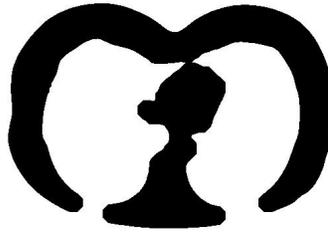
(B)



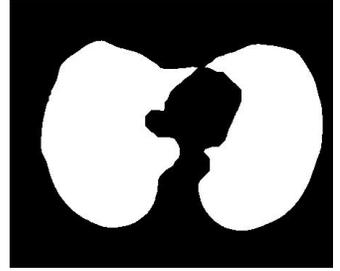
(C)



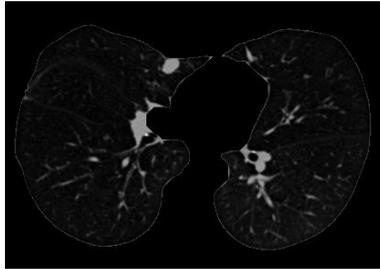
(D)



(E)



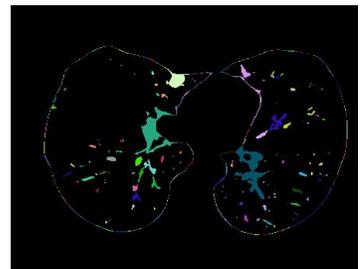
(F)



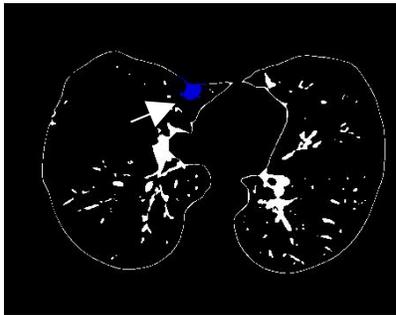
(G)



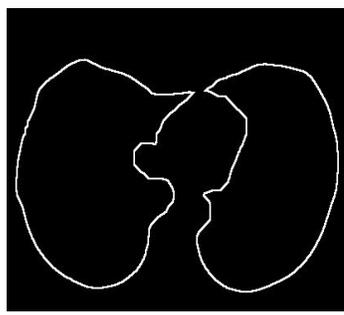
(H)



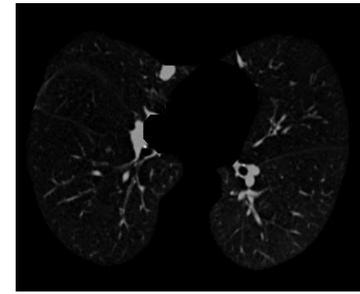
(I)



(J)

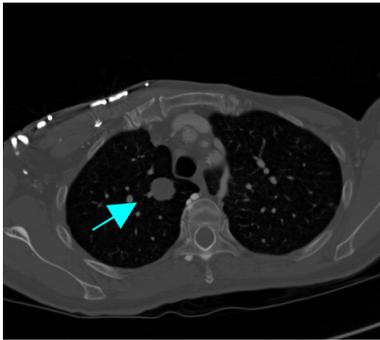


(K)

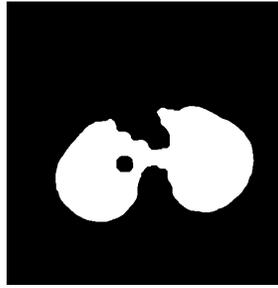


(L)

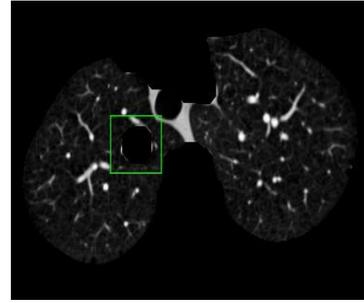
Fig. 5.5 Different stages of lung segmentation



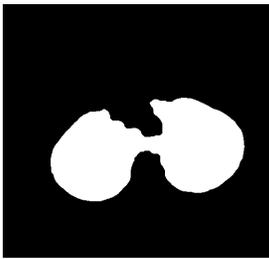
(A)



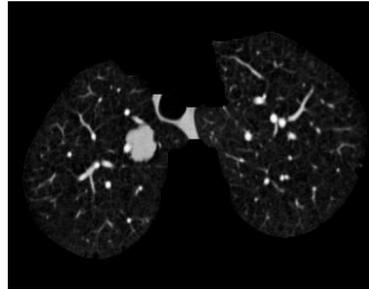
(B)



(C)



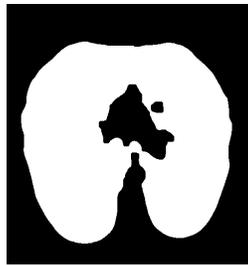
(D)



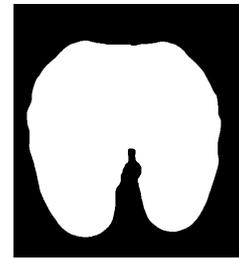
(E)



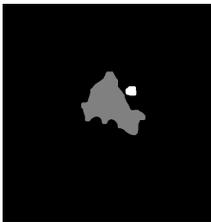
(F)



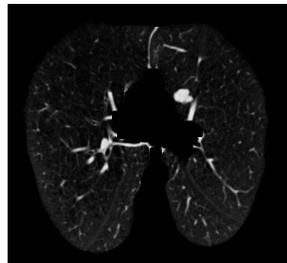
(G)



(H)



(I)



(J)

Fig 5.6 Tackling the challenges in the lung segmentation

3-Filling in holes for internal big nodules

Some big nodules as the one shown in (Fig. 5.6-A) are too big to be closed by a big size kernel in the closure operation (Fig 5.6-B,C), and using very large kernel size will cause unwanted distortion to the image. A way to tackle this problem, is to do a filling in holes operation to the extracted mask of the lungs (Fig 5.6-D, E). Additional treatment will still be needed for special cases as the one shown in (Fig. 5.6-F), where the two lungs got merged together during the closure operation (Fig. 5.6.G) causing an additional hole representing the mediastinum to be formed which will be filled in during hole filling in process (Fig. 5.6-H), an easy and yet efficient way to discern whether the filled in hole is unlikely to be a nodule, is to get its area. Objects have area above certain threshold will not be filled in (Fig 5.6.I), and hence we will get the finally extracted lungs with only holes likely to be nodules restored (Fig. 5.6.J)

5.2.2.3 Challenges still present in the segmentation operation

1-Nodules attached to the mediastinum

For the nodules that are attached to the mediastinum (Fig 5.7-A), if the closure operation failed to isolate the nodule from the mediastinum (Fig 5.7-B), the nodule will be part of the mediastinum (Fig 5.7-C, D), and will be removed during the extraction of the lungs (Fig 5.7-E). For big size nodules that appear in multiple slices, it is very likely for parts of these nodule to survive the segmentation process and be considered as stand alone objects (in 2D approach). An intelligent system might be developed to predict the presence of additional parts of the survived parts in the above or below slices, and start searching for it, and a re-segmentation of the mediastinum might be considered.

2-Small pleural nodules that got removed with the lung wall removal

The thickness of the lung wall to be removed is still a challenging issue that needs more elaborate review. In fact, the lung wall that remains, as in (Fig. 5.5-G) results originally from the closure operation applied, it is not the original pleural which is removed during the thresholding operation. In addition to that, since the roughness of the external outer pixels of each lung depends on the resolution of the slice, it is very expected to have varying thicknesses of the lung wall frames formed. Hence, setting a fixed thickness of a lung wall to be removed with all the scans, might not be a good idea, and will of course lead to the loss of the very small nodules attached to it when it is set to be thicker than needed. As we can see in (Fig. 5.7-F), a relatively small Juxta-pleural nodule will be removed during the lung wall removal. In fact, most of the pleural nodules that this CAD system missed were small nodules (less than 2mm in diameter), and got removed with the lung wall pleural.

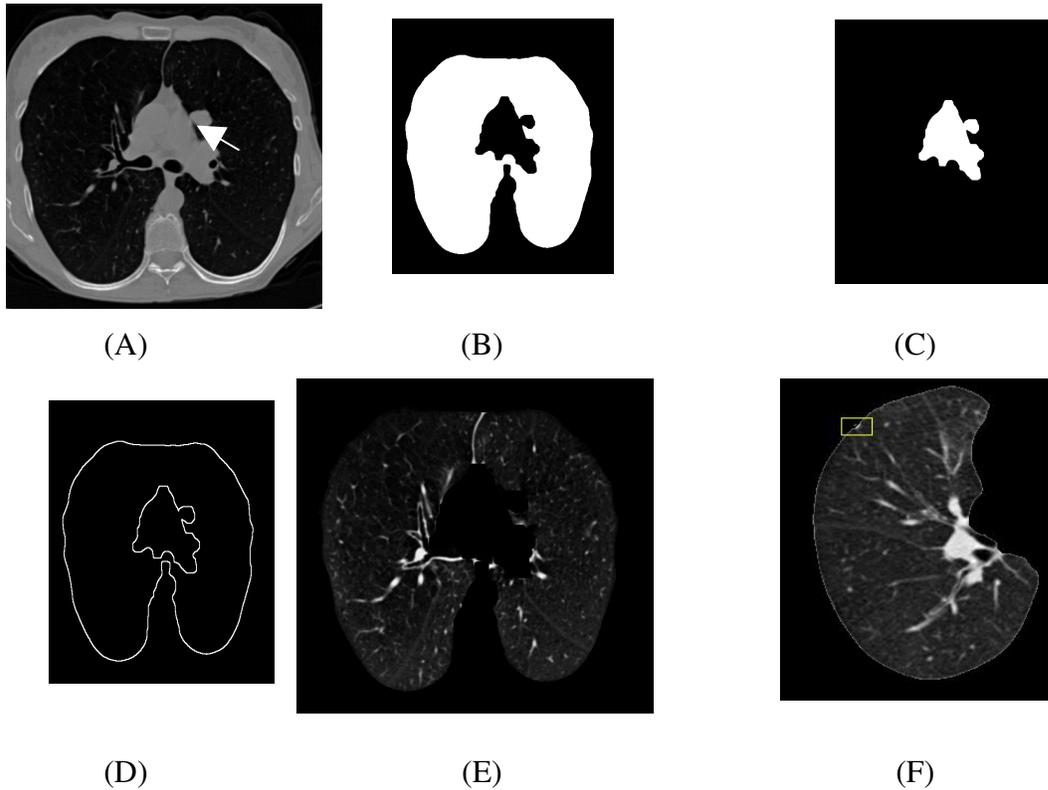


Fig. 5.7 Challenges still present in lung segmentation

5.2.3 Multiple Gray Level Thresholding

The idea of thresholding is basically to set a certain gray level value, where pixels having gray value above it will be assigned the value '1', and pixels below it will be deleted (assigned the value '0').

Most nodules have in general very high gray level value (GLV) relative to the surrounding tissues such as vessels. However, because of the thickness limitation in CT imaging, some parts of the big nodules (on the 3D level) do not occupy the entire section of the slice, and hence, they appear in CT with average gray level value with the tissue above or below it, which is usually the background of the parenchyma. That will eventually lead to these parts appear with low GLV. Moreover, nodules that are smaller than the slice thickness, will also appear in CT with averaged GLV, and will thus have low GLV value. Therefore, there is a necessity to at least have two gray level threshold values, one in the upper band and the other in the lower band, to make sure all the nodules will be included.

In multiple gray level thresholding technique, a threshold, at a relatively high GLV (700), is applied, and the extracted objects are labeled, and then a filtration and region growing is done to the extracted objects to finally get the first group of candidate objects. The process is repeated again, but in the second time the thresholding is applied at relatively low GLV (300), and also labeling, filtration, and region growing were done again to the newly extracted objects (the second group). At the end the two groups of objects are combined together to represent all the objects that could be extracted that have seed points either in the upper band or in the lower band.

As we can see in (Fig. 5.8), thresholding the image (Fig. 5.8-A) at different decreasing gray level values will lead to the extraction of more objects, and also in having some objects got merged together to form bigger objects. We could also notice that, some objects got enlarged in size in the lower GLV thresholds, that means that these objects have gray level distribution that decays towards the outer edges. In fact, most nodules and even vessels have this gray level distribution pattern.

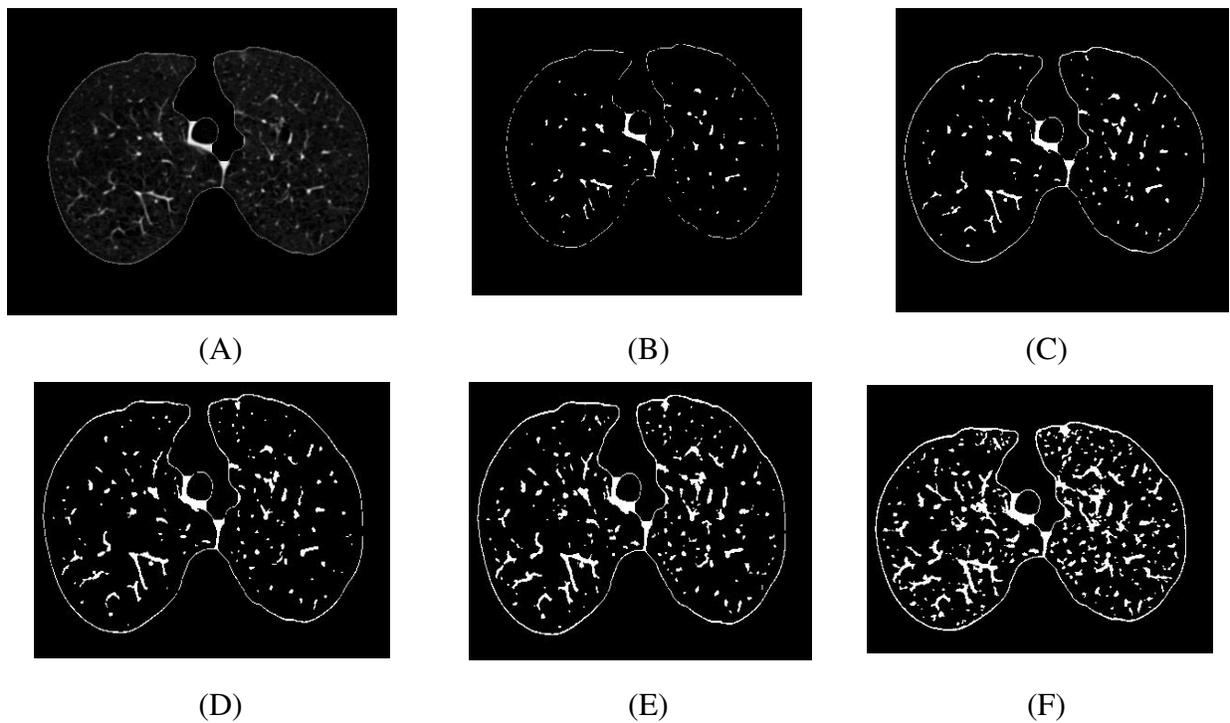


Fig. 5.8 Thresholding at different gray level values

The idea behind applying two different apart gray level thresholds is: First, to make sure that seeds from nodules spanning the high and the low GLV spectrum bands will be included. Second, to make sure that nodules that are attached to the vessels (which usually have lower GLV than nodules) will be detached from them. Third, as per what is known with nodules [6], and even with vessels, the nodules tend to have its central part in high gray level values, which starts to decay exponentially towards the edge, and thus we can easily see in(Fig. 5.8) that many objects had their areas increased in the low GLV.

It is very important to note that, the outcome of the application of each thresholding operation is to just to extract the core part of the nodule (Fig. 5.9-A), which itself alone is very insufficient to have a quantitative measurement of its dimensions, nor a judgment on whether this object has gray level texture that resembles the known one of nodules. Hence, it will be the role of the region growing to incorporate the remaining part of the nodule (Fig 5.9-B).



Fig. 5.9 The core part of nodules extracted at each GLV
A: The central part of the nodule that have the gray level value matching that of the threshold set. B: The role of the region growing to extend the region of this object towards the outer edge.

5.2.3.1 Challenges tackled in the thresholding stage

In this algorithm, seeds are obtained from objects in the high GLV first, and then they are allowed to grow up, meaning they are given a chance to get expanded freely towards the low GLV bands, while preventing them from swallowing other objects or from starting to have irregular geometric shape. Afterward, these objects got removed from the original image because otherwise, their grown parts will appear again in the low gray

level value thresholding. The next step is to apply the low GLV threshold to get the seeds from the objects in the low GLV, which will also be allowed to grow up freely towards the minimum GLV nodules are known to have, which is experimentally found to be 150.

In this way, we will guarantee that; First, objects that are close to each other will be well discerned; Second, nodules that are attached to the vessels will be extracted; Third, all nodules in all GLV bands will be included, Fourth, nodules will have the opportunity to form their final shape.

The values of the high and low threshold values and the minimum GLV nodules are known to have, were all determined experimentally after examining more than fifteen scan sets containing more than 2000 slices. As we mentioned before, the minimum band nodules lie in is because of the slice thickness limitation, otherwise, nodules would only lie in one high band.

It should be noted that, thresholding usually results in the production of some scattered few pixels which are usually deleted when they pass by the filtration stage, and getting rid of them saves significant processing power and time. However, a trade-off is present in this case, since a group of few close pixels might be the seed and the only seed of an object that might eventually grow up in the Region Growing stage to form an object of a considerable size which could be a candidate nodule. More importantly, these few pixels might be an end part of the 3D nodule that has the remaining part in the above or the below slices, and removing these part will cause deformation to the 3D geometric shape of the nodule, which might lead at the end to giving it low chance to survive as a candidate nodule (in the 3D approach). Most importantly, these few pixels might be the onset of a pulmonary nodule at its very beginning stage, that is the utmost valuable objective of having a CAD system, which is to detect nodules at their very early stages, and to start the treatment afterward, the thing that lead to terrific chance of survival.

From the above, a limit value for the number of pixels below which these set of scattered pixels could be considered as noise, and could be safely deleted, has to be determined. Setting this limit value could be based on: First, the resolution of the slices, which is an important factor in determining the area of these small groups of pixels. Second, the minimum size of nodules the algorithm is targeting. Third, the processing limitation of the machine executing the algorithm.

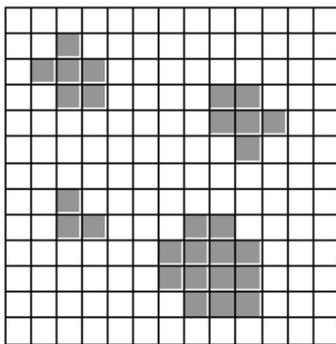
5.2.3.2 Challenges still present in the thresholding stage

For the Juxta-vascular nodules that are in the low GLV spectrum (meaning their initial

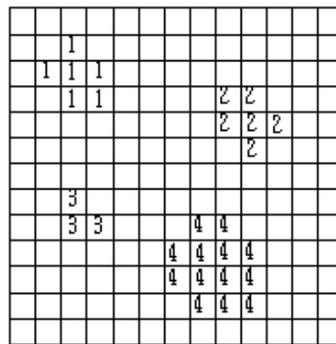
seeds are in the low GLV), it is found that, it is extremely difficult to discern them from nodules. That is why applying more than just two GLV might help at some points to separate them [2], if there is any slight difference in their gray level values. What is really promising is, this problem will not be present with the use CT scans that have smaller slice thickness, because as we mentioned, the reasons that the initial seed of the nodules appear only in the low GLV spectrum is either because they are entirely smaller than the thickness of the slice, and therefore they have their GLV averaged, or because their outer parts are not filling in the entire thickness of the section, and hence these parts are reconstructed in the CT image in low GLV.

5.2.4 Labeling

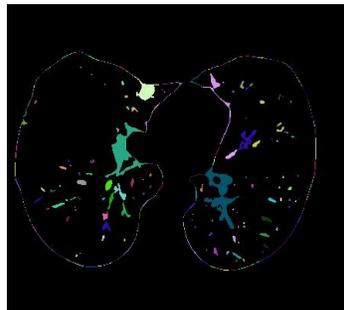
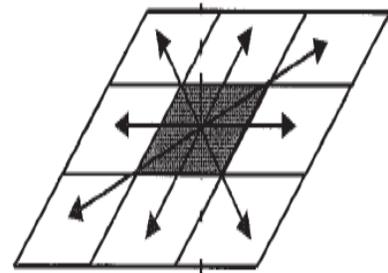
Its basic idea is to assign labels in sequential order to the pixels that are close to each other according to the neighborhood initially defined (Fig. 5.10-A). In this CAD system, 8-neighborhood connectivity in labeling was used, as the one shown in (Fig 5.10-B). The result of the labeling is shown in (Fig. 5.10-C)



(A)



(B)



(C)

Fig 5.10 the Labeling technique

5.2.4.1 Challenges tackled in the labeling stage

The most important one was the fact that pleural nodules have parts of the remaining lung wall labeled with them (Fig 5.11-B,C), threatening them from being rejected. And as mentioned before an acceptable solution was to remove the lung wall. However, the price paid for that was the loss of the very small pleural nodules (Fig. 5.11-A) that have diameter close to the thickness of the wall, and hence were removed during the lung wall removal. Another challenge encountered in this stage was the memory capacity needed by the labeling operation. There was a need to split the stack of slices into smaller groups to overcome the limitation of the memory needed.

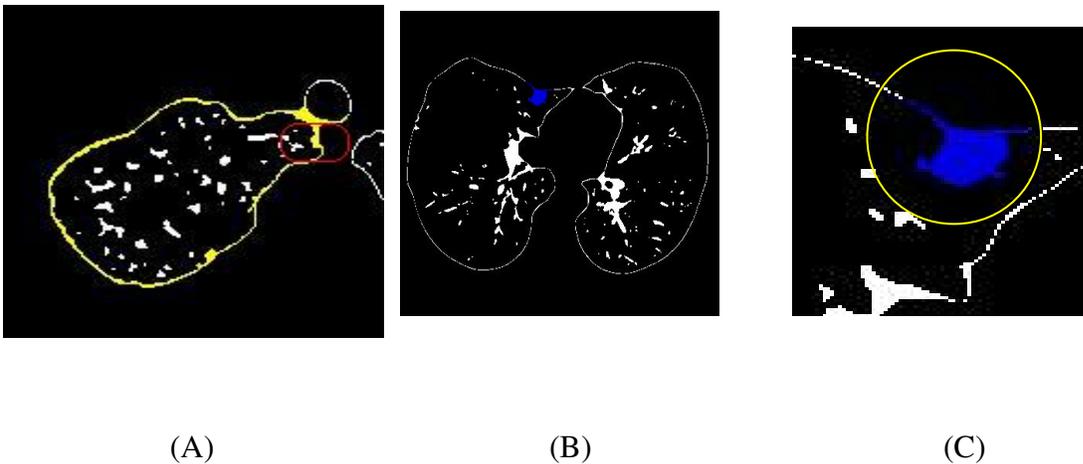


Fig. 5.11 Remaining parts of the lung wall were assigned the same label as the pleural nodule

5.2.4.2 Challenges still present in the labeling stage (Juxta-Vascular Nodules)

Labeling stage is a coronary to what was extracted in the thresholding stage, the stage in which objects are formed. The main problem we can see here is getting juxta-vascular nodules extracted with the part of the vessel (or natural structure) they were attached to, as we can see in (Fig. 5.12-A,B). This problem is a direct consequence to how successful the thresholding operation managed to maintain reasonable separation between the nodule and the vessel (or natural structure) they are close (or attached) to. Otherwise, and as we can see in (Fig. 12-A,B), the nodule and the vessel will be assigned the same label and hence will be treated as a single object, and will most probably get rejected when they pass by the filtration stage, causing a loss to the detection of these nodules.

Juxta-vascular nodules is the most important challenge that face the development of the CAD systems. However, because of the variation in GLV between nodules and the surrounding tissues, the improvement in CT resolution is very promising to lead to an increase in sensitivity of the the extraction of these nodules. In addition, the use of multiple gray level thresholding [2], compared to the use of only two thresholds, will definitely increase the chance of finding a certain threshold at which an acceptable separation between the nodule and the vessel will be found. Apart from using the thresholding approach, many previous researches went for the use of template matching [6], or model-based filter [20] searching for round or oval objects (2D), or spherical or ellipsoid (3D), following the model-based search approach rather than the density-based search approach that has a known limitation in extracting juxta-vascular nodules that have very close GLV to the vessels.

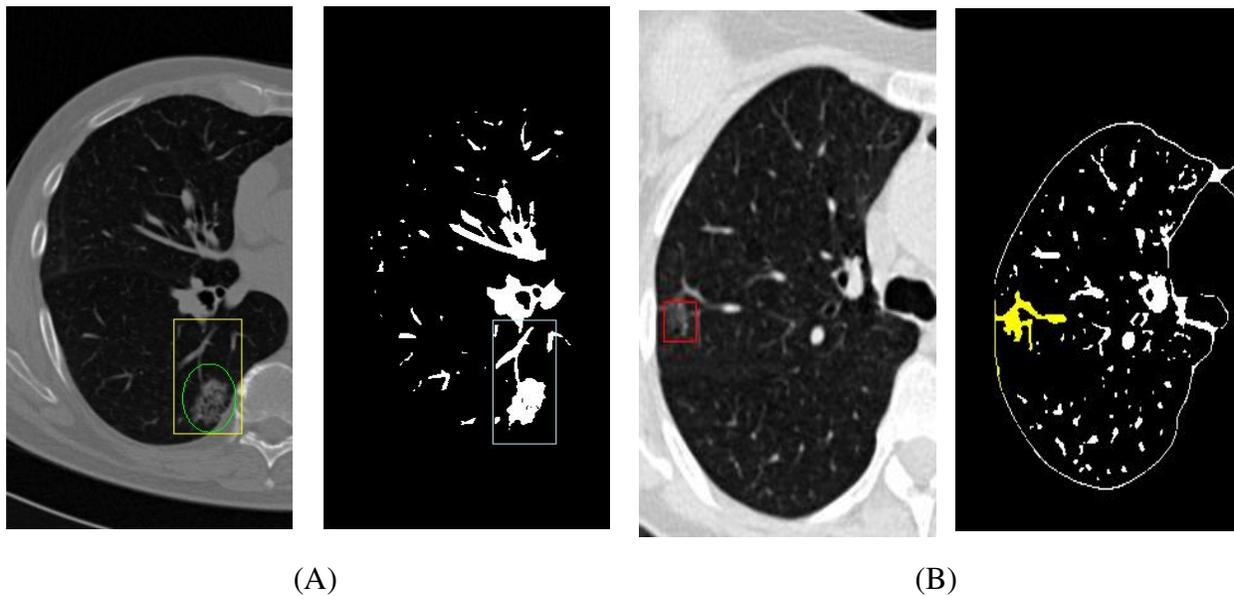


Fig. 5.12 Nodules attached to vessels failed to be extracted during the thresholding operation

5.2.5 Shape Features Extraction (Stage1)

In this stage, the stack of the labeled images are fed into a process to extract some shape features from each extracted object. The features extracted are:

1. **Width** and the **Height** of the object in pixels and in mm

2. **Area, Equivalent Diameter**: Area is the actual number of pixels in the object, while Equivalent Diameter, is a scalar that specifies the diameter of a circle with the same area as the object. Computed as:

$$\text{Equivalent Diameter} = \sqrt{4 * \text{Area} / \pi} \quad (5.2)$$

3. **Eccentricity**: which is a measurement of how circular or ellipsoid the object is (Fig. 5.13-A). It is the ratio of the distance between the foci of the ellipse and its major axis length. The value is between 0 and 1. (0 and 1 are degenerate cases; an ellipse whose eccentricity is 0 is actually a circle, while an ellipse whose eccentricity is 1 is a line segment.)

4. **Extent**: is a scalar that specifies the proportion of the pixels in the bounding box that are also in the region. Computed as the Area divided by the area of the bounding box. In other words, it is the density of the object relative to the bounding box (Fig. 5.13-B). Extent measures how compact the object is.

5. **Density(compactness)**: is exactly the same as Extent, but it is the proportion of the pixels in the object to the bounding square. Density measures how compact and square the object is.

6. **Irregularity**: it is the standard deviation of the distance between the center of the object and all the pixels at the edge (Fig. 5.13-C)

7. **Roundness**: it is a measurement for how round the object is, and is calculated by:

$$\text{Roundness} = 4 * \pi * \text{Areal Perimeter}^2 \quad (5.3)$$

This metric is equal to one only for a circle and it is less than one for any other shape.

8. **Elongation factor**: Ratio of max edge to min edge

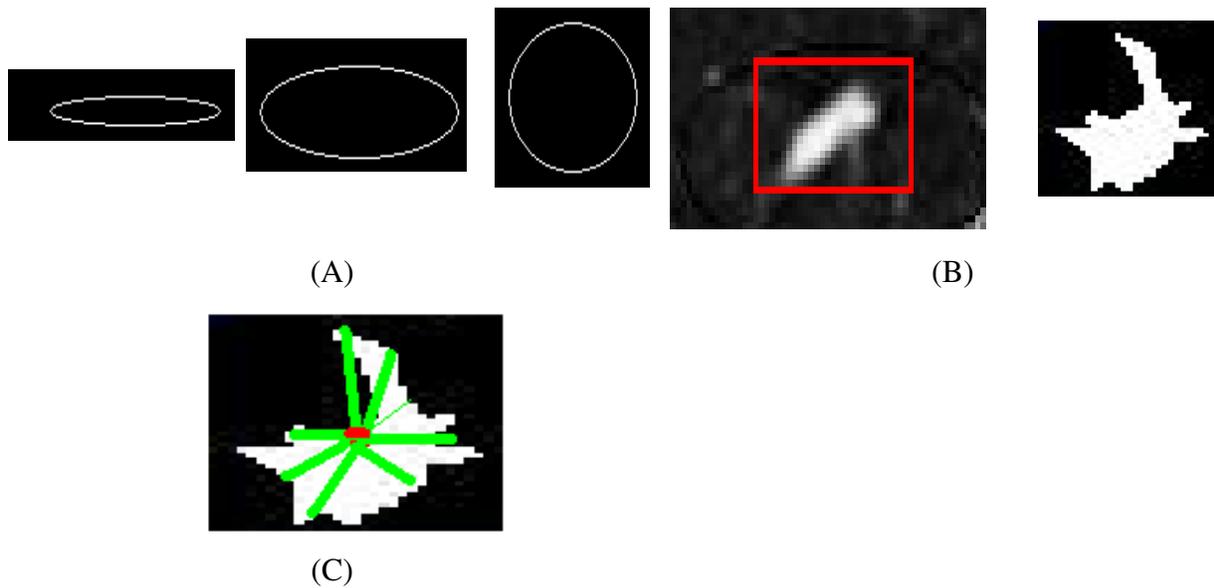


Fig. 5.13 Some Shape Features Extracted

5.2.5.1 Challenges tackled in this stage

An important property to be extracted at this stage for all the objects was a marking flag for whether this object is a pleural object or not; Because, as we will see later, pleural nodules do not tend to have geometric features resemble those of the internal parenchymal objects; Their shape are usually half or part of an ellipse. Moreover, depending on how adhesive the nodule is to the pleural and whether there is a part got removed during the lung wall removal, the final remaining part might be well deformed having irregular shape. Their vulnerable situation to be considered candidate objects exacerbates more during the region growing, because the initially extracted part at the high threshold usually fail to grow as it does not have the valid conditions to grow more. From the above, there was a necessity to tag pleural objects so that they got exempted from being subject to the same filters that other objects will be subject to.

To mark pleural nodules, a stack of frames to the corresponding stack of the segmented slices was constructed, and by doing a binary multiplication to each slice (Fig. 5.14 A:C), the survived objects could be tagged as pleural objects. Getting a frame of the slice is done in a similar way followed during the lung wall removal, by subtracting the original image from the eroded image. Image erosion is done using a kernel of radius shape having a radius of 1, to make it very thin.



(A)

(B)

(C)

Fig. 5.14 Creating an outer frame to be used to label pleural objects

5.2.6 Shape Features Filter (Stage1)

In this filter an experimentally determined limit for the different features were set, and any non pleural object that lies outside this limit will be deleted. As it is mentioned before, the stage1-filter is very lenient in the limits set for the different features, as many objects will have different geometric shape after growing. The main point of extracting the shape feature prior to the region growing and making lenient filtration to the guaranteed not to be candidate objects, is to save processing time, as it is a waste of resources to make region growing to objects that they are very unlikely to have any chance in rectifying their status after region growing. Examples on these objects are objects that have width, height, area, or irregularity that is totally out of the known limits of nodules. Also, objects that have extremely low density, extent, or roundness values.

5.2.7 Region Growing

Region growing is the most complex algorithm in this CAD system and is also the most critical stage because: First, in this stage, the whole external part of the nodule that has GLV lower than the core part needs to be extracted forming one object. Second, extracted objects prior to this stage passed by a very lenient filter in a way to give border line objects an additional chance to rectify their status, by being expanded so that their final geometric shape will be acceptable to pass the successive more aggressive filter. Otherwise, these initially extracted objects will be rejected (deleted).

The primitive idea of region growing as we can see again in (Fig. 5.15) is, to expand the initially extracted part of the objects towards the outer edge to incorporate the area having GLV lower than the core part but is likely to be part of that object. Since nodule tend to have a Gaussian gray level distribution as the one shown in the figure below (Fig. 5.15), we can easily see that the area between the core part and the outer border of the nodule has a varying radially decreasing gray level distribution, hence the task of the region growing algorithm is to make expansion while varying the gray level value of the pixels that could be included. The algorithm should give priority to the expansion that include the pixels of high GLV first, as these pixels have higher likelihood to be part of the that object than neighbor pixels that have lower GLV that could be close noise pixels, such as pixels from normal structure in the lung or vessels.

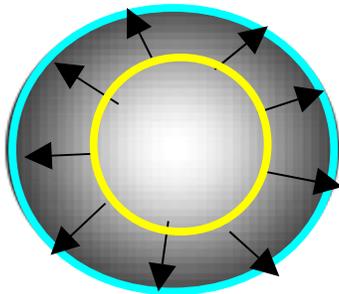


Fig. 5.15 Concept of Region Growing

5.2.7.1 Implementation of Region Growing:

A step decrease in the GLV is set $GLV2$ below the initial threshold $GLV1$, with which the initial core part was extracted. And then a gradual expansion is made to the object using morphological image dilation operation; In every expansion, only pixels that have GLV in that range ($GLV1 - GLV2$) will be included. The operation is repeated until the expansion result in no change in the area of the objects, meaning no more pixels having $GLV > GLV2$. Another step decrease in the GLV below $GLV2$ is set ($GLV3$), and the operation is repeated. The lowest GLV was determined experimentally to be $GLV=150$, and the step is set to equal 50.

Since in this CAD system, the objects are extracted at two different GLVs, one at 700, and the other one at 300; In the first threshold, region growing is done starting from $GLV=700$ till $GLV=150$, while in the second it starts from $GLV=300$ till $GLV=150$, with step decrease=50 in both of them.

The growth of objects are done on the level of the slice, meaning, all objects in the slice are grown together. That has been proven to make significant improvement in the processing time needed. However, a list of a “Not allowed to grow further” objects needed to be maintained, and the algorithm had to restore the previous status of any of these objects to their status before growth. This list contains the objects that their growth yielded unacceptable values of shape features.

After every growth of any object, it is examined by checking some geometric features which are: its area, extent, and roundness. A certain limit is set for a maximum area an object can have, which is determined to be 700 mm^2 , the area is calculate as: $\pi * r^2$, where $r=15\text{mm}$ is the maximum radius nodules are known to have. So much emphasis was put on the roundness feature, so that when objects start to have their roundness value deviate from a certain sharp limit, their growth will stop on the spot.

5.2.7.2 Challenges in the Region Growing Stage

1-Nodules of inhomogeneous texture

In region growing, objects are not allowed to swallow other neighbor objects while growing, even though the finally created object will satisfy the requirements of the accepted objects; from one side, that is so useful to make sure no close vessel will swallow a nodule, or vice versa, maintaining good isolation between them and preventing the loss of the extraction of the juxta-vascular nodules. In this case, each seed extracted during the thresholding stage will end up either being deleted or forming a stand alone object. However, on the other hand, the draw back of this rule is, some nodules have inhomogeneous texture, and they have more than one core region of high gray level value in them, so when they are thresholded, they got split into more than one apart seeds (Fig. 5.16), and will eventually be treated as separate objects, most of the time, these small parts will be rejected altogether, or few parts got left, that are usually cause confusion to the classifier if they were included in the training.

2-Pleural and irregular shape nodules

Since pleural and irregular shape nodule do not satisfy the roundness requirement for the Region Growing operation, they are usually deprived of getting a chance to grown after their first initial part is extracted at the relatively initial high threshold. The consequence of that is, their other part will be left intact to be extracted at the second low threshold, and being considered as another separate object. Hence, we will end up having

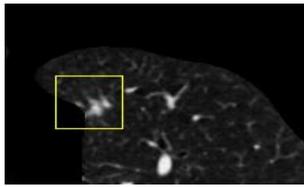
two separate objects (Fig. 5.17). The outer part will usually be deleted afterward, while the central part will be left if it is a pleural, or has acceptable geometrical shape. The left part will cause serious confusion to the classifier if it was a Ground Truth nodule used in training, unless it is deleted manually.

3-Irregular and inhomogeneous shape nodules

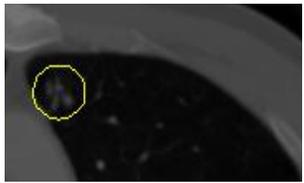
The most difficult types of nodules (Fig. 5.18-A). As a result of being inhomogeneous, they will get extracted in more than one part (Fig. 5.18 B,C,D), and as a result of being irregular, most of these parts will be rejected; and finally some parts might be left which will cause disturbance if used later in either the build up of the 3D shape, or in the classification process, because they have texture that is of no meaning, while these parts will be known in the classifier as Ground Truth and will of course mislead the classifier in the learning stage.

4-Inhomogeneous but regular shape nodules

As we can see in Fig. 5.19, this nodule is inhomogeneous and as a result, the core was extracted at the first high threshold as two parts merged together (Fig. 5.19-B), which of course was rejected to grow and was finally deleted, nevertheless the outer part, got extracted at the second low threshold (Fig. 5.19-C), and was eligible according to its shape features to be considered as a candidate nodule (Fig. 5.19-D). The only problem that will remain is, this finally extracted object (Fig. 5.19-D) lacks the texture a nodule should have, and as a consequence will cause serious confusion to the classifier.

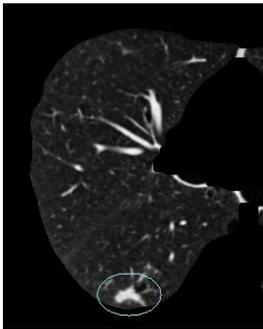


(A)

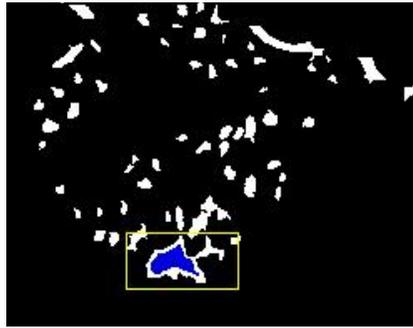


(B)

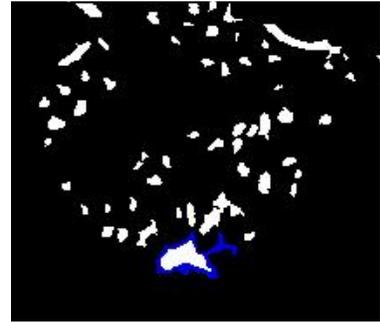
Fig. 5.16 Inhomogeneous objects after being thresholded



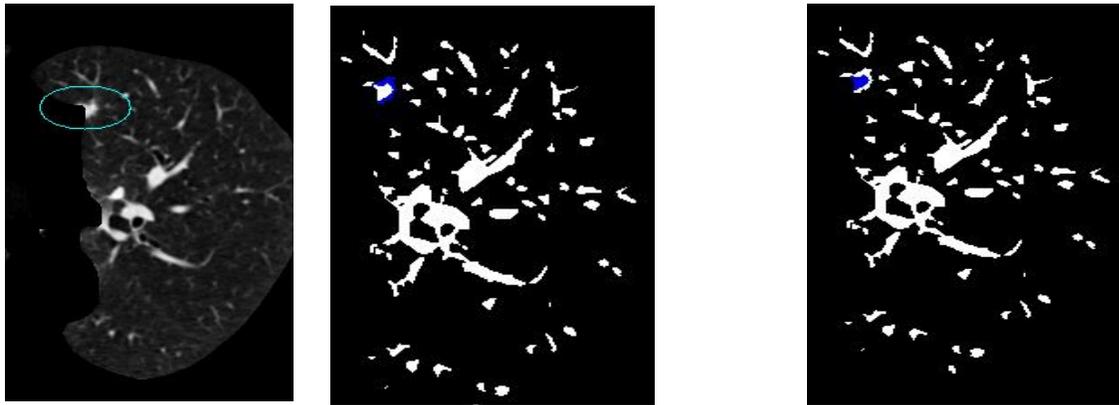
(A)



(B)



(C)



(D)

(E)

(F)

Fig. 5.17 The limitation in region growing to merge the parts of the nodules having irregular shape



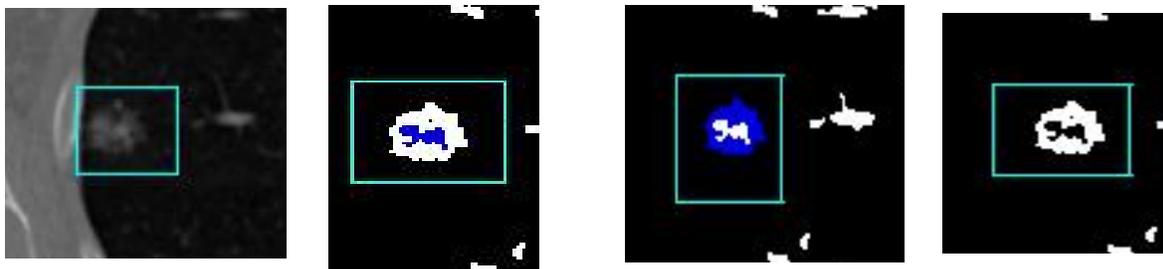
(A)

(B)

(C)

(D)

Fig. 5.18 Inhomogeneous and Irregular shape nodules



(A)

(B)

(C)

(D)

Fig. 5.19 Inhomogeneous but have regular shape nodules

5-Processing time

This stage takes so much processing time, as every single object, from an average of tens of the thousands objects in every scan, is processed separately by being grown at the multiple threshold levels between the upper and lower limit, and in every time, its features are calculated to judge on whether it will be allowed to grow further or not. In case not, it will need to be marked in what is called 'Not to grow more list'. One of the most decisive features to limit the growth operation was the irregularity feature. However, because of a limitation in the processing, it was substituted by the extent feature combined with the roundness, and the efficiency did not differ that much, but a significant reduction in the processing needed was achieved .

5.2.8 Shape Features Extraction (Stage2)

Since Region Growing will make changes to the geometric shape of the objects, their same shape features needed to be extracted again to represent the final shape features of the extracted and grown objects.

5.2.9 Shape Features Filter (Stage2)

This filter is more aggressive than the previous one, because objects were given chance to expand and make their final shape determined.

5.2.10 Texture feature extraction

In this stage several textures are extracted such as:

1-***Fitness***: which is the similarity computed as the convolution of the bounding box of the object with a template generated that has a Gaussian gray level distribution to its pixels [8], with a max gray level value and variance equal to those inside the bounding box of the object.

The template is constructed using the following equation:

$$pv_{x,y} = m \cdot e^{-\frac{(x^2+y^2)}{n}} \quad (5.4)$$

where $pv_{x,y}$ is the pixel value of coordinate (x,y) , and m and n are parameters representing the maximum value and variance of the distribution, respectively.

This value is very useful as a judging parameter on whether this object is likely to be a candidate nodule or not, but that is only if three conditions were satisfied: First, the bounding box encompassing the object is close to a square, otherwise, to have a square size of the object, its bounding box will be padded with lots of zeros, causing significant distortion to it, and hence the fitness computed afterward will be of no meaning. Second, the nodule has a relatively big size, as small size nodules do not give reliable fitness value, as they do not have a well defined Gaussian pattern. Third, nodules should be solid, homogeneous in their texture, and have a Gaussian distribution pattern, meaning have their core with high gray level value that decays towards the edges.

2-MeanGLV, StdGLV, MinGLV, MaxGLV: the mean GLV, Standard deviation of the GLV, Min GLV, and Max GLV. Nodules tend, in general to have mean GLV higher than vessels. However, that highly depends on the resolution of the scan set. For low resolution scans, the value for nodules will not differ a lot than vessels, and using a mix of high resolution and low resolution scans will cause confusion the classifier.

3-The co-occurrence matrix features, which is well known as one of the typical methods in texture analysis to eliminate FPs. Features calculated from the co-occurrence matrix are the **angular second moment, entropy, inverse difference moment** and **contrast**.

These features were calculated in the four directions :0, 45, 90, 135. thus we ended up having 16 features.

4-Entropy

5-Shewness

6-Kurtosis

7-Contrast and **MaxMean:** As defined in [6], contrast is the difference ($|M_c - M_n|$) between the mean CT values in the candidate region (M_c) and the neighboring region (M_n) enclosing it. If this value is high, the candidate has marked contrast and is deemed a FP. The max mean CT value was defined as the mean value of the five maximum pixels in the candidate region.

8-MinGradX, MaxGradX, StdGradX, MeanGradX, MinGradRatio, MaxGradRatio, MeanGradRatio, SkewnessGradX, KurtosisGradX.

The min, max, standard deviation, mean, ratio, skewness, kurtosis of the gradient of the

object in both the X and the Y direction were calculated.

8-Second Moment

5.3 Summary

This chapter covered the main techniques used for developing a CAD system, starting from reading the CT scan sets, then extracting the lungs from the thorax by using a threshold value that separates the lungs from the second dominating object which is the fat and muscles. A certain threshold value in the histogram of the thorax image that lies in the middle distance between the two peaks is the optimum threshold value to select, to guarantee full segmentation of the lungs. Then multiple thresholding was applied to extract the objects, which got labeled, and their features were extracted. Candidate objects were determined to be the objects that share common features values with those for nodules, such as roundness, density, extent, fitness, etc. Some texture features were also extracted, primarily to be used during the classification stage, however, a texture value called “fitness” was used in doing filtration to the extracted objects based on the fitness feature value they have.

Chapter 6

False Positives Reduction

6.1 Introduction

Although the extracted objects have already passed through three filters, and although enormous amount of the extracted objects were filtered out, still the number of the remaining objects, which are called False Positives (FPs) is still high making the usage of the system at this stage not to be feasible at all. Therefore, a complimentary stage is always conducted after any most CAD systems, which is to work in reducing the number of FPs making it as minimal as possible while keeping the accuracy of the extracted true nodule intact as possible as we can. In this chapter, a comparison is conducted between following three different approaches to see which one will lead to the best reduction in the number of FPs. First, a 3D approach was followed, in which objects were filtered according to their 3D geometric shape. Second, extracted objects from the 2D approach obtained before were fed into classifier for classification operation; Third, a combined 2D and 3D approach was followed, in which objects got extracted by 2D and 3D filters.

6.2 Developing a CAD system using 3D approach

On a serial-section lung CT slice, a cylindrical vessel can appear circular, and many vessels in the lung have a diameter that is similar to the lesions of interest. If the CAD systems detect the candidates based on the 2D image features, then thousands of candidates occur. That is exactly what was encountered in the algorithm implemented, a scan set of average 120 slices lead to around 16000 objects, and after passing them through the three filters, the number drops to around 6000, which is still enormous. Unfortunately, and any attempt to reduce this number by making the filters more aggressive will lead to the omission of some true positives. In general, the experienced radiologists look for lung nodules not by independently considering individual image slices, but by searching through the serial images for the characteristics of the 3D appearance that distinguish nodules from vessels. In addition to that, it is very useful to benefit from the difference in the 3D geometric shapes between nodules and regular vessels. Nodules tend to be spherical in shape, with their diameters vary along the Z-axis. That all inspired many CAD developers to build the 3D structure of the objects, extract their features, and trying to make their classification based on that.

6.2.1 The algorithm for the 3D approach

The algorithm used for extracting nodules using 3D features of the objects is shown in Fig. 6.1, which proceeds through the following operations:

6.2.1.1 3D labeling

As we can see in (Fig. 6.1), after the extraction of the initial 2D objects, they undergo the first filtration stage, after which the very unlikely to be candidate objects are removed. Right after this stage, 3D labeling operation is conducted to build 3D objects from the corresponding 2D objects, using 10-neighborhood connectivity (Fig. 6.2-A). The idea behind building the 3D objects at this stage is, first, to get rid of some objects that are going to be deleted anyway during the 3D filtration, and thus we will save unwanted overhead in the processing time of the whole algorithm. It is important to note that the processing of 3D objects far exceeds in time that of the 2D objects. Examples of these objects that got removed at the very beginning could be: remaining parts of the mediastinum, longitudinal vessels running across the slices, big parts of the bronchioles, and so on. Second, building the 3D objects before their growth is important to maintain boundaries between objects, before some of them get very close in a way it will be impossible to differentiate them as separate 3D objects. Examples of that is the successfully extracted juxta-pleural nodules thanks to the thresholding applied, and because of the differences in their gray level values. In the 2D stage, the detached nodule will be assigned a different label, which will be maintained before and after growth. In addition to that, and as we mentioned before, the 2D Region Growing algorithm is designed in a way to prevent close extracted objects to get merged again, or for one object to be swallowed by another one. However, 2D Region Growing algorithm might leave close objects with zero boundaries between them, and as a consequence of that, these objects will be assigned the same 3D label during the 3D labeling stage, unless the 3D labeling is done before the 2D objects underwent the 2D Region Growing.

6.2.1.2 3D Region Growing

After the 3D objects are built from the 2D objects before growth, a reference mapping between each 2D object and its corresponding 3D part is maintained, this reference mapping is used to replace each 2D part in the 3D structure with the new grown 2D part, and hence we finally get grown 3D objects from the corresponding grown 2D objects.

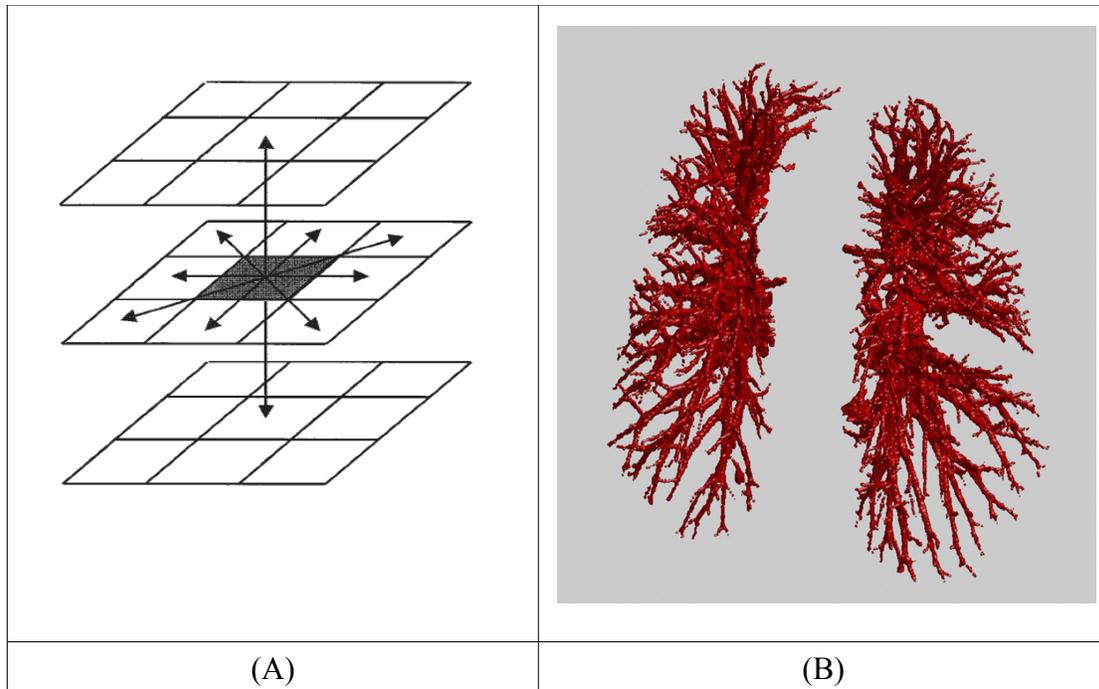


Fig. 6.2

A:10-Neighborhood connectivity labeling 3D objects, B:The vascular structure built using 3D labeling

6.2.1.3 3D features extraction

Many geometric and texture 3D features are extracted from the objects such as:

1-Width, Height, Depth, Volume, Equivalent Diameter, : The width, height, depth, and volume of the 3D bounding box. The Equivalent diameter is calculated as the diameter of the equivalent sphere having the same volume as the that of the object.

2-Compactness (Density): is defined in the same way as in 2D features, it is the ratio of the number of voxels in the object to the number of voxels in the bounding cube.

The bounding cube is calculated as the square of the maximum horizontal edge (MaxHorizEdge) times the depth (D) in pixels.

$$\text{No. of pixels in the bounding cube} = \text{MaxHorizEdge}^2 * D \quad (6.1)$$

3-Extent: is defined as the ratio of the number of voxels in the object to the number of voxels in the bounding box.

$$\text{No. of pixels in the bounding box} = H * W * D \quad (6.2)$$

where H,W, D are the height, Width, and Depth in pixels respectively

4-Ratio (H/W): The max edge of the bounding box to the minimum edge

5-Ratio (Depth/MaxHorizEdge) [Elongation Factor]: Of the bounding Box

6-Standard deviation of the Equivalent diameter of corresponding 2D objects

7-Spherical Disproportion: Which is the ratio of the surface area of the object to the surface area of the corresponding sphere that has the same equivalent diameter. This feature is very useful in measuring the sphericity of the object.

8-Mean GLV, StdGLV, MinGLV, MaxGLV

9-Fitness: Is calculated as the average fitness of the corresponding 2D objects

10-Entropy

11-Max Mean: calculated as in 2D objects, it is the max mean of the highest GLV in five voxels.

12-Contrast: It is the same as in 2D objects, it is the absolute difference between the Mean GLV inside the bounding box of the object and the Mean GLV in the 50% larger in size neighborhood box surrounding the bounding box of the object.

6.2.1.4 3D filter

The 3D filter works on some features where the limit value for nodules are known, and hence filtration of unlikely to be candidate nodules could be removed. These features are:

Width, Height, Depth: Nodules enclosed in a bounding should not have any of this value exceeding 3cm.

Compactness and Extent limit values were set experimentally.

Ratio (H/W) and Ration (D/MaxHorizEdge) are set not to exceed '2', as it is not expected for a nodule to have a height to depth ratio greater than 2

Volume: The focus is made on the nodules that have a diameter ranging from 3mm-30mm, and hence the limits for the volume is set to be between 14.14-14140 mm³

Notes:

1. The width and height were calculated to be those of the bounding box rather than being the max width or high (of the corresponding 2D objects along the vertical slices that constitute the 3D object), that helps in identifying vessels that are running with an angle along the Z-direction, as these 3D objects will have Width to Height proportion problem, and will get filtered out by the Ratio (H/W) feature
2. Vessels also have poor extent and compactness values no matter whether they are running vertically or horizontally.

6.2.1.5 Classification

The features extracted underwent a TTest to get those that reject the null hypothesis and hence will be influential when used by the classifier, and only the found to be influential features were fed into the Support Vector Machine Classifier to make the final judgment on which among the candidate objects are nodules, based on the learning stage the classifier underwent. An SVM package called LIBSVM was used for that.

6.2.2 Major challenges in the 3D approach

1-The removal of an internal 2D slice from the 3D object

The removal of any 2D internal part, during the first 2D filter they pass through before participating in the formation of the 3D object, will result in splitting the original 3D object into two separate 3D objects. The thing that is very likely to happen to any big in size object that spans more than one slice, as it is very rare for a nodule to be found utterly free in the air, or to have all its corresponding 2D slices got successfully detached from any vessels or natural structure they are attached to. As a consequence of that, the two separate 3D objects extracted will usually have very unacceptable geometric proportions, compactness, and extent values, and will most probably be deleted during the 3D filtration stage. A way to workaround that could be to build the 3D objects from the originally created 2D objects before they undergo filtration stage (Fig. 6.1); However, and as mentioned before, the price paid will be: First, getting so many 2D objects, the thing that will

require huge processing power to process all the created 3D objects; Second, still the problem that a 2D part attached to a vessel will not be solved, because what will happen is, this part will not be deleted but will be labeled as a part of the 3D object, and hence the final shape features of the 3D object will be deformed; Moreover, this part might lead to having a section from the vascular tree got labeled as one object with it (Fig. 2-B).

This truly is the main challenge that stands as the main obstacle from getting reliable results from the extracted 3D candidate objects.

2-Slice thickness limitation

It is the second challenge in importance that is present before the wide and reliable use of 3D features in CAD systems. With having pixel size in the range of 0.6mm in the horizontal space, and in the range of 2mm in the vertical space, it is very difficult to infer meaningful proportions of the objects that can help us to judge on whether this object could be a candidate object or not, in particular, with small in size objects, that are hardly seen spanning more than one slice because of the slice thickness limitation. That will eventually lead to the enormous, hard to get rid of, number of Fps. Slice thickness limitation also badly affect the mean gray level of the nodules, in particular, small nodules that got their gray level value averaged with the surrounding tissues in the parenchyma.

3-No Region Growing along the Z-Direction

In my algorithm, the 3D region growing was done along the horizontal direction, neglecting a possible growth to the object in the Z-direction. Implementing a growth along the Z-direction is so complex and will require very high processing power, but will of course lead improvement in the accuracy of detecting the nodules as 3D objects.

4-The wrongly connected objects along Z-direction

Although the 3D labeling was done using only 10-neighborhood connectivity, and because of the limitation in the slice thickness, it is not uncommon for an object of interest to be wrongly connected and labeled with another different object found below or above it in the parenchyma, causing damage to the final geometric and texture features of that object, and threatening it for being deleted.

6.2.3 Summary

The 3D approach, although gave poor results compared to the 2D approach, it made significant reduction in the number of false positives. With the advances taking place in the CT imaging, and the reduction of slice thicknesses, the use of

3D approach will give superior results compared to the 2D approach. A future work to tackle the problem that a 2D middle part of the 3D object could be filtered out could be, to have a smart system that can check the presence of an above or below deleted part, and then try to extract it (or reconstruct it) based on the information present in the parts above and below it. That could be by doing some kind of interpolation, which is the same operation done any way during the reconstruction process in CT imaging. Of course, with small in size slice thicknesses, the accuracy of doing that will be very acceptable.

6.3 Classification of the extracted 2D candidate objects using SVM Classifier

Although the extraction accuracy for the 2D approach was quite pleasing, exceeding 90% in some scan sets, the number of FPs remained represent a big challenge in making this algorithm feasible to be used. Hence, there was a need to go for a classifier to help, based on the numerous texture and shape features extracted, to make some reduction in the number of FPs.

A TTest was first conducted to get the most influential features, and then these objects with the features selected were fed into the Support Vector Machine (SVM) classifier to make final judgment on which of the candidate objects are classified as nodules. I used an SVM package called LIBSVM.

6.4 Combined 2D & 3D candidate objects

To make reduction in the number of FPs, as we can see in Fig. 1, a consultation operation take place between the extracted 2D objects and the extracted 3D objects, in which the corresponding 2D objects of the 3D deleted objects got removed; making significant reduction in the number of FPs, but of course affected the initial accuracy of the extracted 2D objects. That return for the first place to the limitations the 3D algorithm currently has. Then the reduced number of candidate objects obtained are fed into the classifier to get the finally detected nodules.

Chapter 7

Results

The algorithm was applied on 18 scan sets containing 3361 slices, each image is in DICOM format and has 512*512 pixels. The horizontal resolution of all the scans ranged from 0.51mm to 0.76mm. Scans used have slice thickness =2mm except one scan set has slice thickness 2.5mm. The number of slices in each scan set varies from 128 to 159 slices, with an average=143. The nodules examined in this study ranged in diameter from 1.3 mm:18.27 mm (for the 2D approach), and from 1.5mm:18.3 mm (For the combined 2D&3D approach), and from 3.05 mm:19.9 mm) for the 3D approach.

In the 2D approach, the minimum limit set for an accepted diameter was 1.2 mm, but for the 3D approach, it is 3mm. The reason for that was because in 3D, we are limited by the thickness of the slice used which is 2mm, and hence by setting the diameter to be larger than 3mm, we will expect the true positive objects to span more than one slice, and hence we will be able to make considerable reduction in the false positives. Of course, the price paid for that is, the limitation of the 3D algorithm to detect nodule smaller in diameter than 3mm.

The difference in the definition of the nodules between the 2D and the 3D approach should be made clear. Nodules, as we mentioned before, are 3D objects that tend to have spherical shape. By defining a 2D object as a nodule, we are referring to the 2D presence of the nodule in a certain slice. In examining the results of the 2D and the combined 2D&3D approaches, nodules are defined in a way that has no relation between each other in terms of which 3D nodule they represent, meaning 2D nodules are treated as separate entities.

Although the 2D led to enormous number of FPs, it has superior advantages over the other two approaches in that; First, it gave the highest detection accuracy; Second, it works on nodules of very small diameters, starting from 1.3 mm; Third, it gives higher chance of detecting the majority of nodules when they are treated as a 3D object, because the likelihood for that approach to detect any presence of the 3D object in any of the slices it spans is very high; Compared to the 3D approach, that any omission to a 3D object, means the full deletion of any presence of that

object in the output result. The use of the Support Vector Machine, although tested in a limited scale, gave promising results, in terms of significant reduction to the number of FPs, while making reduction to the sensitivity reached in the initial detection of the candidate objects.

7.1 Results obtained before using the classifier:

7.1.1 Result of the 2D approach

From a total of 313 nodules, the system detected 259 nodules (82.75%), with a total number of FPs= 49/Slice.

7.1.2 Result of combined 2D and 3D approach

From a total of 313 nodules, the system detected 202 nodules (64.54%), with a total number of FPs= 16.9/Slice.

7.1.3 Result of the 3D approach

From a total of 196 nodules, the system detected 146 nodules (74.49%), with a total number of FPs= 1.42/Slice.

A summary of these results is depicted in (Table 7.1)

Table 7.1 All results obtained before using the classifier

	2D	Combined 2D with 3D	3D
Accuracy	82.75%	64.54%	74.49%
FPs/Slice	48.95	16.9	1.42

7.1.4 Detected Nodules according to the diameter

Table 7.2 Detected nodules according to the diameter

	Total Detected	d<3mm	5mm>d>3mm	d>5mm
2D	259	26 (10%)	105 (40.5%)	128(49.4%)
Combined 2D&3D	202	4(1.98%)	85(42%)	113(55.9%)
3D	146	0	103(70.5%)	43(29.5%)

7.1.5 Detected Pleural Nodules

Table 7.3 Detected Pleural Nodules

	Total Nodules	Pleural Nodules	Number detected	Accuracy
2D	313	71 (22.6%)	58	(81.6%)
Combined 2D&3D	313	71(22.6%)	52	(73.2%)
3D	209	56(26.79%)	37	(66%)

7.1.6 Missed 2D Nodules

Table 7.4 gives an analysis on all the 2D nodules missed

Table 7.4 Analysis on the missed 2D nodules

	Total Nodules Missed	59
1	Attached to mediastinum and got removed during the lung segmentation stage	1(1.85%)
2	Very small pleural nodules got removed with the plural frame during the lung segmentation stage	13(24.07%)
3	Non homogeneous nodule	1(1.85%)
4	Nodules had irregular shape; some had oval shape, others had steak like shape, others had triangle shape.	9(16.67%)
5	Nodules were attached to vessels (Juxta-Vascular nodules) or normal structure within the lung parenchyma.	14(25.93%)
6	So small, and thus has very low GLV to be extracted even at the low GLV threshold.	8(14.81%)
7	Nodule had low GLV, but was not that small.	3(5.56%)
8	Nodules were like a burst mass or a spectrum with no defined shape at all.	3(5.56%)
9	Nodules had very low contrast relative to the parenchyma. Most probably, there was distortion in the imaging of these scans.	2(3.70%)

7.1.7 Missed 3D Nodules

Out of the 50 3D nodules deleted, 26 were deleted because they have diameter less than 3mm, the others were missed because they gave poor compactness, extent, and/or ratio (width/depth). Most probably because the nodule spanned more than one slice, while one or more slices were filtered by the initial 2D filter, leading to the end to the nodule being split into two or more disfigured 3D parts. Another frequent for the 3D nodule to be deleted is, when it got labeled with any other normal structure in the parenchyma, the thing that will result in an object with deformed structure to be considered as a nodule.

Note

The number of the 3D nodules missed did not include the original 2D nodule missed at the very beginning, before the construction of the 3D objects. Example of these nodule is the 13 nodules that were removed during the lung wall removal process.

7.2 Results obtained after using the classifier:

For all the three different approaches followed in this research, the 18 scan sets used were split into two groups, a training set of 15 scan set, and a testing set of 3 scan set. For the 2D features, a ttest was first applied on a list of 86 features, folded seven times, and picked only 56 features that rejected null hypothesis in the TTest. For the 3D features, a TTest was also applied first on a list of 22 features, and only 16 features were picked that rejected null hypothesis.

For Sensitivity and Specificity defined as:

$$Sensitivity = \frac{TPs}{TPs + FNs} \quad (7.1)$$

$$Specificity = \frac{TNs}{FPs + TNs} \quad (7.2)$$

The sensitivity, specificity, and Fps/Slice calculated from the output of the classifier is depicted in Table 7.5

Table 7.5 Result after applying the SVM

	2D	Combined 2D&3D	3D
Sensitivity	19.78%	16.60%	11.53%
Specificity	98.40%	97.99%	94.30%
FPS/Slice	0.015/Slice	0.02/Slice	0.015FP/Slice

7.3 Discussion

Although SVM resulted in enormous reduction in the number of FPs, the sensitivity got affected severely. However, by making adjustment to the parameters of the classifier, we might end up having better sensitivity, even with an increase in the Fps/Slice. Also, because of the huge number of slices either in the training set or in the testing set, I opt to use only the slices that have nodules in them for the 2D and the combined 2D&3D approaches. For the 3D approach, I selected only the top slices where nodules were found in.

Using more scan sets, with more nodules, will definitely lead to improvement to the final results obtained.

In addition, it might be better to make separate classification to the plural objects. Because these objects apparently causing serious confusion to the classifier, as they do not share neither the geometric nor the texture features with the other parenchymal or vascular nodules.

Chapter 8

Conclusions and Future Work

8.1 Conclusions

This research resulted in the development of algorithms for the thorax and lung segmentation, region growing, multiple gray level thresholding, shape and texture features extractions, building rule-based filters, the use of SVM for classification.

Although my CAD system developed is still preliminary, the results for nodules detection is promising. The system extracted 259 nodules out of 313 (for the 2D approach), and 146 out of the 196 (for the 3D approach) The missed nodules were predominantly smaller than 3mm in the 3D approach, or got removed with the lung wall during the segmentation process.

To ensure that nodules abutting the lung border were not unintentionally excluded from the lung parenchyma, the CAD system does some morphological operation to the image including, dilations, hole filling, lung wall removal. The CAD system accounted for the difficulty of detaching juxta-vascular nodules from the vessels they are attached to, by using multiple gray level thresholding.

Method used in extracting juxta-pleural nodule needs to be developed further to account for the small nodules that got removed during the removal of the lung wall. The high number of false positives is a very challenging issue, and a further analysis to the performance of the classifier is needed. Further work is needed to deal with the nodules not from the perceptive that they are round, have well defined boundaries, have Gaussian distribution gray level values, compact in shape, have gray level value higher than the surrounding tissue. As these all are not the characteristics of the majority of nodules. However, many nodules were missed because their characteristics were not compliant with these ones. Hence, the algorithm needs to deal with all the types, different structures of nodules not with a specific class of them.

The juxta-vascular nodule was and still a big challenge to many researchers, as their gray-level values are hardly different from the vessels they are attached to. Nevertheless, they still retain having a spherical or semi-sphere shape, the thing that strengthen the model-based approaches that applies filter or use template matching, that look for spherical-like shapes. That is probably the most effective way of extracting them.

Further work is needed to improve the classifier results, better selection to a

kernel, reduction in the number of features used, feeding more ground truth nodules for training, and isolation of the pleural nodules. All these will lead to its improvement.

This preliminary computer aided diagnosis system demonstrates the potential for a potentially useful automated nodule detection system. With further development, such a computer system could be applied in real diagnostic cases.

8.2 Future work

Although the CAD system developed in this research showed some promising results in the extraction of the pulmonary nodules, there are still many improvements need to be done to this system to find its way being used in practice. A future work on this system could include the following:

1-Considering the nodules as ellipsoid object [20] rather than spherical object, as it was considered in this system, is highly expected to improve the detection accuracy of the candidate nodules. As it was found from this research, spherical shape of nodule is the ideal model, which is not always case with the majority of the nodules found.

2-Doing a removal of a fixed width lung wall led to the loss of many small juxta-pleural nodules, a development in the system can, based on the resolution of each individual scan, determine a varying thickness of the lung wall to be removed.

3-As it is mentioned in the 3D approach, the main challenge that hinders the free reliance on the 3D approach, is the loss of an middle part of the 3D object, leading to getting two separate apart 3D objects, usually non of them will be eligible to be considered a candidate object. A future work, could try to compare centroids of two vertically aligned 3D objects, and expects the presence of an in-between missed part, and them merge them all forming one part.

4-The way region growing followed in this algorithm was proven to be that efficient, particularly with irregular and inhomogeneous nodules. The idea followed by Serhat et. al.[11] in searching for neighbor objects and to be bound by only distance and densities, might be more efficient that relying mainly on the roundness feature to decide whether to allow for further growth or not.

5-For the classifier to give acceptable results, a feature selection might need to be done to reduce the number of features which might improve its performance, also the number of Ground Truth nodules used for training needs to be increased. It seems also that doing separate classification to the pleural nodules might give better performance, as these nodule have significant difference with other nodules.

References

- [1] Manfred Tillich, “*Detection of Pulmonary Nodules with Helical CT: Comparison of Cine and Film-Based Viewing*”. AJR 1997; v169:1611-1614.
- [2] Armato, S.G., Giger, M.L., Moran, C.J., Blackburn, J.T., Doi, K. and MacMahon, H., “*Computerized detection of pulmonary nodules on CT scans*”. Radiographics 1999; v19: 1303-1311.
- [3] Ko, J.P. and Betke, M., “*Chest CT: automated nodule detection and assessment of change over time-preliminary experience*”. Radiology 2001; v218: 267-273.
- [4] G. Samuel, III, Armato, M. Geoffrey, F. Michael, R. Charles, Y. David, “*Lung image database consortium—developing a resource for the medical imaging research community*”, Radiology 24 (2004) 739–748.
- [5] Van Ginneken B, ter Haar Romeny BM, Viergever MA. “*Computer-aided diagnosis in chest radiography: a survey*”. IEEE Trans Med Imaging 2001; 20:1228–1241.
- [6] Y. Lee, T. Hara, H. Fujita, S. Itoh, and T. Ishigaki, “*Automated detection of pulmonary nodules in helical CT images based on an improved template-matching technique,*” IEEE Trans. Med. Imaging, vol.20, no.7, pp.595–604, 2001.
- [7] Zerhouni EA, Spivey JF, Morgan RH, Leo FP, Stitik FP, Siegelman SS. “*Factors influencing quantitative CT measurements of solitary pulmonary nodules. J Comput Assist Tomogr*” 1982; 6:1075–1087.
- [8] Shiyong Hu, Eric A. Hoffman, Member, IEEE, and Joseph M. Reinhardt. “*Automatic Lung Segmentation for Accurate Quantitation of Volumetric X-Ray CT Images,*” IEEE Trans. Med. Imaging, vol.20, no.6, pp.490–498, June 2001.
- [9] Tuddenham WJ. “*Glossary of terms for thoracic radiology: recommendations of the nomenclature committee of the Fleischner Society*”. AJR Am J Roentgenol 1984; 143: 509–517.
- [10] Travis WD, Colby TV, Corrin B, Shimosato Y, Brambilla E, Sobin LH.

- “Histological typing of lung and pleural tumors: World Health Organization international histological classification of tumours series”*. Berlin, Germany: Springer-Verlag, 1999.
- [11] Ozekes, Serhat; Osman, Onur; Ucan, Osman N. *“Nodule detection in a lung region that's segmented with using genetic cellular neural networks and 3D template matching with fuzzy rule based thresholding”*. Korean Journal Of Radiology 9:1 (2008) 1-9, doi:10.3348/kjr.2008.9.1.1
- [12] Melissa Conrad Stöppler, *“Lung Cancer, causes, types, and treatment”*, retrieved from MedicineNet Inc., http://www.medicinenet.com/lung_cancer, [8 Sep. 2009]
- [13] David C. Dugdale et al., *“Chest radiography”*, retrieved from MedlinePlus, <http://medlineplus.gov/> [8 Sep. 2009]
- [14] Cheng HD, Chen JR, Li J. *“Threshold selection based on fuzzy c-partition entropy approach”*. Pattern Recognition 1998;31:857-870
- [15] American Cancer Society. *“ACS cancer facts and figures 2002”*. American Cancer Society, Atlanta, GA, 2003.
- [16] J. W. Gurney: *“Missed lung cancer at CT. imaging findings in nine patients Radiology”* no. 199, 1996.
- [17] Flehinger BJ, Kimmel M, Melamed MR. *“The effect of surgical treatment on survival from early lung cancer: implication for screening”*. Chest. 1992;101:1013–1018. [[PubMed](#)]
- [18] Sobue T, Suzuki R, Matsuda M, Kuroishi T, Ikeda S, Naruke T. *“Survival for clinical stage I lung cancer not surgically treated. Cancer”*. 1992;69:685–692. [[PubMed](#)]
- [19] L. Ries et al. SEER *“Cancer Statistics Review 1973–1996”*. National Cancer Institution, Bethesda, MD, 1999.
- [20] Martin Dolejší and Jan Kybic *“Automatic two-step detection of pulmonary nodules”*. Proc. SPIE, Vol. 6514, 65143J (2007); doi:10.1117/12.709161.