Computer-Aided Diagnosis of Tuberculosis in Paediatric Chest X-Rays Using Local Textural Analysis

Author:
Andre Mouton (MTNAND005)

Supervisor:
Dr Tania Douglas

SUBMITTED TO THE UNIVERSITY OF CAPE TOWN
In partial fulfilment of the requirements for the degree of
Master of Science in Medicine, in Biomedical Engineering

Faculty of Health Sciences
University of Cape Town

July 2009

Declaration

I, ..., hereby declare that the work on which this dissertation is based is my original work (except where acknowledgements indicate otherwise) and that neither the whole work nor any part of it has been, is being, or is to be submitted for any other degree in this or any other university.

I empower the university to reproduce, for the purpose of research, either the whole or any portion of the contents in any manner whatsoever.

Signature: .....
Date: .....

Acknowledgements

I would like to thank the following for their contribution to this project:

- Dr Tania Douglas for supervising the project.
- Dr Richard Pitcher for analysing and annotating the data.
- Dr Rupesh Daya for his guidance and advice regarding TB study images.
- Dr Bram van Ginneken and Laurens Hogeweg of the ISI research group at the University Medical Centre Utrecht for their guidance and for making my 2 months in Utrecht the most productive period of my project.
- UCT's Lodox programme for their generous financial support.

Abstract

This report presents a computerised tool to analyse the appearance of the lung fields in paediatric chest X-rays to detect the presence of tuberculosis.

The computer aided diagnosis (CAD) tool consists of 4 phases: 1) lung field segmentation; 2) lung field subdivision; 3) feature extraction and 4) classification. Lung field segmentation is performed using a semi-automatic implementation of the active shape model algorithm. Two approaches to subdividing the lung fields into regions of interest are compared. The first divides each lung field into 21 overlapping regions of varying sizes, resulting in a total of 42 regions per image; this approach is called the big region approach. The second approach divides the lung fields into a large number of overlapping circular regions of interest. The circular regions have a radius of 32 pixels and are placed on an 8 x 8 pixel grid. This approach is called the circular region approach. Textural features are extracted from each of the regions using the moments of responses to a multiscale bank of Gaussian filters. Additional positional features are added to the circular regions.

For the big region approach, the regions are classified independently of one another using a k nearest neighbour (kNN) classifier with a leave-one-out strategy. The per region classification results are combined using a weighted multiplier to give an overall abnormality score for the image. The circular regions are classified using a linear discriminant analysis (LDA) classifier in a 5-fold cross-validation framework. The per region results are combined using quantile fusion to give an overall abnormality score for the image.

The system is evaluated on a data set obtained using the Lodox STATSCAN linear slot-scanning digital X-ray machine at the Red Cross Children's Hospital in Cape Town. The data set contains 45 normal images and 60 abnormal images. The classification when using the big region approach has a sensitivity of 0.935 at a specificity of 0.932 and an area under the ROC curve of 0.982. The classification when using the circular region approach has a sensitivity of 0.925 at a specificity of 0.879 and an area under the ROC curve of 0.941.

The system performs most poorly in the perihilar regions of the lungs and struggles to distinguish between normal and abnormal vasculature. The size of the data set and the fact that many normal images were cropped from full-body images are limitations.

Although the performance of the system is very encouraging, testing of the system on a larger database is required before definitive conclusions regarding its utility can be made.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

Despite the availability of effective treatment, tuberculosis (TB) remains one of the leading causes of disease-related deaths in South Africa. According to the World Health Organisation (WHO), the number of TB mortalities in South Africa has increased from 78 per 100 000 population in 1990 to 218 per 100 000 population in 2006. Even more alarmingly, the percentage of deaths attributed to TB among HIV positive people increased from 5% in 1990 to more than 60% in 2006 [4]. Children carry a significant portion of the TB burden with those younger than 5 years being at the greatest risk of developing TB [5]. Childhood TB has, however, been neglected to a large degree due to the challenges faced in its diagnosis [6]. HIV co-infection has further complicated the diagnostic procedure and there is thus an immense need for a more efficient and accurate technique for the diagnosis of TB in children.

In developing countries especially, the chest radiograph is one of the most practical and helpful everyday tools in the diagnosis of pulmonary TB [6]. Despite the recent rise of three-dimensional imaging techniques such as computed tomography (CT) and magnetic resonance imaging (MRI) and a strong likelihood that these modalities will soon replace many current diagnostic and screening techniques, the chest radiograph easily remains the most common radiological procedure, accounting for at least one third of all radiological exams [7]. It is simple, inexpensive and contains a huge amount of information about the condition of the patient, thus possessing tremendous clinical value [8]. However, chest radiographs are extremely difficult to interpret [7]. Poor contrast and complex backgrounds comprised of superimposed anatomical structures (such as large blood vessels and ribs) overlapping with regions of interest make the detection of abnormalities difficult, even for experienced radiologists [9]. The development of computerised systems to aid radiologists in reading images and to harness the clinical value of chest radiographs would ease these difficulties. It is therefore not surprising that computer

aided diagnosis (CAD) of lung diseases in chest radiographs has become a popular area of research in medical imaging and diagnostic radiology [10].

Although much research has been conducted into the use of computer-aided textural analysis for the detection of interstitial lung diseases in chest radiographs, research into its application for the detection and diagnosis of pulmonary TB is very limited. To the author's knowledge, computer-aided detection of pulmonary TB in paediatric chest radiographs has not been attempted.

1.1 Study objectives

The aim of this study was to develop a computerised tool to detect tuberculosis in paediatric chest X-rays obtained using a linear slot-scanning digital X-ray machine. The study forms part of a larger project aimed at devising measures that may be used to detect TB by analysing both the appearance of the lungs, and the size and shape of the airways in chest X-rays. This study, in particular, has focussed primarily on analysing the appearance of the lung fields. The specific objectives were to:

1. Analyse the appearance of the lungs in chest X-ray images to derive measures that may be used to detect tuberculosis.

2. Develop a computer-aided diagnosis tool based on (1).

1.2 Project summary

The project presents a computerised tool to analyse the appearance of the lung fields in paediatric chest X-rays obtained from the Lodox STATSCAN linear slot-scanning digital X-ray machine at the Red Cross Children's Hospital in Cape Town, in order to detect the presence of tuberculosis. The tool was developed in MATLAB.

The CAD tool consists of 4 phases: 1) lung field segmentation; 2) lung field subdivision; 3) feature extraction and 4) classification. The lung field segmentation is performed using a semi-automatic implementation of the active shape model algorithm. Two approaches to subdividing the lung fields into regions of interest

are compared, the first of which divides each lung field into 21 overlapping regions of varying sizes, resulting in a total of 42 regions per image. This approach is called the big region approach. The second approach subdivides the lung fields into a large number of overlapping circular regions of interest; this is referred to as the circular region approach. Once the lung fields have been subdivided, a set of textural features are extracted from each of the regions.

The classification phase is divided into two stages. The first stage involves classifying the individual regions, while the second stage combines the region results into an overall result for the image. The output of the first stage is a probability map indicating the degree of abnormality of each location in the lung fields. This is to be used as a visual aid to a radiologist reading the image, drawing the radiologist's attention to areas displaying high degrees of abnormality and then allowing the radiologist to make the final diagnosis. The output of the second stage is an overall abnormality score for the image, which is to be used as a stand-alone indication of the abnormality of the image or in conjunction with the probability maps (it can also be used in screening applications).

The overall performance of the system is tested using receiver operating characteristic curve (ROC) analysis. Figure 1.1 shows a simple flow chart illustrating the steps involved in applying the CAD system, described above, to an unseen test image.

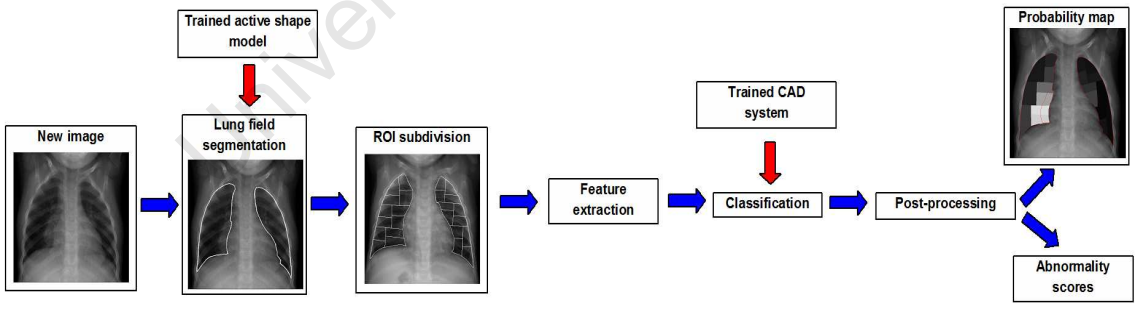


Figure 1.1: Flow chart of the CAD system. The chart illustrates the steps involved in producing a diagnosis for every new test image. Firstly, the lung fields are segmented from the test image using an active shape model. The lung fields are then divided into regions of interest from which a set of textural features are extracted. Using the trained CAD system and the extracted features, the test image is classified. The classification result is used to produce a probability map, showing the degree of abnormality of every region in the lung fields, and an overall abnormality score, giving an indication of the abnormality of the image as a whole.

1.3 Outline of the dissertation

This dissertation is structured as follows:

Chapter 2 presents a review of the relevant literature, covering the characteristics of paediatric pulmonary TB as well as developments in computer-aided tools to detect related pulmonary abnormalities in chest radiographs.

Chapter 3 describes the images analysed during the project.

Chapter 4 provides the reader with a detailed explanation of the methods involved in each phase of the image analysis framework.

Chapter 5 presents and discusses the results of the segmentation phase of the framework established in Chapter 4.

Chapter 6 presents and discusses the results of the feature extraction and classification phases of the framework established in Chapter 4.

Finally, Chapter 7 summarises the results and the most important observations made in the project and makes recommendations for future work.

Chapter 2

Literature Review

2.1 Mycobacterium tuberculosis

Tuberculosis (TB) is an infectious disease caused by micro-organisms, the most common of which is *Mycobacterium tuberculosis* (MTB). It most commonly attacks the lungs in what is known as pulmonary tuberculosis (PTB). Infection occurs via the inhalation of droplets, infected with MTB organisms, which have been expelled by an infected person via coughing, sneezing or speaking. The bacilli are carried by cilia through the respiratory tree and deposit in a bronchiole or alveolus.

In the majority of cases, the host factors, although not removing the bacilli, prevent their multiplication, thereby preventing the infection from progressing to active disease [6]. A person with latent TB infection will be asymptomatic and is incapable of spreading the infection. When the bacteria prevail over the host factors however, the bacilli multiply causing the infection to progress to active TB. A local inflammatory response is triggered, resulting in a parenchymal nodule. This primary nodule is known as the Ghon focus and typically surrounds the lobar fissures in the upper part of the lower lobe or lower part of the upper lobe of the lung [11]. The bacilli may then spread further in a lymphatic or haematogenous manner. Lymphatic spread results in the localisation of bacilli in the lymph nodes; the subsequent enlargement of the lymph nodes together with the Ghon focus is known as the Ghon complex. Haematogenous spread of the bacilli results in the spread of the infection to other organs of the body (extra-pulmonary TB) [11]. The most common sites of TB infection and disease manifestations in the lungs are discussed in Section 2.3 which addresses the radiographic manifestations of TB.

Post-primary TB refers to the reactivation of dormant residual foci in a person who has been previously infected. It generally occurs in one of four ways [12]: 1)

reactivation of bacilli in the lung apices; 2) perforation of a lymph node into a bronchus causing ulceration of the bronchi; 3) directly from the primary focus or 4) by re-infection. It is important to note that children generally develop primary TB, whilst post-primary TB is classically seen in adults.

2.2 Current diagnostic procedure

The current gold standard for the diagnosis of active PTB is an MTB positive sputum culture [13]. The process is however costly and may require up to 4 weeks to produce a result; this time delay is significant as early diagnosis is important in the prevention of the spread of the infection, both in the patient's own body and to other people [6]. More importantly, the method has limitations in the paediatric setting as children generally do not produce sufficient sputum for a reliable test [14]. *Mycobacterium tuberculosis* is isolated in less than 50% of paediatric cases, making the diagnosis of paediatric pulmonary TB a very uncertain process [13]. The diagnosis of paediatric TB is thus often based on the following factors [6, 13]:

- A record of known contact with an infected person.
- A positive tuberculin skin test (TST).
- A chest X-ray suggestive of TB.
- Presentation of a cough for a period of greater than 2 weeks.

Although this procedure is fairly accurate in low burden countries, it is very limited in endemic settings. Children in endemic settings are exposed to adult TB cases at home on a regular basis, and it is difficult if not impossible to keep a record of this. Furthermore, these children are often malnourished and therefore prone to suffer from a large variety of ailments; the presentation of a cough for a period of greater than 2 weeks is therefore not as strong an indicator of TB [6].

HIV co-infection presents the greatest challenge in the accurate diagnosis of PTB in children in endemic settings. The risk of developing active TB in the first year of infection is increased from approximately 3% in HIV negative cases to 30% in HIV positive cases [12]. According to WHO [4], TB is the main cause of death among people infected with HIV. In sub-Saharan Africa, HIV is the single most important factor contributing to the increase in TB, with an estimated 60% of TB infected patients being HIV positive.

The challenges faced in the diagnosis of PTB in children in endemic settings in general are summarised below: [6]:

- Regular exposure to adult TB cases at home.
- Low sensitivity to the TST (often due to malnourishment).
- The large diversity of disease manifestations in children.
- Presentation of non-specific symptoms or the absence of symptoms in children.

The diagnostic procedure is complicated further in HIV positive children due to the following factors:

- Overlap between common HIV and PTB symptoms (e.g. failure to thrive).
- Complication of chest radiographs by HIV related co-morbidity (e.g. pneumonia, lymphocytic interstitial pneumonitis, bronchiectasis etc).

As a result, diagnosis often relies on imaging [14]. The chest X-ray is often the main diagnostic tool for children in endemic settings [6]. However, chest X-rays are extremely difficult to interpret [7]. Poor contrast and complex backgrounds comprised of superimposed anatomical structures (such as large blood vessels and ribs) overlapping with regions of interest make the detection of abnormalities difficult, even for experienced radiologists [9]. The motivation for the development of a computer-aided diagnostic tool to analyse paediatric chest X-rays for the presence of pulmonary TB is clear.

2.3 Radiographic manifestations of TB

Although computed tomography (CT) provides superior views of most radiographic manifestations of PTB, its use in developing countries, where the diagnostic challenges posed by tuberculosis are most pronounced, is inappropriate for a number of reasons, including: limited availability; high radiation doses; high costs; the need for IV contrast mediums and the need for sedation [14]. The chest radiograph is thus still the most practical and helpful everyday tool in the diagnostic procedure for PTB [6].

The manner in which PTB manifests itself radiographically has shown to vary with the age of the patient as well as the stage of the TB infection (primary or

post-primary) [14, 15]. Children generally develop primary TB and so display radiological features typical to primary TB. There are also various progressions of TB which are more typical to particular age groups within the paediatric category (defined as patients younger than 16 years [15]). Those radiological features typical to paediatric cases will be emphasised in this section, but it is important to note that HIV co-infection presents several challenges and irregularities [12].

The radiographic appearance of HIV PTB depends on the level of immunosuppression of the sufferer. In the early stages of infection radiographic findings of HIV positive TB patients are fairly standard, but at severe levels of immunosuppression, 10-20% of chest x-rays are normal or demonstrate findings associated with primary TB - despite known previous exposure to TB [12]. The effect of HIV co-infection is especially significant in sub-Saharan Africa where the majority of TB cases are HIV related [4].

A normal anterior-posterior (AP) chest radiograph is shown in Figure 2.1, with several important anatomical structures labeled. An AP chest radiograph is one where the radiation passes through the patient from front to back. The left side of the image shows the right lung field.

Figure 2.2 shows a number of abnormal chest radiographs illustrating several of the radiographic manifestations of TB which are discussed below.

The radiological hallmark of primary TB is hilar and mediastinal lymphadenopathy, occurring in the vast majority of patients. Lymphadenopathy refers to the enlargement of the lymph nodes caused by tuberculous caseating granulomas [16]. On anterior-posterior (AP) chest radiographs it appears as a lobulated density occupying the hilum and obliterating the hilar point. As a result, the hilum may appear outwardly convex [14]. Lymphadenopathy occurs primarily in the right paratracheal and hilar regions and is rarely seen without associated parenchymal abnormalities [5]. A typical example of right hilar lymphadenopathy is shown in Figure 2.2(b).

Most studies have shown that the prevalence of lymphadenopathy decreases with increasing age. Its prevalence decreases further in developing countries where malnourishment results in a later presentation of the disease [15]. The initial spread of TB bacilli to the central lymph nodes from the Ghon focus (which ultimately results in lymphadenopathy) may result in linear interstitial disease in children manifesting on chest radiographs as prominent linear interstitial markings [16]. It

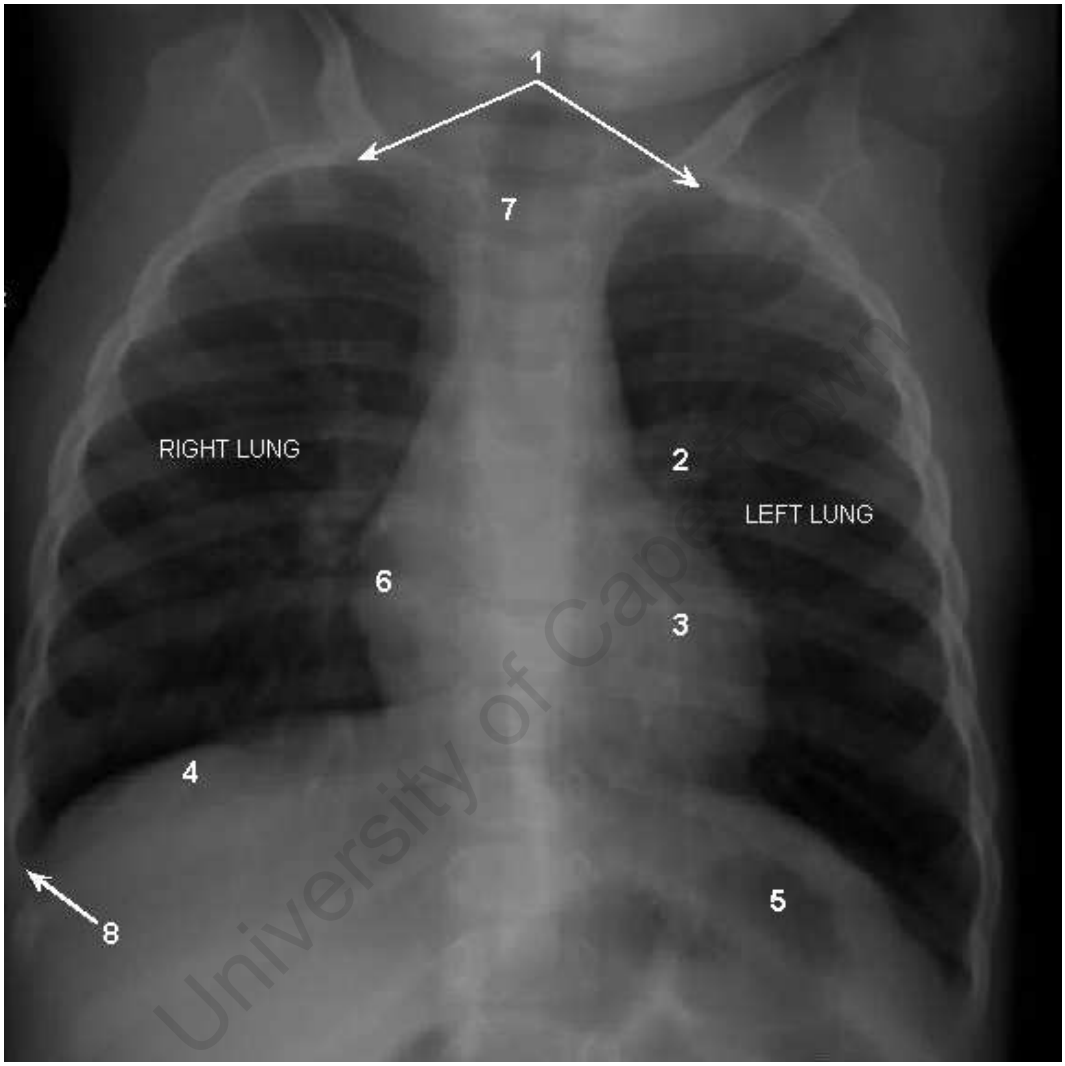


Figure 2.1: A normal anterior-posterior (AP) chest radiograph. The right lung field appears on the left side of the image. The lung fields appear black in the image. 1) The lung apices. 2) The left hilum, where the arteries, veins and bronchi enter the lung. 3) The cardiac shadow. 4) The right hemi-diaphragm begins below the lung field. 5) The left hemi-diaphragm with stomach gases (darker region) visible. 6) The right atrium. 7) The trachea. 8) The right costophrenic angle (where the diaphragm meets the ribs). The area projected between the two lung fields is the mediastinum.

should be noted that the standard plain chest radiograph is somewhat insensitive to the detection of lymphadenopathy as it is often masked by normal structures in the lung and mediastinum [14].

Tuberculous consolidation or infiltration, the hardening or filling of a previously collapsible region of the lung, is a common radiological manifestation of primary PTB [15]. It most commonly occurs in the acini of the lungs (the respiratory bronchioles, alveolar ducts, alveolar sacs and alveoli) in the apex of the upper and lower lobes and is usually bilateral [17]. Radiographically it manifests as an opacification within the lung parenchyma, appearing as patches of brighter grey levels in the usually dark regions of the healthy lung fields [16, 17]. The opaque patches are most commonly located in the upper and middle zones of the lung field [17]. Infiltration most commonly occurs in conjunction with lymphadenopathy and is seen in approximately two thirds of all TB cases [12]. An example of infiltration is shown in Figure 2.2(c).

In a large number of cases the only radiographical sign of primary TB is the Ghon complex consisting of the Ghon lesion and an accompanying enlargement of lymph nodes (lymphadenopathy) [11, 12].

The hallmark radiographic manifestation of post-primary PTB is cavitation [5]. A cavity is caused by the degeneration of a region of infiltration [17]. Cavities appear as lucent (darkened) areas in the lung parenchyma and are usually surrounded by annular shaped regions of consolidation (brighter grey levels) [16]. A fully formed cavity has a well defined dark region within a brighter region of consolidation [17]. In Figure 2.2(d) a cavity is clearly visible in the right middle lobe. Although cavitation is generally uncommon in primary PTB cases, its prevalence increases in immunocompromised children, a statistic which is significant in HIV endemic areas [5].

Exudative pleuritis occurs in both primary and post-primary PTB cases [12]. It results from an accumulation of excess fluid in the pleural cavities and manifests itself as pleural effusion on chest radiographs. Figure 2.2(e) displays extensive pleural effusion in the right lung field. Although uncommon in young children, its prevalence increases in older children (4 to 15 years) [5]. In primary TB cases effusion usually appears 3-6 months after infection [11]. In children especially, the blunting of the costophrenic angle may be indicative of fluid in the pleural space [16].

The miliary spread of TB is a classical occurrence in paediatric TB cases [11]. It occurs as a result of an acute haematogenous dissemination of bacilli in the lungs. A huge number of small tuberculous granulomas develop appearing as many micrococondular infiltrates diffusely distributed throughout both lungs [12]. Miliary TB only manifests radiographically after a few weeks and appears as an enormous number of tiny opacities, 1-2 mm in diameter [16]. Figure 2.2(f) clearly displays the diffuse nature of miliary TB. The prevalence of miliary TB increases significantly in HIV co-infected patients.

Tuberculomas, well defined granulomatous lesions in the lung parenchyma, are sometimes the only radiographic sign of post-primary TB. They usually occur as a result of healed TB and may be accompanied by airspace consolidation [12].

Cicatrisation atelectasis, a collapsed lung or deflated alveoli, is a parenchymal lesion which occurs predominantly in post-primary TB patients [15]. Approximately 40% of post-primary TB cases exhibit a marked fibrotic response manifested as atelectasis of the upper lobe [11]. As a result the hilum retracts and a mediastinal shift towards the fibrotic lung and compensatory hyperinflation of the lower lobe occurs [12].

In general, computed tomography provides a clearer view of the abovementioned abnormalities, but due to the drawbacks of CT there is a need for a system to evaluate standard chest radiographs more effectively.

2.4 Computer-aided diagnosis in chest radiography

Arguably the primary objective of developing the digital computer was to create a system which could aid or replace humans in performing tasks which traditionally relied entirely on human intelligence. It is thus not surprising that soon after the development of the first digital computer in the 1940's, research began surrounding the topic of computerised analysis of radiographic images [7]. Early studies investigating the computerised detection of lung abnormalities in chest radiographs, published in the 1970's, generally assumed that computers would ultimately replace radiologists in the diagnosis procedure [10]. At this point it is important to distinguish between fully automated computer diagnosis (FACD) and computer

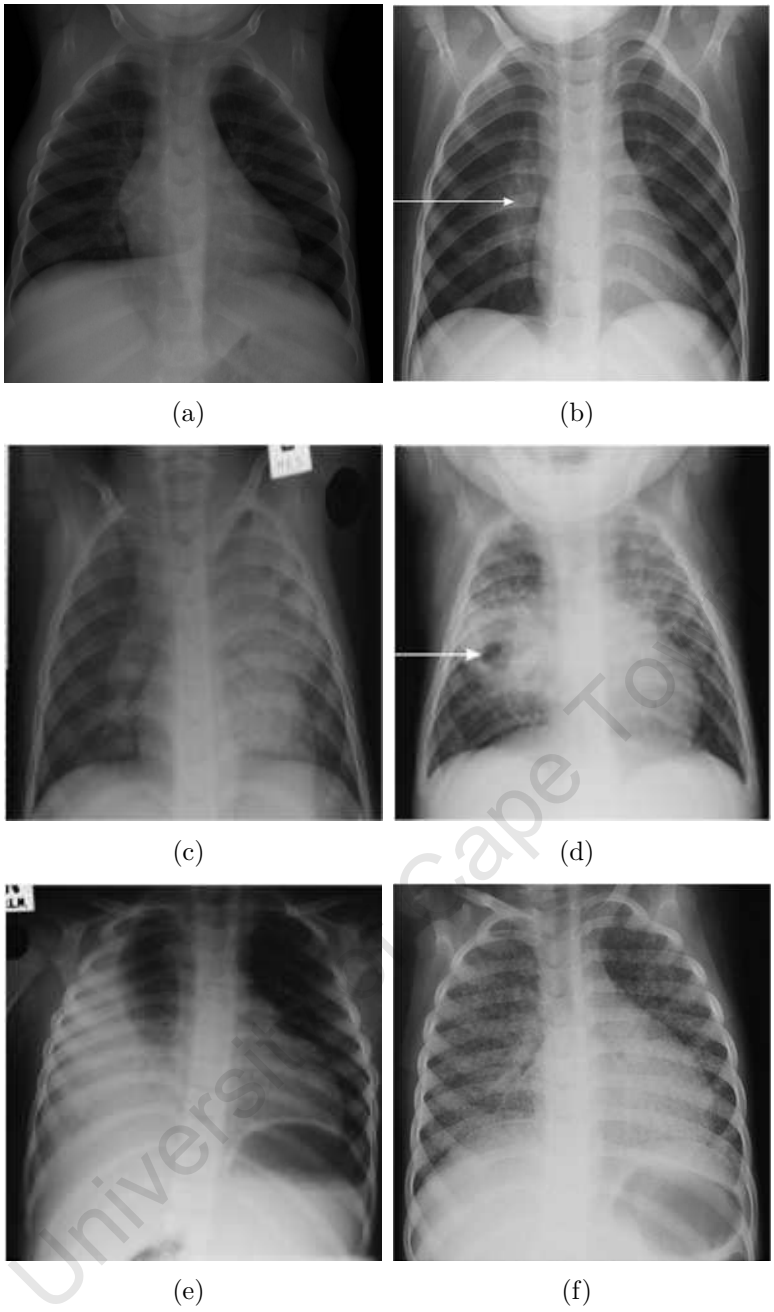


Figure 2.2: (a) A completely healthy chest X-ray. (b) Typical hilar lymph gland enlargement in the right lung field. (c) Extensive infiltration (consolidation) of the left lung field; note the brighter appearance in comparison with the right lung field. (d) A cavity in the right middle lobe. (e) Extensive pleural effusion in the right lung field. (f) Microcondular infiltrates diffusely distributed throughout both lung fields (Miliary TB). Abnormal images obtained from [18].

aided diagnosis (CAD).

Although FACD and CAD, in the context of radiography, are both based on the development of computer algorithms for the quantitative analysis of digital images, the difference between the two lies in the utilisation of the output of the algorithms. An FACD system involves no external input, rather the digital ra-

diograph is independently examined by the algorithm and the computer's output is the final diagnosis. The accuracy of the system is thus based entirely on the performance and accuracy of the algorithm. The output of the algorithm used in a CAD system, on the other hand, is merely an assistance tool to the radiologist. It is used as a second opinion and it is the radiologist who makes the final diagnosis. The performance of the system is measured by evaluating the accuracy of the radiologist's diagnosis when using the computer output as an aid. The accuracy of the algorithm thus does not have to be nearly as high as it does in an FACD system [10].

As with the majority of early enthusiasm surrounding the capabilities of artificial intelligence, the expectations of computerised detection schemes for lung diseases have subsided [7]. The primary focus of current research is directed at developing CAD systems to aid radiologists in making improved diagnoses of chest abnormalities. The purpose of these systems is to improve the diagnostic accuracy and image reading time of radiologists [9]. Literature reveals that the topic can be separated into three predominant areas: 1) image pre-processing; 2) segmentation techniques and 3) detection of abnormalities [7]. The literature reveals further that the most widely researched examination of chest abnormalities (in terms of CAD) are [9]:

- nodule detection
- detection of interstitial diseases
- differential diagnosis
- evaluation of interval changes
- detection of asymmetric opacities

Although lung nodule detection has been the most studied problem in computer-aided diagnosis [8], the detection of TB in chest radiographs is most closely related to the detection of interstitial diseases. Attention, in this review, is thus focussed on this area of computer-aided diagnosis.

2.5 Image pre-processing and segmentation

A crucial stage in the computerised evaluation of chest radiographs is the suppression of background anatomical structures in order to enhance abnormalities.

Furthermore, chest radiographs often display poor contrast and so image manipulation for optimal display is common [8]. As discussed in the subsequent sections of this review, enhancement, subtraction and lung field segmentation techniques are the most common solutions to these two problems.

In [8], the automatic segmentation of lung fields is considered a compulsory procedure before computerised analysis of a chest radiograph can occur. The problem has received considerable attention and is a topic on its own. A fairly comprehensive evaluation of lung segmentation techniques is presented and discussed in [2]. The study presents a comparison between three common segmentation approaches: 1) active shape models, 2) active appearance models and 3) a multi-resolution pixel classification method.

The active shape model (ASM)approach is a segmentation technique that was first developed by Cootes *et al.* [19]. The scheme consists of three elements [2]: 1) a global shape component, 2) a multi-resolution grey-level appearance component and 3) a multi-resolution search algorithm. The technique requires a set of annotated training images, in which a set of corresponding points has been marked, from which the model can be built. Statistics about the model's global shape and appearance, in the vicinity of each model point, are then captured. This statistical model is then used in an iterative search to locate particular structures in unseen test images [19].

Similarly to ASMs, active appearance models (AAMs) build statistical models of a particular object's shape and texture to be matched to unseen test images. AAMs use the same set of annotated training images as used by ASMs as the input to the system [20]. The main difference in the AAM approach, lies in the fact that it considers the texture across the entire object (and not only in the vicinity of the model points) in its description of appearance [20]. Synthetic images are built as linear combinations of the texture and shape of the training images. These synthetic images are then used to find the model parameters that most reliably represent the object in a particular unseen test image. In order for the model to converge to the correct result, additional processing is usually necessary to appropriately position the model on a test image [20].

Pixel classification (PC) methods involve assigning a class to every pixel in a particular image (in this case one of two classes, namely lung or not lung). Classification is achieved using a classifier trained on feature vectors extracted from every pixel in a set of training images [2]. Performance is measured using an over-

lap measure (Ω) defined as the intersection of the segmentation results with the gold standard, divided by the union. The study shows that the pixel classification approach outperforms the ASM and AAM methods in the task of segmenting the lung fields, but that the results yielded by the ASM ($\Omega = 0.903$) and AAM ($\Omega = 0.847$) are, nonetheless, comparable to the PC results ($\Omega = 0.938$) [2].

2.6 Textural analysis and detection of interstitial lung disease

The lung interstitium is the connective tissue between the blood vessels and alveoli of the lungs. An interstitial lung disease (ILD) may refer to any disease causing damage to this interstitium. As a result, *interstitial lung disease* is the common term used for more than 200 different lung diseases [21]. Nonetheless, the detection of interstitial lung diseases remains one of the most challenging tasks in radiology [8]. A damaged interstitium results in a change in the textural appearance of the lungs and the difficulty in diagnosis lies in the fact that a particular ILD can result in a large variety of textural changes to the lungs while many different ILDs may cause almost identical textural changes to the lungs [22]. Additionally, there exists a large degree of ambiguity in the distinction between normal and abnormal lung textures even for radiological experts [23].

Interstitial lung abnormalities are generally diffuse in nature and so the textural analysis of the lung tissue forms the basis of the majority of ILD detection techniques presented here.

Extensive work in the area of ILD detection has been published by Katsuragawa *et al.* [24] and Kido *et al.* [22]. Evaluation of these schemes reveals a common structure to the solution of the task: Multiple regions of interest (ROIs) are manually or automatically selected within the lung fields. Features, in the form of textural measurements, are then extracted from each of these ROIs. The ROIs are classified using rule-based and/or pattern recognition techniques resulting in soft labels (posterior probabilities) for each of the regions in the lung fields. The results are combined to obtain an overall diagnosis for the image, indicating whether or not it contains any interstitial abnormalities [22, 23]. This structure is only implemented subsequent to the pre-processing and segmentation steps discussed earlier.

Arguably the foremost challenge in the textural analysis of the lungs is the presence of superimposed anatomical structures [8]. Dividing the lung fields into small ROIs allows for each ROI to be classified individually with a classifier trained with features extracted from the specific region only, thereby capturing common anatomical characteristics of particular lung regions in the feature vectors, reducing the effect of background structures [23]. In many cases, however, this is not sufficient, and the detection of abnormalities can still be complicated considerably by the visibility of the rib cage in particular. A knowledge of the location of the rib cage in an image allows for compensatory measures to be taken and aids the diagnoses considerably.

The task of segmenting or suppressing the rib cage in chest X-rays is a challenging one. Most approaches attempt to fit parabolas to rib border pixels detected by Hough transforms or by template matching [25, 26]. The main shortcoming of these approaches is that rib borders are often inaccurately detected, resulting in the need for heuristic rules to determine the true rib borders [27]. Van Ginneken *et al.* [27], present a rib segmentation algorithm that fits the global rib cage directly to the image, eliminating the need to detect rib borders locally and reducing the number of parameters considerably. The results presented in these studies are promising, but the task is far from solved.

Most commonly, textural features are calculated using filter response methods, but a variety of methods exist, including: Fourier spectrum based methods [9], geometric features [24] and fractal dimensions [28].

Filter response features are calculated by finding the moments of histograms of images filtered with a bank of Gaussian filters [29]. The method is effective as filtering an image with a bank of Gaussian filters is equivalent to finding the Taylor expansion of that image, where the filters represent the Taylor coefficients; the filters are thus approximations of local image structure [23]. In most cases the filters are applied at multiple scales so that textural elements of varying sizes may be represented. Commonly the second, third and fourth central moments (standard deviation, skew and kurtosis) as well as the mean are used as textural features. Additional features are usually included in this feature set including: difference features calculated as the differences between corresponding regions in the left and right lung fields (as most abnormalities are asymmetric) [23], and position features describing the locations of the ROIs [29].

A common challenge faced when calculating filter response features in the lungs

is the distortion in the filter output in regions near the lung borders due to the often distinct differences in texture inside and outside the lung fields. In [29], this issue is dealt with by mirroring pixels with respect to the lung borders prior to filtering so that pixel values outside the lung are replaced with their counterparts inside the lung, thereby eliminating the drastic change in appearance near the lung borders.

The classification stage is generally performed using two-class classifiers (i.e. trained on normal and abnormal data). Various classification techniques are used, most commonly: artificial neural networks [30], k-nearest neighbours (kNN) classifiers [23], discriminant analysis [29], support vector machines (SVM) [29] and threshold techniques [9]. In order to train and test such classifiers, it is necessary that data be fully labeled (i.e. the locations of the abnormalities must be known).

An artificial neural network (ANN) is a nonlinear statistical data modelling tool which is based on the networks within the brain. In its simplest form an ANN is fed with input data (in this case the extracted textural features) which are then fed through two successive nonlinear calculations in the internal (hidden) and output layers, producing an output. An ANN consists of a training and testing stage. During training the internal parameters of the neural network are iteratively adjusted to minimise the difference between the actual and desired outputs of the network. By iterating through this process with the same input and output sets, the ANN learns the relationship between the training input data and the desired output [30]. At completion, the ANN is capable of making distinctions between different input data.

The kNN approach is a nonparametric method which is one of the simplest classification techniques. The method involves the estimation of the posterior probabilities for each of the classes by the fraction of training examples within the k-nearest neighbours of a particular test sample. The test sample is assigned to the class that is most common amongst its k nearest neighbours [29]; k is a free parameter which is determined experimentally.

Discriminant analysis is based on the assumption that the data in each of the classes are normally distributed. Linear discriminant analysis (LDA), assumes equal covariance matrices for each distribution while quadratic discriminant analysis (QDA) does not. LDA seeks to find the linear combination of features which best separates the two classes. QDA is a more general version of LDA, which seeks to find the best separation of two classes using a quadric surface [31].

Support vector machines are a group of supervised learning algorithms belonging to a family of linear classifiers. Input data are viewed as two sets of vectors in n-dimensional space. In its simplest form, the SVM finds an optimal discriminative plane between the two sets of vectors in that space. This is achieved by maximising the margin between the nearest samples of both classes. The objective is to obtain a plane which exhibits large distances to the neighbouring data points of each class. In finding the plane, consideration must be given to the tradeoff existing between the classification error and the margin maximisation. This tradeoff is represented as a penalty parameter [29].

In [29], several classifiers are compared in the task of detecting interstitial lung disease in chest radiographs. Filter response textural features are used in each of the classifiers (kNN, LDA, QDA and SVM) and performance is measured using the area, A_z, under the receiver operating characteristic (ROC) curve. It is shown that although the SVM ($A_z = 0.78$) and LDA ($A_z = 0.78$) approaches yield the most satisfactory results, the kNN classifier ($A_z = 0.76$), although much simpler to implement, still yields comparable results.

Van Ginneken *et al.* [21] investigate classification techniques for small ROIs and present an interesting one-class classifier approach. The motivation for using a one-class classifier is that no abnormal data is needed for training the classifiers. Due to the wide variety of abnormal data in the real world (especially in the case of ILDs) it is difficult to train conventional two-class classifiers efficiently. This scheme fits well into medical imaging analysis as classes are often poorly sampled [21]. The one-class case simply estimates a standard model for a healthy lung (which is well defined) and calculates the deviation of an instance from this model to determine its degree of abnormality. Although the method performs poorly (relative to the two-class method) in this study, it has much potential and appeal for future developments. The method could be especially useful for detecting abnormalities which are not common and hence not well documented.

As mentioned, ILD is a term used for more than 200 different diseases, and so the differential diagnosis of ILDs is an important but difficult task. Some CAD schemes provide a likelihood measure that a particular case (after it has been diagnosed as an ILD) belongs to a particular type of ILD [9]. The most common approach is to use a combination of radiological and clinical measures in an ANN to distinguish between a group of common ILDs. The schemes yield performances comparable to or better than that of radiologists alone [32, 33].

The results obtained by the various schemes (all evaluated using receiver operating curve (ROC) analysis), show that they do effectively assist radiologists in diagnosis of ILDs, but the problem is still largely unsolved, especially in terms of more subtle and uncommon abnormalities. A widespread shortcoming of the CAD research field is the complete absence of large, publicly available datasets to allow for valuable comparative studies [8]. Furthermore, although much research has been conducted into the use of textural analysis for the detection of interstitial lung diseases, research into its application to the detection and diagnosis of pulmonary TB is limited.

Sarkar and Chauduri [17] use pattern recognition techniques to detect and measure the progression of infiltration and cavitation in digital chest radiographs. Infiltration (and subsequently cavitation) is detected by using an iterative segmentation based on grey levels within the images. The scheme follows a similar framework to the general three-step ILD detection framework discussed earlier.

Koeslag [34] investigates the problem of computer-aided diagnosis of miliary TB in chest x-rays. Various methods for the enhancement of lung textures are investigated and template matching is used to detect the lesions associated with miliary TB. Although encouraging results are produced, the final system is unable to differentiate between miliary TB and other ILDs resulting in similar lung texture distortions.

Perhaps the most comprehensive work on the computer-aided detection of TB is presented by van Ginneken *et al.* [23]. In this work, the lung fields are subdivided into 42 overlapping regions of interest, from which filter bank features are extracted and used in a kNN classification framework. Although the system performs reasonably well on TB data, results are considerably better when the same system is tested on an ILD data set.

More recently, Arzhaeva *et al.* [35] have presented a dissimilarity based classification framework to detect TB (and ILDs) in chest X-rays. The methods presented are suited particularly to cases where the locations of the lung pathologies are unknown, making the use of local feature-based classifiers impossible. On a fully labeled tuberculosis data set (i.e. one in which the locations of the abnormalities are known), the system performs comparably to the best region classification method.

To date, no studies have been published that deal with the computer aided detection of paediatric tuberculosis (or interstitial lung diseases).

2.7 Summary

This review has provided the reader with a brief introduction to tuberculosis infection in children and the challenges faced in its diagnosis. In high burden countries, especially in sub-Saharan Africa, HIV-co-infection has contributed significantly to the difficulties faced in diagnosing TB in children. The radiographic manifestations of TB were presented and discussed and it was shown that many TB pathologies result in diffuse textural changes in the lung tissue.

The field of computer-aided diagnosis was introduced. Despite the fact that the detection of interstitial lung disease via textural analysis of the lungs has been widely researched and has shown considerable success, there has been very limited work presented in the computer-aided detection of TB, especially in the field of paediatric TB, where, to date, no CAD systems have been developed. There is, therefore, a clear need for further development of computer-aided TB detection schemes. Given that PTB often manifests itself similarly to many interstitial lung diseases, textural analysis of the lung fields as an approach to the detection of PTB in chest radiographs, bears much potential.

Chapter 3

Materials

The study makes use of images obtained from scans performed on children at the Red Cross Children's Hospital in Cape Town using the Lodox STATSCAN linear-slot-scanning digital X-ray machine. All the children were aged between 0 and 5 years at the time the scans were taken. The final data set consists of 60 fully labeled abnormal images and 45 normal images.

3.1 The Lodox STATSCAN unit

The STATSCAN system was originally developed for use as a rapid, low-dose X-ray surveillance system in the South African mining industry. The system has since been developed for use in emergency medical settings and has been approved for diagnostic use by the US Food and Drug Administration as well as the European Union [36]. The uniqueness of the system lies in its ability to produce rapid high definition X-rays at considerably lower radiation doses than conventional and computed radiography systems.

In a recent study [37] it has been shown that the effective radiation dose delivered to children ranging between 1 and 5 years by the STATSCAN digital radiography unit is less than that of a standard computed radiography unit. As a result, it is concluded that the STATSCAN radiography unit is highly suitable for use in paediatric applications. In another recently published study [36], the STATSCAN system produced images with improved visibility of the tracheo-bronchial tree; this is of particular interest due to the susceptibility of the airways to TB-related pathologies in children.

The machine consists of an X-ray tube providing collimated fan-beam radiation

and a detector array made up of scintillator arrays linked to charge coupled devices. The X-ray tube and detector are mounted on opposite ends of a C-arm which can be rotated around the patient up to 100°, allowing for different views of the patient. An anterior-posterior scan (as used in this project) is taken with the C-arm rotated so that the X-ray tube is positioned directly above the patient. The fan-beam is collimated to 1 mm or less and travels across a patient at speeds of 35 mm/s, 70 mm/s or 140 mm/s, allowing an adult full-body scan to be completed in approximately 13 seconds. The image resolution is variable from 1 line pair per millimetre (lp/mm) to 5 lp/mm. The X-ray machine is interfaced to a dedicated workstation where the user can choose the size of the patient and the type of scan to be taken; the technique factors are then automatically determined. A paediatric category, where technique factors are adjusted for children, is available [36]. The user can also view and manipulate the resulting image at the workstation. The images used in this project are described in the following sections.

3.2 Abnormal images

The abnormal images were obtained from scans performed on suspected TB patients, who have been enrolled in a related TB study aimed at identifying airway narrowing and lymph node enlargement. Antero-posterior (AP) chest scans of the patients were taken at a spatial resolution of 4.16 lp/mm, using the Lodox STATSCAN machine. At the time of the project described here, a total of 77 children had been enrolled in the TB study.

The images were examined by a single radiologist. Of the 77 images, 11 were excluded by the radiologist due to poor image quality (e.g. excess rotation). Furthermore, 6 of the 77 images were diagnosed as normal by the radiologist, leaving a total of 60 images with some form of TB-related lung pathology. To allow for the training of classifiers to distinguish between normal and abnormal regions, the abnormal regions in the images were outlined by the examining radiologist on a computer monitor using a mouse pointer (Figure 3.1). Additionally, the radiologist assigned a number to each abnormal region denoting its degree of subtlety; a description of each of the levels is given in Table 3.1. In this way, the performance of a classifier on abnormalities which human experts detect with varying degrees of difficulty, could be tested. In total, 135 abnormal regions are present in the 60 abnormal images, of which 89 occur in the right lung field and 46 in the left lung field. Table 3.2 shows the spread of the data across the 5 subtlety levels.

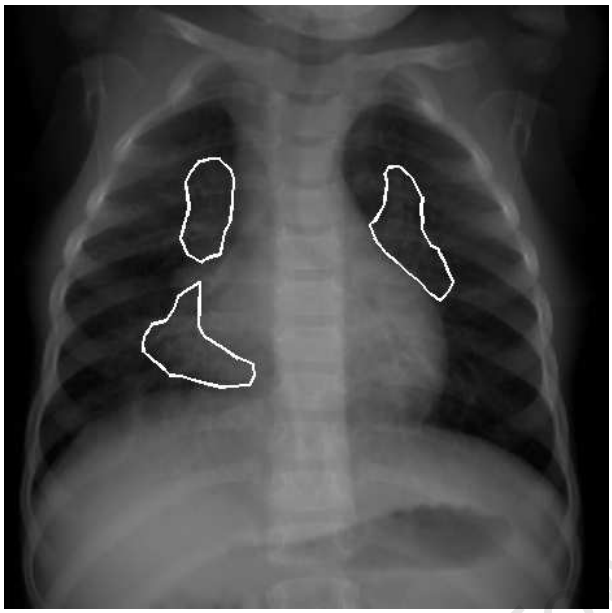


Figure 3.1: An example of the delineation of the abnormal regions performed by the examining radiologist.

Number	Subtlety	Description
1	Obvious	Abnormality detection is easy
2	Relatively obvious	Abnormality detection is relatively easy
3	Intermediate	Abnormality has average detection difficulty
4	Subtle	Abnormality detection is difficult
5	Very subtle	Abnormality detection is very difficult

Table 3.1: Subtlety of abnormalities. Every abnormality was assigned a subtlety grading by the radiologist, denoting its detection difficulty.

Subtlety	Right Lung Field	Left Lung Field
1	5	2
2	17	8
3	28	16
4	26	12
5	13	8
total	89	46

Table 3.2: The number of abnormal regions for each subtlety level in the left and right lung fields.

3.3 Normal images

Of the 77 images from the TB study, the radiologist diagnosed 6 as entirely normal (i.e. no lung pathology) and these formed part of the normal data. The remaining 39 normal images were obtained from routine AP full-body scans performed on trauma patients using the Lodox STATSCAN digital X-ray machine.

The TB study which produced the abnormal and some normal images for the present study obtains diagnostically useful chest radiographs of children suspected of having TB. Exposing healthy young children to the radiation associated with a chest X-ray for the purposes of the present study would have been inappropriate and difficult to justify to the Research Ethics Committee of the Faculty of Health Sciences. This motivated the decision to use the readily available full-body scans obtained in the Trauma Unit to construct the normal data set. Although the full-body scans were taken at a different spatial resolution to the chest scans (2.08 lp/mm as opposed to 4.16 lp/mm), subsampling all the images (normal and abnormal) to a fixed resolution compensates for these differences; this subsampling is explained in further detail in Chapter 4.

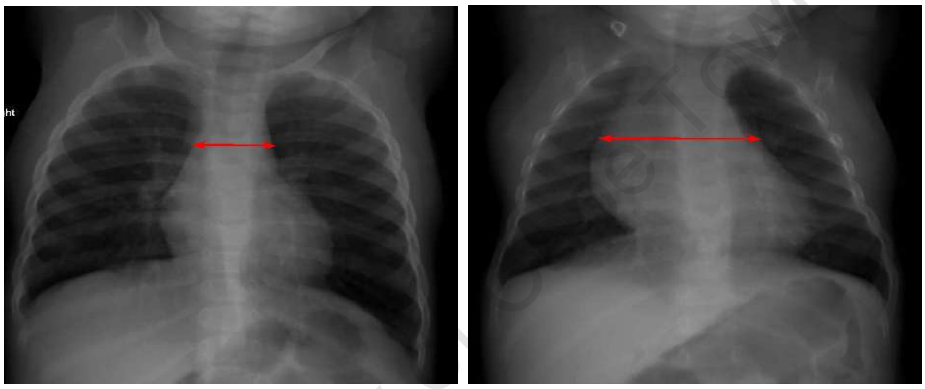
The full-body images were examined by the same radiologist to confirm that no lung pathologies existed and were then cropped to contain the same anatomical information as the chest scans (Figure 3.2).

(a) (b)

Figure 3.2: An example of a full-body scan in (a) which is cropped to contain only the chest in (b).

3.4 Paediatric chest X-rays

This is the first study aimed at developing a computer-aided diagnostic tool for TB in children. The lung fields of a young child ($\leq$ 5yrs) can exhibit somewhat inconsistent characteristics [38]. The heart size and mediastinal width is often exaggerated by the AP projection. The thymus gland often results in a very prominent anterior mediastinal shadow, with a sail-like appearance. The size and appearance of the thymus gland can vary significantly between different children, as shown in Figure 3.3. The thymus gradually becomes less evident with age, eventually disappearing from the frontal chest X-ray. The impact of this variation on the CAD tool is discussed in Chapter 5.



(a) (b)

Figure 3.3: Two healthy lung images showing very different lung shapes. The thymus gland is barely visible in the image in (a), while the image in (b) has a very wide, but healthy, mediastinum, mainly due to the thymus gland. The red arrow indicates the width of the mediastinum in each case.

Chapter 4

Methodology

The following chapter provides the reader with a thorough explanation of the methods used in the development of the CAD scheme for this thesis. In broad terms the system consists of four stages: 1) lung field segmentation, 2) region of interest subdivision, 3) feature extraction and 4) classification. Finally, the framework for evaluating the performance of the each of these components is discussed. All of the methods discussed here were implemented in MATLAB (The Mathworks, Natick, MA). The majority of the MATLAB code was developed by the author; a freely available ASM tool [39, 40] was used as the basis for the lung field segmentation and a Gaussian derivative tool was used to compute the kernels with which to filter the images [41].

4.1 Lung field segmentation

The segmentation of the lung fields is a necessary task in the computer-aided diagnosis of any lung disease. In the preceding literature review an overview of the most common approaches to the automatic segmentation of the lung fields was presented. Perhaps the single most comprehensive overview of these methods is presented in [2], where a comparison between three supervised methods (active shape models, active appearance models and multi-resolution pixel classification) is presented. The study shows that all three methods perform well for lung field segmentation, with pixel classification returning the best results.

As the name indicates, pixel classification involves assigning a class to every pixel in a particular image (in this case one of two anatomical classes, namely lung or not lung). Classification is achieved using a classifier trained on feature vectors extracted for every pixel from a set of training images. A limitation of this

method is that it fails to classify lung portions with extensive pathology as lung tissue because the diseased pixels appear so different to healthy lung pixels. As the objective of this study is to develop a textural analysis framework to distinguish between healthy and unhealthy (TB infected) lung tissue, it is essential that the parts of the lung fields containing the pathologies are included in the segmentation results. It was therefore decided to use a shape based approach instead. In [2], the active shape and active appearance models perform comparably. Therefore, due to its simplicity (in comparison to active appearance models) the active shape model was chosen as the method of choice for the segmentation of the lung fields. For a description of the difference between active shape and active appearance models, the reader is referred to Section 2.5. A detailed explanation of the segmentation algorithm follows. Throughout the segmentation process, the definition of the lung field given in [2] is adopted: any pixel for which radiation has passed through the lung, but not through the mediastinum, heart, structures below the diaphragm and the aorta.

4.1.1 Active shape models

The active shape model implementation used here is based on that developed by Cootes *et al.* [19]. Active shape models (ASMs) build statistical models of a particular object's shape and grey-level landscape to be matched to unseen test images. The technique requires a set of annotated training images from which the model can be built. Statistics about the model's shape and appearance (in the vicinity of each landmark point) are then captured. This statistically based flexible model then moves iteratively towards structures, similar to those on which the model was trained, in unseen test images.

The statistical model, known as a point distribution model, consists of the following 3 components:

1. The position of the object in the image.
2. The shape of the object.
3. The appearance of the object.

The following subsections describe the algorithm and how it has been applied to the task of segmenting the lung fields in this project. The left and right lung fields are segmented independently of one another.

4.1.2 The position component

Each image in the training set is represented by a set of points. Therefore, for every training image the borders of the lung fields need to be delineated manually. This is a crucial step in the algorithm as the method works by modelling how different labeled points tend to move together as the shape varies [19]. It is thus essential that every lung border consists of the same number of points and that these points correspond. The following approach is used to achieve this:

The lung contours in all of the chest images in the data set are manually outlined by clicking points along the lung boundaries on a computer monitor, using a mouse pointer. These points are connected by straight line segments and every lung therefore requires a different number of points to ensure a smooth border around the lung. The next step is to convert these borders to a fixed number of corresponding points. As is done in [2], several additional, distinguishable points on the lung border are clicked. These distinct landmark points are assumed to correspond. For both the left and right lung fields the following 3 landmark points are used:

1. The bottom lateral corner of the lung.
2. The lung apex.
3. The bottom medial corner of the lung.

Once these three points have been located in a particular lung field all the original border points are discarded, leaving only the three landmark points and the lung border (created by the straight line segments). The border points to be used in the training of the active shape model are obtained by equidistantly sampling a certain fixed number of points between the three landmark points, thereby ensuring a fixed number of corresponding points for every lung field [2]. In this application, 65 points are sampled along the border of the right lung and 60 along the border of the left lung. The ASM method works by examining the statistics of the coordinates of these labeled points over a particular set of training images [19]. The process of defining the landmark points described here is illustrated in Figure 4.1.

4.1.3 The shape component

Shape may be defined as all the geometric information that remains when location, scale and rotational effects are filtered out from an object [42]. According to

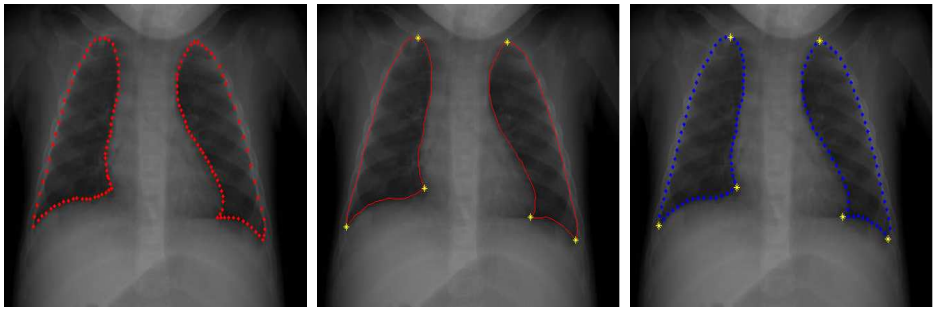


Figure 4.1: Object delineation process: The image on the left shows the boundary points positioned manually. The image in the middle shows the resulting boundary (created by straight line segments) as well as the distinctive landmark points marked in yellow. The image on the right shows the points interpolated between the anatomical landmarks marked in yellow.

this definition, in order to obtain a true representation of the shape of an object it is necessary to filter out the effects of pose (position, scale and rotation). This is achieved by establishing a coordinate reference to which all shapes are aligned—i.e. normalising the shape of the set of annotated training images with respect to pose and then representing the shape of every object as the difference from the mean shape of the entire training set.

This coordinate reference or mean shape is obtained using a procedure known as Procrustes analysis. The procedure is composed of the following 4 steps, which are illustrated in Figure 4.2 [42]:

1. Compute the centroid of each of the two shapes.
2. Normalise the sizes of the two shapes (i.e. remove effects of size).
3. Align the two shapes with respect to position at their centroids (i.e. remove the effects of position).
4. Align the two shapes with respect to orientation, by rotation.

A shape, $\mathbf{x}_1$, in an image is represented as a set of n ordered 2-D points:

$$\mathbf{x}_1 = \begin{pmatrix} x_{a1} & y_{a1} \\ x_{a2} & y_{a2} \\ \vdots & \vdots \\ x_{an} & y_{an} \end{pmatrix}$$

The centroid $(\overline{x}, \overline{y})$ of the shape is computed as:

$$(\overline{x}, \overline{y}) = \left(\frac{1}{n} \sum_{j=1}^{n} x_j, \frac{1}{n} \sum_{j=1}^{n} y_j \right)$$

The sizes of the two shapes are normalised by normalising the centroid sizes, $S(\mathbf{x})$, of the two shapes:

$$S(\mathbf{x}) = \sum_{j=1}^{n} \sqrt{(x_j - \overline{x})^2, (y_j - \overline{y})^2}$$

A position and scale normalised shape, $\mathbf{x}_n$, is thus defined as:

$$\mathbf{x}_n = \frac{\mathbf{x} - (\overline{x}, \overline{y})}{S(\mathbf{x})}$$

Given two position and scale normalised shapes, $\mathbf{x}_1$ and $\mathbf{x}_2$, step 4 removes the rotational differences between the two shapes by applying a singular value decomposition (SVD) technique:

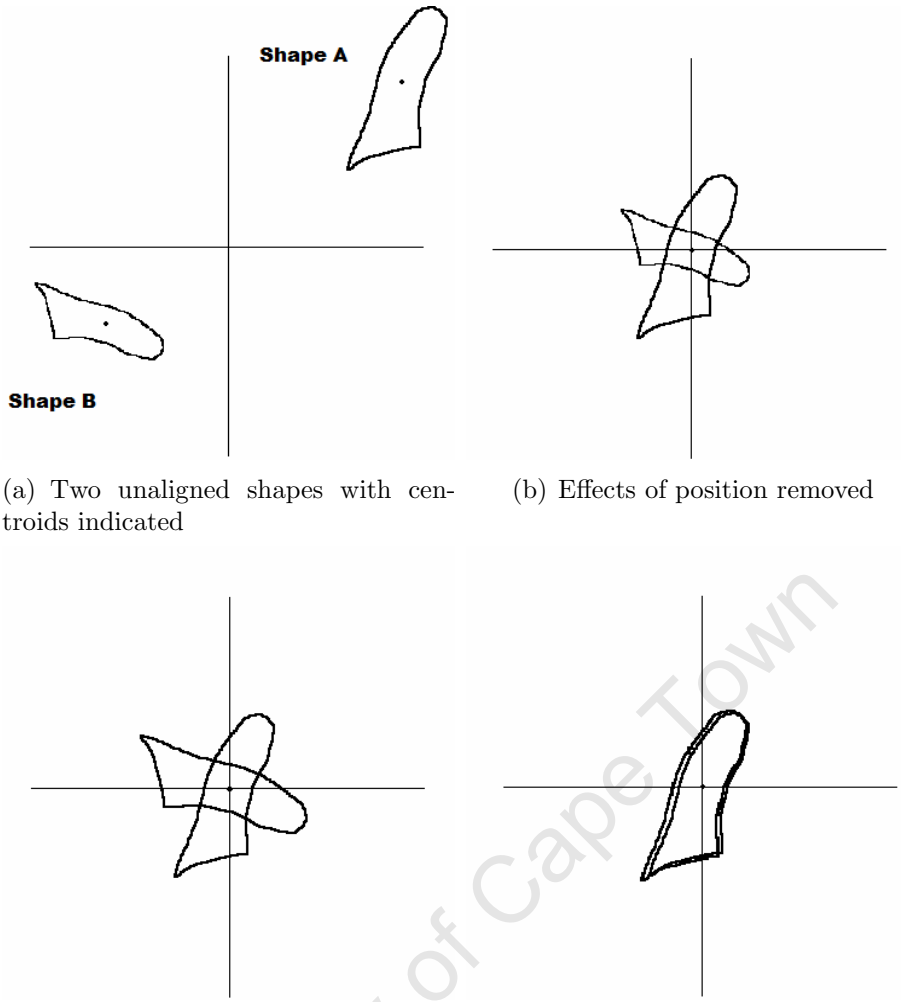
1. Represent $\mathbf{x}_1$ and $\mathbf{x}_2$ as nxk matrices.
2. Compute the SVD, $\mathbf{UDV^T}$ of $\mathbf{x}_1^T \mathbf{x}_2$.
3. Apply the transform, $\mathbf{VU^T}$ to optimally rotate and align $\mathbf{x}_1$ to $\mathbf{x}_2$.

where $\mathbf{U}$ and $\mathbf{V}$ are unitary matrices containing the set of basis direction vectors for $\mathbf{x}_1^T \mathbf{x}_2$ and $\mathbf{D}$ is a diagonal matrix containing the singular values of $\mathbf{x}_1^T \mathbf{x}_2$.

In $\mathbf{R}2$, the rotation matrix, $\mathbf{VU^T}$ is given by:

$$\begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix}$$

A least-squares type measure known as the Procrustes distance can now be computed as the difference between the two shapes $\mathbf{x}_1$ and $\mathbf{x}_2$. The squared Procrustes distance is defined as:



(a) Two unaligned shapes with centroids indicated

(b) Effects of position removed

(c) Effects of scale removed

(d) Effects of rotation removed. The shapes are now aligned

Figure 4.2: Shape alignment by Procrustes Analysis. The four steps of the alignment procedure are illustrated here.

$$P_d^2(\mathbf{x}_1, \mathbf{x}_2) = \sum_{j=1}^{n} (x_{j1} - x_{j2})^2 + (y_{j1} - y_{j2})^2$$

The mean shape is computed using the following iterative algorithm:

1. Set the first shape in the training set as the estimated mean shape.
2. Align the remaining shapes to this mean to minimise the distance metric P.
3. Re-calculate the estimated mean shape from the aligned shapes.
4. Repeat steps 2-3 until the mean shape converges on a stable shape.

Finally, the shape component for every instance in the training set is calculated as the difference between the mean shape and the pose normalised shape for that particular instance.

Once the training data has been aligned the next step of the process is to capture the statistics of these aligned shapes. The purpose of the Point Distribution Model is to model the codependent variation of the corresponding points of all the shapes in the training set.

Every example in the training set can be represented entirely by a set of n, 2 dimensional points. Collectively, a training set consisting of N examples results in a cluster of $2n$ dimensional points contained within a region of space that Cootes *et al.* [19] call the Allowable Shape Domain. In this domain, every training example is represented by a single point in a $2n$ dimensional space. The objective of this stage of the algorithm is to model the shape of this cluster, thereby defining the relationships between the locations of the individual landmark points. Cootes *et al.* [19] make the assumption that the Allowable Shape Domain cluster is approximately ellipsoidal and proceed to perform a principal component analysis (PCA) on the data.

PCA provides a linear orthogonal transformation M that, when applied to a data matrix x, produces an output y, representing the transformation of x onto a new coordinate system, where the major variations in the columns of x are expressed in the first few rows of y. In other words, the principle axes of the Allowable Shape Domain ellipsoid represent the modes of variation of the training data or the manner in which the landmark points tend to move together as the shape is varied within the Allowable Shape Domain. Theory shows that this variation can be expressed adequately by the first few modes of variation (i.e. the largest few principle axes), allowing for a considerable reduction in the dimensionality of the problem. The principal component analysis is performed as an eigenanalysis of the covariance matrix of the aligned shapes and consists of the following steps [19]:

1. Compute the mean shape of the training data.
2. For each shape in the training set, compute its deviation from the mean.
3. Compute the $2n$ x $2n$ covariance matrix of the shapes in the training data.
4. Compute the unit eigenvectors of the covariance matrix (yielding the principle axes).
5. Select the number of modes of variation to retain.

Given a set of N aligned shapes, the mean shape, $\overline{\mathbf{a}}$, is computed as:

$$\overline{\mathbf{x}} = \frac{1}{N}\sum_{i=1}^{N}\mathbf{x}_i$$

The deviation of shape $\mathbf{x}_i$ from the mean is given by:

$$d\mathbf{x}_i = \mathbf{x}_i - \overline{\mathbf{x}}$$

The maximum likelihood estimate of the covariance matrix can then be computed as:

$$\mathbf{\Sigma}_s = \frac{1}{N}\sum_{i=1}^{N} d\mathbf{x}_i d\mathbf{x}_i^T$$

The principle axes of the ellipsoid, defining the modes of variation, are computed as the eigenvectors, $\mathbf{\Phi}_s$, of the covariance matrix:

$$\mathbf{\Sigma}_s\mathbf{\Phi}_s = \mathbf{\Phi}_s\mathbf{\Lambda}_s$$

where $\mathbf{\Lambda}_s$ is a diagonal matrix of eigenvalues corresponding to the eigen vectors of the columns of the covariance matrix $\mathbf{\Phi}_s$.

$$\mathbf{\Lambda}_s = \begin{bmatrix} \lambda_i & & \\ & \ddots & \\ & & \lambda_{2n} \end{bmatrix}$$

As mentioned, most of the variance in the data can be explained by the first few modes of variation. In order to determine the exact number of modes to retain,

a trade-off between accuracy and simplicity of the model needs to be considered. It has been shown that the variance along each eigenvector is equal to the corresponding eigenvalue. In order to retain p percent of the variation in the training data, t modes can be retained as follows:

$$\sum_{i=1}^{t} \lambda_i \geq \frac{p}{100} \sum_{i=1}^{2n} \lambda_i$$

The parameter p is then a value which is specified during the construction of the ASM and is generally of the order of 90-99%.

Any point in the Allowable Shape Domain (or equivalently any allowable shape instance) can now be generated by deforming the mean shape by adding a linear combination of the retained eigenvectors:

$$\mathbf{x} = \overline{\mathbf{x}} + \mathbf{\Phi}_s \mathbf{b}_s$$

where $\mathbf{\Phi}_s$ is the matrix of the first t eigenvectors, and $\mathbf{b}_s = (b_1, b_2, \cdots, b_t)^T$ is a vector of shape model weights or parameters. By varying these parameters within some predefined limits, new shapes, similar to those in the training set, can be generated. Cootes *et al.* [19], use the fact that the variance of b_i is equal to λ_i and the fact that most of the population lies within three standard deviations of the mean to define the limits of b_i as:

$$-3\sqrt{\lambda_i} \leq b_i \leq 3\sqrt{\lambda_i}$$

4.1.4 The appearance component

In addition to the shape model discussed above, an appearance model, describing the typical image structure around each landmark point, is required. This is obtained by sampling pixel profiles around each landmark, perpendicular to the object contour at that point. Cootes and Taylor [3], describe the process as follows:

To compute the direction perpendicular to the edge on which the landmark (x_n, y_n) lies in the i^{th} training image, a vector between (x_{n-1}, y_{n-1}) and (x_{n+1}, y_{n+1}) is rotated through 90°. The objects used in this project are closed contours, so for the first landmark, (x_1, y_1), the last landmark and the second landmark are used. Once the direction has been found in this way, k pixels are sampled on either side of the landmark along this direction using a fixed step size. This gives a profile of length $2k + 1$ pixels, which is placed in a vector $\mathbf{g}_i$. To reduce the effects of global intensity changes, the normalised first derivatives of these profiles are used to build the appearance model. The derivatives are computed using finite differences between the $(j-1)^{th}$ and $(j+1)^{th}$ points in the profile. The sample is then normalised by dividing through by the sum of the absolute element values:

$$\mathbf{g}_i \rightarrow \frac{1}{\sum_j |g_{ij}|}\mathbf{g}_i$$

This process is repeated for every training image, yielding a set of normalised samples $\{\mathbf{g}_i\}$ for each landmark. By making the assumption that these samples are normally distributed, their mean $\overline{\mathbf{g}}$ and covariance $\mathbf{S}_g$ can be estimated, giving a statistical model for the grey-level profile about the landmark. This process is repeated for every landmark in the Point Distribution Model. The Mahalanobis distance between some new profile, $\mathbf{g}_s$, and the corresponding profile model can then be computed as:

$$f(\mathbf{g}_s) = (\mathbf{g}_s - \overline{\mathbf{g}})\mathbf{S}_g^{-1}(\mathbf{g}_s - \overline{\mathbf{g}})$$

Minimising this distance metric is equivalent to maximising the probability that $\mathbf{g}_s$ originates from the distribution.

4.1.5 Image search

Given a rough starting position, the active shape model can be fitted to an unseen test image. Cootes *et al.* [43] present an iterative algorithm to achieve this. The starting position will be discussed in more detail in the next section. The search involves finding the shape and pose parameters that result in the model coinciding

with the object of interest in the test image. An instance $\mathbf{X}$ of the model can be created in the test image coordinate frame by defining 1) the starting position, 2) the orientation and 3) the scale of the model:

$$\mathbf{X} = M(s, \theta)[\mathbf{x}] + \mathbf{X}_c$$

where,

$\mathbf{x}$ defines the shape of the model instance in a model centred coordinate frame.
$\mathbf{X}$ defines the shape of the same model instance in the test image coordinate frame.
$\mathbf{X}_c = (X_c, Y_c, X_c, Y_c, \cdots, X_c, Y_c)^T$ is an n x 2 matrix where (X_c, Y_c) is the centre of the model in the test image coordinate frame.
$M(s, \theta)[.]$ is a rotation by θ and a scaling by s.

Briefly, the iterative process consists of the following steps (which are repeated until there is no significant change in the result) [43]:

1. Examine a region of the image around each model landmark to determine a displacement of the landmark required to move it to a better location.
2. Transform these displacements into adjustments in the pose, scale and shape parameters.
3. Update the model parameters.

In order to ensure that the shape of the model remains similar to those in the training set, global shape constraints are applied by applying limits to the shape parameters.

Step 1 above is achieved by sampling a profile perpendicular to the boundary at each landmark point and finding the best match (minimum Mahalanobis distance) to the appearance model. This is done using the same techniques used in the construction of the appearance model. The suggested movement for the current model point is then in the direction of the point on the sampled profile that returned the best match. Once this has been completed for every model point (or landmark) the collective adjustments can be represented as an adjustment vector, $d\mathbf{X}$:

$$d\mathbf{X} = (dX_0, dY_0, \cdots, dX_{n-1}, dY_{n-1})^T$$

Step 2 is to transform this adjustment vector into an equivalent change in the pose and shape parameters of the model, so that the model points can be moved from their current locations, $\mathbf{X}$, as near to the new suggested locations ($\mathbf{X} + d\mathbf{X}$) as possible.

Recall that the current estimate of the model is centred at (X_c, Y_c), and has an orientation θ and scale s. Cootes *et al.* [19] use a least squares approach to find the desired changes in pose $(dX_c, dY_c, ds, d\theta)$. For the details of this procedure the reader is referred to [19]. In order to compute the required changes in the shape parameters of the model the following procedure is used:

Recall that the initial position of the model points in the test image frame is given by:

$$\mathbf{X} = M(s, \theta)[\mathbf{x}] + \mathbf{X}_c$$

The objective is to derive a means for finding the suggested movements, $d\mathbf{x}$, to the points $\mathbf{x}$ in the local model coordinate frame, satisfying:

$$M(s(1 + ds), (\theta + d\theta))[\mathbf{x} + d\mathbf{x}] = (M(s, \theta)[\mathbf{x}] + d\mathbf{X}) - (\mathbf{X}_c + d\mathbf{X}_c)$$

It is worth emphasising the difference between $\mathbf{X}$, $d\mathbf{X}$, $\mathbf{x}$ and $d\mathbf{x}$ again. The lower case, $\mathbf{x}$ and $d\mathbf{x}$, refer to the current model position ($\mathbf{x}$ is an n x 2 matrix of model point locations) and displacement vector in the model centred coordinate frame, while $\mathbf{X}$ and $d\mathbf{X}$, refer to the equivalent model position and model point displacement vector in the test image coordinate frame. After determining the optimal model point displacements, $d\mathbf{X}$, in the image coordinate frame, the objective is to determine the equivalent displacements in the model centred coordinate frame, so that the corresponding model parameter adjustments can be calculated.

Solving for $d\mathbf{x}$ in the above equation yields:

$$d\mathbf{x} = M((s(1 + ds))^{-1}, -(\theta + d\theta))[\mathbf{y}] - \mathbf{x}$$

where,

$$\mathbf{y} = M(s, \theta)[\mathbf{x}] + d\mathbf{X} - d\mathbf{X}_c$$

To ensure that only 'legal' shapes (i.e. shapes within the Allowable Shape Domain) are generated by the deformations $d\mathbf{x}$, it is necessary to transform $d\mathbf{x}$ into model parameter space, yielding $d\mathbf{b_s}$, which represents the changes in the model parameters, $\mathbf{b_s}$, that will adjust the model points as closely to $d\mathbf{x}$ as is allowed. Cootes *et al.* [19] again achieve this using a least-squares approximation such that:

$$d\mathbf{b_s} = \mathbf{\Phi_s}^T d\mathbf{x}$$

where,

$$\mathbf{x} = \overline{\mathbf{x}} + \mathbf{\Phi_s} \mathbf{b_s}$$

The model parameters are updated in this way in an iterative manner until some criterion is reached:

$$X_c \rightarrow X_c + w_t dX_c$$

$$Y_c \rightarrow Y_c + w_t dY_c$$

$$\theta \rightarrow \theta + w_\theta d\theta$$

$$s \rightarrow s(1 + w_s ds)$$

$$\mathbf{b_s} \rightarrow \mathbf{b_s} + \mathbf{W}_b d\mathbf{b_s}$$

Where w_t, w_s and w_θ are scalar weights and $\mathbf{W}_b$ is a diagonal matrix of weights of size t x t containing one weight for each of the t modes. As discussed earlier, limits are placed on the values of b_k to ensure only allowable shapes are obtained in each iteration. The procedure (computing and updating new pose and shape parameters) is repeated until no significant change in the resulting shape occurs.

4.1.6 Mutli-resolution framework

Cootes *et al.* [3], introduce an extension of this framework to a multi-resolution active shape model; a similar approach is used here. The reason for this extension is that it makes the search more efficient and robust. The multi-resolution approach involves first searching a coarse, low resolution version of the test image and then refining the search on finer, higher resolution versions of the test image. This is done by constructing appearance models (as discussed earlier) for each resolution. These images are generated by creating a multi-resolution image pyramid containing Gaussian smoothed versions of the original image at each level. The finest resolution (level 0) uses the original image and a step size of 1 pixel when sampling the grey-level profile. The next resolution (level 1) is the image observed at a scale $\sigma = 1$ and a step size of 2 pixels. Subsequent levels are constructed in this way by doubling the image scale and step size. The number of resolutions used is denoted by L_{max}, and is a user defined parameter.

An unseen test image is searched, using the multi-resolution approach, as follows. The search is started at the coarsest or lowest resolution image (i.e. the highest level in the pyramid) by running a number of iterations of the ASM using the models trained at that level. Each iteration of the search is performed as described earlier; by examining a certain number of pixels (n_l) on either side of the current model point to find a better position for that point. The search at each level is repeated either for a certain number of maximum iterations (N_{max}) or until convergence is reached (N_{max} is a user defined number). The search is then changed to a finer resolution (i.e. moved one level down in the pyramid). The entire process is run until N_{max} iterations have been completed on the original (bottom level) image. Convergence here is defined as follows [43]:

At every model point, the fit of the appearance model is assessed at n_l pixels either side of the model point, along a profile perpendicular to the model boundary at that point (as described in the Section 4.1.4). The proportion of points, p_{close}, for which the best fit points lie within $n_l/2$ of their previous positions is computed. If this proportion is greater than some predefined value (e.g. 95%) then the search has converged and the process moves to the next level in the pyramid. If convergence is never reached, the process moves to the next level after N_{max} iterations have been completed.

4.1.7 ASM implementation

For the purposes of segmenting the lung fields, the ASM approach explained in the preceding sections was implemented in MATLAB. A freely available, working MATLAB ASM tool was used as a basis for this implementation [39, 40]. A significant modification, involving the initial registration of the model search, was made to this tool to make it suitable for the required task.

The ASM search requires an initial starting point for the model search. In most applications the mean shape is a sufficient initialisation point as it can be shown that using the multi-resolution approach makes the search more robust, and less likely to converge to the incorrect solution [1]. Initial experiments still showed however, that a more accurate initial registration of the model yielded considerably better and more consistent results. This can most likely be attributed to the inconsistencies in the shape of paediatric lung fields as discussed in Chapter 3. A semi-automatic approach to the ASM scheme is thus used here, where the user is required to provide a rough starting point for the model search. This is done by clicking a mouse on three landmark points in the lung field to be segmented (recall that the two lung fields are dealt with independently of one another). The landmarks required are:

1. The bottom lateral corner of the lung.
2. The lung apex.
3. The bottom medial corner of the lung.

The model is then moved to approximately this location using an affine warp (to ensure that the model shape is not compromised) and the search is started.

In simple terms, an affine transformation in Euclidean space is one that preserves the collinearity between points and the ratios of distances between points [44]:

Affine transformations consist of combinations of linear transformations followed by a translation. To ensure that the transformation is shape preserving, it can be composed only of combinations of linear rotations and uniform scalings and a subsequent translation. The transformation takes the following form [45]:

$$x \rightarrow Ax + b$$

where,
x is the set of points (model points in this case) to transform.
A is the affine transformation matrix (consisting only of rotations and uniform scalings).
b is a vector representing the translation.

As is given in [1], a summary of the parameters required for the active shape model is shown in Table 4.1.

4.2 Regions of interest

As was discussed in Chapter 2, the majority of CAD schemes for detecting lung pathologies contain a step where the lung fields are subdivided into specific regions of interest (ROIs). Two approaches are implemented and compared here. The first approach is based on that implemented in [23] and the second on that implemented in [35]. For the sake of simplicity, the first approach will be referred to as *Big ROIs* and the second as *Circular ROIs*. At this stage, all the images (as well as the segmentation results) are subsampled to a working resolution of 700x700 pixels, and are used in this format for the remainder of the analysis.

4.2.1 Big regions of interest

When referring to different regions in the lung fields, it is common practice amongst radiologists to divide the lung fields into an upper, middle and lower part and to subdivide each of these into a medial and lateral part. An attempt to simulate this approach is implemented here. The process is automated in the following way [23]:

The mean lung shapes are computed from the training data used to train the active shape models (again, the left and right lung fields are dealt with separately). The boundaries defining the regions are found using a simple search algorithm, with equal area as the objective. By computing horizontal lines that divide each of the lungs into 3 regions of equal area, the upper, middle and lower regions of the lungs are defined. Vertical lines that divide each of these 3 regions (per lung field) into two regions of equal area are then computed; defining the medial and lateral regions of the lungs. Finally, these regions are again subdivided horizontally, each into 2 regions of equal area giving a total of 21 overlapping regions per lung and

Position component	
N	number of training images used to construct active shape model
Shape component	
n	number of landmark points representing active shape model
t	number of modes of variation to retain in the shape model
p	percentage of variance to retain in the training data (controls t)
m	bounds on eigenvalues λ_i
Appearance component	
k	number of points to examine on either side of grey-level profile
Image search	
n_s	number of positions to examine on either side of current landmark point during each iteration
L_{max}	number of multi-resolution levels
N_{max}	maximum number of iterations per resolution level if convergence is not reached

Table 4.1: Active shape model parameters [1]

42 regions in total. This defines the region maps illustrated in Figure 4.3, which are used to automatically position the ROIs in subsequent test images.

Subsequent images, with segmented lung fields, are warped to these region maps to define the ROIs in the specific image. This warping is done using radial basis functions with multiquadrics and interpolation [46]. The procedure is explained below:

The objective here is to derive a function $f(\mathbf{x})$ which, when applied to a pixel location, $\mathbf{x}_i = (x_i, y_i)$, in the current lung field, will produce a displaced or warped pixel location corresponding to some location in the ROI map, giving the region number to be assigned to the original input pixel. The problem of defining the function, $f(\mathbf{x})$, falls in the category of scattered data interpolation. In general terms such problems can be defined as follows [46]:

Input: n data points $(\mathbf{x}_i, \mathbf{y}_i) \in \Re^2$; $i = 1, \cdots, n$.
Output: A continuous function $f : \Re^2 \rightarrow \Re^2$ interpolating the given data points,

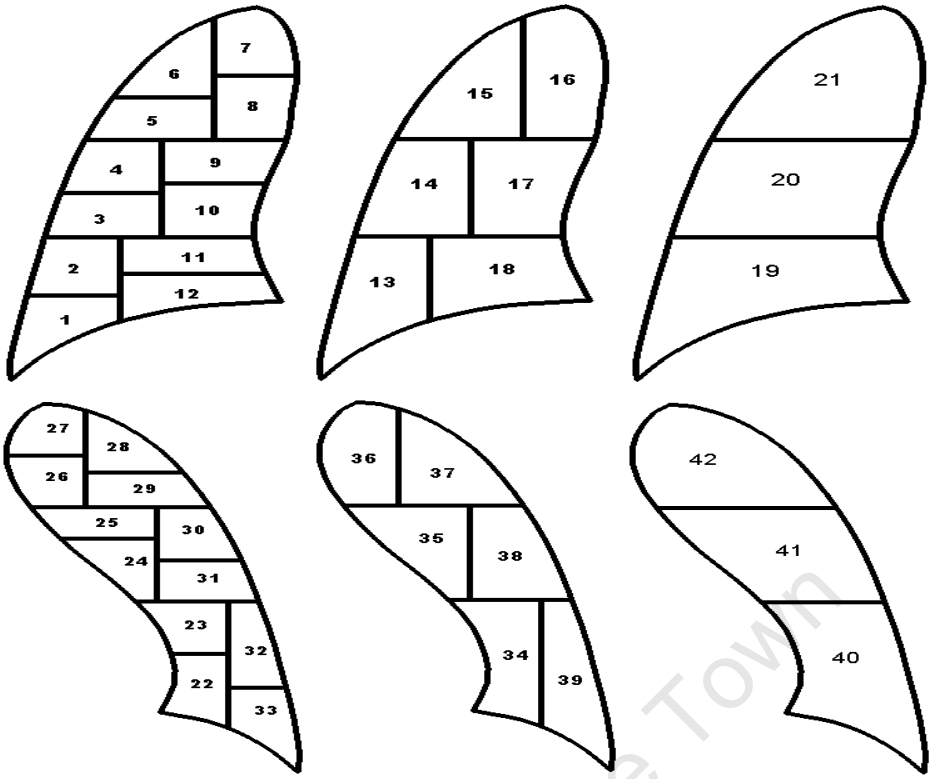


Figure 4.3: The subdivision of the lung fields into 42 overlapping regions of interest. The numbering scheme shown here is used throughout the remainder of this report.

i.e. $f(\mathbf{x}_i) = \mathbf{y}_i, i = 1, \cdots, n.$

The function $f(\mathbf{x})$ is called the interpolant and the n data points are the points for which the transformation is known (in this case the lung borders - it is known that the lung border points in the test image must be warped to the lung border points in the region maps so $\mathbf{x}_i$ will be the lung border in the test image and $\mathbf{y}_i$ the lung border in the region map). These datapoints are used to solve for or define the interpolant.

In [46], several approaches to the scattered data interpolation problem are discussed, of which the radial basis function approach is used here. An interpolant constructed in this way consists of a linear combination of basis functions, where each basis function has some weight or coefficient attached to it. The values of the basis function depend only on the distance from the data point n_i, hence the name 'radial'. A radial basis function interpolant takes the following form:

$$f(\mathbf{x}) = \sum_{i=1}^{n} \alpha_i R(d_i(\mathbf{x})) + p_m(\mathbf{x})$$

The basis function coefficients, α_i, are computed by substituting the known data points, n, into the above equation and solving the resulting set of linear equations using a least squares method. As mentioned, the known points in this application are the lung borders ($\mathbf{x}$ = test image lung border and $f(\mathbf{y})$ = ROI map lung border). For the left and right lung fields this will yield a system of 60 and 65 linear equations respectively.

The term $p_m(\mathbf{x})$ is a polynomial of m^{th} degree that assures a polynomial precision of degree m. A linear ($m = 1$) p_m is used here by setting $p_m(\mathbf{x}) = \mathbf{x}$.

The term, R, represents the basis functions. Multiquadrics are used as basis functions here; they take following form [46]:

$$R(d) = (d^2 + r^2)^{\frac{\mu}{2}}$$

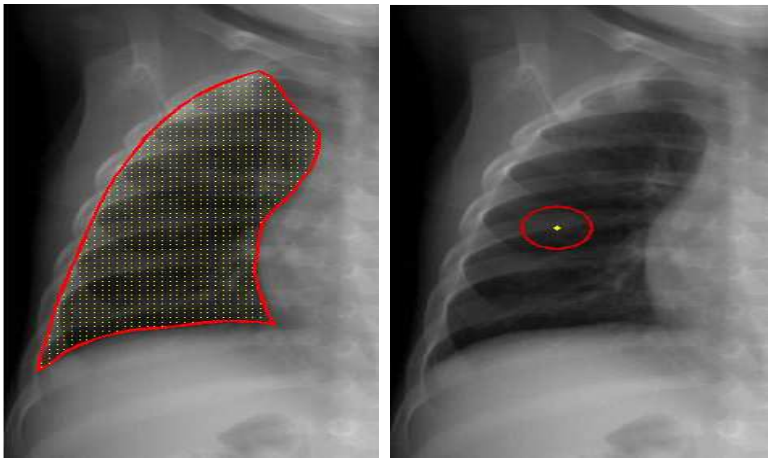
where,

$d = d_i(\mathbf{x})$ is the distance from the current pixel to the i^{th} data point.
r is a characteristic radius that determines the smoothness of the interpolation.
μ is an exponent which is set to 1.

The selection of r is critical for good results and Ruprecht and Muller [46] suggest that individual values for r should be chosen for each control point. This approach is necessary to avoid unwanted effects such as unevenness (r too small) or foldover (r too big) in the interpolation results. The value of r_i is computed as the distance from the data point $\mathbf{x}_i$ to the next nearest data point:

$$r_i = min_{i \neq j} d_i(\mathbf{x}_j)$$

Once the coefficients, α, have been computed, the interpolant can be applied to all the remaining points, $\mathbf{p}$, in the lungs fields:



(a) 8x8 ROI grid (b) One circular ROI

Figure 4.4: Circular regions of interest. The image in (a) shows the grid defining the centres of the circular ROIs and the image in (b) shows an example of one such ROI with its centre marked. The images are of size 700 x 350 pixels.

$$f(\mathbf{p}) = \sum_{i=1}^{n} \alpha_i((d_i(\mathbf{p}))^2 + r_i^2)^{\frac{\mu}{2}} + \mathbf{p}$$

$f(\mathbf{p})$ then represents the location in the region map to which the original point, $\mathbf{p}$, is warped and from this location the corresponding region number can be found. After this warping, the borders between the regions in the test image will no longer necessarily be perfectly horizontal or vertical.

4.2.2 Circular regions of interest

In this approach the lung fields are subdivided into a large number of overlapping circular ROIs. As is done by Arzhaeva *et al.* [35], an 8x8 pixel spacing is used to define the centres of the circular ROIs, yielding a large number of ROIs for each lung field (of the order of 1000). The ROIs all have a radius of 32 pixels. An example is shown in Figure 4.4. The motivation for using this approach is that it is expected that by extracting features from such a large number of smaller regions a more detailed description of the lung tissue will be obtained, possibly making it easier to detect smaller, more subtle abnormalities. Obviously, due to the large number of regions, the process may become computationally expensive - a factor that will be discussed at a later stage.

4.3 Feature extraction

In [47], van Ginneken and ter Haar Romeny conducted experiments comparing the performance of filter bank methods with that of co-occurrence matrix methods in the task of textural analysis. In these experiments, the filter bank methods, where moments of local histograms are extracted as texture features, outperformed the co-occurrence matrix method, where features are extracted from the Fourier spectrum. The filter bank method has also been used successfully for the automatic detection of textural abnormalities in chest X-rays in a number of previous systems [23, 29, 35]. This has motivated the choice to use the filter bank approach for the purpose of extracting local texture features.

4.3.1 Image pre-processing

Prior to the extraction of the image features, several pre-processing steps are completed. Firstly, the left lung field is flipped about the centre line of the image. This is done so that the left lung will resemble the right lung in orientation of anatomical structures such as the ribs and vessels, as well as several textural elements [35].

Due to the large difference in appearance inside and outside the lung fields, the filter outputs at the lung borders are expected to contain major, unwanted distortions. To deal with this, Arzhaeva *et al.* [35], mirror pixel values outside the lungs symmetrically with respect to the lung borders. In other words, for each pixel outside the lung field, its value is substituted by its counterpart inside the lung field using the nearest pixel on the lung border as the axis of symmetry. As this process can become computationally expensive it is only necessary to perform the mirroring for a radius of 32 pixels around the lung border (since this is the radius of the circular ROIs).

4.3.2 Filter bank

In simple terms, filter bank methods capture the statistics of filtered versions of an image as textural features. The first step is to decide upon the types of filters to use. A large variety of filter banks are described in the literature, but a bank of Gaussian filters is used here. An attractive property of Gaussian filters is that they are analogous to the Taylor expansion for functions. In other words, they

describe the local structure of the image, where the coefficient of every term in the approximation is given by the derivative of the image. Images are thus filtered with a multiscale filter bank of Gaussian derivatives.

The filter bank consists of Gaussian derivatives of orders, 0, 1 and 2 at five scales, $\sigma = 1, 2, 4, 8, 16$. Without going into the details of scale space, it is widely accepted and has been shown in numerous studies, that using a multi-resolution approach (i.e. multiple scales for the Gaussian filters) improves performance considerably as it enables the properties of textural elements of different sizes to be captured [48]. By varying the degree of the filter (i.e. using multiple orders) more of the local image structure can be captured [49].

For each input image, 30 filtered versions of the image are computed, where the derivative of the image, $L(\mathbf{x})$, computed at an inner scale of σ is denoted by: $L_{n_x n_y}(\mathbf{x}; \sigma)$, with n_x and n_y denoting the order of the derivatives in the x and y directions. For each of the 5 scales the following 6 filtered images are thus computed: $L_{00}, L_{10}, L_{01}, L_{20}, L_{02}, L_{11}$.

A Gaussian kernel with which to filter the input image is computed in the following way [50]:

Consider a Gaussian function:

$$G(x, y; \sigma) = \frac{1}{2\pi\sigma^2} e^{-\frac{x^2+y^2}{2\sigma^2}}$$

Its n^{th} order derivative is computed by multiplying the n^{th} order Hermite polynomial by the Gaussian function:

$$\frac{\delta^n G(\mathbf{x}; \sigma)}{\delta \mathbf{x}^n} = (-1)^n \frac{1}{(\sigma\sqrt{2})^n} H_n\left(\frac{\mathbf{x}}{\sigma\sqrt{2}}\right) G(\mathbf{x}; \sigma)$$

where,

$H_n(\mathbf{x})$ is the n^{th} order Hermite polynomial.

The derivative is then normalised and convolved with the input image to give the filtered version of the image.

4.3.3 Feature extraction

The most widely used approach when using filter bank methods is to extract the moments of pixel intensity histograms as textural features; this approach has been used successfully in the detection of lung pathologies in [23], [35] and [29]. The first four central moments (mean, standard deviation, skew and kurtosis) of the pixel intensity distribution, $H(i)$, from each ROI in the filtered and original images are computed as textural features. These moments, m_1, m_2, m_3, m_4, are defined as:

$$m_1 = \int iH(i)di$$

$$m_2 = \sqrt{\int (i - m_1)^2 H(i)di}$$

$$m_3 = \frac{1}{m_2^3} \int (i - m_1)^3 H(i)di$$

$$m_4 = \frac{1}{m_2^4} \int (i - m_1)^4 H(i)di - 3$$

The skew, m_3, is an indication of the symmetry of the distribution and the kurtosis, m_4, is a measure of how much the distribution deviates from a Gaussian distribution (for a Gaussian distribution the kurtosis will be 0).

As these features are computed for all 30 filtered versions of the input image, as well as the original, unfiltered image, the process results in each ROI being described by a feature vector containing 124 features.

Due to the fact that normal lung tissue appears slightly different depending on the area of the lung it comes from, it is expected that adding a position component to the feature vectors will improve performance [29]. Of course, this is unnecessary for the approach using the big ROIs as corresponding ROIs can be assumed to come from corresponding regions in the lung field. Without an intrinsic position component for the circular regions however, it is impossible to determine where they originated from. The position features are thus computed in the following way:

The coordinates of a particular lung field are normalised using the segmentation

results. Firstly, the centre of gravity of the lung field is computed and set as the origin, $(0, 0)$, of the coordinate system. The lung apex is then set to $y = +1$ or equivalently $(0, 1)$. The lung bottom is set to $y = -1$ or $(0, -1)$. Similarly the left most and right most points in the segmentation are set to $(-1, 0)$ and $(1, 0)$ respectively. This effectively gives a universal coordinate system for the lung radiographs. Using this approach, an x and a y coordinate (ranging between -1 and $+1$) corresponding to the centre of the current ROI is appended to every feature vector, resulting in feature vectors containing 126 features describing every circular region.

For the big ROI approach each image is thus described by a matrix of size 124x42, where column i ($i = 1, \cdots, 42$) is a feature vector extracted from region number i in that image. For the circular region this matrix size varies, depending on the number of circular ROIs in that image. Depending on the size of the lung fields, the number of circular ROIs per image varies in the range of 1400 to 4000.

During the feature extraction process a region abnormality check is also performed (as this is required during the training and testing phases). For the big regions, a region is deemed abnormal if at least 10% of that region overlaps with an abnormal area marked out by the radiologist. For the circular regions, a region is deemed abnormal if the centre pixel of the region falls in an abnormal area marked out by the radiologist. An abnormality flag is (1 = abnormal, 0 = normal) is appended to the feature vectors.

4.4 Classification

The classification procedure (training and testing) is slightly different for the big regions and the circular regions due to the difference in the feature matrices describing the images in each of the approaches. These differences are explained below.

4.4.1 Big regions

As mentioned in Section 2.7, one of the foremost challenges in the textural analysis of the lung fields is the complexity introduced by the presence of normal anatomical background structures. Van Ginneken *et al.* [23] claim that training separate

classifiers for each of the 42 ROIs, thereby allowing them to be examined independently of one another, incorporates anatomical variation into the description of every region (similar reasoning to why position features should be included in the circular regions). A further advantage of training separate classifiers for each of the regions is that it effectively increases the size of the data set - the number of normals and abnormals in the dataset is no longer the number of normal or abnormal images but rather the number of normal and abnormal regions. This is a particularly attractive and important approach in the present study, where the data set is relatively small. Separate classifiers are thus used for each of the regions.

A k-nearest-neighbour (kNN) classifier is used (for a brief description of the method the reader is referred to Section 2.6). This classifier was chosen mainly because of its simplicity - it does not require training, results are repeatable and there is only parameter (k). kNN classifiers have also been shown to perform comparatively to more complex classifiers (such as artificial neural networks and support vector machines) on difficult classification tasks. The value of k has to be determined empirically.

Due to the small size of the data set, a leave-one-out method is used where the training set for a particular classifier consists of the feature vectors for the current region, i ($i = 1, \cdots, 42$), from all the images in the data set except the one to be classified. In the training set, each feature is translated and scaled to have zero mean and unit standard deviation. The same normalisation factors are then applied to the test vector. The k nearest neighbours (based on Euclidean distance) of the test feature vector are extracted from the training set. Each neighbour votes for the region to be abnormal or normal (depending on its abnormality flag). This process is repeated until feature vector i from every image has been used once as a test feature (i.e. classified once). The process is then repeated for the feature vector for the region $(i + 1)$, etc, until every region in every image has been classified. Note that the output of the classification process is a soft label for each feature vector and not a binary normal/abnormal label. This is done to allow for a receiver operating characteristic (ROC) curve analysis to be performed, which is discussed in the subsequent section. The classifier soft labels are determined in the following way [49]:

If the k nearest neighbours are situated at distances $d_1, \cdots, d_k$ from the feature vector to be classified, and have classes $c_1, \cdots, c_k$ where $c = 0$ for normal and $c = 1$ for abnormal regions, the classification output for the current feature vector is:

$$c = \frac{\sum_{i=1}^{k} exp\left(\frac{-d^2}{\sigma^2}\right) c_i}{\sum_{i=1}^{k} exp\left(\frac{-d^2}{\sigma^2}\right)}$$

where σ is the mean of $d_1, \cdots, d_k$.

The value of c falls in the range $0 \rightarrow 1$ where 0 = normal and 1 = abnormal.

The classification outputs can be used in two possible ways: 1) per region or 2) per image. The per region results will be used interactively by a radiologist during image reading in the form of probability maps. These maps will attract the radiologist's attention to a region of abnormality and then allow the radiologist to make the final diagnosis regarding that area. The per image results provide an indication of the abnormality of the image as a whole, which can be used as a stand-alone indication of abnormality or in conjunction with the per region results. It may also be useful in screening applications.

The per region results are used to create probability maps for the chest images where every pixel in the lung fields is replaced by the average of the soft labels of all the regions it falls in - giving a posterior probability of that pixel being normal or abnormal. This is used as a visual tool.

In order to combine the region results into an overall label for the image, the method presented in [23] is used. The method combines the classification results for all 42 regions using a weighted multiplier to give an overall abnormality score for the image. The method is motivated by the fact that an image is abnormal if any region in that image is abnormal. The area A_z under the ROC curve for each region is used to weigh the probabilities for that region since it gives an indication of the classification performance for that region (ROC curve analysis is discussed later in this chapter). The region is only considered if A_z is above some predefined threshold, T_{A_z}, which is determined experimentally. Therefore, if the framework performs poorly on a particular region, the results for that region will not be considered in the classification of the overall image. The weighted probability that an image is abnormal is given by:

$$c = 1 - \prod_{i=1}^{N} (1 - f_i c_i)$$

where,

c_i = the classification result for the i^{th} region (0 = normal, 1 = abnormal),
N = the number of regions,
f_i is the weighting factor for region i:

$$f_i = \text{Max}\left(\frac{A_{zi} - T_{A_z}}{1 - T_{A_z}}, 0\right)$$

4.4.2 Circular regions

The notion of training individual classifiers for each of the regions when considering the circular regions does not apply because each image has a different number of regions per lung field. Instead, one classifier is constructed from all of the regions in the training set; this classifier is then used to classify every region in the test set individually. All experiments in this approach are therefore performed using 5-fold cross validation. The complete data set of 105 images (45 normals and 60 abnormals) is divided into 5 non-overlapping folds. Each fold contains 12 abnormal images and 9 normals. The different subtlety levels are also approximately evenly divided among the 5 folds.

A k-fold cross validation framework consists of k iterations. In each iteration, one of the k folds is used as the test set while the remaining $k-1$ folds are used as the training set. The classification results from the k iterations are then averaged to give an estimate of the performance of the classifier on the entire data set. This ensures optimal use of the available data as well as preventing any overlearning as there is never any overlap between the training and testing data.

Experiments conducted in [29] show that the best classification performance is achieved when using only obviously and relatively obviously abnormal samples in the training set. Testing is done on all abnormal categories. The classifier in this project is therefore trained with only obviously and relatively obviously abnormal samples and normal samples from both normal images as well as normal regions in abnormal images. During training, all the feature vectors from the training folds (left and right lungs) are pooled together. From this pool all the obviously and relatively obviously abnormal samples are extracted and counted. For the sake of computational time, an equal number of normal samples are extracted randomly from the training pool. This then constitutes the training set for the current fold.

This training set is used as before to compute normalisation factors which are then applied to all the training and testing data. The training data is used to train the classifier.

Due to the large number of ROIs per image, using a kNN classifier is computationally expensive. For that reason, a linear discriminant analysis (LDA) classifier is used. The LDA classifier is discussed in further detail in Section 2.6. Again, the output of the classifier is a soft label giving an indication of the probability of each ROI being normal or abnormal (0 = normal, 1 = abnormal). As the ROIs overlap, it is necessary to perform some post-processing to convert these results into a probability map (i.e. a posterior probability for each individual pixel). Again, for each pixel this value is computed as the average of the classifier outputs of all the ROIs within which it lies. More specifically, for pixel i, its posterior probability, p_i is computed as:

$$p_i = \frac{1}{N_{R_i}} \sum_{r \in R_i} p_r^c$$

where,

R_i is the neighbourhood of i defined by the ROIs in which i lies,
p_r^c is the classifier output for region $r, r \in R_i$ and
N_{R_i} is the total number of ROIs in R_i.

This produces equivalent probability maps to those discussed for the big regions, which can be used as a visual tool by a radiologist. The next step is to combine these posterior probabilities into an overall abnormality score for the entire image. Loog *et al.* [51] present a probability combination scheme to do this.

The scheme uses the quantiles (points situated at regular intervals on a cumulative distribution function of a random variable) of the posterior probabilities to fuse the local posteriors into a global posterior probability. The scheme is based on the order statistics of the pixel posterior probabilities. Loog *et al.* [51] motivate this choice with the following observations: the fusion rule is insensitive to the exact behaviour of the pixel posterior probabilities and it allows for deviation from the perfect classifier, where an image contains pathology if and only if there is at least one pixel with a posterior probability higher than 0.5.

For a given $\alpha \in [0, 1]$, the global probability, P_α, which gives a population estimate

of the α-quantile over all per-pixel posterior probabilities, is computed as:

$$P_\alpha = \frac{p_{(\lfloor \alpha n \rfloor)} + p_{(\lceil \alpha n \rceil)}}{2}$$

where,

$\lfloor . \rfloor$ and $\lceil . \rceil$ indicate floor and ceiling respectively,
$p_{(i)}$ is the i^{th} order statistic,
n is the total number of pixels in the lung field.

The i^{th} order statistic referred to above is obtained by sorting all the n posteriors in the lung in ascending order and selecting the i^{th} entry from the list [51]. The value of α is determined empirically.

4.5 Performance evaluation

4.5.1 Lung segmentation

Since the segmentation of the lung fields can be seen as a two-class segmentation problem, the following overlap measure can be used to measure the performance of the system [2]:

$$\Omega = \frac{TP}{TP + FP + FN}$$

where,

TP = true positive area (area correctly classified as lung tissue),
FP = false positive area (non-lung tissue incorrectly classified as lung tissue),
FN = false negative area (lung tissue incorrectly classified as non-lung tissue).

Ω equal to 1, indicates a 100% correct segmentation.

This measure is used, as opposed to the distance between the true and detected landmark points, because the latter is not sensitive to the shifts in the landmark points along the true contour (i.e. the resulting lung border can still be correct if the landmark points are at different positions on that border) [49].

A further indication of the performance of the segmentation algorithm can be obtained by comparing the classification results for the true (manual) segmentations to the classification results for the ASM segmentations. Obviously, large differences in these figures will indicate large differences in the segmentations. This comparison will also give an indication of the system's dependence on good segmentations.

4.5.2 Classification

As is general practice, the performance of the various classification tasks is measured by performing receiver operating characteristic (ROC) curve analysis. An ROC curve plots sensitivity as a function of (1 - specificity) at varying confidence thresholds. The ROC curve can also be described as a plot of the true positive rate vs the false positive rate of a classifier. The area, A_z, under the ROC curve is then used as a classification performance measure. A_z is equal to the probability that a classifier will rank a random positive instance higher than a random negative instance, in other words it gives an indication of the reliability of any random classification result. The perfect classifier has an $A_z = 1$. A_z is computed using the trapezoidal rule [52].

ROC curves are plotted and analysed for both the region results (for the big regions) and the image results (as described previously). Furthermore, separate ROC curves are computed for the different subtlety levels, to give an indication of the system's performance on the different levels. To compute an individual ROC curve for a particular subtlety level, only abnormal samples of that subtlety level are used as positives, while all normals are used as negatives.

Chapter 5

Lung Field Segmentation

This chapter deals with the segmentation phase of the CAD system. Section 5.1 presents the results of the active shape model segmentation of the right and the left lung fields. Section 5.2 discusses these results in further detail and highlights the most important observations.

5.1 Results

As was discussed in Chapter 4, the left and right lung fields were segmented independently of one another. The results given here will show the parameters used in the training of the active shape models, present the performance of the models in terms of the overlap measure discussed and provide illustrations of the best and worst segmentation results in the data set. The impact of the ASM segmentations on the final classification results will be presented and discussed in Chapter 6.

The active shape models were trained using 25 images for the right lung field and 20 images for the left lung field. 65 landmark points were defined along the lung borders in each of the right lung images, and 60 landmark points in each of the left lung images, using the technique presented in Section 4.1.3. The proportion of variation represented in the shape model was set to $p = 0.98$. When the shape model was fitted to a test image the bounds on the eigenvalues were set to $m = 3$. The appearance component was constructed using a profile with a length of 7 pixels ($k = 3$) at 3 levels of resolution. During each iteration of an image search, 5 positions ($n_s = 2$) were evaluated around each landmark for a maximum of $N_{max} = 40$ iterations.

These parameters are summarised in Table 5.1. Pilot experiments showed these

values to return the most consistent results, although no significant changes in performance resulted from varying these parameters within the range of values shown in the last column of Table 5.1.

		Right Lung	Left Lung	Tested
General				
N	Number of training images	25	20	
Shape				
n	Number of model points	65	60	
p	Proportion of variance represented in shape model	0.98	0.98	0.95-0.99
m	Eigenvalue limits	3	3	2-3
Appearance				
k	Number of profile points	3	3	1-7
Image search				
n_s	Number of new positions to evaluate in each iteration	2	2	1-7
L_{max}	Number of resolution levels	3	3	1-5
N_{max}	Maximum number of iterations/level	40	40	5-40

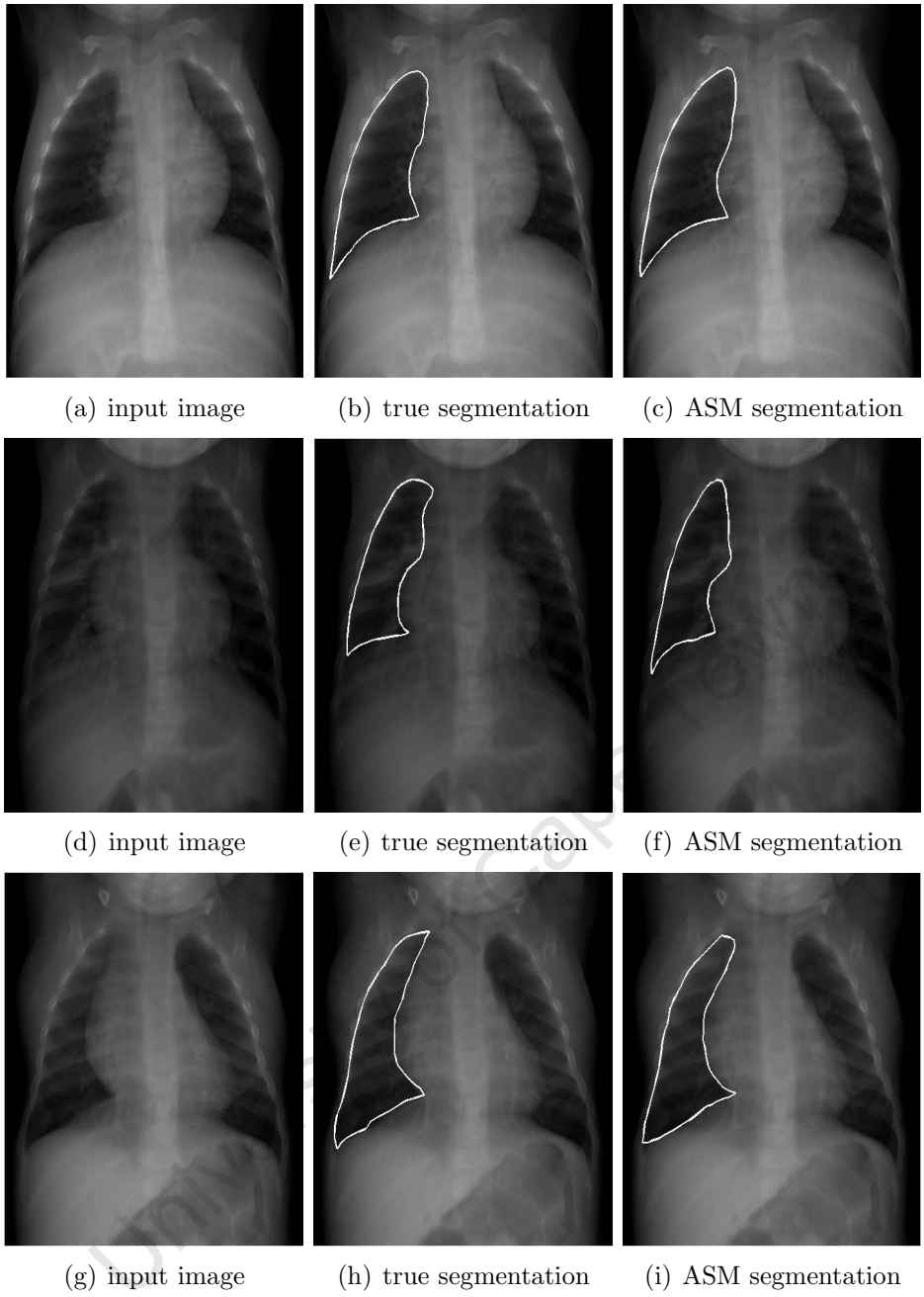
Table 5.1: The parameter values used in the training of the two active shape models.

The results of the active shape models, in terms of the overlap measure defined in Section 4.5.1, trained on the left and right lung fields are summarised in Table 5.2. The right and left lung models yielded average overlaps of 0.933 and 0.919 respectively. Apart from the minimum (worst) score for the two lungs, the segmentation results for the left and right lung field are almost identical.

	$\mu \pm \sigma$	Minimum	Median	Maximum
Right lung field	0.933 ± 0.024	0.858	0.935	0.971
Left lung field	0.919 ± 0.046	0.708	0.931	0.972

Table 5.2: Mean, standard deviation, worst, median and best results for the segmentation experiments performed on the left and right lung fields.

Typical results for the right and left lung field segmentations are shown in Figures



(a) input image (b) true segmentation (c) ASM segmentation

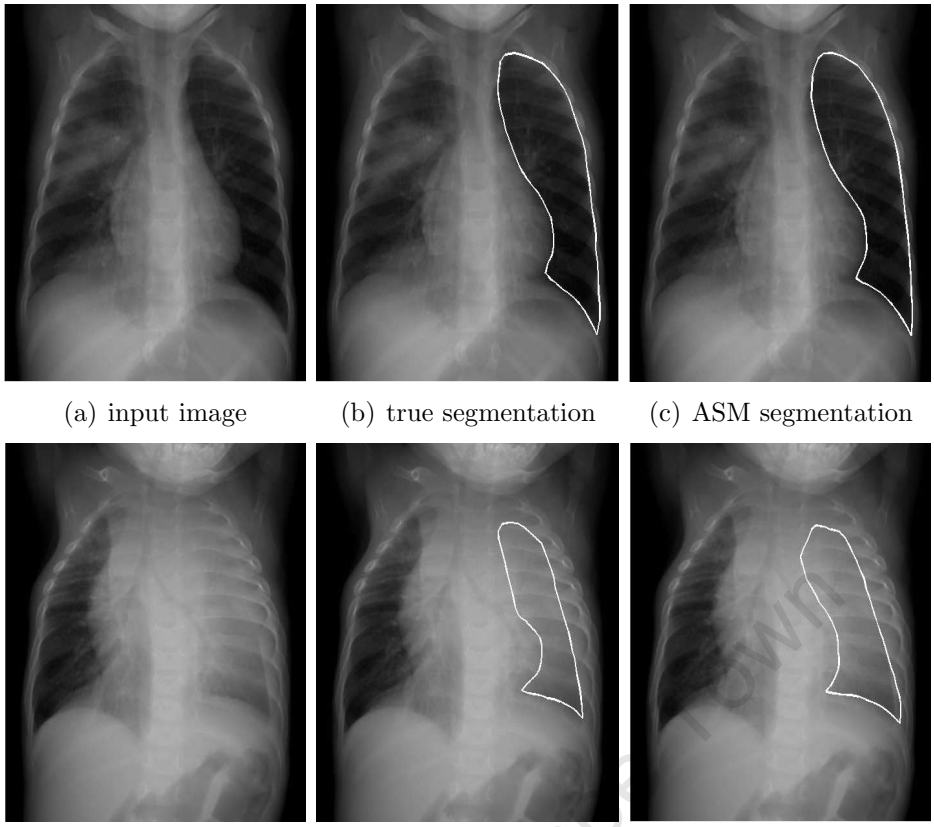
(d) input image (e) true segmentation (f) ASM segmentation

(g) input image (h) true segmentation (i) ASM segmentation

Figure 5.1: ASM segmentation results for right lung field. The top row shows the best segmentation result in the data set. The middle row shows the worst segmentation result—note the large amount of pathology and the irregular shape of the right lung field returning the worst result. The bottom row shows an example of where the model succeeded in segmenting an irregularly shaped right lung—note the wide mediastinum.

5.1 and 5.2 respectively. For each lung the best and the worst segmentation results are shown, along with the input image and the true segmentation. Additionally, an example of where the system succeeded on an irregularly shaped right lung field (irregular due to the visibility of the thymus gland) is shown.

As is explained in Section 4.2.1, the segmentation results are used to warp the



(a) input image (b) true segmentation (c) ASM segmentation

(d) input image (e) true segmentation (f) ASM segmentation

Figure 5.2: ASM segmentation results for left lung field. The top row shows the best segmentation result in the data set and the bottom row the worst. Note, again, the large amount of pathology for the lung field returning the worst result.

region maps in Figure 4.3 (Chapter 4) to the test images. Typical results of the warping of the ROI maps to normal and abnormal test images are shown in Figures 5.3. The top row shows two normal test images. The image in Figure 5.3(c) is an example of a test image with a diffuse nodular pattern in both lungs and the image in Figure 5.3(d) displays abnormally prominent vascular markings in the perihilar regions of both lungs. These images merely show the way the region maps change according to the lung shape. As explained in Section 4.2.1, after warping, the region boundaries are no longer perfectly horizontal or vertical, but the regions are of approximately equal area.

The results of the ROI warping procedure can be used to determine the failure rate of the segmentation scheme on different regions in the lung field. This statistic is valuable as it can be used to provide insight into the performance of the classification algorithm. If it is known that the segmentation scheme always fails to detect the apex of the lung correctly, it can be expected that the classification scheme will perform poorly in detecting abnormalities in that region. In [23], van

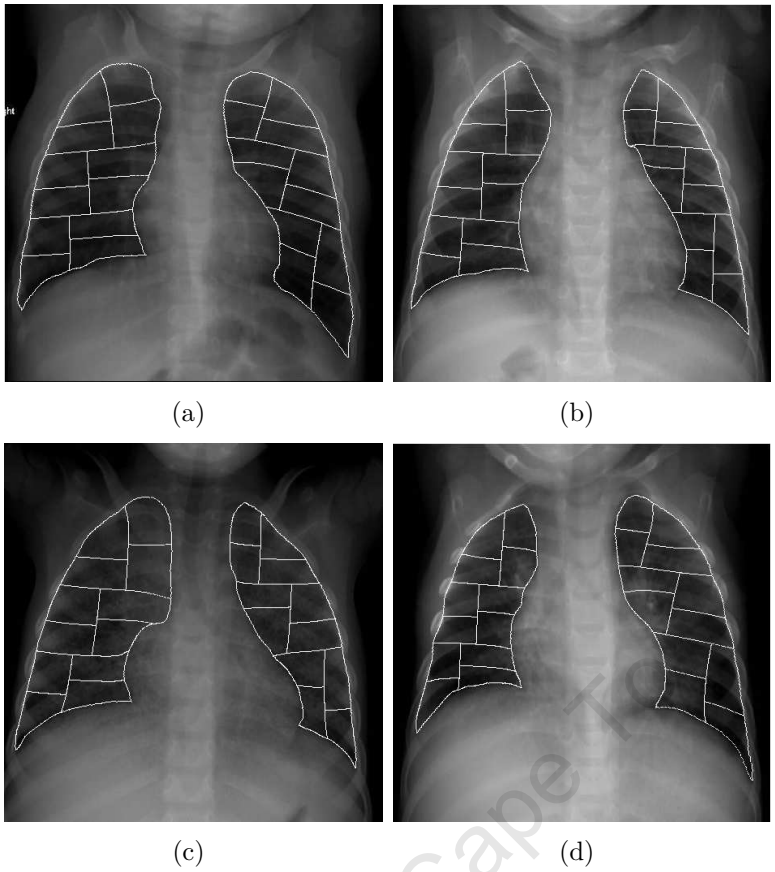


Figure 5.3: Typical results of the ROI warping procedure. The images in (a) and (b) are two normal test images, while the images in (c) and (d) are two abnormal test images. Note that the region boundaries are no longer perfectly horizontal or vertical, but the regions are still of approximately equal area.

Ginneken *et al.* do this by counting the number of failed segmentations on a particular region in the lung. A failed segmentation is defined as one where at least one of the 24 smallest regions into which the lung fields are divided overlaps with less than 50% (in terms of pixels) of the corresponding region when defined using the true segmentation.

Applying similar reasoning to that given for using the overlap measure instead of model point displacements as a performance measure for the ASMs (Section 4.5.1), this approach to assessing the performance of the ASMs on different lung regions can be misleading because it is not sensitive to the shifts in the landmark points along the true contour (i.e. the resulting lung border can still be correct if the landmark points are at different positions on that border). Therefore, the 12 medium sized regions were examined in a similar way.

The small and medium ROIs corresponding to particular lung regions are given

in Table 5.3. For the ROI numbering scheme used in Table 5.3 refer to Figure 4.3 in Chapter 4.

As expected, the mediastinum is the most poorly classified region of the lung (Table 5.4). This is largely due to the fact that the shape of a young child's mediastinum can vary considerably making it difficult to train an active shape model to incorporate this variation, especially with such a small data set. Nonetheless, the results are still encouraging, especially considering the improvement when examining the bigger regions.

Lung Region	small regions	medium regions
Right rib cage	$1 \rightarrow 6$	$13 \rightarrow 15$
Right lung apex	6,7	15,16
Right mediastinum	$7 \rightarrow 12$	$16 \rightarrow 18$
Right diaphragm	1,12	13,18
Left rib cage	$28 \rightarrow 33$	$37 \rightarrow 39$
Left lung apex	27,28	36,37
Left mediastinum	$22 \rightarrow 27$	$34 \rightarrow 36$
Left diaphragm	22,33	34,39

Table 5.3: The ROI numbers that make up particular regions of the lung used to test the failure rate of the segmentation scheme in these lung regions.

Lung Region	small regions	medium regions
Right rib cage	3	0
Right lung apex	7	0
Right mediastinum	12	0
Right diaphragm	0	0
Left rib cage	7	3
Left lung apex	7	4
Left mediastinum	14	3
Left diaphragm	6	2

Table 5.4: Failure rate of segmentation scheme on different lung regions. A failure is defined as one where the segmentation result for at least one of the ROIs comprising the lung region is less than 50% correct.

Table 5.5 shows the number of abnormal regions for each of the 42 regions in the

entire data set; a region is defined as abnormal if at least 10% of that region overlaps with an abnormal region marked out by the radiologist. The results shown in the table correspond to the regions resulting from the ASM segmentation; the equivalent numbers for the true segmentations are shown in brackets in the table. The fact that the numbers for the ASM and true segmentation regions are almost identical is encouraging, as this is an indication that the segmentation scheme has captured all/most of the abnormalities correctly in the segmentation. The impact of these statistics on the final classification results are discussed Section 6.2.1.

region	#	region	#	region	#	region	#
1	3 (3)	**12**	35 (39)	**23**	12 (12)	**34**	14 (14)
2	6 (6)	**13**	5 (5)	**24**	11 (10)	**35**	10 (9)
3	13 (12)	**14**	13 (13)	**25**	6 (7)	**36**	10 (9)
4	10 (10)	**15**	20 (18)	**26**	12 (13)	**37**	14 (15)
5	22 (22)	**16**	38 (37)	**27**	7 (6)	**38**	24 (26)
6	14 (15)	**17**	53 (50)	**28**	5 (6)	**39**	24 (24)
7	15 (16)	**18**	46 (46)	**29**	19 (18)	**40**	23 (23)
8	39 (39)	**19**	39 (41)	**30**	23 (23)	**41**	25 (25)
9	46 (45)	**20**	43 (41)	**31**	23 (23)	**42**	14 (12)
10	48 (47)	**21**	34 (35)	**32**	19 (19)		
11	47 (48)	**22**	14 (15)	**33**	24 (24)		

Table 5.5: Region abnormality count: the number of abnormal regions for each of the 42 regions resulting from the ASM segmentations. The equivalent values for the true segmentations are shown in brackets.

5.2 Discussion

Several aspects of the segmentation results are worth discussing. The training of the ASM for the right lung required a slightly larger training set to yield comparable results to that of the left lung. This is as a result of the greater variance in the shape of the right lung, mainly due to the varying appearance of the thymus gland in some of the images. The thymus gland is discussed in further detail in Section 3.3.

The segmentation results for both the left and the right lung fields are excellent. The overlap values in Table 5.2 are comparable, if not slightly better than those achieved in similar experiments presented in [2] and [49]. In those experiments, higher overlap values were only achieved by applying several modifications to the standard active shape model. In [2], the left and right lung fields are segmented

together, along with the heart and clavicles. The overlap value for the lung segmentation using default ASM parameters (defined in [3]) was 0.903. By finding the optimal parameter values, the performance was improved to 0.927. The results were improved further by forming a hybrid segmentation technique by combining ASMs and the pixel classification approach, giving an overlap value of 0.934. In [49], the left and right lung fields are segmented independently of one another using ASMs. Using the standard ASM technique (as used in this study) the overlap values for the right and left lung fields were 0.882 and 0.861 respectively. The standard ASM approach was modified to be steered by optimal local features, as opposed to normalised first order derivative profiles. This modification improved the segmentation results on the right and left lungs to 0.929 and 0.887 respectively. A summary of these comparisons is shown in Tables 5.6 and 5.7.

The high overlap values presented here, can be attributed largely to the fact that separate models are trained for the left and right lung fields, as this immediately simplifies the search considerably. Furthermore, an accurate starting point for the search is provided by the user, which (as mentioned in Section 4.1.7), improved results in pilot experiments considerably.

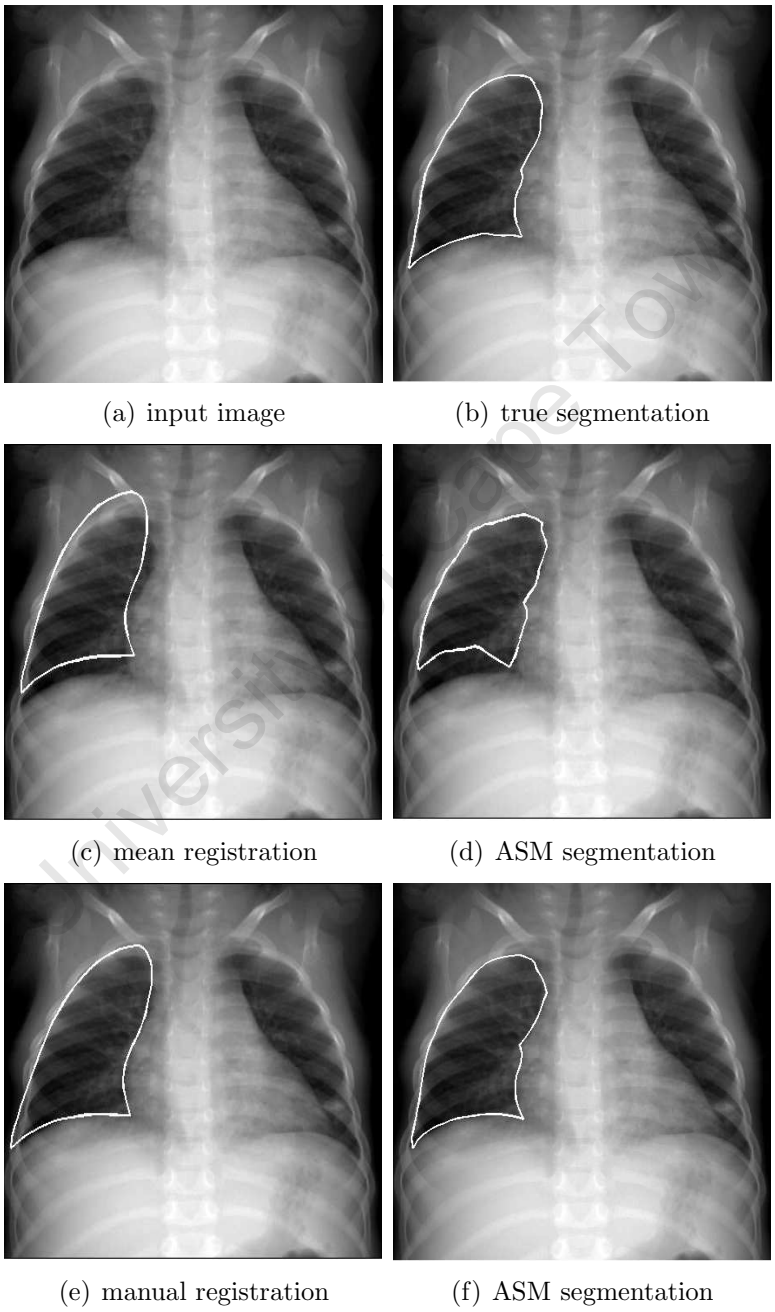
To illustrate the impact of having the user provide an approximate starting point for the model, an example of the search performed on a right lung field initiated at the mean and then initiated manually is shown in Figure 5.4. The improvement due to the manual registration is clear. As pilot testing immediately showed considerable improvements in performance in almost all instances, when using manual registration, a complete comparison between the results of the two approaches is not provided.

Method	Overalp value
ASM (default)	0.903
ASM (tuned)	0.927
ASM/PC Hybrid	0.934

Table 5.6: Lung field segmentation performance presented in [2]. The two lung fields were segmented together. ASM (default) is the standard ASM approach using the default parameters suggested in [3]. ASM (tuned) is the standard ASM approach using the optimal paramters. ASM/PC Hybrid as a hybrid approach, combining the ASM and pixel classification approaches.

Method	Right Lung	Left Lung
ASM	0.882	0.861
ASM with optimal features	0.929	0.887

Table 5.7: Lung field segmentation performance presented in [1]. The lung fields were segmented independently of one another. ASM is the standard ASM approach. ASM with optimal features is a modified ASM approach that is steered by optimal local features.



(a) input image (b) true segmentation

(c) mean registration (d) ASM segmentation

(e) manual registration (f) ASM segmentation

Figure 5.4: Comparison between ASM results for mean and manual registration. Beginning the model search at a manually positioned location produces considerably better results than beginning the search at the mean model position in the image.

5.3 Conclusion

The segmentation results presented here are satisfactory. The performance of the ASM algorithm was improved considerably by employing the semi-automatic approach, whereby the user defines the starting point of ASM search in each test image. User initiation of the search is acceptable in a computer aided diagnostic tool, as the aim is to assist the radiologist in locating abnormalities, rather than producing a fully automatic solution, as discussed in Section 2.4.

Although the overlap values presented here are high, and compare well with those of previous studies, the true measure of the performance of the segmentation phase of any lung CAD system is the impact that the segmenation has on the final classification results. This comparison is presented and discussed in Section 6.2.3.

Chapter 6

Feature Extraction and Classification

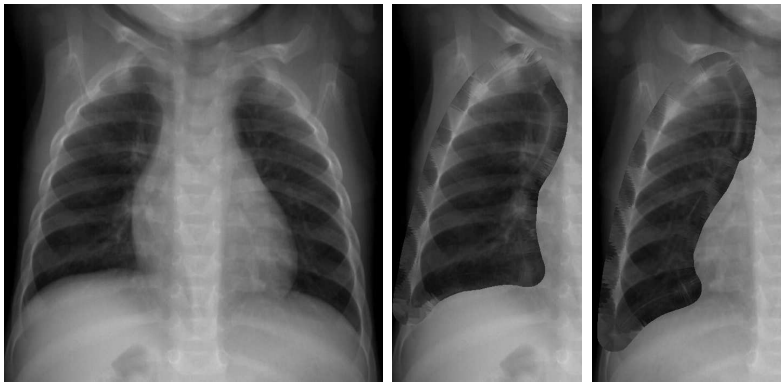
This chapter presents and discusses the results of the feature extraction and classification phases of the system. Section 6.1 presents the results of the pre-processing and feature extraction processes, as well as the classification results of the big and circular region approaches. The results are then discussed in detail in Section 6.2. Finally, the conlusions and most important observations of the chapter are summarised in Section 6.3.

6.1 Results

6.1.1 Image pre-processing and filtering

Typical results for the pre-processing steps are shown in Figure 6.1. The left lung field was flipped and the pixels around the borders of the left and right lung fields were mirrored to resemble the inside of the lung, to avoid distortions in the filter outputs along the lung borders.

The results of filtering the left lung field from Figure 6.1 with a bank of Gaussian filters are shown in Figure 6.2. All 30 filtered versions of the lung are shown. Note the different structures that are prominent in each filtered image. The impact of using multiple scales is clear when comparing the first row of images (scale $\sigma = 1$) to the last row of images ($\sigma = 16$). At a scale of 1, the textural detail in the image is much finer (ribs visible) compared to the more global appearance of the images at scale 16 (only general lung shape visible). Similarly, the additional information



(a) (b) (c)

Figure 6.1: The preprocessing steps on an example image: (a) the input image; (b) right lung field with border pixels mirrored and (c) the left lung field flipped, and with border pixels mirrored.

obtained by using multiple orders and directions for the filters is clear when comparing the individual images in each row.

6.1.2 Feature extraction and classification

Big regions of interest

Using a 9-NN classifier and threshold $T_{A_z} = 0.8$, the final classification results (in terms of area, A_z, under the ROC curve) of the system when using the big regions of interest were as follows: a per region score of $A_z = 0.752$ averaged over the 42 regions and a combined image score of $A_z = 0.982$ over the entire data set. The results are presented in further detail below.

The number of k nearest neighbours and the threshold T_{A_z} (used to combine region scores into an overall image score) were determined empirically during the classification stage. The value for k was chosen by first examining the average per region A_z for varying values of k and then examining the image A_z values for varying k and T_{A_z}. These two experiments were conducted for the entire data set, using the leave-one-out approach discussed in Section 4.4.1. From these two experiments, the best values could be chosen. The resulting curves are shown in Figure 6.3. Varying k between 9 and 39 yielded very little change in performance for the average per region scores. The plot showing the covariance of k and T_{A_z} again shows very little change in results for k values in the range 9 to 39 and T_{A_z} values in the range 0.6 to 0.9. The optimal operating point in Figure 6.3(a) nonetheless, occurred for $k = 9$ and $T_{A_z} = 0.8$. The legend on the right of Figure

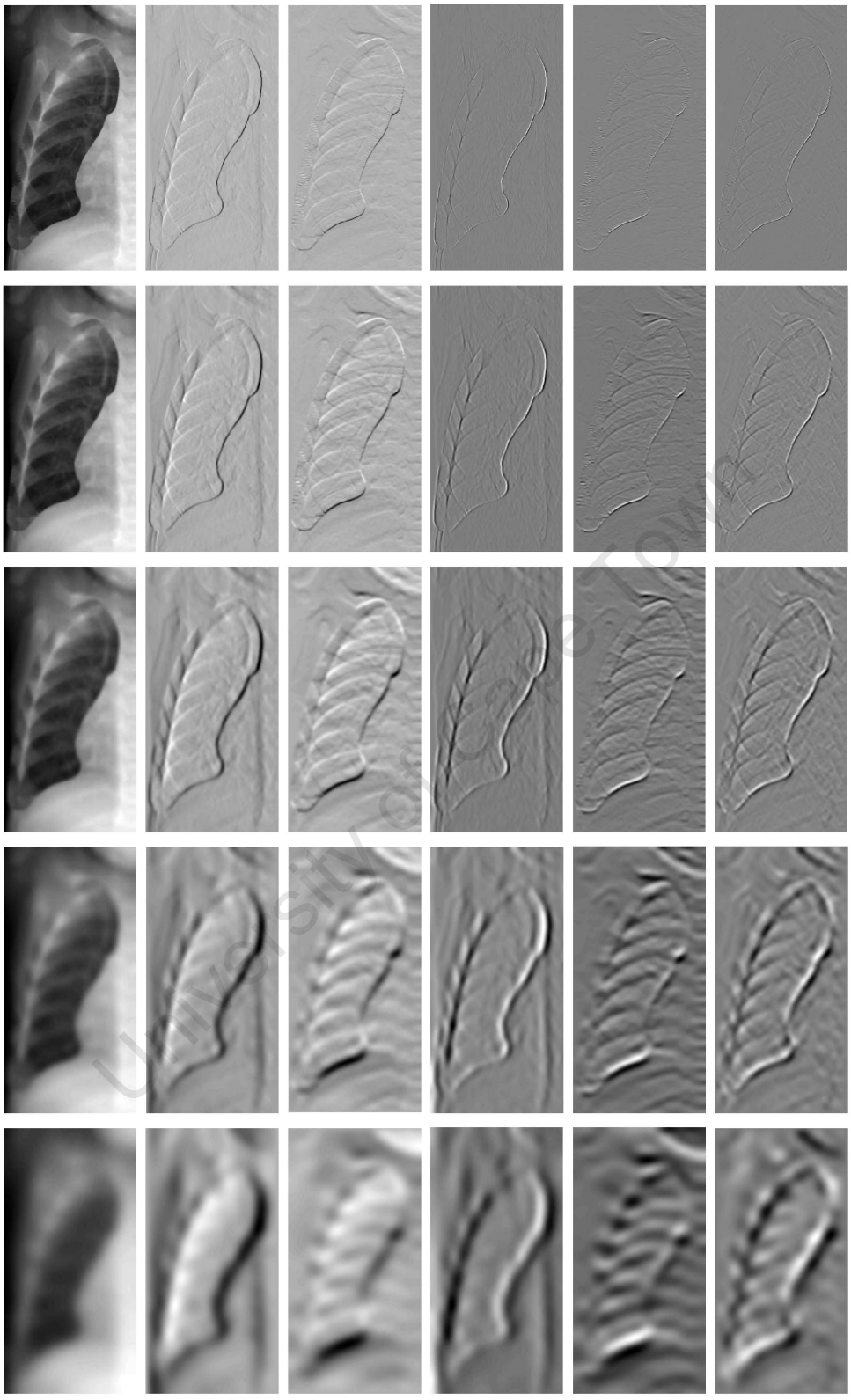
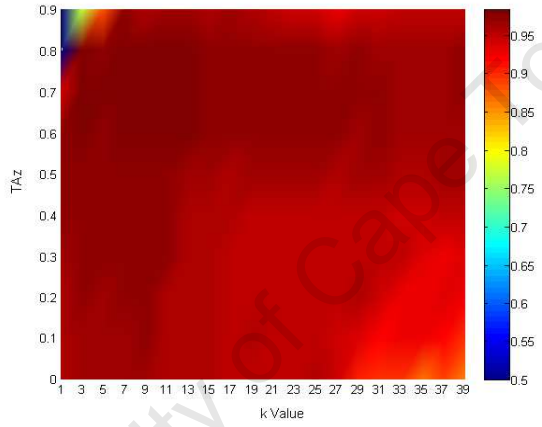


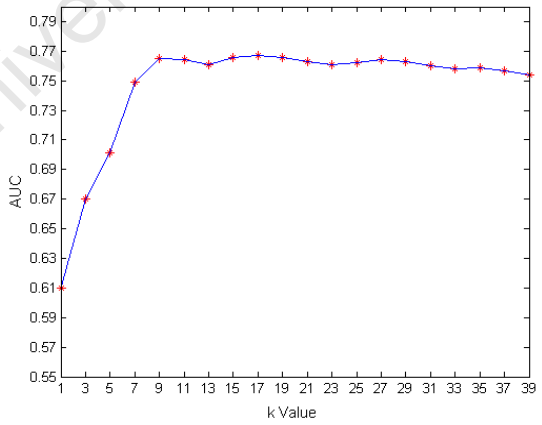
Figure 6.2: The 30 filtered versions of the left lung (flipped) from Figure 6.1. Each row shows the filtered images $L_{00}, L_{10}, L_{01}, L_{20}, L_{02}, L_{11}$ at one scale (rows $1 \rightarrow 5$ correspond to scales $\sigma = 1 \rightarrow \sigma = 16$).

6.3(a) shows the colour corresponding to every A_z value; the region in the plot displaying the darkest red corresponds to the optimal operating point. As this value for k corresponds well with the curve in Figure 6.3(b), these parameter values were chosen.

The resulting A_z values for each of the 42 regions are shown in Table 6.1. To clarify these results, Figure 6.4 shows lung subdivisions with the corresponding A_z inside the region. When considering the region scores, the system performed most poorly on the bottom lateral corner of the right lung and the apical and medial regions of the left lung. In general the results for the right lung were slightly better than those for the left lung.



(a)



(b)

Figure 6.3: Experimental determination of k and T_{A_z} values. The image in (a) shows the covariance of k and T_{A_z} with the corresponding A_z value. The curve in (b) shows the area, A_z or AUC, under the ROC curve, averaged over the 42 regions for varying k. The optimal values for T_{A_z} and k in (a) are 0.8 and 9 respectively, giving a per image A_z of 0.982. In (b) the optimal k value of 9 yields an average per region A_z of 0.752.

region	A_z	region	A_z	region	A_z	region	A_z
1	0.436	**12**	0.774	**23**	0.591	**34**	0.682
2	0.593	**13**	0.929	**24**	0.546	**35**	0.679
3	0.744	**14**	0.866	**25**	0.636	**36**	0.561
4	0.741	**15**	0.843	**26**	0.590	**37**	0.699
5	0.820	**16**	0.832	**27**	0.687	**38**	0.791
6	0.787	**17**	0.906	**28**	0.335	**39**	0.925
7	0.702	**18**	0.871	**29**	0.768	**40**	0.813
8	0.905	**19**	0.856	**30**	0.794	**41**	0.731
9	0.862	**20**	0.906	**31**	0.774	**42**	0.626
10	0.831	**21**	0.773	**32**	0.911		
11	0.900	**22**	0.692	**33**	0.871		

Table 6.1: A_z values for individual regions.

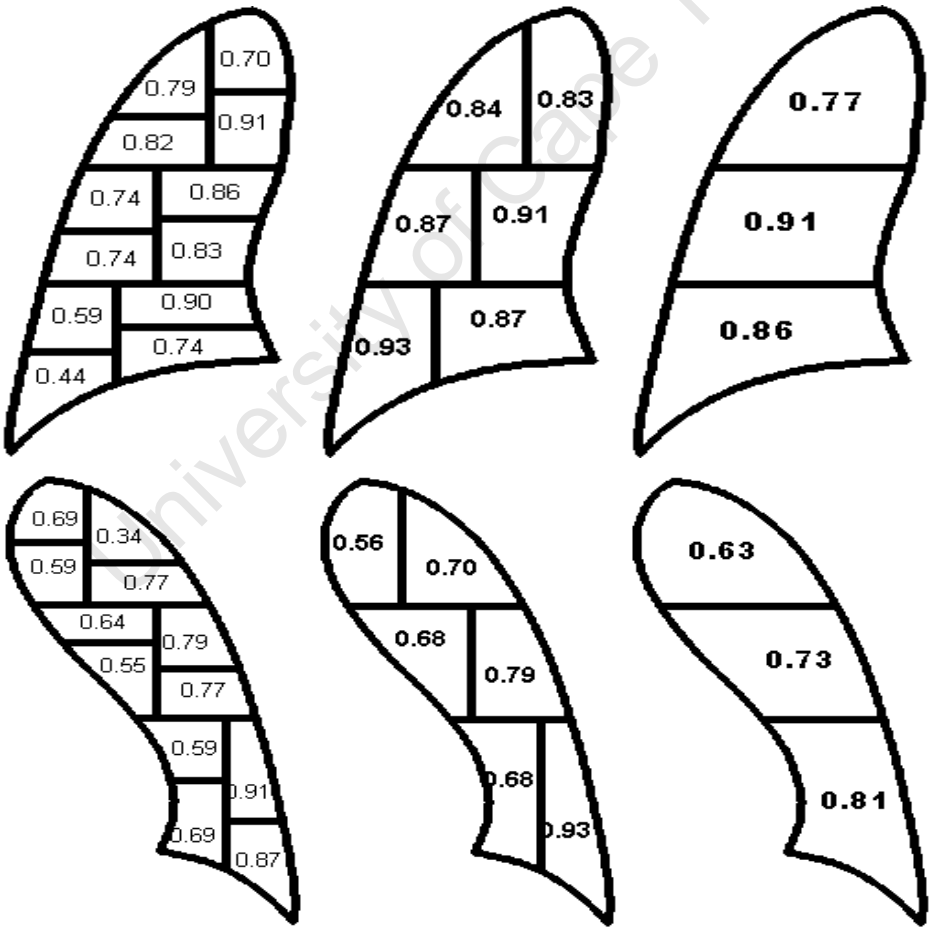


Figure 6.4: The A_z values for each of the 42 regions in the lungs fields. The top row represents the right lung, and the bottom row, the left. The A_z value is indicated inside the corresponding ROI.

The objective of a probability map is to indicate the degree of abnormality of every region in the lung field. The colour overlays on the lungs in Figures 6.5 to 6.7 define this degree of abnormality—a very white region will indicate extreme abnormality, while a greyish-white region will indicate that some subtle abnormalities could be present. A value of 0 (black) corresponds to 100% normal, while a value of 1 (white) corresponds to 100% abnormal. The manner in which the probability maps are obtained, is explained in Section 4.4.1. The abnormalities as well as the lung borders in each of the images are outlined in red.

Figures 6.5 to 6.7 show probability maps produced by the CAD system for the following 3 situations: 1) probability maps which accurately indicate abnormalities of varying subtleties, 2) probability maps which accurately depict healthy cases and 3) the most common failures of the system.

In Figure 6.5, the probability maps produced by the system for images containing abnormalities in each of the five subtlety levels are shown. Each abnormal image shown here contains abnormalities from only one of the five subtlety levels. The images show examples of where the system produced probability maps that accurately depict the region and degree of abnormality. As is expected, the probability maps for the lower subtleties are more definitive (whiter in the region of the abnormality) than those for the more difficult cases. Nonetheless, although the abnormality in the final image is ranked as extremely subtle, the probability map still showed some response in the abnormal region, enough perhaps to attract the attention of a radiologist.

The images in Figure 6.6 illustrate the performance of the system on normal images. The three cases shown correspond to a normal image cropped from a full body image, a normal image from the TB study (i.e. a chest scan) and an abnormal image from the TB study in which the regions of abnormality are small and most the lung tissue is normal. In all three cases, the probability maps accurately and clearly depict the normality of the healthy lung tissue.

Finally, the probability maps in Figure 6.7 illustrate the most common failures of the system—both in terms of false positives and false negatives. The most frequent errors occurred in the basal and perihilar regions of the lungs, especially those of the left lung. The final example in Figure 6.7, is an interesting case. The image shown contains a diffuse nodular pattern throughout both lung fields and is quite clearly abnormal, and was thus assigned a subtlety ranking of 1 (obviously abnormal). The resulting probability map however, shows little indication of ab-

normality in the left lung field.

The region scores were combined using a threshold, $T_{A_z} = 0.8$. The classification results for the different subtlety levels are shown in Table 6.2 and the corresponding ROC curves are shown in Figure 6.8. Due to the lack of data in some of the subtlety levels, subtleties 1 and 2 were combined and subtleties 4 and 5 were combined.

When tested on all of the data, the system achieved an area under the ROC curve of, $A_z = 0.982$ at a sensitivity of 0.935 and a specificity of 0.932. The values are consistently high for all the subtlety levels.

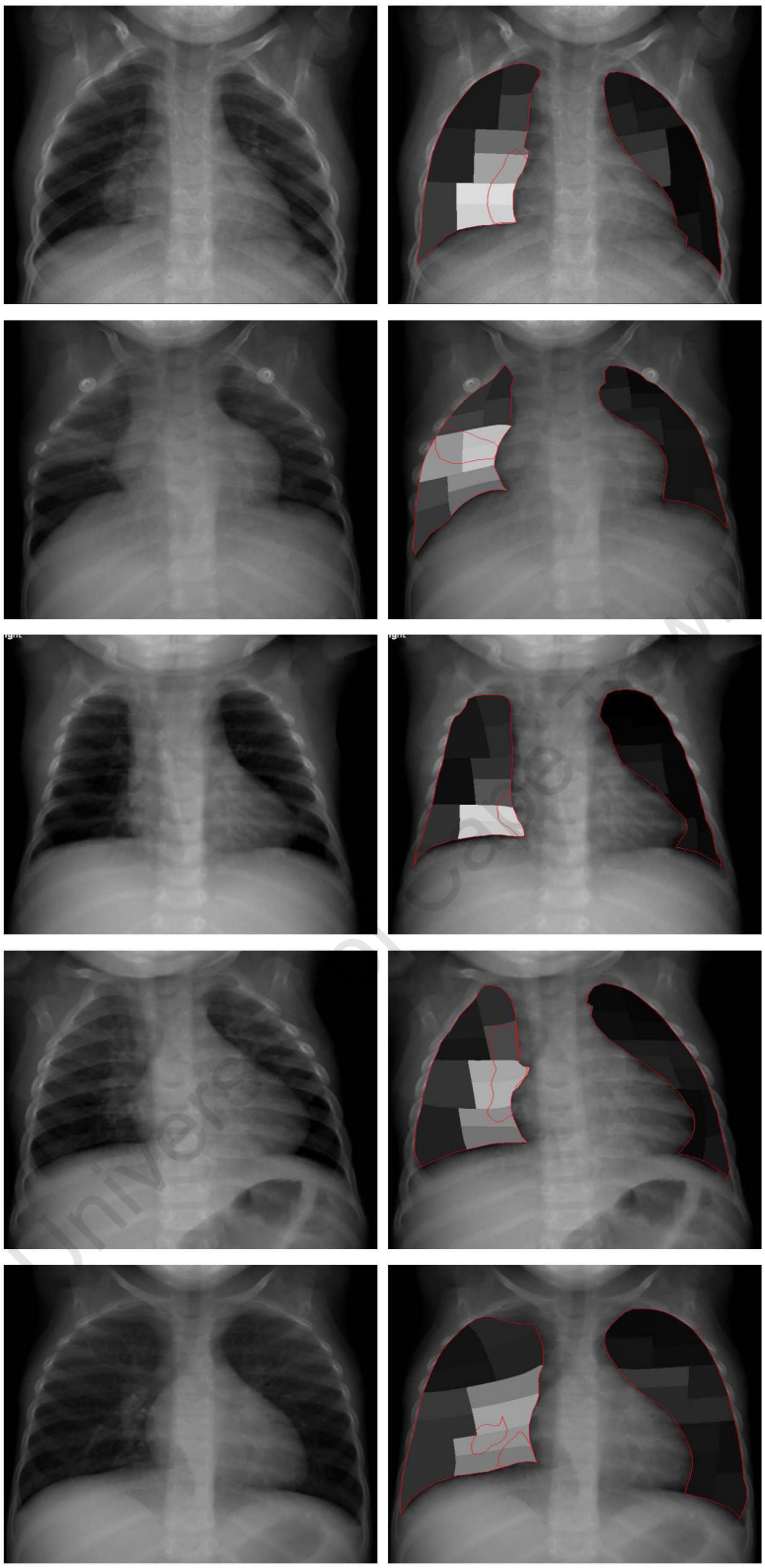


Figure 6.5: Successful probability maps for the big regions for each subtlety level. The first row shows the input and output for subtlety 1 and the last row for subtlety 5. In each case the probability map indicates the abnormality in the appropriate region (brighter grey values = more abnormal), and accurately depicts the rest of the lung fields as normal (darker grey levels).

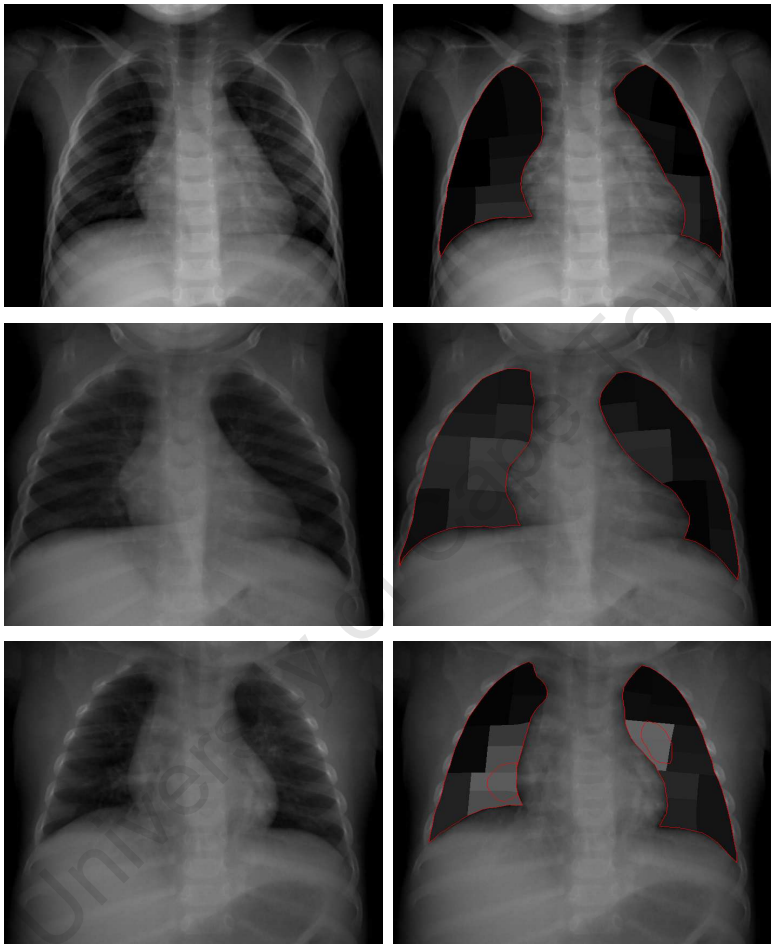


Figure 6.6: Successful normal probability maps for big regions. The first row shows a normal image from a full-body scan, the second row shows a normal image from the TB study and the final row shows an abnormal image with large regions of normal tissue. In each case the healthy lung tissue is accurately depicted by darker grey values.

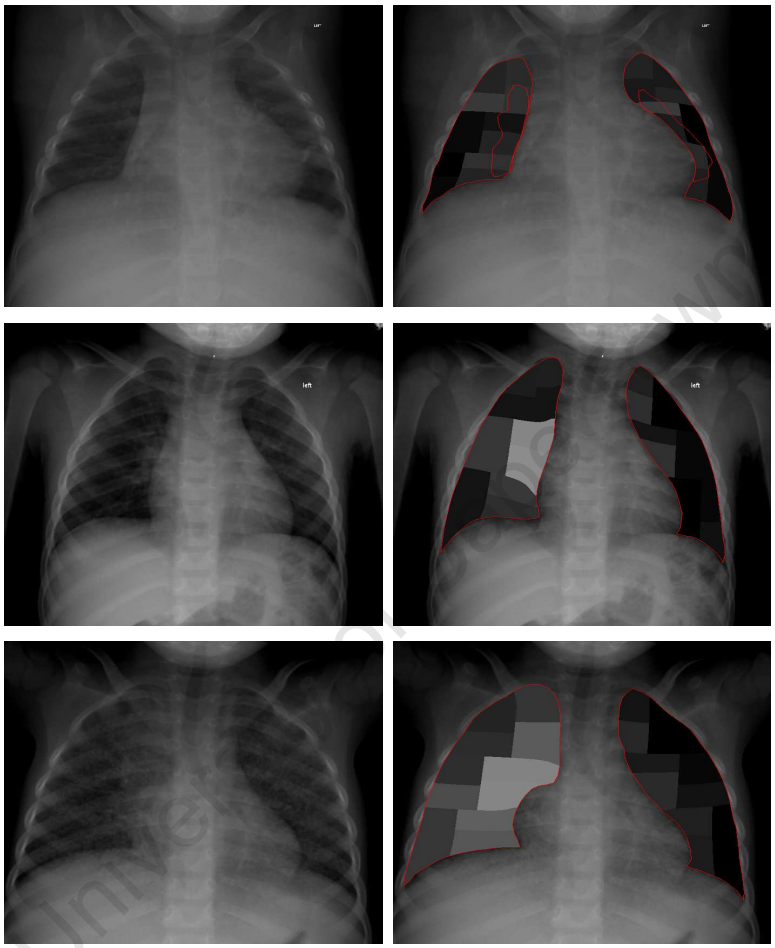


Figure 6.7: Common errors for big regions. The first two rows show an example of how the system struggled in the perihilar regions of the lungs. The first row is an abnormal image (increased perihilar vascular markings in the left and right lung field) and the second row is a normal image. Note that the probability map for the normal case indicates a greater degree of abnormality in the right perihilar region. The final row shows an input image with a diffuse nodular patter in both lungs which is only depicted in the right lung in the probability map.

Abnormality Subtlety	A_z	Sensitivity	Specificity
Obvious and Relatively Obvious	0.990	0.978	0.929
Intermediate	0.998	1	0.941
Subtle and Very Subtle	0.959	0.848	1
All categories	0.982	0.935	0.932

Table 6.2: A_z values for different subtlety rankings as well sensitivities and specificities for each subtlety level.

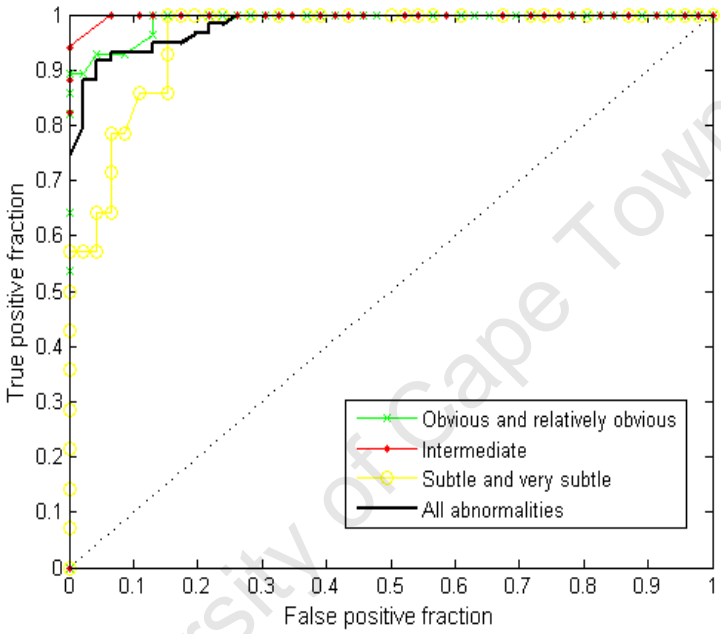


Figure 6.8: ROC curves for the evaluation of per image results for big regions using 9NN classifier. The curves have been plotted for the different abnormality subtlety levels.

The process of computing the filtered versions of a test image, extracting the 124 features from each of the 42 regions in the image, training the classifier and classifying the image took on average $10 \rightarrow 15$ seconds per image (depending on the sizes of the lung fields in that image) on a 1.73GHz Intel Core Duo processor with 1G of RAM.

Circular regions of interest

Using a linear discriminant analysis (LDA) classifier in a 5-fold cross-validation framework and a quantile fusion scheme with $\alpha = 0.89$ to combine pixel posteriors into an overall image score, the final classification performance of the system when

using the circular regions of interest was $A_z = 0.941$.

Figures 6.9 and 6.10 show the probability maps illustrating the performance of the CAD system using circular regions. Due to the large number of overlapping regions in every lung field, the probability maps appear more diffuse than those for the big regions. The lung borders and the abnormalities are again outlined in red.

The images in Figure 6.9 show examples of where the CAD system produced accurate probability maps. Examples from each of the 5 subtlety levels are shown, as well as an example of a normal image. The images in Figure 6.10 show the most common shortcomings of the system.

Although the system performed well on a large number of cases, it appeared to have more shortcomings than the big region approach. As with the big region approach, performance in the perihilar region was fairly poor. The system also, however, performed inconsistently in areas which were not problematic for the big region approach (Figure 6.10). Firstly, the bottom lateral corners of both lungs tended to be classified incorrectly as abnormal in a large number of images and false positives also occurred fairly frequently in the apices of the lungs. Finally, the system also seemed to perform more poorly on normal images, especially those originating from full-body scans—probability maps for several of these images displayed fairly large, diffuse areas of abnormality, especially in the perihilar regions of the lungs.

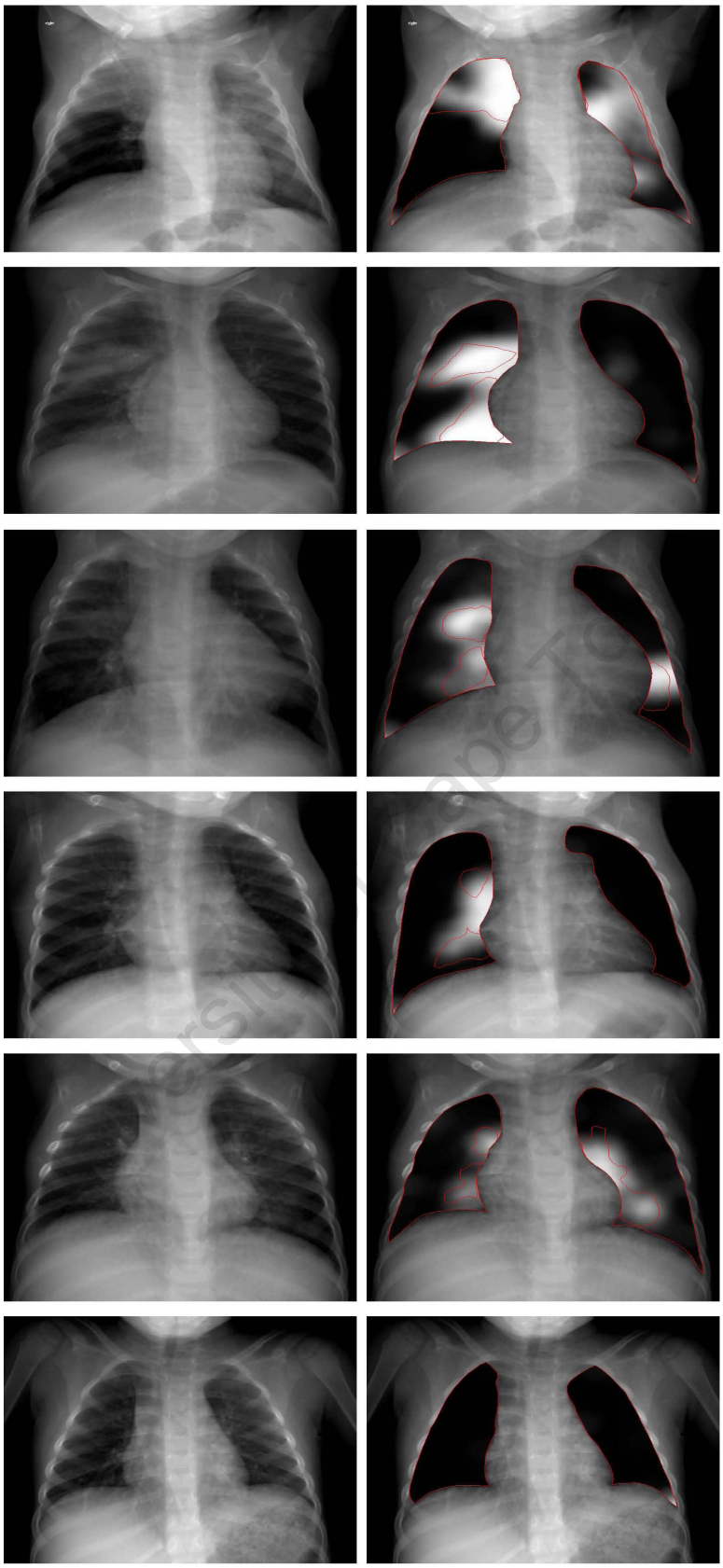


Figure 6.9: Successful probability maps for circular regions. The first 5 rows show examples of images falling in each of the 5 subtlety levels (Row 1 = subtlety 1 etc...). The final row shows an example of a normal image. In all of the cases the abnormalities are clearly and distinctively indicated by the probability maps and the healthy lung tissue is accurately depicted by the darker grey levels.

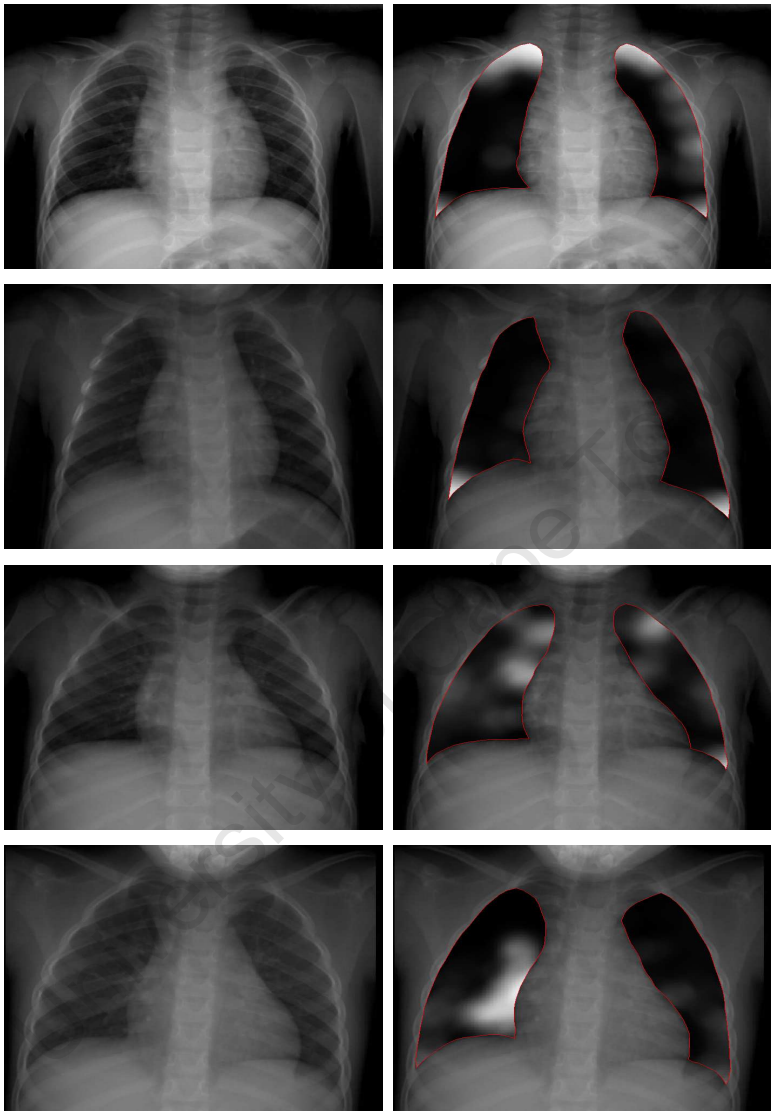


Figure 6.10: Common errors for circular regions. The first row shows an example of where the system failed in the apices of a normal image. The second row shows an example of where the system failed in the bottom lateral corners of a normal image. The third row shows a false positive case caused by the visibility of the clavicles and sternoclavicular junctions and the final row shows a diffuse indication of abnormality in a normal image, due to prominent vascular markings in the right perihilar region. All these normals originated from full-body scans.

The pixel posteriors used to create the probability maps were combined into overall abnormality scores for the whole images using quantile fusion (Section 4.4.2) with $\alpha = 0.89$. The final classification performance of the system, in terms of area, A_z, under the ROC curve, on all the data was 0.941.

The optimal value for α was determined experimentally—the curve in Figure 6.11 shows the area under the ROC curve for varying values of α. The performance increased significantly with increasing values for α. For a small range of values around the optimal value ($0.79 \leq \alpha \leq 0.99$) there was very little change in performance–this is the stable region of the system. As expected, the classification performance drops when α appoaches 0 and 1.

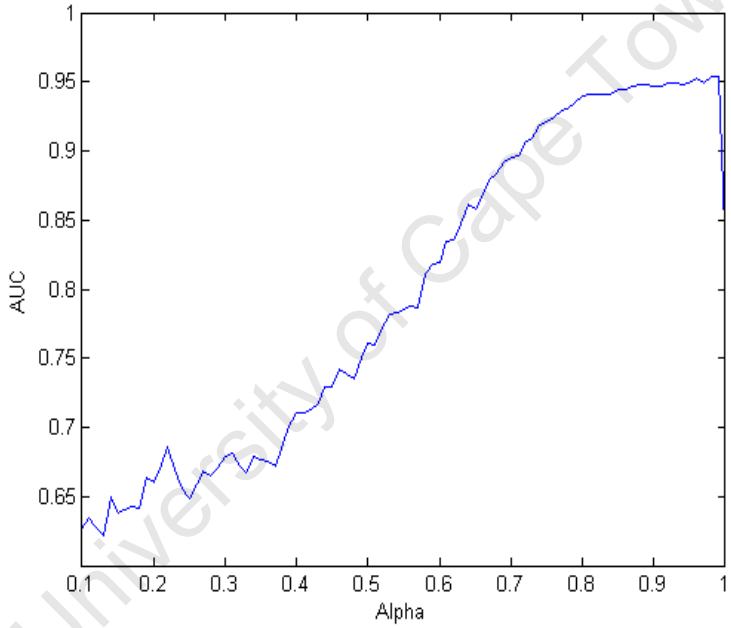


Figure 6.11: The area under the ROC curve (indicated as AUC in the image) for varying α. The curve is fairly unstable for $\alpha < 0.75$, but small deviations from the optimal value of 0.89 do not cause a significant change in A_z. The curve drops off as α tends to 0 and 1.

The classification results for the various subtlety levels are shown in Table 6.3 and the corresponding ROC curves in Figure 6.12. When tested on the data from all of the subtlety levels, the system yielded an area under the ROC curve of $A_z = 0.941$, at a sensitivity of 0.925 and a specificity of 0.879. Performance remained high when tested on subtlety levels of 3 and lower, but dropped considerably for the subtle and very subtle cases with $A_z = 0.846$ at a sensitivity of 0.800 and a specificity of 0.786.

The process of computing the filtered versions of a test image, extracting the 126 features from the circular regions in the image, training the classifier and classifying the image took considerably longer for the circular regions, due to the large number of regions (2000) to be dealt with in every image. For an image with 2000 circular regions (1000 per lung) the feature extraction process took approximately 6 minutes. The cross validation testing procedure took on average 10 minutes per fold (30 seconds per image).

Abnormality Subtlety	A_z	Sensitivity	Specificity
Obvious and Relatively Obvious	0.986	0.925	0.966
Intermediate	0.953	0.925	0.882
Subtle and Very Subtle	0.846	0.800	0.786
All categories	0.941	0.925	0.879

Table 6.3: A_z values for different subtlety rankings for the circular regions, as well as the sensitivity and specificity of the classification at each subtlety level.

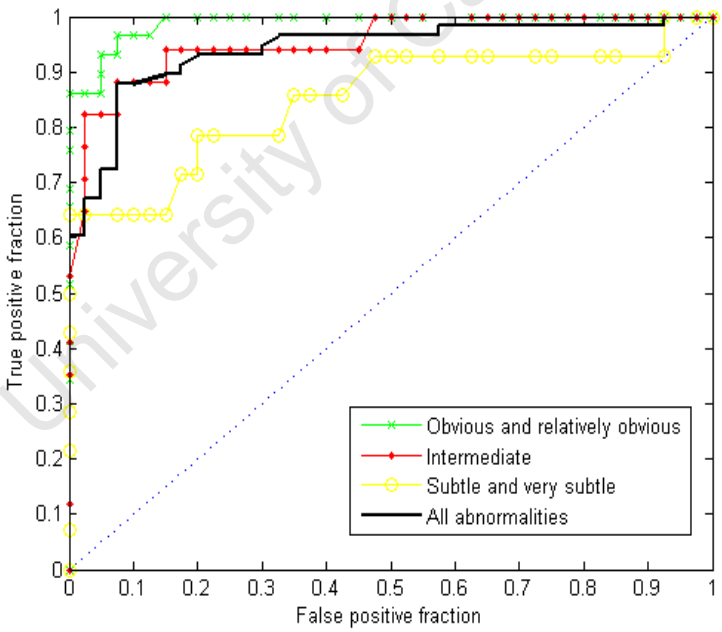


Figure 6.12: ROC curves for the evaluation of per image results for circular regions using LDA classifier. The curves have been plotted for the different abnormality subtlety levels.

6.2 Discussion

The results obtained for the big region and the circular region approaches are discussed in further detail below.

6.2.1 Big regions of interest

Table 6.1 shows that the system performs most poorly in regions 1 and 2 in the right lung field and regions 23 to 28 in the left lung field. Tables 5.4 and 5.5 in Chapter 5, provide insight to these results.

Table 5.4 shows the failure rate of the ASM segmentations for different regions in the lung fields. Table 5.5 shows the number of abnormal regions for each of the 42 regions in the entire data set; where a region is defined as abnormal if at least 10% of that region overlaps with an abnormal region marked out by the radiologist. The results shown in the table correspond to the regions resulting from the ASM segmentation; the equivalent numbers for the true segmentations are shown in brackets in the table.

Table 5.5 shows that regions 1 and 2 in the right lung field only contain 3 and 6 abnormalities respectively—in other words, one false negative alone in region 1 would result in an error rate of 33%. Such sparse data makes its difficult to obtain a true reflection of the system's performance. Similarly, regions 23 to 28 (especially the apical regions 27 and 28) are also very sparsely populated (in terms of number of abnormalities). Furthermore, Table 5.4 shows that the segmentation results were poorest along the mediastinal border of the left lung (regions 22-27). The regions which are most densely populated in terms of abnormalities correspond directly with the regions which were most accurately classified and it appears as if the classification results are proportional to the number of abnormalities in that region—in fact, all regions containing 15 or more abnormalities returned an $A_z > 0.7$.

The figures in Table 3.2 in Chapter 3 (showing the number of abnormalities for each subtlety level) clarify these results further. The table shows that in the entire data set 89 abnormalities occur in the right lung and only 46 in the left lung. Furthermore, of the 46 abnormalities in the left lung, 36 have a subtlety grading of 3 or higher. It is therefore understandable that the CAD system performs better on

the right lung field than the left—as the right lung field has both more and easier data to train and test on. These factors clearly highlight the need for a larger, more diverse database.

Note that in Table 6.1 the performance of the system in the bottom lateral corner of the right lung is very poor for the smallest regions but increases considerably for the medium and larger regions. This indicates that using overlapping ROIs of varying sizes compensates somewhat for the low scoring individual regions. Recall that the final output of the system (for both the probability maps and the combined image scores) uses a combination of the scores for the overlapping regions. Since the probability maps for the big regions are obtained directly from combining the per region classification results (as explained in Section 4.4.1), the effects of combining the results over the various overlapping regions can be seen most clearly in the probability maps.

The probability maps in Figures 6.5 and 6.6 show that the system generally performs accurately for the full range of abnormality subtleties as well as for the normal images. A potential problem with the data set, however, is that most of the normal images originate from full body scans and may thus have different characteristics than the normal images or regions from the TB study, due to the different spatial resolutions at which they were taken. This would result in a false indication of the classification performance of the system as it would become too simple for the system to distinguish between normal data from full body images and abnormal data from chest only images. The results shown in Figure 6.6 are very encouraging in this sense. The system appears to treat the normal images from both the TB study and the full-body scans, as well as the normal regions in the abnormal images similarly.

The probability maps in Figure 6.7 show that the most frequent errors occurred in the basal and perihilar regions of the lungs, especially those of the left lung. These errors may largely be attributed to: 1) the sparsity of the data (in terms of number of abnormals) in those regions of the lungs (discussed above) and 2) the prominence of the vasculature in the lung fields.

Vascular markings in general (both normal and increased markings), proved problematic for the system. In many images, where the radiologist indicated increased vascular markings (always with an associated abnormality ranking of 4 or 5), the resulting probability maps displayed some, but not considerable, indication of abnormality; while in many normal images (especially those originating from

full body scans) the resulting probability maps showed a very similar (sometimes greater) indication of abnormality in the regions containing vasculature. Vascular markings thus led to several false positives as well as several false negatives. An example of this is shown in Figure 6.7—note the high visibility of the vascular markings in both the normal and the abnormal images.

A possible cause of this ambiguity is the fact that the radiologist may have had some bias when analysing the images from the two data sets, as they were examined at different times and for different purposes. The suspected TB images were examined with the intention of outlining TB related lung pathologies and it was known to the radiologist that the images came from suspected TB patients. When the full-body images were examined, however, the intention was to confirm that the images contained no lung pathologies and it was known that the images came from trauma patients with no suspicion of lung disease. It is therefore possible that, at the time of examining the suspected TB images, the radiologist was biased to confirming the presence of lung pathology. The probability maps for normal images (and normal regions in abnormal images) from the suspected TB images is thus a better indication of the system's performance on normal data.

The final example in Figure 6.7 shows an example of where the probability map indicates diffuse abnormality in the right lung field but not in the left lung field for an input image containing a very obvious diffuse nodular pattern in both lung fields. The reason for this failure can be attributed to the fact that only two images containing this pattern in both lungs exist in the data set, while several more exist displaying the pattern in the right lung only. The system is thus very poorly trained on such abnormalities for the left lung and since individual classifiers are trained for every region the results are poor—the performance on the second left lung image displaying this abnormality was equally poor. This again highlights the need for a larger, more diverse database.

Combining the per region scores into overall abnormality scores for every image improved the performance of the system considerably. Similar results were obtained by van Ginneken *et al.* [23], where the improvement was attributed to statistical averaging—most abnormal images contain more than one region of abnormality (due to the overlapping regions) and combining several estimates always improves accuracy.

The important values when considering the reliability of the system as a standalone diagnostic tool (i.e. determining if the lung fields contain pathology or

not) or as a screening tool are the sensitivity and specificity of the classification at the different subtlety levels. The optimal operating point on the ROC curve for all the data occurs at a sensitivity of 0.935 and a specificity of 0.932. This means that when the system is used as a stand alone diagnostic tool, 93.2% of normal images will be correctly classified as normal. Only 6.5% of abnormals will be missed by the system. These values are similarly high for all the subtlety levels.

The performance values obtained are extremely high—classification is near perfect for the subsets containing abnormal images with subtleties of 3 (intermediate) or less. The performance remains very high on the subtle and very subtle subsets too. These values are considerably higher than those achieved in a previous study using the same approach to detect TB [23]. The reason for the high performance level must therefore have something to do with the data. A few scenarios are possible: the abnormalities in this data set could be more obvious and therefore more easily detectable in general than those in the data set used in [23]; the manner in which TB manifests in paediatric chest X-rays may be more suited to detection by textural analysis than adult cases; the X-ray machine used (Lodox STATSCAN) may produce images which present textural pathologies in the lung fields more clearly (superiority of the system for visualising the airways in chest X-rays has been demonstrated [36]). The reality is most likely a combination of these factors. Further studies, using a larger, more diverse database obtained from the same machine, is clearly necessary and would clarify these questions. Provisionally, however, these results are very promising.

6.2.2 Circular regions of interest

The successful cases produced by the circular region approach and shown in Figure 6.9 are very attractive in the sense that the predicted regions of abnormality are depicted very clearly and distinctly in the probability maps. It is encouraging that the system performed so well for both the abnormal and normal regions in the TB images. The system yielded accurate probability maps over the full range of subtleties. The final example in 6.9 shows an example of a probability map for a normal image cropped from a full-body scan—again the result is very satisfactory.

In Chapter 4 it was mentioned that one of the possible advantages of using small circular regions instead of the larger regions is that the extracted feature vectors would provide a more detailed description of the lung fields, possibly improving performance on the more subtle abnormal cases. It is apparent that this more

detailed description of the lungs has in fact contributed to the shortcomings of the approach. As explained, the system failed mostly in terms of false positives as opposed to missed detections. Since the system extracts features from fairly small, localised areas in the lungs, it becomes more vulnerable to background anatomical structures, such as the ribs and clavicles, as individual regions in normal images could easily, for example, consist predominantly of rib shadows (which appear similar to abnormalities). The system also appeared even more sensitive, than classification using the big regions, to vascular markings in the lungs, resulting in a large number of false positives. These failures occurred most frequently in the normal images originating from the full-body scans.

When a full-body scan is taken the patient is situated in the supine position—the sternoclavicular junction as well as parts of the clavicles are therefore often visible in the lung fields, appearing as regions of brighter grey levels. All the patients from the TB study, on the other hand, had their arms raised above their heads during their chest scans—the clavicles are thus not visible in the chest images. Since the clavicles are not consistently visible in all the normal images, the training phase would not incorporate it as a normal background anatomical structure in the lung fields, and they were thus often misclassified as pathology. The data set was too small to exclude the normal images where the clavicles are visible. Furthermore, judging from the fairly substantial areas of abnormality in many of the probability maps for the normal images, it is likely that the ribs have also contributed to the false positives in this way.

An obvious solution to this problem would be to employ some form of rib suppression or segmentation, as discussed in the literature review in Chapter 2. This, however, is an extremely difficult step to implement, and was beyond the scope of this project.

The fact that the system consistently struggles in common regions of the lungs (apices and bottom lateral corners) is an indication that the simple (x, y) position component of the feature vectors is not a sufficient description of the region's position. The bottom lateral corners of the lungs often contain mostly rib shadows and since this is a very common presentation in chest images, it is quite a clear indication that the position components used are somewhat ineffective. A possible solution to this problem would be to subdivide the lung fields in the same way as was done for the big regions and then to use the circular regions (without the position components) from a particular big region as the description for that part of lung field. Individual classifiers for each of the big regions could then be

trained, providing a more robust description of position. This would differ from the big region approach presented previously, in that each big region would not be described by a single feature vector, but rather by all the feature vectors from the circular regions falling in that big region—thereby combining the benefits of a more detailed description of the lung tissue provided by the circular regions with the robust description of lung position offered by the big regions.

The fact that the system performs better on the normal images and normal regions from the TB study than those cropped from full-body scans supports the hypothesis proposed earlier that the normal images from the TB study are more appropriate for the CAD studies performed in this project.

The drastic improvement in performance for $0 \leq \alpha \leq 0.75$ in Figure 6.11(a) is slightly worrying, as this indicates that, after applying the quantile fusion to combine pixel posteriors into overall image abnormality scores, the image posterior probabilities for all the images (normal and abnormal) fall within a fairly small range—allowing very little room for distinguishing between abnormal and normal. This result can be attributed to the fact that a large number of normal images contained pixel posteriors with fairly high values (hence the probability maps discussed earlier). A more stable curve would indicate a more reliable and consistent system.

Again, for the same reasons as before, combining the pixel scores into overall abnormality scores for the images improved results considerably. Despite the unstable curve in Figure 6.11, the classification results are still very impressive. The impact of the instability, however, is evident when looking at the relatively low scores for the subtle and very subtle cases. The area under the ROC curve is 0.846 at a sensitivity of 0.800 and a specificity of 0.786; these values are certainly not high enough for practical applications. The performance on these cases is so much lower because of the small range of stable α values—after fusion there will have been a large degree of overlap (in terms of image posterior probabilities) between the normal images that had high pixel posteriors and the subtle abnormal cases, obviously making classification difficult. This problem is not a result of the fusion scheme, but rather due to limitations in the features extracted from the circular regions (as discussed earlier).

Although the classification performance of the system as a whole ($A_z = 0.941$) is high, there is clearly room for improving the reliability and consistency of these results. Again, this will require a larger and more diverse database, and possi-

bly some adjustments to the manner in which the features are used to train the classifier—especially in terms of their description of lung position.

6.2.3 Impact of lung field segmentation on classification

The impact of the segmentation results on the final classification scores has not been discussed in great detail thus far. Recall that an effective means by which to test the quality of the segmentations is to compare the classification results using the true segmentations with the classification results using the ASM segmentations. Tables 6.4 and 6.5 show these comparisons. The results using the true and ASM segmentations are almost identical, confirming that ASM segmentations are indeed very accurate, and improving them will have little effect on the classification results.

Abnormality Subtlety	ASM	True
Obvious and Relatively Obvious	0.990	0.990
Intermediate	0.998	0.997
Subtle and Very Subtle	0.959	0.967
All categories	0.982	0.983

Table 6.4: Comparison between classification results for true and ASM segmentations for big regions.

Abnormality Subtlety	ASM	True
Obvious and Relatively Obvious	0.986	0.989
Intermediate	0.953	0.954
Subtle and Very Subtle	0.846	0.852
All categories	0.941	0.939

Table 6.5: Comparison between classification results for true and ASM segmentations for circular regions.

6.2.4 Feature selection

Feature selection is known to improve the performance of learning models in a number of ways, such as speeding up the learning process through a reduction

of dimensionality by removing irrelevant or redundant features, improving the interpretability of the model and improving classification accuracies [53]. Among the most common feature selection techniques are sequential feature forward and sequential feature backward selection [53]. These techniques, however, can be computationally expensive as tests need to be performed repeatedly until the optimal feature subset is found.

The performance of the system using the big regions was extremely good, with regard to classification accuracy as well as speed—it is unlikely therefore that feature selection would improve results significantly. The desired improvements in the speed of the classification system using circular regions is unlikely to be attained using feature selection, since the main contributing factor to the slow speed is the number of regions per image and not the number of features. Furthermore, the very fact that the feature extraction and classification processes for the circular regions *are* so slow makes feature selection (especially greedy searches such as forward and backward selection) an extremely computationally expensive task. Due to these factors, feature selection techniques were not investigated.

6.3 Conclusions

The most important results and observations presented in this chapter are summarised below.

The outputs of the CAD system are: 1) a visual tool to aid radiologists in locating possible regions of abnormalities in the lung fields and 2) an overall abnormality score for each image, which assigns to a particular chest X-ray a score denoting its degree of abnormality. The reliability of the per image abnormality score is determined by the optimal operating point on the ROC curve which gives the sensitivity and specificity of the system. The sensitivity indicates the percentage of abnormals that will be correctly classified, while the specificity indicates the percentage of normals that will be correctly classified.

The probability maps produced by both the big and circular region approaches were accurate in most cases—performing well for the full range of abnormality subtleties. Both approaches, however, produced inconsistent results in the perihilar regions of the left lung and struggled to distinguish between normal and abnormal vasculature. The main contributing factors to these shortcomings were the lim-

ited amount of abnormal data found in the left lung and possibly an inconsistent analysis of the vascular markings by the radiologist. Additionally, the circular region approach performed poorly in the bottom lateral corners of the lungs, the lung apices and several of the normal images cropped from full-body scans. These shortcomings were attributed to the fact that the circular regions, because of their size, were affected more by normal background anatomical structures and details in the lung fields—the visibility of the clavicles and sterno-clavicular junctions in several of the normal images thus proved problematic. It was recommended that rib suppression or segmentation algorithms be explored as a solution to this problem. It was also concluded that the position component of the features extracted from the circular regions did not contain sufficient information to capture common properties of particular lung regions effectively. It was recommended that the circular region approach be modified to include a more robust description of the lung region in each of the feature vectors.

Combining the results into overall abnormality scores for whole images improved the results of both approaches considerably. The performance of the big region approach, in particular, was very impressive, with an area under the ROC curve of 0.982 at a sensitivity of 0.935 and a specificity of 0.932 for the entire data set and consistently high results for the individual subtlety levels. The fact that the results were so high for every subtlety level (near flawless for subtleties of 3 and lower) is very promising. The performance of the circular region approach was also impressive, with $A_z = 0.941$ at sensitivity of 0.925 and a specificity of 0.879 for the entire data set. The approach struggled slightly on the more subtle cases, however, and showed much room for improvement. These results (especially for the big regions) are better than those achieved by previous studies on TB images—the improvement is believed to be as a result of a combination of the nature of the images produced by the Lodox Statscan machine, the fact that the study restricted to paediatric X-rays and the high rate of obvious abnormalities in the data set.

The classification results obtained using the ASM and true segmentations were almost identical, indicating that the ASM segmentations were satisfactory and improving them would have little effect on the classification results.

In terms of speed, the big region approach performed considerably better than the circular region approach. Feature extraction and classification took approximately 15 seconds per image for the big region approach and several minutes ($\pm$ 7 minutes) for the circular regions. A more efficient software implementation of the system (for example in C++ as opposed to MATLAB) would likely speed the

process up considerably. Although this difference appears considerable (and is in the context of training and testing for experiments in this study), in reality, taking a few minutes to produce a diagnostic result might not detract from the practical utility of the method; this would have to be tested in a clinical implementation of the system.

Perhaps the most important observations in this chapter pertain to the data set used. The data impacted the final classification results in a number of ways. Several regions in the lung fields, especially in the left lung, contained very few abnormalities—per region classification results were shown to be proportional to the number of images with abnormalities in that region. Many of the normal images used were cropped from full body scans and thus appeared slightly different to the chest images, one of the main differences being the visibility of the clavicles and sternoclavicular junctions, which affected the circular region approach in particular. In general, both classification schemes performed better on the normal images from chest only scans. Several abnormalities occurred only a few times in the data set (e.g. a diffuse nodular pattern), making it difficult to train classifiers on these types of abnormalities; the full-body scans were analysed at different times and under different circumstances to the suspected TB images and this may have affected the results—the fact that classification results were generally better for the normal images from the TB study support this. All of these factors lead to the same conclusion—the need for a larger, more diverse database of images obtained in a consistent manner.

The results presented in this chapter are very encouraging and show much potential both as a visual and diagnostic aid for radiologists and as a screening tool. Although the performance of the system using the big regions is better than any previous CAD systems for TB images, a much larger database is needed before any definitive conclusions can be made.

Chapter 7

Conclusions and Future Work

The following chapter summarises the results obtained in this project, highlights the most important observations made and makes recommendations for future developments of the CAD tool.

7.1 Summary

Chapter 2 presents a review of the literature on both paediatric pulmonary tuberculosis, as well as computer analysis of chest X-rays. Paediatric TB, especially in high burden regions such as South Africa, is shown to present many challenges in its diagnosis. The current gold standard for the diagnosis of active TB is a positive sputum culture, but the method is limited in the paediatric setting. Children often do not produce sufficient sputum for a reliable test, resulting in an estimated less than 15% of TB infected children yielding a positive sputum culture. The diagnostic procedure is complicated further in children in endemic settings, largely due to HIV co-infection. As a result, the chest X-ray is often the main diagnostic tool for children in endemic settings. Considerable research has been done in the field of computer-aided diagnosis in chest radiography, of which the most relevant to this project is the development of CAD tools for the detection of interstitial lung diseases. These tools commonly consist of the following components: 1) lung field segmentation, 2) lung field subdivision, 3) textural feature extraction and 4) classification. The predominant limitation of all the CAD schemes discussed is the lack of large publicly available databases on which to test and compare different systems. A very limited amount of research has been done in the development of CAD tools for TB, and nothing for paediatric TB in particular. The review shows a clear need for developments in this area.

Chapter 3 describes the data set used in this project. The data set consists of 105 antero-posterior chest images of which 60 are abnormal and 45 are normal. The abnormal images were obtained from chest scans performed on suspected TB cases. Of the 45 normal images, 6 were obtained from the suspected TB study. To avoid having to expose healthy children to radiation to obtain normal images, the remaining 39 normal images were cropped from routine full-body scans performed on trauma patients. All of the scans (abnormal and normal) were taken using the Lodox STATSCAN linear slot-scanning digital X-ray machine at the Red Cross Children's Hospital in Cape Town. The images were examined by a single radiologist, who defined both the location and subtlety of the abnormalities, if present. The children were all aged between 0 and 5 years at the time the scans were taken; their lung shapes were variable, due to the varying visibility of the thymus gland in children in this age bracket, which can result in a prominent anterior mediastinal shadow.

In **Chapter 4**, a detailed explanation of the methods used in the development of the CAD tool is presented. The scheme consists of the following four stages: 1) lung field segmentation, 2) lung field subdivision, 3) feature extraction and 4) classification.

The lung field segmentation is performed using separate active shape models for the left and right lung fields. The ASM search is manually initiated in each test image, as this was found to improve performance considerably. After segmentation, all of the images are subsampled to a working resolution of 700 x 700 pixels.

Two approaches to the subdivision of the lung fields are considered. The big region approach subdivides each lung field into 21 overlapping regions of varying size, giving 42 regions per image in total. The circular region approach divides each lung field into a large number of small overlapping circular ROIs with a radius of 32 pixels and placed on an 8 x 8 grid. This results in approximately 2000 regions per lung field (depending on the size of the lung).

The left lung field is flipped and the pixels around the borders of both lung fields are mirrored to avoid distortions in the filter outputs. Textural features are extracted from each region, using the moments of responses to a multi-scale Gaussian filter bank. Additional positional features are added to the circular regions, to give an indication of their location in the lung fields.

For the big region approach, separate classifiers are trained for each of the 42

regions. The regions are classified using a kNN classifier, with a leave-one-out strategy. These classification results are used to produce probability maps giving an indication of the degree of abnormality in every location in the lung fields. The per region results are combined using a weighted multiplier, to give overall abnormality scores for the images.

A linear discriminant analysis (LDA) classifier is used for the circular regions. 5-fold cross validation is used to train and test the classifier. Posterior probabilities for every pixel in the lung fields in a particular image are computed using the classifier outputs. These posterior probabilities are used to create probability maps and are combined using quantile fusion to give overall abnormality scores for each of the images.

The performance of the segmentation algorithm is measured using an overlap measure and the classification performance of the system is tested using receiver operating characteristic (ROC) curve analysis.

Chapter 5 presents the results of the segmentation phase of the framework established in Chapter 4. The segmentation results for both the right and left lung fields are good, yielding overlap measures of 0.93 and 0.92 respectively. These values are comparable to those found in other studies using similar approaches to segmenting the lung fields. The semi-automatic approach, whereby the user defines the starting position of the ASM search in a test image, improves the performance of the algorithm considerably. The segmentation results are poorest along the mediastinal borders of the lungs, which is expected due to the varying size and shape of the thymus gland.

Chapter 6 presents and discusses the results of the feature extraction and classification phases of the framework.

When using the big region approach, the results were excellent, both in terms of the probability maps and the overall image scores. The probability maps were consistently accurate for the full range of abnormality subtleties, struggling only in the perihilar regions of the lungs. The per image results were excellent with an area under the ROC curve of 0.982 at a sensitivity of 0.935 and a specificity of 0.932. These values indicate that the system has a very high reliability as a stand-alone diagnostic tool; 93.2% of normals will be correctly classified and only 6.5% of all abnormal cases will be missed. The values also indicate strong potential for use as a screening tool.

The circular region approach yielded less consistent, but nonetheless impressive results. The approach was affected more by normal anatomical structures such as the rib cage and the clavicles. The positional component of the feature vectors is shown to be somewhat ineffective. Although the performance of the approach on a per image basis is good for the entire data set, with an area under the ROC curve of 0.941 at a sensitivity of 0.925 and a specificity of 0.879, it struggles on the more subtle cases, yielding results which are not high enough for practical implementation.

In terms of speed, the big region approach performed considerably better than the circular region approach. Feature extraction and classification took approximately 15 seconds per image for the big region approach and several minutes ($\pm$ 7 minutes) for the circular regions.

The classification results of the system using the ASM segmentations and the true segmentations are almost identical, confirming that the segmentation results are satisfactory. It also indicates that the classification performance of the system will not be improved significantly by improving the segmentations.

Perhaps the most important observations in this chapter involved the data set used. The data impacted the results in a number of ways, all of which lead to the same conclusion—the need for a larger, more diverse database.

7.2 Methods in context

A summary of the methods used in this project are given below. The author's contribution to the various existing methods is explained.

The lung field segmentation is performed using the multi-resolution active shape model method presented by Cootes *et al.* [43]. Active shape models have been used to segment the lung fields in a number of previous studies [1,2,23]. In these studies the ASM search is initiated at the mean model position in each test image. This approach is not used in this project. Instead the initial position of the ASM model in a test image is defined by the user, who is required to define three anatomically distinct locations in each lung field (the bottom lateral corner, the lung apex and the bottom medial corner) using a mouse pointer on a computer monitor. An

affine warp is then used to position the ASM model at approximately that location without compromising the shape of the model. This approach is shown to improve the segmentation results considerably. As is done in [1], the left and right lung fields are segmented independently of one another.

Two approaches to subdividing the lung fields into regions of interest are compared: 1) the big approach and 2) the circular region approach. The big region approach was developed by van Ginneken *et al.* [23] and the circular region approach by Arzhaeva *et al.* [29, 35].

The big region approach uses the segmentation results to subdivide the lung fields into 21 overlapping regions of interest per lung field (42 regions per image). Textural features are extracted from each of these 42 regions using the moments of responses to a multiscale bank of Gaussian filters. Individual kNN classifiers are then trained for each of the 42 regions using a leave-one-out strategy. The per region classification results are used to create probability maps for every test image, giving an indication of the degree of abnormality of every location in the lung fields. The probability maps are to be used as a visual aid by a radiologist reading the image. The per region results are then combined into an overall abnormality score for every image using a weighted multiplier; this score gives an indication of the abnormality of the image as a whole.

The big region approach uses the methods presented in the TB CAD system published by van Ginneken *et al.* [23]. The approach presented here differs from that presented by van Ginneken *et al.* [23] in the following ways: the segmentation phase differs in the ways described above (manual registration and separate ASM models for the left and right lung fields); only the moments of responses to the Gaussian filter bank are used as textural features, while additional difference features are used in [23]; in this study the per region results are used to produce probability maps for the lung fields, this is not done in [23]; the output of the system presented in this project is to be used as a diagnostic tool to aid radiologists reading the image—in [23] the output is intended for use in screening applications.

The circular region approach uses the segmentation results to subdivide the lung fields into a large number of overlapping circular regions of interest. The circular regions have a radius of 32 pixels and are placed on an 8 x 8 pixel grid. Again, textural features are extracted from each of the regions using the moments of responses to a multiscale bank of Gaussian filters. Additional positional features, giving the location of the ROIs relative the centre of mass and extremities of the

lung fields, are also used. Images are then classified using a linear discriminant analysis classifier, giving a posterior probability for every circular region in the image. The per region classification results are used to compute pixel posterior probabilities, giving an indication of the abnormality of every pixel in the lung fields. The pixel posterior probabilities are used to a create probability map for the image. The pixel posterior probabilities in a particular image are combined using a quantile fusion scheme to give an overall abnormality score for that image.

The circular region approach is based on the interstitial lung disease CAD system developed by Arzhaeva *et al.* [29]. The approach presented here differs in the following ways: in [29] segmentation is performed manually and not using active shape models; the positional features used here are computed in a different way, but ultimately contain the same information; an overall abnormality score is computed for every image using a quantile fusion scheme, while the overall abnormality scores for images are not considered in [29] (the quantile fusion scheme that is used is described by Loog *et al.* [51] where it is used in a different interstitial lung disease CAD system); finally the most important difference is the fact that this project considers the detection of tuberculosis and not interstitial lung diseases.

This is the first study where a direct comparison between the big and circular region approaches has been presented. More importantly, this is the first study to present a CAD system for the detection of TB in paediatric chest X-rays.

7.3 Conclusion

The research objectives, presented in Chapter 1, have been met. A computer-aided diagnostic tool for TB, that analyses the textural appearance of the lungs in paediatric chest X-rays has been successfully implemented. This is the first study that has implemented such a system for paediatric cases alone. The system performed well when tested on a set of images obtained from the Lodox STATSCAN digital X-ray machine at the Red Cross Children's Hospital in Cape Town. Although the results are extremely encouraging, the system needs to be tested on a larger, more diverse database before definitive conclusions can be made.

7.4 Recommendations for future work

The most important areas in which the CAD tool implemented in this study can be bettered, are summarised below:

- The need for a larger, more diverse database has been emphasised. As is the case for the vast majority of similar studies, this is the single most important recommendation for improving both the understanding and the performance of the system presented in this study.
- The data should be analysed by more than one radiologist. This will result in a more objective diagnosis for each of the images. It is important, especially in the case of subtle abnormalities, that ground truth (i.e. the diagnosis of the radiologist) is consistent.
- The effects of normal background anatomy, such as the rib cage, should be dealt with more directly than merely training individual classifiers for each of the lung regions. It is likely that, when the system is tested on more images, the impact of background anatomy will be more obvious. An accurate implementation of some form of rib segmentation or suppression can only be beneficial to any computerised analysis of the lungs.
- It may be worth investigating the implementation of a hybrid scheme that combines the big region and circular region approaches. The fact that the circular region approach yielded a more detailed description of the lung texture is clear and it is expected that, if a more descriptive representation of the lung region can be incorporated into the resulting feature vectors, performance will improve significantly, perhaps beyond the current performance of the big region approach.
- Due to the large variety of ways in which TB can manifest in paediatric chest X-rays, it may be worth investigating the use of a one-class classifier similar to that described in Section 2.6 (instead of the two-class classifier used in this project). Although this approach may improve performance on rare abnormalities, a large, diverse database will still be required to test the system effectively.

Bibliography

[1] van Ginneken B, Frangi AF, Frangi RF, Staal JJ, ter Haar Romeny BM, Viergever MA. Active Shape Model Segmentation with Optimal Features. IEEE Transactions on Medical Imaging. 2002;21:924–933.

[2] van Ginneken B, Stegmann MB, Loog M. Segmentation of Anatomical Structures in Chest Radiographs using Supervised Methods: A Comparative Study on a Public Database. Medical Image Analysis. 2006;10(1):19–40.

[3] Cootes TF, Taylor CJ. Statistical Models of Appearance for Medical Image Analysis and Computer Vision. In: Proceedings of SPIE Medical Imaging; 2001. p. 236–248.

[4] WHO. World Health Organization - Data and Statistics: Global Tuberculosis Database; 2008. http://www.who.int.

[5] Kim WS, Choi J, Cheon J, Kim I, et al. Pulmonary tuberculosis in infants: radiographic and CT findings. American Journal of Radiology. 2006;187:1024–1033.

[6] Marais BJ, Madhukar P. New approaches and emerging technologies in the diagnosis of childhood tuberculosis. Paediatric Respiratory Reviews. 2007;8:124–133.

[7] Shiraishi J, Li Q, Suzuki K, Engelmann R, Doi K. Computer-aided diagnostic scheme for the detection of lung nodules on chest radiographs: Localised search method based on anatomical classification. Medical Physics. 2006;33:2624–2653.

[8] van Ginneken B, ter Haar Romeny BM, Viergever MA. Computer-aided diagnosis in chest radiography: A survey. IEEE Transactions on Medical Imaging. 2001;20.

[9] Katsuragawa S, Doi K. Computer-aided diagnosis in chest radiography. Computerized Medical Imaging and Graphics. 2007;31:212–223.

[10] Doi K. Computer-aided diagnosis in medical imaging: Historical review, current status and future potential. Computerized Medical Imaging and Graphics. 2007;31:198–211.

[11] Fonseca-Santos J. Tuberculosis in children. European Journal of Radiology. 2005;55:202–208.

[12] Dyck PV, Vanhoenacker FM, Van den Brande P, De Schepper AM. Imaging of pulmonary tuberculosis. European Radiology. 2003;13:1771–1785.

[13] Salazar GE, Schmitz TL, Cama R, et al. Pulmonary tuberculosis in children in a developing country. Pediatrics. 2001;108:448–453.

[14] Andronikou S, Wieselthaler L. Modern imaging of tuberculosis in children: thoracic, central nervous system and abdominal tuberculosis. Paediatric Radiology. 2004;34:861–875.

[15] Leung AN, Mller NL, Pineda PR, FitzGerald JM. Primary tuberculosis in childhood: radiographic manifestations. Radiology. 1992;182:87–91.

[16] Sutton D. Textbook of radiology and medical imaging, 5th ed. Churchill Livingstone. 1993;.

[17] Sarkar S, Chaudhuri S. Evaluation and progression analysis of pulmonary tuberculosis from digital chest radiographs. Computerized Medical Imaging and Graphics. 1998;22:145–155.

[18] Gie R. Diagnostic Atlas of Intrathoracic Tuberculosis in Children. A Guide for Low Income Countries. International Union Against Tuberculosis and Lung Disease; 2003.

[19] Cootes TF, Taylor CJ, Cooper DH, Graham J. Active shape models—their training and application. Comput Vis Image Underst. 1995;61(1):38–59.

[20] Cootes TF, Taylor CJ. Statistical models of appearance for computer vision. Image Science and Biomedical Engineering, University of Manchester; 2004.

[21] Arzhaeva Y, Tax DMJ, van Ginneken B. Improving computer-aided diagnosis of interstitial disease in chest radiographs by combining one-class and two-class classifiers. SPIE Medical Imaging. 2006;p. 6144.

[22] Kido S, Tamura S, Nakamura H, Kuroda C. Interstitial lung diseases: evaluation of the performance of a computerized analysis system versus observers. Computerized Medical Imaging and Graphics. 1999;23:103–110.

[23] van Ginneken B, ter Haar Romeny BM, Katsuragawa S, et al. Automatic detection of abnormalities in chest radiographs using local texture analysis. IEEE Transactions on Medical Imaging. 2002;21:139–149.

[24] Katsuragawa S, Ishida T, Kobayashi T, MacMahon H, Doi K. Computerized analysis of interstitial disease in chest radiographs: improvement of geometric-pattern feature analysis. Medical Physics. 1997;24:915–924.

[25] Vogeslang F, Weiler F, Dahmen J, Kilbinger MW, Wein B, Gunther RW. Detection and compensation of rib structures in chest radiographs for diagnose assistance. In: SPIE. vol. 3338; 1998. p. 774–785.

[26] Yue Z, Goshtasby A, Ackerman LV. Automatic detection of rib borders in chest radiographs. IEEE Transactions on Medical Imaging. 1995;14(3):525–536.

[27] van Ginneken B, ter Haar Romeny BM. Automatic delineation of ribs in frontal chest radiographs. In: Medical Imaging 2000: Image Processing, Proceedings of the SPIE. vol. 3979; 2000. p. 825–836.

[28] Kido S, Tamura S. Computerized classification of interstitial lung abnormalities on chest radiographs with normalized radiographic index and normalized fractal dimension. European Journal of Radiology. 2007;37:184–189.

[29] Arzhaeva Y, Prokop M, Tax DMJ, De Jong PA, Schaefer-Prokop CM, van Ginneken B. Computer-aided detection of interstitial abnormalities in chest radiographs using a reference standard based on computed tomography. Medical Physics. 2007;34:4798–4809.

[30] Asada N, Doi K, MacMahon H, Montner S, Giger ML, Abe C, et al. Potential usefulness of artificial neural network for differential diagnosis of interstitial lung diseases: a pilot study. Radiology. 1990;177:857–860.

[31] Mclachlan GJ. Discriminant Analysis and Statistical Pattern Recognition (Wiley Series in Probability and Statistics). Wiley-Interscience; 2004.

[32] Shiraishi J, Abe H, Engelmann R, Aoyama M, et al. Computer-aided diagnosis for distinction between benign and malignant solitary pulmonary nodules in chest radiographs: ROC analysis of radiologists performance. Radiology. 2003;227:469–474.

[33] Gurney JW, Swensen SJ. Solitary pulmonary nodules: Determining the likelihood of malignancy with neural network analysis. Radiology. 1995;196:823–829.

[34] Koeslag A. Computer-aided diagnosis of millary tuberculosis in chest x-rays. University of Cape Town; 2002.

[35] Arzhaeva Y, Tax DMJ, van Ginneken B. Dissimilarity-based classification in the absence of local ground truth: application to the diagnostic interpretation of chest radiographs. Pattern Recognition. 2009;42:1768–1776.

[36] Pitcher RD, van As AB, Sanders V, et al. A pilot study evaluating the STATSCAN digital x-ray machine in paediatric polytrauma. Emergency Radiology. 2008;15:35–42.

[37] Maree GJ, Irving BJ, Hering ER. Paediatric dose measurement in a full-body digital radiography unit. Paediatric Radiology. 2007;37:990–997.

[38] Arthur R. Interpretation of the paediatric chest X-ray. Paediatric Respiratory Reviews. 2000;1:41–50.

[39] Hamarneh G, Abu-Gharbieh R, Gustavsson T, Abu-gharbieh R, Gustavsson T. Review - Active Shape Models - Part I: Modeling Shape and Gray Level Variations. In: Proceedings of Swedish Symposium on Image Analysis; 1998. p. 125–128.

[40] Abu-Gharbieh R, Hamarneh G, Gustavsson T, Hamarneh G, Gustavsson T. Review - Active Shape Models - Part II: Image Search and Classification. In: Proceedings of Swedish Symposium on Image Analysis; 1998. p. 129–132.

[41] Klein A; 2008. www.mathworks.com/matlabcentral/fileexchange/19696.

[42] Stegmann MB. Active appearance models - theory, extensions and cases. Technical University of Denmark; 2000.

[43] Cootes T, Taylor CJ, Lanitis A. Active Shape Models: Evaluation of a Multi-Resolution Method for Improving Image Search. In: Proceedings of the British Machine Vision Conference. BMVA Press; 1994. p. 327–336.

[44] Lang S. Linear algebra. Springer; 1987.

[45] Fisher R, Perkins S, Walker A, Wolfart E. Affine Transformations; 2003. http://homepages.inf.ed.ac.uk/rbf/HIPR2/affine.htm.

[46] Ruprecht D, Muller H. Image warping with scattered data interpolation. IEEE Computer Graphics and Applications. 1995;15(2):37–43.

[47] van Ginneken B, ter Haar Romeny BM. Multi-scale texture classification from generalized locally orderless images. Pattern Recognition. 2002;36(2):899–911.

[48] Lindeberg T. Scale-space: A framework for handling image structures at multiple scales; 1996.

[49] van Ginneken B. Computer-aided diagnosis in chest radiography. University Medical Center Utrecht, Netherlands; 2001.

[50] ter Haar Romeny BM, Florack LMJ, Salden AH, Viergever MA. Higher order differential structure of images. Image and Vision Computing. 1994;12:317–325.

[51] Loog M, van Ginneken B, Nielsen M. Detection of interstitial lung disease in PA chest radiographs. In: Jaffe MJ, Flynn MJ, editors. SPIE Medical Imaging: Physics of Medical Imaging. vol. 5368; 2004. p. 848–855.

[52] Fawcett T. An introduction to ROC analysis. Pattern Recognition Letters. 2006;27:861–874.

[53] Guyon I, Elisseeff A. An introduction to variable and feature selection. Journal of machine learning research. 2003;3(1):1157–1182.