CHAPTER 1

INTRODUCTION

Lung is the most important part of the respiratory system that takes oxygen from the air into the bloodstream and releases carbon dioxide from the body. Lung disease is any disease or disorder that occurs in the lungs or that disallows the normal functioning of the lungs. Lung disorders are one of the major threats to the health of individuals all over the world especially in South-East Asia. According to a WHO report (WHO 2010), an estimated 1.3 million people died from Tuberculosis (TB) in 2008; the highest number of deaths was in South-East Asia and the number of new cases arising each year is still increasing globally in the WHO regions of Africa, the East Mediterranean and South-East Asia. In India, 331000 deaths (WHO 2009) have been reported per year and it has been estimated that the death rate would be 21 per 100000 of the population per year in 2015 due to TB. It is estimated that 440000 people had Multidrug-resistant TB (MDR-TB) worldwide in 2008 and that a third of them died. The impact of the epidemic was the highest in Asia. Almost 50% of MDR-TB cases worldwide are estimated to occur in China and India. According to a report by UNICEF, pneumonia and diarrhoeal diseases together account for 37 percent of all deaths in children under five.

The mortality rate due to lung disorders can be reduced if the disorder is diagnosed at an early stage. This has become possible with the tremendous growth of imaging technology and information processing
techniques which has led to the development of CAD Systems to assist the radiologist in performing better diagnosis. A CAD system is one which is capable of analyzing the data related to the patient under investigation and suggesting a diagnosis based on the analysis. In this research work the data is in the form of CT images of the chest and hence image analysis is done to perform diagnosis.

1.1 TYPES OF LUNG DISEASES

There are three main types of lung disease, namely, airway diseases, lung tissue diseases and pulmonary circulation diseases. Airway diseases are those diseases that affect the tubes (airways) that carry oxygen and other gases into and out of the lungs. These diseases cause a narrowing or blockage of the airways. They include asthma, emphysema, and chronic bronchitis. Lung tissue diseases are those diseases that affect the structure of the lung tissue. Scarring or inflammation of the tissue makes the lungs unable to expand fully ("restrictive lung disease"). It also makes the lungs less capable of taking up oxygen (oxygenation) and releasing carbon dioxide. Pulmonary fibrosis and sarcoidosis are examples of lung tissue diseases. Pulmonary circulation diseases are those diseases that affect the blood vessels in the lungs. They are caused by clotting, scarring, or inflammation of the blood vessels. They affect the ability of the lungs to take up oxygen and to release carbon dioxide. These diseases may also affect heart function. Most lung diseases actually involve a combination of these categories.

Lung cancer is a disease of uncontrolled cell growth in tissues of the lung. It is the most common cause of cancer-related death in men and women. It is responsible for 1.3 million deaths worldwide annually (WHO 2006). The symptoms of lung cancer that are clearly and directly related to lung function are persistent cough, bloody sputum, shortness of breath, chest pain, hoarseness, wheezing or vibrating breathing noises, recurrent
pneumonia or bronchitis and difficulty swallowing (dysphagia). Lung cancer may be seen on chest radiograph and Computed Tomography. The diagnosis is confirmed with a biopsy. This is usually performed by Bronchoscopy or CT-guided biopsy.

Bronchiectasis is a disease state defined by localized, irreversible dilation of part of the bronchial tree. The radiological features of bronchiectasis are signet ring sign, tram-tracks, string of beads, circles filled with air or air and fluid, tubular and branching opacities, bronchi visible within 1 cm of the pleura.

1.2 IMAGING TECHNIQUES USED FOR DIAGNOSIS OF LUNG DISORDERS

The advancements in the field of imaging hardware and information processing have led to the use of several imaging techniques for diagnosis of lung disorders. A few imaging techniques worth mentioning are Radiography, CT, Magnetic Resonance Imaging (MRI), Ultrasound and Positron Emission Tomography (PET). X-ray procedures and Nuclear Medicine procedures use ionizing radiations whereas MRI and Ultrasound do not. Procedures that use radiation are necessary for accurate diagnosis of disease and injury. They provide important information about the patients’ health to the doctor and help ensure that the patients receive appropriate care. Physicians and technologists performing these procedures are trained to use the minimum amount of radiation necessary for the procedure. Benefits from the medical procedure greatly outweigh any potential small risk of harm from the amount of radiation used.

1.2.1 Radiography

The most common imaging procedure used for diagnosis of lung disorders is the chest x-ray, because of its low costs and weak radiation
exposure. It is the first diagnostic imaging test suggested by physicians to diagnose the disease in the presence of symptoms like shortness of breath, persistent cough, chest pain or fever. In x-ray procedures, x rays pass through the body to form pictures on film or on a computer or television monitor, which are viewed by a radiologist. Physicians use this chest x-ray to evaluate the lungs, heart and chest wall, especially in diagnosing or monitoring the treatment for conditions like pneumonia, heart problems, emphysema and lung cancer.

1.2.2 Computed Tomography

CT, which allows the radiologist to look at the lung in cross-section rather than from the outside, has helped them to understand physiologic abnormalities in the lungs and make more accurate diagnosis. Thoracic CT alternatively called Chest CT, CT scan of lungs or CT scan of chest is a computed tomography scan of the chest and upper abdomen. Thoracic CT may be recommended when there is a need for examination of the structures inside the chest. It is noninvasive and poses less risk than invasive procedures (such as angiography or exploratory surgery). High Resolution Computed Tomography (HRCT) of the chest often provides more information for diagnostics of Interstitial Lung Diseases (ILDs).

Thoracic CT may show many disorders of the heart, lungs, or chest area, including lung cancer, pneumonia, tuberculosis, emphysema, bronchiectasis, inflammation or other diseases of the pleura - the membrane covering the lungs, diffuse ILD, enlarged lymph nodes (lymphadenopathy), abnormalities of the structure or position of the heart, lungs, or blood vessels, the stage of some lung tumors or esophageal cancer, aortic aneurysm (thoracic) and accumulations of blood or fluid.
A CT scan is one of the best ways of looking at soft tissues such as the heart and lungs. The benefits of a CT scan usually far outweigh the risks. CT scans are used routinely to detect lung cancer and locate the most appropriate site for biopsy, if necessary, establish how advanced the cancer is and whether it has spread to other parts of the body, help physicians and patients decide on courses of treatment that are tailored to patients’ individual conditions and needs, determine whether medical treatments are working as intended and to detect whether the disease has recurred after treatments are completed.

Patients are generally advised to undergo thoracic CT scan to further examine abnormalities found on conventional chest x-rays, when there is a chest injury, when a tumor or mass (clump of cells) is suspected, to determine the size, shape, and position of internal organs or to look for bleeding or fluid collections in the lungs or other areas. Additional conditions under which the test may be performed include Alcoholic cardiomyopathy, Asbestosis, Atrial myxoma, Cardiac tamponade, Coarctation of the aorta, Dilated cardiomyopathy, Echinococcus, Heart failure, Histoplasmosis, Hypertensive heart disease, Idiopathic cardiomyopathy, Infective endocarditis, Ischemic cardiomyopathy, Left-sided heart failure, Mesothelioma (malignant), Metastatic cancer to the lung, Mitral regurgitation; acute, Mitral regurgitation; chronic, Mitral valve prolapse, Pericarditis; bacterial, Pericarditis; constrictive, Pericarditis; post-MI, Peripartum cardiomyopathy, Pulmonary edema, Restrictive cardiomyopathy, Senile cardiac amyloid, SVC obstruction.

1.2.3 Magnetic Resonance Imaging

MRI is also a noninvasive medical imaging technique used in radiology to visualize detailed internal structure. MRI does not use any ionizing radiation. Rather, it uses a powerful magnetic field to align the nuclear magnetization of (usually) hydrogen atoms in water in the body. MRI
provides much greater contrast between the different soft tissues of the body than CT does. Hence it is especially useful in neurological (brain), musculoskeletal, cardiovascular, and oncological (cancer) imaging.

1.2.4 Choice of Diagnostic Imaging Techniques

X-rays usually have no side effects in the diagnostic range, but excessive exposure to radiation might lead to cancer (Lapp 2004). However, the benefit of an accurate diagnosis outweighs the risk. Stark (2009) states that chest radiography could provide a confident diagnosis in only 23 percent of cases, and those confident diagnoses proved correct only in 77 percent of cases. Husen et al (1971) in their study have concluded that the presence of negative chest film does not rule out the presence of tuberculosis. Depeursinge (2006) have quoted that chest x-rays are negative in a large portion of diseases and often unspecific.

Unlike conventional x-rays, CT scanning provides very detailed images of many types of tissue as well as the lungs, bones, and blood vessels. Because CT scans are able to detect even very small nodules in the lung, chest CT is especially effective for diagnosing lung cancer at its earliest, most curable stage. MRI reveals more details compared to CT in most cases. Cheung et al (1992) in their article have stated that MRI is superior to CT in the depiction of tenosynovitis and peritendinitis, tendonitis, tendon rupture, and tendon dislocation and subluxation. These abnormalities can be diagnosed by CT also, but it is difficult to differentiate the accompanying scar tissue, early changes of tendon degeneration and small amounts of inflammatory fluid. They have also stated that CT is superior for demonstrating calcifications, convex retromalleolar groove, bone fragments or spurs that complicate tendon dislocation and rupture. Tendons and ligaments around the shoulder and knee are best seen by the physics used in MRI. This is due to the density of the tissues that compose the tendons and ligaments.
The choice between CT and MRI depends on the region of the body to be analyzed and the purpose of the scan. Bleeding in the brain, especially from injury, is better seen on CT than MRI. But a tumor in the brain is better seen on MRI. Broken bones and vertebral bodies of the spine are better seen on CT but injury to the spinal cord itself is displayed on MRI far better than CT. CT is far superior at visualizing the lungs and organs in the chest cavity between the lungs. CT is a good tool for examining tissue composed of elements of a higher atomic number than the surrounding tissue. Hence, it is useful for imaging bone and calcification within the body or of structures. MRI is best suited for non-calcified tissue.

1.3 COMPUTER AIDED DIAGNOSIS SYSTEM

Interaction between engineering, computing, physics, and clinical science has become much closer than it has ever been before due to the rapid advancement of computing technology and imaging hardware design. The revolution in the field of Information Technology along with digital imaging in the medical domain facilitates the generation and storage of large collection of images which are an important source of anatomical and functional information for the accurate diagnosis of diseases. Advances in these fields led to the development of CAD systems.

CAD System is a software system for interpreting digital images or laboratory tests to provide a diagnosis. It aids the physicians in performing diagnosis by giving a “second opinion” based on known similar cases by applying rule-based inference techniques. A CAD system can infer similar cases based on the similarity between the features of the “unknown” case and the features of previous cases using techniques from the Artificial Intelligence (AI) community and Content Based Image Retrieval (CBIR). This will be of great help for medical practitioners.
1.4 NEED FOR COMPUTER AIDED DIAGNOSIS SYSTEM

Although skilled radiologists have a high degree of accuracy in diagnosis of lung disorders using advanced CT imaging technology, there remain challenges that cannot be overcome even by high levels of clinical skill and expertise. This is especially true in such lung diseases like ILDs that are characterized by specific abnormal findings and frequently confusing symptoms as discussed by Depeursinge et al (2008). Often, the first symptom of ILD is difficulty in breathing during exercise. Dry cough is the other common symptom. Both of these symptoms are associated with various types of ILD and can be mild or severe. In addition, different forms of ILD appear the same on the x-rays taken to identify what type of ILD a patient has. It can be difficult for doctors to identify the cause of ILD if a thorough patient history is not available. For example, the symptoms of Farmer’s lung include shortness of breath, a dry irritating cough, a sudden general feeling of sickness, fever and chills, a rapid heart rate, and rapid breathing which is often confused with pneumonia.

The clinical importance of CT scans of the chest, and their complicated nature induces the development of CAD system to assist radiologists in reading the chest CTs. A CAD system may support the medical practitioners in diagnosing diseases by giving them an inference based on quantitative analysis of the image, so that if the diagnosis prompted by the system is the same as the diagnosis made by the physician, the physician can confirm the diagnosis without second thought whereas if the diagnoses made by the CAD system and the physician differ, the physician can perform a more careful diagnosis before beginning the therapy. CAD will thus help increase the diagnostic accuracy of radiologists by providing objective data obtained from the diagnostic studies beside subjective data given by the patients.
1.5 ARTIFICIAL INTELLIGENCE TECHNIQUES USED FOR COMPUTER AIDED DIAGNOSIS

The use of AI in diagnostic expert systems aims at making the systems to mimic the decision-making process of human experts. However, such systems can be used only to provide a “second opinion” to the physicians in improving diagnostic accuracy and cannot be considered as a replacement for a physician. AI techniques that are most frequently used in CAD systems in the literature are Neural Networks, Genetic Algorithms, Fuzzy Logic and hybrid approaches.

1.5.1 Neural Networks

Neural Network is based loosely upon the cellular structure of the human brain. The system uses an algorithm to “learn” from examples. After sufficient learning, it has the ability to identify or classify “similar” cases, although it may have never seen the combination of characteristics before. It has “stored” the knowledge of the examples it has seen. Neural nets are an inductive reasoning method.

Neural networks have been used in for lung nodule detection (Penedo et al 1998, Coppinini et al 2003). Penedo et al (1998) have used a two-level artificial neural network, the first level is used for detecting suspicious regions and the second level is used for positive detection. This network combination has achieved 89% to 96% sensitivity depending on the size of nodules. Baemani et al (2008) in their work have used Artificial Neural Network (ANN) for detection of respiratory abnormalities from lung function test results. They have achieved an accuracy of 92%. Suzuki et al (2005a) in their work have used Massive Training Artificial Neural Network (MTANN) for distinguishing between benign and malignant nodules in chest
radiographs. Their scheme has identified 100% malignant nodules as malignant and 45% of benign nodules as benign.

1.5.2 Genetic Algorithms

Genetic Algorithms are based on the biological concept of genetic combination to produce offspring. Boroczky and Zhao (2006) in their work have used genetic algorithms in feature subset selection for false positive reduction in computer aided diagnosis of lung nodule. They have proved that a support vector machine classifier trained with the optimal feature subset resulted in 100% sensitivity and 56.4% specificity. Lee et al (2001) in their work have developed a technique for computer aided diagnosis of pulmonary nodules in helical CT images using a technique based on template matching and genetic algorithm. They have achieved a detection rate of 72%. Jaffar et al (2009) in their work have used genetic algorithm for choosing the threshold used in lung segmentation. To extract the ROIs they have computed the threshold from the histogram of the extracted lung part and thresholded the segmented lung. They have tested their work with 50 patient’s data set and have found 97% ROIs contain nodules.

1.5.3 Fuzzy Logic

Fuzzy Logic is based on the idea that propositions are not necessarily true or false but rather have a degree to which they are true, represented by a number between 0 and 1, inclusive. It is especially useful in medical expert systems because patients suffer from a particular disease to different degrees. From the literature it is found that Fuzzy Logic has been used to develop medical expert systems since 1980. Ye et al (2009) in their work have used fuzzy logic for detection of pulmonary nodules. They have validated the proposed method on a clinical dataset of 108 thoracic CT scans. The experimental results indicate an average detection rate of 90.2%.
Tschirren et al (2005) in their work have used fuzzy logic for airway segmentation. They have tested their approach in 22 low-dose scans. Only in 2 of these the algorithm has returned a somewhat inferior result. Dehmeshki (2008) in their work have used fuzzy connectivity map for segmentation of pulmonary nodules. They have evaluated their result with the help of a radiologist. They state that 84% of the result was considered accurately segmented. The other 16% were part of vascularised nodules or nodules very close to the lung wall or diaphragm. The method has been tested with 22 volumetric CT scans. The RMS error between the automatically segmented and manually segmented fissures is reported to be $1.96 \pm 0.71$ mm.

### 1.6 CONTENT BASED IMAGE RETRIEVAL

CBIR is a technique used in CAD systems based on images. The goal behind CBIR is to retrieve images similar to the query image from an image database based on visual contents. Early work on image retrieval (late 1970s) was based on textual annotation of images. This involves manual annotation of the image with text followed by a search using a text-based approach based on traditional database management systems. Later (1990s), it was found that properties that are inherent in the images can be used to represent and index images more efficiently. On this basis, Image Retrieval techniques of recent days extract features that are representative of the visual content of the image and retrieve images based on distance between the feature vectors formed from the extracted features.

### 1.7 ROLE OF IMAGE PROCESSING IN CAD SYSTEMS

Image processing plays a vital role in diagnosis of diseases using CAD systems based on images. This research focuses on CAD systems for diagnosis of lung disorders. As with any system employing image analysis and image understanding, segmentation is one of the primary processes in a
CAD system that works with images. Segmentation of lung tissue is a challenging task especially in the presence of peripherally placed PBRs. The regions of interest can be extracted only if the lungs are correctly segmented; if the segmentation is incorrect it may increase the false positives and false negatives thereby affecting the diagnostic accuracy of the system. Hence, it is important to use a robust and reliable technique for segmentation.

In addition, such CAD Systems rely on image processing algorithms to extract measures of image properties such as color, texture and shape and represent them as feature vectors. The diagnosis made by CAD systems is based on the features extracted from the image. Thus, medical imaging approaches provide physicians with suggestions essential for effective and efficient diagnosis of various diseases.

The radiological features of diseases are represented using image properties namely, color, shape and texture. As color is of less importance in medical images than in stock photography, the texture and shape features gain importance.

1.7.1 Color

Color is the most extensively used visual content for image retrieval. 3-dimensional values make its discrimination potentially superior to the 1-dimensional grayscale values of images. Before selecting the appropriate color description color space must be determined. Commonly used color spaces are RGB, CMY, CIEL*ab*, CIEL*u*v*, HSV and opponent color space. The commonly used color descriptors are color moments, color histogram, color coherence vector and color correlogram. Color moments are used when the image contains just the object. Color histogram is an effective representation if the color pattern is unique compared with the rest of the dataset. Color coherence vector provides better
retrieval performance than color histogram because of additional spatial information. Color correlogram characterizes not only the color distribution of pixels but also the spatial correlation of pairs of colors. Hence, this provides the best retrieval results compared to color histogram and color coherence vector, however, it is the most computationally expensive one. But color is of less significance in chest CT. Hence color feature is not used in this work.

1.7.2 Shape

The measures of shape commonly used in literature are curvature, area, eccentricity and elongatedness. Curvature is the rate of change of slope. It can also be defined as the ratio between the length and the number of boundary pixels where the boundary direction changes significantly. The smaller the number of boundary changes the straighter is the boundary. Area is the number of pixels the region consists of. Eccentricity is the ratio of major axes length to minor axes length. Elongatedness is the ratio of area of the region and the square of its thickness, where thickness is the number of erosion steps that can be applied before the region totally disappears. In this research work the shape features are represented using the measures discussed herein and the statistical moments.

1.7.3 Texture

Texture is a function of the spatial variation in pixel intensities and is an important approach to regional description. There are three major approaches to describe the texture of a region, namely, statistical, structural and spectral. Statistical approaches yield characterizations of textures as smooth, coarse, grainy and so on. Structural approaches deal with the arrangement of image primitives such as description of texture based on regularly spaced parallel lines. Spectral approaches are based on properties of the Fourier spectrum and are primarily used to detect global periodicity in an
image by identifying high energy narrow peaks in the spectrum. Of these the most frequently used approach is the statistical approach which can further be divided into measures based on histogram and those which are based on the relative position of pixels. Statistical moments, uniformity and entropy fall into the former category while the features extracted from the GLCM fall in the latter category. In this research work, the measures of texture feature of the ROIs are extracted using statistical approaches.

1.8 MOTIVATION

Three issues motivated the research proposed in this thesis. First, even experienced radiologists have trouble in distinguishing the normal pattern of blood vessels and nodules that indicate lung cancer. Quantitative analysis of ROIs helps discrimination of cancerous nodules from normal patterns. Second, currently available CAD systems use segmentation techniques that eliminate the peripherally placed PBRs along with the outer chest region. Hence such CAD systems detect the PBRs in the interior regions of the lung but tend to miss the peripherally placed PBR. Efficient segmentation of lung parenchyma in the presence of peripheral PBR is a challenging task. It is made possible with the help of approaches based on the shape characteristics of the lung. Third, inspite of the usage of CT technology, small nodules are still common to be missed by radiologists in their early stages. Using image processing techniques it is possible to identify even small nodules. Hence this research work aims at improving the performance of the CAD system for diagnosis of lung disorders thereby increasing the diagnostic accuracy by handling these issues.

1.9 RESEARCH OBJECTIVES

The research aims at developing a model for computer aided diagnosis of lung disorders from chest CT with improved performance. In this
work, approaches have been developed to improve the inference mechanism used for computer aided diagnosis of lung disorders. The improvement is achieved in two ways. First, the number of ROIs missed is decreased by enhancing the approaches available for extracting the lung parenchyma from the chest CT. Second, the quantitative analysis of the ROIs is strengthened by considering features from different domains. This improves the performance of the CAD System for diagnosis of lung disorders.

1.10 SOCIAL CONTRIBUTION

The CT scan of the chest of patients with a lung disorder, viz. lung cancer, pneumonia or tuberculosis, exhibit PBR in the periphery of the lung parenchyma which may not be detected by most of the existing CAD systems. In some cases, the PBR could be very small that it is missed by the radiologist. The approaches discussed in this work would help in detecting such PBRs thereby aiding the physicians in reducing the false negative rate; hence, improving the diagnostic accuracy and reducing the mortality rate of patients with lung disorders.

1.11 THESIS ORGANIZATION

The rest of the thesis is organized as follows: Chapter 2 provides a comprehensive review of the research already carried out on various aspects of Computer Aided Diagnosis of lung disorders and brings out the contributions of this research work. Chapter 3 describes the model proposed for computer aided diagnosis of lung disorders. The five different works contributed in accordance with the proposed model to improve the diagnostic accuracy of CAD systems that use chest CT for diagnosis of lung disorders are discussed in Chapters 4 through 8. Chapter 4 discusses the work proposed for detection of cancerous tissues from chest CT images. Chapter 5 presents the contribution of this research work in diagnosis of bronchiectasis from
chest CT. Chapter 6 discusses a contribution of this research work for segmentation of lung parenchyma based on convex hull. Chapter 7 presents the second contribution of this research work for segmentation of lung parenchyma based on graphical methods. Chapter 8 discusses the approach proposed for diagnosis of lung cancer based on analysis of the slice chosen as suitable for analysis of a nodule. Chapter 9 presents the conclusion derived from the work and the enhancements that can be carried out in future.