CHAPTER 1

INTRODUCTION

This chapter mainly focuses on lung cancer, its causes, symptoms, stages and clinical tests for diagnosis. It deals with the necessity of Computed Tomography (CT) images and Computer-Aided Diagnostic (CAD) systems for early detection of lung cancer and explains the areas such as machine learning and image processing. It also introduces the general methodology applied in the present research work. The processes involved in the proposed CAD systems, the phases and techniques involved in the proposed approaches for the lung cancer detection are explained in this chapter.

1.1 LUNG CANCER

Lung cancer is the foremost cause of cancer-related deaths worldwide (Siegel et al. 2012; Jermal et al. 2009). It affects 100,000 Americans of the smoking population every year of all age groups, particularly those above 50 years of the smoking population (American cancer society 2009). In India, 51,000 lung cancer deaths were reported in 2012, which include 41,000 men and 10,000 women (Behera 2012). It is the leading cause of cancer deaths in men; however, in women, it ranked ninth among all cancerous deaths (Statistics on lung cancer 2012).

Lung cancer is difficult to diagnose because of the soft nature of lung tissues. The process called lung biopsy (taking a cancerous tissue out for microscopic analysis) is really painful (Senthilkumaret al. 2016). Hence,
physicians will not suggest for biopsy, unless there is a strong evidence for lung cancer.

Lung nodules are the indicator of cancer and are the tiny mass inside the lung. Most of the lung nodules are non-cancerous (benign), but about 40% of them are cancerous (Kuruvilla and Gunavathi 2014). Therefore, it is a real challenge for the researchers to quantify and describe the lung nodules. Hence it is essential to develop an efficient algorithm to make an accurate decision on the cancerous nature of the nodules.

CT scans are usually preferred by the physicians to diagnose the cancer from patients who come up with the lung cancer symptoms. To assess the growth of the cancer, the patients should undergo several CT scans at regular intervals. The new CT scanner produces nearly 300 scans for a complete lung cross section for every patient (Chheang and Brown 2013). The radiologists’ job become tough as they need to analyse all these cross-sectional images of lungs, hence there is high possibility of error.

The most recent statistics provided by World Health Organization (WHO) indicates that around 7.6 million deaths worldwide each year is because of lung cancer. It is expected to continue rising to 17 million worldwide deaths in 2030 (Taher and Sammouda 2011). The only way to control the lung cancer death rate is diagnosis of cancer in its early stage (Lo et al. 1995). So the necessity of techniques for lung cancer detection at an early stage is increasing.

1.1.1 Causes of Lung Cancer

The research shows that smoking is the leading cause of lung cancer deaths (80%) in the world. The people who don’t smoke (non-smokers), also get lung cancer, however the chances of risk are 10 times
lesser than the smokers (Jemal et al. 2010). People exposed to radon gas, asbestos, air pollution, second hand smoke, diesel exhaust or certain other chemicals can also be affected by lung cancer. Other factors like heredity likely plays a role as well (Goswami et al. 2015).

1.1.2 Symptoms of Lung Cancer

The most common symptoms of lung cancer are persistent cough, shortness of breath, persistent fatigue, cough up blood/blood in sputum, loss of appetite, weight loss, persistent chest pain and pain in bone/shoulder/neck/arm. Less common symptoms are: Hoarseness of voice, wheezing, difficulty in swallowing, frequent/unexplained fever, swelling in the face/neck/feet and frequent headache/dizziness/seizures (Quadrelli et al. 2015).

1.1.3 Clinical Tests for Lung Cancer Diagnosis

In order to diagnose the lung cancer, physician may recommend the following tests:

- **Imaging tests:** An X-ray image of lungs discloses an abnormal nodule or mass. Even small lesions in the lungs can be detected by the Computed Tomography (CT) images which could not be detected in the X-ray images. It is proven that CT offers better contrast than the X-ray images between nodule and background with no overlapping structures and detects earlier stage nodules with higher sensitivity.

- **Sputum cytology:** Looking at the sputum produced by cough, under microscope, sometimes may give the presence of lung cancer cells.
● **Tissue sample (biopsy):** A section of unusual cells may be removed in a procedure called biopsy. Biopsy can be performed in number of ways:

  ✓ **Bronchoscopy:** Examining the abnormal areas of lungs using a lighted tube that is passed into the lungs through the throat.

  ✓ **Mediastinoscopy:** Incision is made at the base of neck and surgical tools are inserted behind the breastbone to take the tissue section from the lymph nodes.

  ✓ **Needle biopsy:** X-ray or CT images to guide a needle through the chest wall and into the lung to collect doubtful cells.

### 1.1.4 Stages of Lung Cancer

Once lung cancer has been diagnosed, doctor will work to determine staging of the cancer. It refers to the severity (spread) of the cancer in the human body (Yang et al. 2010). Staging helps the doctor to determine the appropriate treatment. The stages of lung cancer are divided into:

- **Stage I:** Cancer is restricted to the lungs and has not extended to the lymph nodes. Generally, tumor is smaller (>3 mm and ≤20 mm size).

- **Stage II:** Tumor has grown larger than 20 mm (21 mm to 30 mm size) and also may have extended to the nearby lymph nodes.
- **Stage III:** Tumor has grown very large and extends to other organs near the lungs. This stage may have larger tumor (31 mm to 70 mm) and extended further away from the lungs.

- **Stage IV:** Cancer has extended outside the affected lung to the other lung or other distinct areas of the body (> 71 mm).

The stages I and II are called initial stage or early stage of cancer. The stages III and IV are called advanced stage or later stage of cancer. Depending on the staging, symptoms of lung cancer differ. Patients’ survival rate with cancer depends on its staging and treatment. Patients with early stage have better chances of survival than its later stage.

### 1.2 COMPUTED TOMOGRAPHY IMAGES

In the present clinical world, lung nodule detection in chest CT images has become very essential for lung cancer detection and analysis. CT is a dominant tool which allows very quick creation of X-ray images of the body with high-resolution cross-sectional imaging. The quick detailed result has made CT very valuable, especially in the Emergency Department (ED). High-quality, cross-sectional images are available in a short time that helps to define the medical status of patients. CT makes use of combinations of many X-ray images taken from different angles around the human body to produce cross-sectional (tomographic) images (virtual ‘slices’) and views of tissues and organs.

#### 1.2.1 Sequential CT Scanner

In sequential CT scanner, a cross-sectional image is obtained by scanning crosswise section of the human body at various angular positions, while the tube and detector rotate 360° around the patient with stable table. The image is reconstructed from the resulting projection data. If the patient moves during the image acquisition, then the data attained from the various
angular positions are no longer constant. The image is degraded by motion artifacts and may be of limited diagnostic value. The tomographic approach is appropriate only to a limited degree of diagnosis of anatomical regions with automatism functions (i.e. the unintentional functioning of an organic process, particularly muscular, without apparent neural stimulation) such as the heart or the lung.

1.2.2 Spiral/Helical CT Scanner

Spiral CT/Helical CT scanner is also called “volume scanner”. It is clearly different from that of a sequential CT. Spiral CT uses a different scanning principle. The patient on the table is moved constantly through the scan field in the z direction while the gantry performs multiple 360° rotations in the same direction. The X-ray draws a spiral rotation around the body and constructs a data volume. This volume is created from a multitude of Three Dimensional (3-D) picture components, i.e. voxels. The movement of the table in the ‘z’ direction usually produces inconsistent sets of data, which causes every image reconstructed directly from a volume data set to be degraded by artifacts. Software applications facilitate the use of spiral CT even for regions which are subject to involuntary movements.

1.2.3 CT in General Clinical Use

Most of the modern hospitals currently use spiral CT scanners have greatly reduced the scan time compared to sequential CT. Spiral CT is very much useful when examining patients who are unable to cooperate. The complete volume in spiral CT is scanned faster and without gaps so that motion artifacts due to various respiratory conditions during the acquisition are minimized. Multiple scans due to breathing, during the acquisition are not needed. So, patient’s dose is also minimized. Figure 1.1 shows a CT scan image showing the cross sectional view of lungs.
1.2.4 Advantages of Spiral CT scan

The advantages of spiral CT scan in clinical use are:

i) Full coverage of organs e.g., brain to leg in a single respiratory position

ii) Reduced scan times

iii) Better diagnostic information due to improved resolution and 3-D visualization in routine operation.

1.2.5 Benefits of CT Scan

The benefits of spiral CT scan are as follows:

- CT imaging provides both good soft tissue resolution (contrast) as well as high spatial resolution.
- CT scan can easily detect the causes of abdominal pain with extremely high accuracy, enabling quicker treatment.
- CT imaging of the head and brain can detect tumors, show blood clots, blood vessel defects and enlarged ventricle images (caused by a building up of cerebrospinal fluid) and
other abnormalities such as those of the nerves and muscles of the eye.

- CT scanning provides thorough views of many types of tissues, including the lungs, bones, soft tissues and blood vessels.
- CT scanning is painless, non-invasive and accurate.
- CT examinations are very simple and rapid.
- Detection and diagnostics done by CT eliminates the need for invasive exploratory surgery and surgical biopsy.
- CT scan identifies both normal and abnormal structures, making it a useful tool to guide radiotherapy, needle biopsies etc.,
- With the advent of spiral CT, the continuous acquisition of complete CT volumes can be used for the diagnosis of blood vessels with CT Angiography.

### 1.3 COMPUTER-AIDED DIAGNOSIS

Computer-Aided Diagnosis (CAD) is a currently applied computerized analysis of medical images and is widely used as a tool in detection and diagnosis of abnormalities in medical imaging. In clinical practice, CAD has become a major research subject in clinical radiology for the detection of pulmonary nodules in chest CT images (Wook-Jin and Tae-Sun 2013).

In radiology, CAD system supports the radiologist, who uses the output from a computerized analysis of medical images as a second opinion to assess the detection of abnormalities, quantification of disease progress, and differential diagnosis of lesions (El-Baz et al. 2013; Muller et al. 2004;
Quellec et al. 2010). Thus, CAD system is a very significant tool for early detection of lung cancer.

1.3.1 Generic Architecture of CAD Systems for Medical Images

Typical architecture of CAD used for finding the abnormalities from medical images consists of four main modules which are shown in Figure 1.2. They are:

i) Segmentation of Region(s) of Interest (ROI)

ii) Extraction and selection of features

iii) Formation of diagnosis rules and

iv) Classification of the selected ROI

1.3.1.1 Region(s) of interest segmentation

The first step after image acquisition is the image segmentation, which sub divides the complex image into small sub-images. The segmentation should be stopped when the Region(s) of Interest (ROI) of a particular application is isolated. The ROI in CT image is usually a cluster of pixels that appears to be a dense mass. Common methods used for ROI segmentation include thresholding, template matching, region growing, active contouring, fuzzy clustering, level set algorithm, morphological approaches etc. Both normal and abnormal anatomical structures, appearing in patients’ images may be defined using

a. Manual or semi-automatic methodologies, where the user interacts with the system and

b. Fully automated methodologies
Figure 1.2 Schematic diagram of CAD system for medical diagnosis

An example of semi-automatic method is the seeded region growing which is used for the ROI segmentation in medical images (Banik
et al. 2009). Active contour approaches have been widely used as fully automatic methods in tracking of anatomical contours in 2-D medical images due to their ability to accurately detect the random shapes of organ boundaries (Joshi et al. 2010). In this thesis, Auto-seed Region Growing and Morphological Masking (ARGMM) and Fuzzy Auto-seed Cluster Means Morphological (FACMM) segmentation approaches are used for early detection of lung cancer.

1.3.1.2 Extraction and selection of features

Feature extraction refers to the extraction of several quantitative measures of medical images for decision making on the pathology of a structure or tissue. Feature extraction can be carried out in either spatial domain or spectral domain.

Selection of a subset of the most robust features is essential from the extracted features, which improves the classification accuracy and reduces the overall complexity.

1.3.1.3 Formulation of diagnosis rules

Set of diagnostic rules are formulated (according to the ROI) using the extracted features in order to focus only the suspected abnormal tissues in medical images. This step minimizes unnecessary analysis of healthy tissues thereby reducing the time taken by the classifier during classification (normal or abnormal) process.

1.3.1.4 Classification of the selected ROI

Pattern classification plays vital role in the classification of a set of features into the proper class in medical image analysis. Classification can be either supervised or unsupervised. In supervised classification, the feature set
is recognized as a member of a predefined class, while in unsupervised classification, it is allocated to an unknown class.

Supervised classification can be done using the statistical classifiers like decision trees, K-nearest neighbor and Bayesian classifiers. C-Means, Kohonen Map are examples of the widely used unsupervised classification methods. Several supervised and unsupervised classification techniques have been comparatively assessed for many tumor diagnoses.

1.4 MACHINE LEARNING

Machine learning is a type of Artificial Intelligence (AI), which explores the study and construction of algorithms that provides computers with the ability to learn without being explicitly programmed. It focuses on the advancement of computer programs that can train themselves to grow and change when exposed to new data.

1.4.1 Types of machine learning algorithms

Various types of machine learning algorithms are given below:

- Supervised learning: It generates a function that maps inputs to desired outputs. One standard formulation of the supervised learning task is the classification problem: the learner is required to learn (to approximate the behavior of) a function which maps a vector into one of several classes by looking at several input-output examples of the function.

- Unsupervised learning: It models a set of inputs; labelled examples are not available.
Semi-supervised learning: It combines both labelled and unlabelled examples to generate an appropriate function or classifier.

Reinforcement learning: This algorithm learns a policy of how to act to a given observation of the world. Every action has some impact in the environment, and the environment provides feedback that guides the learning algorithm.

Transduction: It is similar to supervised learning, but does not explicitly construct a function: instead, tries to predict new outputs based on training inputs, training outputs, and new inputs.

1.4.2 Machine Learning Algorithms

Most widely used algorithms in machine learning approaches include decision tree learning, artificial neural network, fuzzy system, neuro-fuzzy system, support vector machine, clustering, Bayesian network and genetic algorithms.

1.4.2.1 Decision tree learning

Decision tree learning uses a decision tree as an analytical model that maps annotations about an item to conclusions about the item's target value. It is one of the predictive modeling approaches used in statistics, data mining and machine learning. Tree models where the target variable can take a finite set of values are called classification trees. In these tree structures, leaves represent class labels and branches represent conjunctions of features that lead to those class labels. Decision trees where the target variable can take continuous values (typically real numbers) are called regression trees.
In decision analysis, a decision tree can be used to visually and explicitly represent decisions and decision making. In data mining, a decision tree describes data but not decisions; rather the resulting classification tree can be an input for decision making.

1.4.2.2 Artificial neural network

Artificial Neural Network (ANN) or Neural Network (NN) is a family of statistical learning model which is inspired by biological neurons (the central nervous systems of humans, in particularly the brain) and used to estimate or approximate functions that can depend on a large number of inputs and are generally unknown. ANNs are generally presented as systems of interconnected "neurons" which exchange messages between each other. The connections have numeric weights that can be tuned based on experience, making neural nets adaptive to inputs and capable of learning.

For example, a neural network for handwriting recognition is defined by a set of input neurons which may be activated by the pixels of an input image. After being weighted and transformed by a function (determined by the network's designer), the activations of these neurons are then passed on to other neurons. This process is repeated until finally, an output neuron is activated. This determines which character was read.

It has been used to solve a wide variety of tasks that are hard to solve using ordinary rule-based programming, including computer vision and speech recognition.

1.4.2.3 Fuzzy system

Fuzzy system proposes a mathematic calculus to translate the subjective human knowledge of the real processes. It has been found to be
suitable applications for many areas ranging from control theory to AI. Fuzzy system uses fuzzy logic, which is a mathematical methodology (and a philosophical ideology) that is similar in construct to Boolean Algebra and similar in appearance to probability, but more general than both in fundamental ideas. This generality is the freedom for truth variables to hold any value between 0 and 1 (inclusive), and fuzzy logic proponents assert this generality permits greater flexibility, freedom, compactness and accuracy when representing real world situations. It is proved to be a potential tool for decision making systems such as pattern recognition and expert systems. It has had a great deal of success where it has been applied in the real world, and is often touted as a means of making machines smarter.

1.4.2.4 Neuro-fuzzy system

Neural networks are brain like systems and excellent at recognising patterns but not at clearing up how they arrive at the decisions. On the other hand, fuzzy systems are good at clearing up what they arrived at the decisions but they cannot attain the rules themselves to make the decisions. Neural networks and fuzzy system could be combined to form neuro-fuzzy system, which overcomes their individual drawbacks. Neuro-fuzzy hybridization results in a hybrid intelligent system that synergizes these two techniques by combining the human-like reasoning style of fuzzy systems with the learning and connectionist structure of neural networks. Neuro-fuzzy hybridization is widely termed as Hybrid Neuro-Fuzzy System (HNFS). Thus in neuro-fuzzy system, neural networks introduce its computational characteristics of learning in the fuzzy systems, which could be helpful in designing accurate decision support system.
1.4.2.5 **Support vector machine**

Support Vector Machine (SVM) is a supervised learning model with associated learning algorithms that analyze data and recognize patterns, used for regression and classification analysis. For a set of training examples, every one marked for belonging to one of two categories, an SVM training algorithm makes a model that assigns new examples into one category or the other, making it a non-probabilistic binary linear classifier. An SVM model is an illustration of the examples as points in space, assigned so that the examples of the different categories are separated by an apparent gap that is as large as possible. New examples are then assigned into that same space and estimated to belong to a category depending on which side of the gap they fall on.

SVMs can effectively perform a non-linear classification in addition to performing linear classification, using kernel trick, implicitly assigning their inputs into high-dimensional feature spaces.

1.4.2.6 **Clustering**

Cluster analysis or clustering is the task of grouping a set of objects in such a way that objects in the same group are more similar to each other than to those in other groups (clusters). It is a main task of exploratory data mining, and a common technique for statistical data analysis, used in many fields, including machine learning, pattern recognition, image analysis, information retrieval and bioinformatics.

Cluster analysis itself is not one specific algorithm, but the general task to be solved. It can be achieved by various algorithms that differ significantly in their notion of what constitutes a cluster and how to efficiently find them. Popular notions of clusters include groups with small distances
among the cluster members, dense areas of the data space, intervals or particular statistical distributions. Clustering can therefore be formulated as a multi-objective optimization problem. The appropriate clustering algorithm and parameter settings (including values such as the distance function to use, a density threshold or the number of expected clusters) depend on the individual data set and intended use of the results. Cluster analysis as such is not an automatic task, but an iterative process of knowledge discovery or interactive multi-objective optimization that involves trial and failure. It will often be necessary to modify data preprocessing and model parameters until the result achieves the desired properties. Various clustering algorithms used for image processing are K-means, C-means, modified K-means, modified C-means, fuzzy C-means, probabilistic C-means, spatial fuzzy C-means, etc.

1.4.2.7 Bayesian network

Bayesian network is a probabilistic graphical model (a type of statistical model) that symbolizes to a set of random variables and their conditional dependencies through a Directed Acyclic Graph (DAG). For instance, a Bayesian network could symbolize the probabilistic relations between symptoms and diseases. Given symptoms, the network could be capable of computing the possibilities of the occurrence of various diseases.

Formally, Bayesian networks are DAG’s whose nodes assign random variables in the Bayesian sense: they may be latent variables, observable quantities, hypotheses or unknown parameters. Edges correspond to conditional dependencies; nodes that are disconnected correspond to variables that are conditionally independent of each other. Each node is linked with a probability function that takes, as input, a specific set of values for the node's parent variables, and gives output probability of the variable represented by the node.
Effective algorithms exist that carry out implication and learning in Bayesian networks. Dynamic Bayesian networks are the Bayesian networks that model sequences of variables (e.g. speech signals or protein sequences). Influence diagrams are the Bayesian networks that can represent and solve decision problems under uncertainty.

1.4.2.8 Genetic algorithm

Genetic Algorithm (GA) is an exploration heuristic model that mimics the process of natural selection in the field of artificial intelligence. This heuristic is characteristically used to produce useful solutions to search and optimization problems. GAs belong to the bigger class of evolutionary algorithms, which produce solutions to search and optimization problems using techniques motivated by natural evolutions, such as, mutation, inheritance, selection, and crossover.

1.4.3 Applications of Machine Learning

One measure of progress in machine learning is its significant real-world applications, such as those listed below:

- Brain-machine interface
- Object recognition
- Medical diagnosis
- Speech and handwriting recognition
- Information retrieval
- Computer vision etc.

As machine learning approaches play a vital role in many of the fields, they can be used in detection of many types of cancer.
In this thesis, artificial neural network, neuro-fuzzy system and support vector machine approaches are used for the detection of lung cancer.

1.5 PROPOSED METHODOLOGY

Lung cancer is the main cause of cancerous death in both men and women. The early detection of lung cancer can be helpful in curing the disease completely. So the necessity of techniques for lung cancer detection at an early stage is increasing.

Various techniques are available for the lung cancer detection, but many approaches provide less accuracy and more false positives. So a new approach to detect cancer in early stage is necessary. This research primarily deals with the following techniques for cancer detection. They are

- Lung cancer and its stage-wise detection based on the observed symptom values of the patients using hybrid neuro-fuzzy system.
- Segmentation of suspected nodules from the two dimensional (2-D) CT lung images by employing ARGMM.
- Segmentation of suspected nodules from the 2-D CT lung images using FACMM.
- Classification of suspected nodules as either malignant or benign by employing ANN and SVM.
- 3-D visualization of malignant nodules using Materialise’s Interactive Medical Image Control System (MIMICS) software.
The main aim of these proposed approaches is to accurately detect the lung cancer by analyzing the observed symptoms and CT images of the patients.

1.6 OBJECTIVES OF THE RESEARCH

The main objectives of this research are:

1. To develop a CAD system for lung cancer detection from the observed symptoms
2. To develop CAD systems for the detection of lung cancer from the chest CT images
3. To develop a 3-D CAD visualization system to find the location of the cancerous nodules. Also to measure cancerous nodules width, height, breadth and volume.

1.7 METHODOLOGY

Six phases are included in the proposed CAD systems for early detection of lung cancer is shown in Figure 1.3 are as follows:

**Phase 1:** Prediction of lung cancer and its stage based on the symptom values

**Phase 2:** Segmentation of suspected lung nodules from chest CT images

**Phase 3:** Feature extraction for the suspected lung nodules

**Phase 4:** Formulation of diagnosis rules from the extracted features

**Phase 5:** Classification of suspected nodules and

**Phase 6:** 3-D visualization of the cancerous nodules
Techniques used

PHASE 1
Lung cancer & its stage-wise detection based on symptom values
(a) Neural network
(b) Fuzzy logic

PHASE 2
Segmentation of suspected lung nodules from the chest CT images
(a) Auto-seed region growing with morphological masking
(b) Fuzzy auto-seed cluster means morphology

PHASE 3
Feature extraction for the suspected nodules
(a) Shape/size features
(b) Centroid shift
(c) Texture features

PHASE 4
Formation of diagnosis rules
(a) Rule 1: Threshold values $T_1$ and $T_21$ - $T_22$
(b) Rule 2: Threshold value $T_3$
(c) Rule 3: Threshold values $T_4$, $T_5$ and $T_6$

PHASE 5
Classification of suspected nodules
(a) Artificial neural network
(b) Support vector machine

PHASE 6
Three dimensional visualization of the malignant nodules
(a) Thresholding
(b) Region growing
(c) Edit mask
(d) Dynamic region growing
(e) 3-D view calculation

Figure 1.3 Research methodology
In the first proposed scheme, hybrid neuro-fuzzy system is used to predict lung cancer and its stages from the lung cancer patients based on the observed symptom values. HNFS approach is used because of its better accuracy. In the second proposed scheme, the suspected nodules from the lung CT images are segmented using Auto-seed Region Growing and Morphological Masking (ARGMM). ARGMM is selected because of stable segmentation results. In the third proposed scheme, Fuzzy Auto-seed Cluster Means Morphology (FACMM) is used for the segmentation of suspected nodules from the lung CT images. FACMM is preferred for further analysis, because it takes comparatively less segmentation time than ARGMM. The fourth proposed scheme uses FACMM for segmentation and Artificial Neural Network (ANN) for classification. ANN is chosen because of its automatic detection of cancerous nodules. The fifth proposed scheme uses FACMM for segmentation and Support Vector Machine (SVM) for classification. SVM is utilized mainly for its simplicity, short training time and accuracy.

The final proposed scheme provides 3-D visualization to locate the actual malignant nodules from a stack of 2-D CT slices using software-Materialise’s Interactive Medical Image Control System (MIMICS). MIMICS is selected for its simplicity.

1.7.1 Performance Parameters

In medical diagnosis, to evaluate the system performance, analysis can be carried out on every classifier for sensitivity (recall), specificity, accuracy, precision, recall and F-measure. The sensitivity and specificity are the measures of proportion of actual positives and actual negatives which are correctly identified. The accuracy is the measure of true results in a diagnostic test. Precision measures how many of the positively classified were relevant.
F-measure (F-score) is a measure of a diagnostic test's accuracy. It considers both the precision and the recall of diagnostic test to compute the score.

The proposed CAD systems performances for the lung cancer detection are analysed using the following parameters:

- Sensitivity
- Specificity
- Accuracy
- Precision
- Recall and
- F-measure

The most important measures of the proposed approaches are the detection accuracy of True Positives (TP) and True Negatives (TN). TP is condition of correct malignant nodule diagnosis when nodule is actually at malignant risk. TN is the situation where the nodule is diagnosed as benign and has sound benign in reality.

1.8 ORGANIZATION OF THE TESIS

The primary objective of this research work is to develop an efficient and competent lung cancer detection technique from the observed symptom values and CT images of patients, which can provide a high level of accuracy and performance. The thesis is organized as follows:

Several researchers have addressed the problem of cancer detection in lungs. The various available techniques related to the present research work are given in Chapter 2, as the literature survey. Chapter 3 explains the first
proposed methodology namely hybrid neuro-fuzzy system for prediction of lung cancer and its stages based on the observed symptoms. Chapter 4 deals with the second proposed approach namely segmentation of suspected lung nodules from CT scan images by auto-seed region growing with morphological masking. Chapter 5 elaborates on the third proposed scheme namely segmentation of suspected lung nodules from the CT scan images by fuzzy auto-seed cluster means morphology. Chapter 6 discusses the classification of suspected nodules by employing artificial neural network and support vector machine. Chapter 7 deals the three dimensional visualization of malignant nodules. The concluding remarks with future scope of work are described in Chapter 8.

The next chapter provides a brief review of literature on various lung cancer detection methods to identify the research problem and design a suitable experimental framework for further performance analysis.