

# A Review on CT Scan Screening for Early Lung Cancer Detection in India

Neeraj Syal  
Assistant Engineer  
Indus International University  
Bathu

Niti Syal  
Editor  
Smarttech Technologies  
Mohali

Urvashi Chaudhary  
B.tech Scholar  
Punjab Technical University  
Chandigarh

## ABSTRACT

Among various diseases cancer has become major threat in India. As per Indian population due to cancer the mortality rate was very high. Cancer is the second most disease responsible for death in India. In view of these facts, this paper will reviews that how CT scan screening technology is helpful for cancer prevention in early stages. In the end this paper will also discusses which techniques has to be adopted for better accuracy of cancer prevention system.

## General Terms

Lung Cancer, SVM, Clustering Algorithms

## Keywords

Cancer, prevention, CT Scan Screening, Machine learning algorithms.

## 1. INTRODUCTION

The high prevalence of lung cancer leads to its early prevention. The introduction of CT technology helps lot in increasing the mortality rate of the lung cancer patients due to its detection at early stages. Determine whether a pulmonary nodule is a being tumor or not in the early stages is very important. But determination of the presence of tumors in small nodules is very difficult. With the rapid advancement of the technology, the interaction between physics, engineering and computing science has become closer than ever before. In recent years many efforts has been made towards the development of CAD technology. The very first image taken by CAD technology is of brain tumor in 1971. With its increasing hit rates, the popularity of the CAD technology is increasing continuously. So CAD also solves this problem by reducing the double reading of images by radiologists (Maryellen et. al, 2002). Image processing and CAD research is needed to develop radiology display applications that implement computer aids, such as cueing, overlays, and annotation, into a broadly supportive workstation. Computer Assists, including decision support, data mining capabilities and access to reference libraries, should be incorporated. Human machine systems for image-based diagnosis need to take advantage of both human and machine capabilities, creating a system, which as a whole will be greater than the sum of its parts (Katherine et.al, 2003).

The rest of the paper is organized as: Section II presents the lung cancer scenario in India, Section III shows the CAD system methodology for prevention of lung cancer and finally, Section IV shows the conclusion and future scope of the paper.

## 2. LUNG CANCER SCENARIO IN INDIA

In India the vast majority cases (90%) of lung cancer is due to exposure to tobacco smoke (Wikipedia). About 10 % of cancer occurs in that people who never smoked. These cases are often caused due to the genetic effects. Lung cancer is the most common cause of death in India and was responsible for 1.56 million deaths annually, as per survey in 2012 and in 1991 around 60, 9000 people was effected by lung cancer.

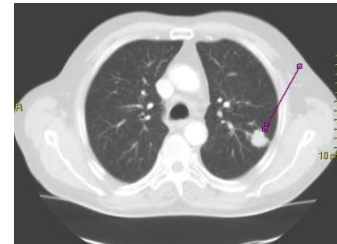


Figure. 1 CT scan showing a cancerous tumor in the left lung [3]

Lung cancer staging is an assessment of the degree of spread of the cancer from its original source. It is one of the factors affecting the prognosis and potential treatment of lung cancer (Hornet.al, 2012). Below chart shows the reasons of death in India. From graph it is clearly seen that Lung cancer is at second most place.

The initial evaluation of non-small-cell lung carcinoma (NSCLC) staging uses the TNM classification. This is based on the size of the primary tumor, lymph node involvement, and distant metastasis (Hornet.al, 2012).

Table.1 TNN Lung Cancer Classification [4]

T: Primary tumor	
TX	Primary cancer cannot be evaluated
T0	No indication of primary tumor
T1	Tumor size less than or equal to 3 cm across, enclosed by lung or visceral pleura, deprived of invasion proximal to the lobar bronchus
T2	Tumor size more than 3 cm but less than or equal to 7 cm across
T3	Separate tumor nodule in the same lobe

T4	Separate tumor nodule in a different lobe of the same lung
----	--

Using the TNM descriptors, a group is assigned, ranging from occult cancer, through stages 0, IA (one-A), IB, IIA, IIB, IIIA, IIIB and IV (four). This stage group assists with the choice of treatment and estimation of prognosis (Rami et.al, 2009).

### 3. METHODOLOGY

Diagnostic task in radiology generally comprises of the two steps: the detection task and the classification task. First step is to find the abnormality of lung. Then after finding the abnormality of the lung, the radiologist must give an appropriate action for cause. After viewing the literature survey in papers (Metinet.al, 2002; Wiemkerl et.al, 2003; Binsheg et.al, 2003; Jin Mo et.al, 2003) it has been found out that the common steps in lung cancer classification as following.

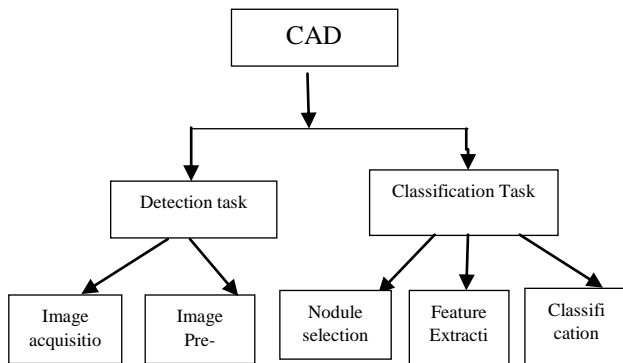


Figure.2 Components of CAD System for lung cancer detection

#### 3.1 Image acquisition

For CAD systems one has to specify the number of CT images and slice of each image. These image can be act as training and testing images as well. Image acquisition of both images normal as well as abnormal are required so that on the basis of feature extraction, abnormal images can be detected. Access to real medical images for experimentation is a very difficult undertaking due to privacy issues and due to the fact that very few hospitals maintain medical images (Maria et.al, 2001).

#### 3.2 Image preprocessing

Image processing generally known as the enhancement of the CT image. Also removal of unwanted part from the image. Morphological operations, edge detection, noise removal methods mainly used for pre-processing of images. (Metin et.al, 2002) used K-Means clustering method for segmentation of main part.

#### 3.3 Nodule Selection

Once lungs has been segmented from whole image then nodule selection is done. This can be achieved by region growing methods (Binsheget.al, 2003), connected component method (Jin Mo et.al, 2003) etc.

#### 3.4 Feature extraction

There are two types of nodules, benign nodules and malignant nodules, this type of nodule is basically responsible for lung cancer. After detecting the nodule candidates from the image, further we need to differentiate between nodules (Actual

cancerous regions) and non-nodules (vessels, benign nodules, airways etc.). For example in case of Breast Cancer Detection, relevant features can be the type of the tissue (dense, fatty and fatty-glandular etc.) and / or the position of the breast: left or right. (Brown et.al, 2013). Various features that can be extracted from lung cancer can be follows:

$$\text{Mean} = \frac{1}{N} \sum_{(i,j) \in R} g(i, j)$$

$$\text{Variance} = \frac{1}{N} \sum_{(i,j)} [g(i, j) - \text{mean}]^2$$

$$\frac{1}{N} \sum_{(i,j)} [g(i, j) - \text{mean}]^3$$

$$\text{Skewness} = \frac{\frac{1}{N} \sum_{(i,j)} [g(i, j) - \text{mean}]^3}{\left( \frac{1}{N} \sum_{(i,j)} [g(i, j) - \text{mean}]^2 \right)^{3/2}}$$

### 4. CLASSIFICATION

Machine learning algorithms helps a lot in decision making. Most popular techniques are neural network and support vector machine. Support Vector Machine (SVM) (<http://www.support-vector-machines.org/>) also called Support Vector Networks are supervised learning models that analyze data and recognize patterns. SVM models represent examples as point in space mapped in manner that separate categories examples are divided by a gap thereby performing linear classification. Apart from this SVMs can also perform nonlinear classification using Kernel trick. The main idea of SVM is that; it finds the optimal separating hyper plane such that error for unseen patterns is minimized. Consider the problem of separating the set of training vectors belonging to two separate classes.

$$x_1, x_2, \dots, x_n$$

Which are vectors in  $\mathbb{R}^D$ .

We consider a decision function of the following form:

$$y = w^T \phi(x) + b$$

Attached to each observation  $x_i$  is a class label,  $t_i \in \{-1, +1\}$ . Without loss of generality, we must construct a decision function such that,  $y x_i > 0$  for all  $i$  such that  $t_i = +1$ , and  $y x_i < 0$  for all  $i$  such that  $t_i = -1$ . We can combine these requirements by stating,

$$t_i y x_i > 0 \quad \forall i$$

The idea is to extend it to multi-class problem is to decompose an M-class problem into a series of two-class problems.

Also, neural networks (V.V Thakre et.al, 2010) are composed of simple elements which operate parallel. A neural network can be trained to perform a particular function by adjusting the values of the weights between elements. Network function is determined by the connections between elements. There are activation functions used to produce relevant output.

Training can be either supervised or unsupervised. In supervised training system learns by trying to predict outcomes for known examples. System compares its predications with the known results and learns from its mistakes. In unsupervised training system no output or result is shown as part of training process. With the delta rule, as with other types of back propagation, 'learning' is a supervised process that occurs with each cycle or 'epoch' (i.e. each time the network is presented with a new input

pattern) through a forward activation flow of outputs, and the backwards error propagation of weight adjustments. Simply, when a neural network is initially presented with a pattern it makes a random 'guess' as to what it might be. It then sees how far its answer was from the actual one and makes an appropriate adjustment to its connection weights. Within each hidden layer node is a sigmoidal activation function which polarizes network activity and helps it to be stable in nature.

The various advantages of using neural network are given below:

- Unlike rule based systems or programmed systems neural networks are flexible in changing environment. Rule based systems are limited to the situation for which they have been designed
- Since the system is developed through learning and not by programming, neural nets teach themselves the pattern. It can be implemented in any application and without any problem

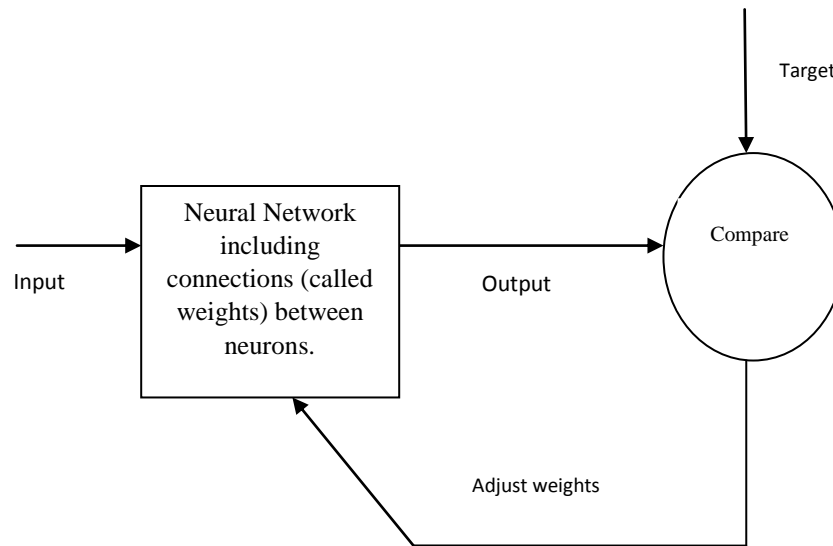


Figure.3 Neural Net Block Diagram

## 5. CONCLUSION AND FUTURE SCOPE

The early detection of the lung cancer is the challenging task. The manual analysis of CAD images is very time consuming, inaccurate etc. The two methods named Neural Network and Clustering Method both are suitable to classify the images of lung cancer.

So, in this paper we describe the general model of CAD lung cancer system model along with lung cancer scenario in India. To overcome the problem of the ever increasing burden of image data to be examined, CAD systems are required. But for better results, better performance is needed and it is necessary to effectively study the machine learning algorithms. Early CAD approaches are not suitable for thin image slices, but it is now possible using CAD system.

Future scope lies in the use of the web mining as well as machine learning algorithms, so that lung cancer detection can be reduced as, both methods are designed to classify the image of N pixels among M classes or regions.

## 6. REFERENCES

[1] Maryellen L. Cigcr, "Computer Aided Diagnosis in Radiology", Received and accepted November 200 1, Acad Radiol 2002; 9: 1-3.

[2] Katherine P. Andriole, "Addressing the Coming Radiology Crisis: The Society for Computer Applications in Radiology, Transforming the Radiological Interpretation Process (TRIP.) initiative." Position Paper from the SCAR TRIPTM Subcommittee of the SCAR Research and Development Committee, November 2003

[3] [http://en.wikipedia.org/wiki/Lung\\_cancer](http://en.wikipedia.org/wiki/Lung_cancer)

[4] Chheang, S; Brown K (June 2013). "Lung cancer staging: clinical and radiologic perspectives". *Seminars in Interventional Radiology* **30** (2): 99–113.

[5] Rami-Porta, R; Crowley JJ; Goldstraw P (February 2009). "The revised TNM staging system for lung cancer". *Annals of Thoracic and Cardiovascular Surgery* **15** (1): 4–9.

[6] Horn, L; Pao W; Johnson DH (2012). "Chapter 89". In Longo, DL; Kasper, DL; Jameson, JL; Fauci, AS; Hauser, SL; Loscalzo, J. *Harrison's Principles of Internal Medicine* (18th ed.). McGraw-Hill.

[7] Metin N. Gurcan, Berkman Sahiner, et.al, "Lung nodule detection on thoracic computed tomography images: Preliminary evaluation of a computer-aided

- diagnosis system”, *Medical Physics*, Vol. 29, No. 11,(2552 - 2558), November 2002.
- [8] R. Wiemkerl, P. Rogalla2, T. Blaffertl et al, “Computer Aided Detection and Volumetry of pulmonary nodules on high resolution CT data”, *Mediamundi*, November 2003
- [9] Binsheg Zhao, Goradon Gasmu, Michelle S. Ginsberg et al, “Automatic detection of small lung nodules on CT utilizing a local density maximum algorithm. *Journal of Applied Clinical Medical Physics*, Volume 4 Number 3, Summer 2003.
- [10] Jin Mo Goo, Jeong Won Lee, Hyun Ju Lee, et al, “Automated Lung Nodule Detection at Low-Dose CT Preliminary Experience”, *Korean Journal of Radiology*, Volume 4, December 2003.
- [11] Maria-Luiza Antonie, Osmar R. Zdyane, Alexandru Coman, “Application of Data Mining Techniques for Medical Image Classification”. *Proceedings of the second international workshop on Multimedia Data Mining*, San Francisco, USA, August 24,2001.
- [12] <http://www.support-vector-machines.org/>
- [13] V.V. Thakare, P. Singhal, “Neural network based CAD model for the design of rectangular patch antennas,” *JETR*,vol. 2(7), 2010.