International Journal of Electrical and Electronics Research ISSN 2348-6988 (online)
Vol. 3, Issue 2, pp: (175-180), Month: April - June 2015, Available at: www.researchpublish.com

# FEATURE BASED DETECTION OF LUNG TUMOR USING FUZZYC-MEANS CLUSTERING AND CLASSIFYING USING PROBABILISTIC NEURAL NETWORKS

[1]Dr. P.V.Ramaraju, [2]G.Nagaraju, [3]K.Poojitha, [4]K.B.V.S.P.Babu, [5]Ch.Balasubramanyam,

[1]Professor, [2]Asst.Professor, [3, 4, 5]B.Tech Students, Department of ECE, SRKR Engineering College, Bhimavaram, A.P.,

***Abstract:*** **In our daily life, cancer is well-known disease that causes of death in both men and women and understand about the survival rate of lung cancer which is extremely poor. To increase this survival rate of cancerous patient, it is primarily to detect at premature stage which enables many new options for the cancer treatment without risk. In this paper, we represent Lung Cancer Detection System for finding of lung cancer with the help of image processing mechanisms. This paper presents a neural network based approach to detect lung cancer from raw chest images. These extracted features are considered as the inputs of neural network to train and to verify whether the extracted nodule is a malignant or non-malignant. This research work concentrate on detecting nodules, early stages of cancer diseases, appearing in patient's lungs. Most of the nodules can be observed after carefully selection of parameters. The training dataset of CT images are processed in three stages to attain more quality and accuracy in the processed examination. For segmentation purpose FCM technique is used.**

***Keywords:*** **probabilistic neural network, image processing, segmentation, Fuzzy c-means clustering, lung tumor.**

## 1. INTRODUCTION

Lung Cancer [4] is a malignant lung tumor characterized by uncontrolled or abnormal growth of cells in tissues of the lung. These cells do not carry out the functions of normal lung cells. These cells will form tumor as they grow up and will be disrupt functions of the lung. At later stage of cancer these cells can traverse to other parts of the body, this process is called metastasis. Primary lung cancer originates in the lungs and cells are abnormal lung

cells. Sometimes cancer cells will traverse to lung from the other parts of body. This is called Secondary lung cancer. Cancer represents 13% of the global deaths and among them lung cancer is number one cancer killer. There are 2 major types of lung cancer exist.

- Non-small cell lung cancer (NSCLC)
- Small cell lung cancer (SCLC)

About 10% to 15% of all lung cancers are small cell lung cancer (SCLC), named for the size of the cancer cells when seen under a microscope.

About 85% to 90% of lung cancers are non-small lung cancer (NSCLC), the cells in this differ in size, shape when looked at under a microscope.

X-ray computed tomography (X-ray CT) is a technology that uses computer-processed X-rays to produce tomographic images (virtual 'slices') of specific areas of a scanned object, allowing the user to see inside the object without cutting. Digital geometry processing is used to generate a three-dimensional image of the inside of the object from a large series of two-dimensional radiographic images taken around a single axis of rotation. Medical imaging is the most common application of X-ray CT. Its cross-sectional images are used for diagnostic and therapeutic purposes in various medical disciplines. The rest of this article discusses medical-imaging X-ray CT; industrial applications of X-ray CT are discussed at industrial computed tomography scanning.

As X-ray CT is the most common form of CT in medicine and various other contexts, the term computed tomographyalone (or CT) is often used to refer to X-ray CT, although other types exist (such as positron emission tomography [PET] and single-photon emission computed tomography [SPECT]). Older and less preferred terms that also refer to X-ray CT are computed axial tomography (CAT scan) and computer-aided/assisted tomography. X-ray CT is a form of radiography, although the word "radiography" used alone usually refers, by wide convention, to non-tomographic radiography. CT produces a volume of data that can be manipulated in order to demonstrate various bodily structures based on their ability to block the X-ray beam. Although, historically, the images generated were in the axial or transverse plane, perpendicular to the long axis of the body, modern scanners allow this volume of data to be reformatted in various planes or even as volumetric (3D) representations of structures. Although most common in medicine, CT is also used in other fields, such as nondestructive materials testing. Another example is archaeological uses such as imaging the contents of sarcophagi. Individuals responsible for performing CT exams are called radiographers or radiologic technologists and are required to be licensed in most states of the USA.

Usage of CT has increased dramatically over the last two decades in many countries. An estimated 72 million scans were performed in the United States in 2007. One study estimated that as many as 0.4% of current cancers in the United States are due to CTs performed in the past and that this may increase to as high as 1.5 to 2% with 2007 rates of CT usage; however, this estimate is disputed, as there is not a scientific consensus about the existence of damage from low levels of radiation. Kidney problems following intravenous contrast agents may also be a concern in some types of studies.

CT scan can be used for detecting both acute and chronic changes in the lung parenchyma, that is, the internals of the lungs. It is particularly relevant here because normal two-dimensional X-rays do not show such defects. A variety of techniques are used, depending on the suspected abnormality. For evaluation of chronic interstitial processes (emphysema, fibrosis, and so forth), thin sections with high spatial frequency reconstructions are used; often scans are performed both in inspiration and expiration. This special technique is called high resolution CT. Therefore, it produces a sampling of the lung and not continuous images.

An incidentally found nodule in the absence of symptoms may raise concerns that it might represent a tumor, either benign or malignant. Perhaps persuaded by fear, patients and doctors sometimes agree to an intensive schedule of CT scans, sometimes up to every three months and beyond the recommended guidelines, in an attempt to do surveillance on the nodules. However, established guidelines advise that patients without a prior history of cancer and whose solid nodules have not grown over a two-year period are unlikely to have any malignant cancer. For this reason, and because no research provides supporting evidence that intensive surveillance gives better outcomes, and because of risks associated with having CT scans, patients should not receive CT screening in excess of those recommended by established guidelines.

A probabilistic neural network (PNN) is a feed forward neural network, which was derived from the Bayesian network and a statistical algorithm called Kernel Fisher discriminant analysis. It was introduced by D.F. Specht in the early 1990s. In a PNN, the operations are organized into a multilayered feed forward network [8] with four layers:

- Input layer
- Hidden layer
- Pattern layer/Summation layer
- Output layer

PNN is often used in classification problems. When an input is present, the first layer computes the distance from the input vector to the training input vectors. This produces a vector where its elements indicate how close the input is to the training input. The second layer sums the contribution for each class of inputs and produces its net output as a vector of probabilities. Finally, a complete transfer function on the output of the second layer picks the maximum of these probabilities, and produces a 1 (positive identification) for that class and a 0 (negative identification) for non-targeted classes.

In fuzzy clustering, every point has a degree of belonging to clusters, as in fuzzy logic, rather than belonging completely to just one cluster. Thus, points on the edge of a cluster, may be in the cluster to a lesser degree than points in the center of cluster. An overview and comparison of different fuzzy clustering algorithms is available. Any point x has a set of coefficients giving the degree of being in the kth cluster $w_k(x)$. With fuzzy c-means, the centroid of a cluster is the mean of all points, weighted by their degree of belonging to the cluster:

$$c_k = \frac{\sum_x w_k(x)^m x}{\sum_x w_k(x)^m}.$$

The degree of belonging, $w_k(x)$, is related inversely to the distance from x to the cluster center as calculated on the previous pass. It also depends on a parameter m that controls how much weight is given to the closest center. The fuzzy c-means algorithm is very similar to the k-means algorithm:

- Choose a number of clusters.
- Assign randomly to each point coefficients for being in the clusters.
- Repeat until the algorithm has converged (that is, the coefficients' change between two iterations is no more than $\varepsilon$, the given sensitivity threshold) :
- Compute the centroid for each cluster, using the formula above.
- For each point, compute its coefficients of being in the clusters, using the formula above.

The algorithm minimizes intra-cluster variance as well, but has the same problems as k-means; the minimum is a local minimum, and the results depend on the initial choice of weights. Using a mixture of Gaussians along with the expectation-maximization algorithm is a more statistically formalized method which includes some of these ideas: partial membership in classes.

Another algorithm closely related to Fuzzy C-Means is Soft K-means. Fuzzy c-means has been a very important tool for image processing in clustering objects in an image. In the 70's, mathematicians introduced the spatial term into the FCM algorithm to improve the accuracy of clustering under noise.

The techniques used here are the fuzzy c-means segmentation for segmentation purpose, artificial intelligence of neural networks (PNN) [13], a collection of CT lung images as database and discrete wavelet transformation technique. Focusing on these parameters we aim a automating the lung tumor using multilevel wavelet and neural networks in MATLAB.

## 2. PROPOSED ALGORITHM

The automated disease identification system is not a single process. This system consists of various modules the success rate of each and every step is highly important to ensure the overall high accurate outputs. Step by step procedure is shown in Fig 1.

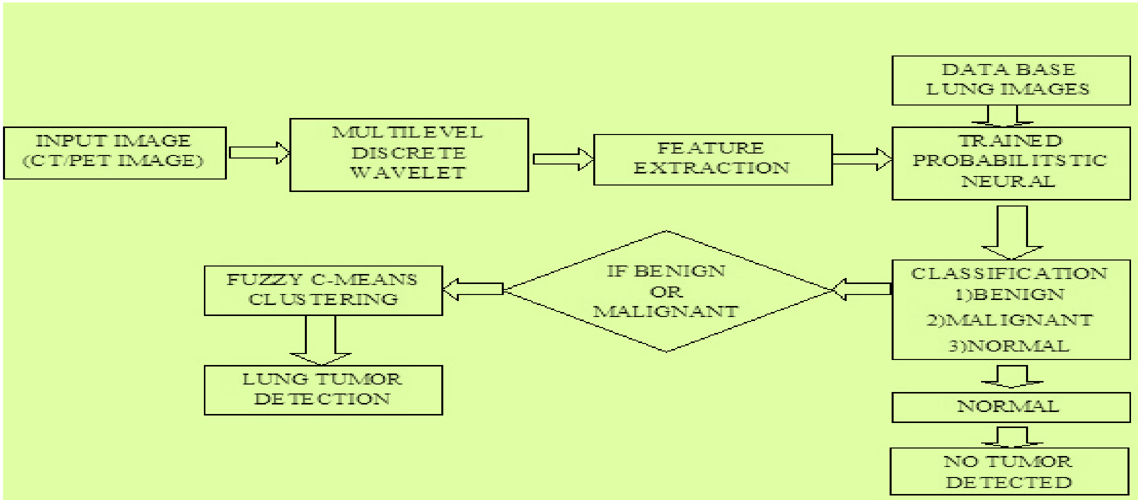



**Fig. 1**

### 2.1 Input Image:

The Input Image is the image on which we will perform the research using the models in the database. The data base images can be CT scan images. This algorithm is mainly focused on the lung tumor images which can be used for analysis of lung tumor.

### 2.2 Data Base Image:

The database is the collection of various image samples of CT and PET of different stages of lung tumor images. It includes various severity levels of lung tumors. Live tumor samples are collected from scanning centers in Bhimavaram. These images are considered as reference images for the analysis of lung tumor. The effective tumor analysis depends upon the number of data base images.

### 2.3 Multi-level Discrete Wavelet Transform:

Wavelet Transform is a type of signal representation that can give the frequency content of the signal at a particular instant of time or spatial location. The Lifting based wavelet transform decomposes the image into different sub band images, It splits component into numerous frequency bands called sub bands. They are LL, LH, HL, and HH sub bands. A high-frequency sub band contains the edge information of input image and LL sub band contains the clear information about the image. If the above decomposition will be done for more than two times it is called as a multi-level wavelet transformation. To extract the texture features effectively which are used in classification using PNN we need this multi-level transformation.

### 2.4 Features Extraction:

Features extraction is a special form which is used to reduce the dimension of the images. When the given input image is too large to process then this technique is used. Then for this processing we covert the image into a reduced set of features called features vector and this conversion is called feature extraction. These reduced features represent the exact information of the input image. By these we can find the parameters like energy, contrast, entropy and correlation.

### 2.5 Probabilistic Neural Network:

Performance of the PNN classifier was evaluated in terms of training performance and classification accuracies. Probabilistic Neural Network gives fast and accurate classification and is a promising tool for classification of the tumors. Existing weights will never be alternated but only new vectors are inserted into weight matrices when training. So it can be used in real-time. Since the training and running procedure can be implemented by matrix manipulation, the speed of PNN is very fast.

The network classifies input vector into a specific class because that class has the maximum probability to be correct. In this paper, the PNN has three layers: the Input Layer, Radial Basis Layer and the Competitive layer. Radial Basis Layer evaluates vector distances between input vector and row weight vectors in weight matrix. These distances are scaled by Radial Basis Function nonlinearly. Competitive layer finds the shortest distance among them, and thus finds the training pattern closest to the input pattern based on their distance.

## 3. SEGMENTATION

In computer vision, **image segmentation** [15] is the process of partitioning a digital image into multiple segments (sets of pixels, also known as super pixels). The goal of segmentation is to simplify and/or change the representation of an image into something that is more meaningful and easier to analyze. Image segmentation is typically used to locate objects and boundaries (lines, curves, etc.) in images. More precisely, image segmentation is the process of assigning a label to every pixel in an image such that pixels with the same label share certain characteristics. In our project we used the Fuzzy c-means segmentation for segmentation purpose.

**Fuzzy c-means segmentation:** Clustering is the process of organizing objects into clusters whose members are similar in some way and items in different clusters are as dissimilar as possible. Different types of similarity measures are used to place each point to a cluster. Fuzzy C means clustering allows one piece of data to two or more clusters. Membership level is associated with each data point. It indicates the strength of the association between data element and particular cluster. Each iteration updates membership level and cluster centers.

$\mu_{ij}=1/\sum_{k=1}^{c}(d_{ij}/d_{ik})^{(2/m-1)}$

$v_j=\frac{\sum_{i=1}^{n}(\mu_{ij})^m x_i}{\sum_{i=1}^{n}(\mu_{ij})^m}$ , $\forall_j=1,2,\ldots\ldots..c$

n = number of data points

$v_j$= $j^{th}$ cluster centre

m=fuzziness index

c=number of cluster centers

$\mu_{ij}$= membership of ith data to jth cluster centre

$d_{ij=}$ Euclidean distance between $i^{th}$ and $j^{th}$ cluster

## 4. CLASSIFICATION OF TYPE OF CANCER

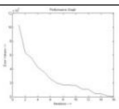
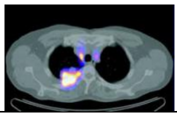
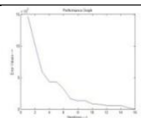
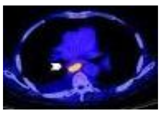
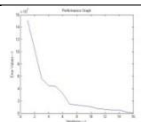
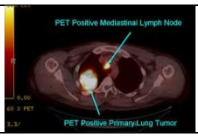
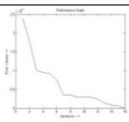
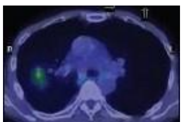
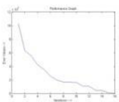
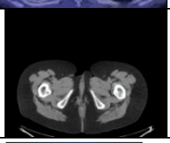
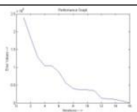
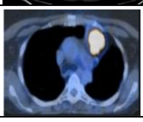
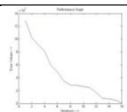
After applying the image to the PNN then it will get classified into a type of cancer based on the parameters calculated. There are mainly three types based on the severity they are benign, malignant and a normal lung.

## 5. RESULTS

The following figures show the results by specifying detection, classification and area calculation to detect and analyze the lung tumor.

The various samples of performance graphs, area of tumor and type of cancer for lung tumor are shown in Table 1.

**Table.1**

| Test Images | Performance Graph | Area of Tumor in mm.sq | Type of Cancer |
|---|---|---|---|
|  |  | No tumor area detected | NORMAL |
|  |  | 18.2351 | BENIGN |
|  |  | 56.0967 | BENIGN |
|   |  | 50.067 | MALIGNANT |
|  |  | No tumor area detected | NORMAL |
|  |  | 7.145 | BENIGN |
|  |  | 19.236 | MALIGNANT |

## 6. CONCLUSION

In this approach, it is very simple to identify the stage and the affected tumor area in the lung tissue. This can be made further effective by improving the database number by including all the types of tumor images into the database. The above results displayed are by using CT and PET image of the organ and they are widely useful for the oncologists in the detection of tumor.

## REFERENCES

[1] Dr.P.V.Ramaraju, ShaikBaji, “Brain Tumor classification, detection and segmentation using digital image processing and probabilistic neural network techniques”, 11-12 October 2014, International conference on Electrical, Electronics Engineering trends, communication optimization and sciences (E3cos), Vijayawada, Andhra Pradesh, India.pp.111-116.

[2] S.S.Kumar, R.S.Moni, I.Rajeesh, "Automatic liver and lesion segmentation: a primary step in diagnosis of liver diseases", Signal, Image and Video Processing DOl: I 0.1007/s11760-011-0223-y, March 31, 2011

[3] E-Liang Chen, Pau-CHoo Chung, Ching-Liang Chen, Hong-Ming Tsai, Chein I Chang, "An Automatic Diagnostic system for CT Liver Image Classification", IEEE Transactions Biomedical Engineering,Vol. 45, pp. 783-794,1 998

[4] Dr.P.V.Ramaraju, Satti Praveen, “Classification of Lung tumor using geometrical and texture features of chest X-ray images”, October 16, 2014. International conference on current innovations in Engineering and Technology, Andhra Pradesh, India. Pp 126-131.

[5] Yu-Len Huang, Jeon-Hor Chen, Wu-Chung Shen, "Diagnosis of hepatic tumors with texture analysis in non enhanced computed tomography images", Academic Radiology,Vol. 13, p p 713-720,2006.

[6] Yu-Len Huang, Jeon-Hor Chen, Wu-Chung Shen, "Computer-Aided Diagnosis of Liver Tumors in Non-enhanced CT Images", Journal of Medical Physics,Vol. 9,p p 141-150,2004.

[7] Int. J. of Recent Trends in Engineering & Technology, Vol. 11, June 2014 Image Processing and Classification Techniques for Early Detection of Lung Cancer for Preventive Health Care: A Survey Prashant Naresh1, Dr. Rajashree Shettar

[8] L. L. Wu, M. S. Yang, "Alternative c-means clustering algorithms", Pattern Recognition Vol.35,p p. 2267-2278,2 002.

[9] Selvaraj.D and Dhanasekaran.R, “MRI BrainTumour Detection by Histogram and Segmentation by Modifie GVF Model”, International Journal of Electronics and Communication Engineering & Technology(IJECET),Volume 4, Issue 1, 2013, pp. 55 - 68, ISSN Print: 0976- 6464, ISS N Online: 0976 –6472.

[10] B.Venkateswara Reddy, Dr.P.Satish Kumar, Dr.P.Bhaskar Reddy and B.NareshKumarReddy, “Identifying Brain Tumour from MRI Image using Modified F CM and Support Vector Machine”, International Journal of Computer Engineering & Tec hnology (IJCET), Volume 4, Issue 1, 2013, pp. 244 - 262, ISSN Print: 0976 – 6367, ISSN Online: 0976 – 6375.

[11] Information Engineering Letters, ISSN: 2160-4114 Volume 2, Number 1, March, 2012 Lung Cancer Cell Classification Method Using Artificial Neural Network Azian Azamimi Abdullah, Syamimi Mardiah Shaharum

[12] d. lin and c. yan, “lung nodules identification rules extraction with neural fuzzy network”, ieee, neural information processing, vol. 4, 2002.