

Lung Cancer Detection Using Artificial Neural Network & Fuzzy Clustering

Masaood A. Hussain¹, Tabassum M. Ansari², Prarthana S. Gawas³, Nabanita Nath Chowdhury⁴

Student, EXTC Engineering, MHSSCOE, Mumbai, India ^{1,2,3}

Assistant Professor, EXTC Engineering, MHSSCOE, Mumbai, India⁴

Abstract: To evaluate the performance of Computer Aided Diagnosis (CAD) for Lung Cancer using artificial neural intelligence on CT scan images. Lung Cancer can be summarized by evaluating region of interest using maximum entropy and supervised learning. Lung cancer seems to be the common cause of death among people throughout the world. Early detection of lung cancer seems to be the only factor which can increase the chance of survival among people.

Index Terms: ANN, CAD, GLCM, NSCLC, SCLC, SE

I. INTRODUCTION

Lung Cancer is the most threatening sort in one of the deadliest malignancies form of cancer. In the last few years the occurrence of harmful tumor has constantly expanded, on the grounds that the cure of the disease depends on its initial judgment. There are two major types of lung cancer, Non-small cell lung cancer (NSCLC) and Small cell lung cancer (SCLC). The lungs are usually large in size hence tumors can grow in them for a long time before they are found. Even when the symptoms such as coughing and fatigue occur, people think they are due to other causes. Hence for this reason, the early-stage lung cancer i.e. stages I and II are difficult to detect. Many people having lung cancer are diagnosed at stages III and IV. Doctors utilize a some methods to diagnose lung tumor, for example, X-rays, CT Scan, PET scan etc

1. Coughing, especially if it persists or becomes intense
2. Pain in the chest, shoulder, or back unrelated to pain from coughing
3. A change in color or volume of sputum
4. Shortness of breath
5. Changes in the voice or being hoarse
6. Harsh sounds with each breath
7. Recurrent lung problems, such as bronchitis or Pneumonia.
8. Coughing up phlegm or mucus, especially if it is tinged with blood

I. SCOPE¹

With the use of same technique another type of cancer can also be used to detect in major applications of cancer detection, At present the image processing is carried on in the grayscale image, future improvements can be directly on the colour images. The million order dataset can be selected and image classification can be done on larger

dataset. With increased size of dataset various issues such as uploading data, managing feature set, increased execution time of classification algorithms etc. can be considered. More image features can be extracted for better classification. Various combinations of preexisting features can be used to correctly classify medical data.

II. OBJECTIVE

The main objective is to detect the cancer in early stage. Lung cancer is considered to be as the main cause of cancer death worldwide, and it is difficult to detect in its early stages because symptoms appear only at advanced stages causing the mortality rate to be the highest among all other types of cancer

III. TECHNICAL SPECIFICATIONS

A. Basic ANN Architecture

An artificial neural network is an information processing system which has certain performance characteristics similar to biological neural networks. Neural network consists of a large number of simple processing elements called neurons. Each neuron connects to other neurons by means of directed communication links, each with an associated weight. The weights represent information used by the network to solve a problem. Below Fig. shows the basic design of a simple multilayer neural network. In general, multilayer ANNs can have more than one hidden layer.

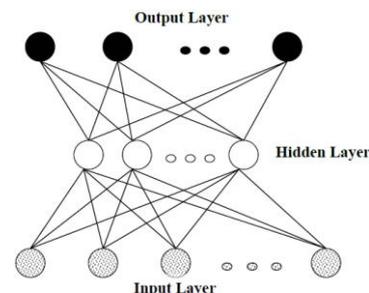


Fig. A basic multilayer neural network.

A. Modes of Operation

There are two modes of operation in a neural network:

- Training mode
- Testing mode.

IV. BLOCK DIAGRAM

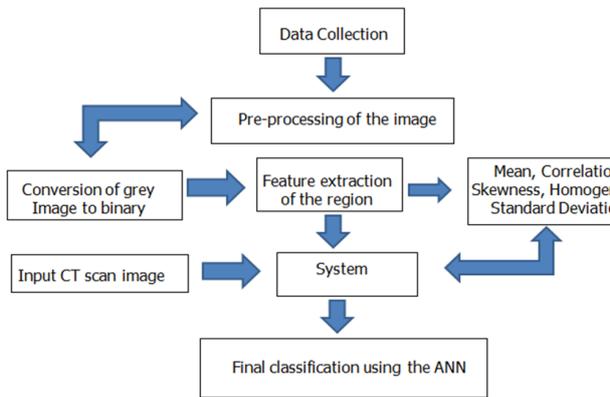


Fig. Basic block diagram

V. METHODOLOGY

A. Data collection

A total number of 100 data managed to be collected. In which 50 are the cancerous and the remaining 50 are non-cancerous. No additional information on the CT scan images are obtained, thus the medical status and the background of the images are left as the task to be analyzed in order to choose the suitable images to be used for the project.

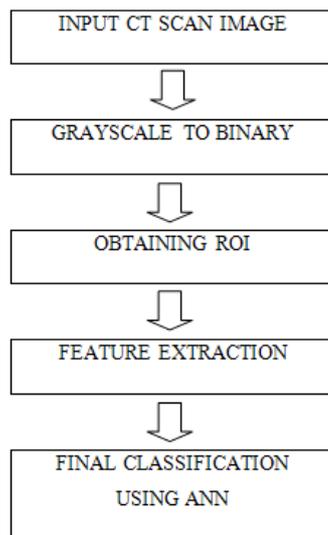


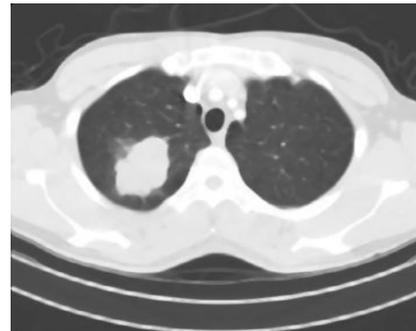
Fig. Implementation steps

B. Noise Removal

Many times image is corrupted by noise as the image is collected from many other sources thus initially noise is removed by making use of median filter which contains highest signal to noise ratio



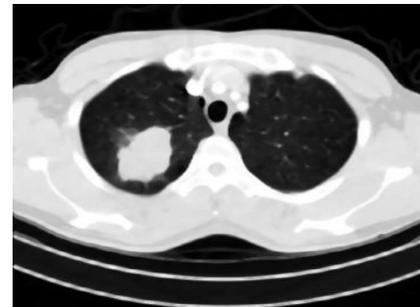
(a) Input image



(b) Image without Noise

C. Image Enhancement

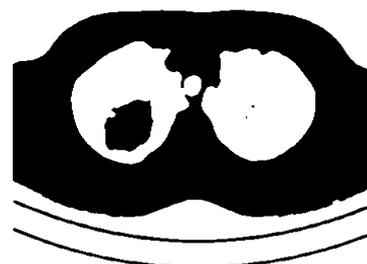
The quality of the image can be enhanced in order to increase the intensity and thus can be proper for feature extraction



(c) Enhanced Image

D. Gray scale to Binary conversion

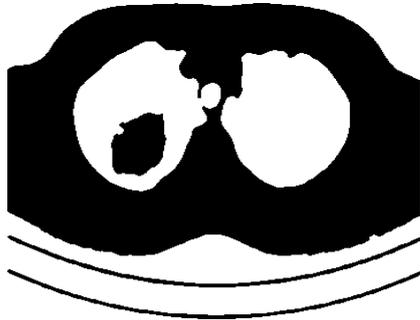
Image binarization is a subclass of image segmentation as it divides an image into segments based on the value of pixels to a threshold value. The simplest of all the thresholding techniques is partition of the image using a single global threshold. With the use of this threshold value the upper half intensities is converted in full white and lower half into black.



(d) Binary Image

E. Removal of erroneous part

Sometimes some of the erroneous part is present which is not required in further process and thus need to be removed



(e) Removal of erroneous part

F. Image Segmentation

The image segmentation is obtained by forming a gray level co-occurrence matrix of the image this matrix is used for feature extraction The gray comatrix function creates a gray-level co-occurrence matrix by calculating how often a pixel with the intensity i.e. gray-level value *i* occurs in a specific spatial relationship to a pixel with the value *j*. The spatial relationship is defined as the pixel of interest and the pixel to right i.e. horizontally adjacent, you can specify other spatial relationships between the two pixel. Each element i.e. *i,j* in the resultant is simply the sum of the number of times that the pixel with value *I* occurred in the specified spatial relationship to a pixel with value *j* in the input image.

VI. CALCULATION AND FEATURE EXTRACTION

Once the segmentation is done on the lung region, the Features can be obtained from it and the diagnosis rule Can be designed detect the Lung Cancer.

1. Mean

The mean intensity value indicates the average intensity value of all the pixels that belong to the same region, add all of the samples together, and divide by N.

$$\frac{1}{N} \sum_{k=1}^N \text{Intensity}(k).$$

2. Correlation

Correlation is the degree and type of relationship between adjacent pixels

$$\frac{\sum_x \sum_y (A_{xy} - A)(B_{xy} - B)}{\sqrt{(\sum_x \sum_y (A_{xy} - A)^2)(\sum_m \sum_N (B_{xy} - B)^2)}}$$

..... 4.1.4(b)

Where *A_{xy}* is the pixel intensity at point (x,y) and *B_{xy}* is the grayscale value at point (x,y)

3. Standard deviation

It is a measure of how much that gray levels differ from mean, defined by.

$$\frac{1}{N} \sum_{k=1}^N (\text{Mean}(k) - \text{Intensity}(k))^2$$

4. Skewness

Skewness is a measure of the asymmetry of the data around the sample mean. The skewness of a distribution is defined as

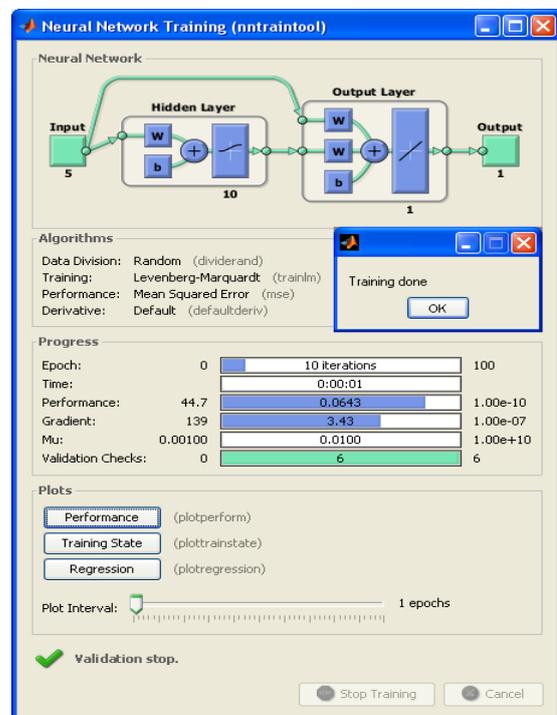
$$\frac{E(x - \mu)^3}{\sigma^3}$$

Where μ is the mean of *x*, σ is the standard deviation of *x*, and *E(t)* represents the expected value of the quantity *t*. Skewness computes a sample version of this population value.

5. Homogeneity

Homogeneity is the measure of similarities in two adjacent pixels Homogeneity is a local information that reflects the uniformity of a region; when a region is homogeneous, data in the homogeneous region are likely to have similar characteristics.

VII. RESULT



```

Command Window
homogeneity
1.2307e+05

Mean_sig =

2.0194e+03

Var_sgl =

1.0281e+17

sd_sgl =

1.0208e+04

skew =

403.0332

fx >>

```

VIII. CONCLUSION

We will design a computer aided diagnosis (CAD) system to detect cancerous cell in CT Scan images. This CAD system comprises four stages: pre-processing, evaluating region of interest, feature extraction, final classification using ANN. The pre-processing stage consists of various image enhancement techniques (contrast enhancement, thresholding, noise removal) to improve the visibility of tumors in CT scan images. We had taken out the ROI and GLCM matrix is formed from the enhanced image feature is extracted and give them as input to the next stage. We will use artificial neural networks in the classification stage.

After training is completed testing will be done in which the same steps will be applied to the testing image and the parameters of it will be compared to the matrix formed after training. A distance will be calculated with each and every row of the matrix this distance will give us the difference and the minimum distance will be the minimum difference and considered as the best match. And final result will be showed that whether the image is cancerous or not.

REFERENCES

- [1] Disha Sharma, Gangadeep Jindal, "Identifying Lung Cancer Using Image Processing Technique", International Conference of Computational Techniques and Artificial Intelligence, 2011
- [2] Atiyeh Hashemi, Abdol Hamid Pilevar, Reza Rafeh, "Mass Detection in Lung CT Images Using Region Growing Segmentation and Decision Making Based on Fuzzy Inference System and Artificial Neural Network", I.J. Image, Graphics and Signal Processing, 6, pp: 16-24, 2013
- [3] Anam Tariq, M. Usman Akram and M. Younus Javed, "Lung Nodule Detection in CT Images using Neuro Fuzzy Classifier", Fourth International Workshop on Computational Intelligence in Medical Imaging (CIMI), pp:49-53, 2013
- [4] Fatma Taher, Naoufel Werghi and Hussain Al-Ahmad, "Bayesian Classification and Artificial Neural Network Methods for Lung Cancer Early Diagnosis", IEEE, pp: 773-776, 2012
- [5] Varalakshmi. K," Clasification of Lung Cancer Nodules using a Hybrid Aproach," Proceeding in the Journal of Emerging Trends in Computing and Information Sciences, Vol. 4, No. 1 ISSN 2079-8407, Jan 2013
- [6] S.K. Vijai Anand, "Segmentation coupled Textural Feature Classification for Lung Tumor Prediction", ICCCT, pp: 518-524, 2010.
- [7] JIA Tong, ZHAO Da-Zhe, YANG Jin-Zhu, WANG Xu, "Automated Detection of Pulmonary Nodules in HRCT Images", IEEE, 2007.

BIOGRAPHIES



Name: **Mr. Masaood A. Hussain**
Designation: Student
Department: EXTC
Qualification: B.E. (EXTC) Pursuing



Name: **Ms. Tabassum M. Ansari**
Designation: Student
Department: EXTC
Qualification: B.E. (EXTC) Pursuing



Name: **Ms. Prarthana S. Gawas**
Designation: Student
Department: EXTC
Qualification: B.E. (EXTC) Pursuing



Name: **Mrs. Nabanita Nath Chowdhury**
Designation: Asst. Prof.
Department: EXTC
Qualification: M.Tech (EXTC)