

Lung Cancer Detection Using Grey Level Co-Occurrence Matrix

Mahpara Alam¹, Bhavika Chaudhari², Shamal Bangar³, Pratiksha Desale⁴, Prof. Chaitali Patil⁵

^{1,2,3,4}Students, Department of Computer Engineering, K.K.W.I.E.E.R, Maharashtra, India

⁵Professor, Department of Computer Engineering, K.K.W.I.E.E.R, Maharashtra, India

Abstract - Lung cancer is curable disease if detected in its early stages. In recent years, Image processing techniques are widely used in medical domain for image improvement in earlier detection. For diagnosis of lung cancer, CT image of lung is one of the most common imaging modalities. Manually identifying tumor from hundreds of CT images for any patient may prove to be very tedious and time consuming task. Hence this system aims at detection of cancer through an automated process to minimize human error and making the process more accurate and hassle-free. Lung carcinomas are categorized by the size and appearance of the malignant cells. The growth of epithelial cells on lung parenchyma region shows in CT image which can be used in the system. In this proposed approach image pre-processing techniques are used like noise removal, enhancement and image segmentation then the feature extraction recognize the region of interest of tumor. SVM (Support Vector Machine) algorithm can be used for classification. GLCM (Grey Level co-occurrence Matrix) based textual features detection can be used to search the lung parenchyma region and extract the features.

Key Words: GLCM, CT images, SVM, Feature extraction, Image processing

1. INTRODUCTION

Manually identifying tumours from hundreds of MRI slices for any patient may prove to be very tedious and time consuming task. Cancer is curable disease if it detected in its early stage. To reduce the death rates due to occurrence of lung cancer we need some methodology to detect the cancer in its early stage. Not only the detection of disease in its early stage may reduce the death rate but also proper diet and doctor treatment can be done on time.

Detecting Cancer is still challenging for the doctors in the field of medical. Even now the actual reason and complete cure of cancer is not invented. For better solution we are going to design and implement lung cancer detection system to minimize human error and making process more accurate and hassle-free.

2. LITERATURE SURVEY

Mokhled S. AL-TARAWNEH [1], The study on the image processing techniques is carried out for detection of cancer. According to author many techniques are available to detect cancer earlier, but none of the technique gave good result

because of time factor. Image quality and accuracy was the core factor of this research. Image quality assessment as well as improvement depends on the enhancement stage where pre-processing techniques was used.

Ritika Agarwal, Ankit Shankhadhar, Raj Kumar Sagar [2], Many techniques are available to determine the existence of nodule in the lung in its early stage but none of the technique gave the best result. Therefore they implemented the system to give more accuracy and best result they propose a system called as Computer Aided diagnosis System CAD for identifying the presence of nodule in lung.

Khinn Mya Mya Tun [3], The implementation of image pre-processing and image segmentation is carried out to obtain the diagnosis result. Median filter was used for image pre-processing. In feature extraction Gray Level Co-Occurrence matrix (GLCM) was used. In classification, feed-forward neural network was used to classify the lung cancer stages.

Mr. Akshay Bhor, Mr. Aditya Likhari, Mr. Azhar Maner, Mr. Aditya Patil, Prof. Dipti Chaudhari [4], The main aim of the system was to automate the classification method for the first detection of carcinoma. It included classification algorithmic program i.e. Neural Network and for improvement GA (Genetic Algorithm) was employed.

Tanushree Sinha Roy, Neeraj Sirohi, Arti Patle[5], focuses their study mainly on the classification of lung images as affected by cancer disease or not. In their system lung CT scan image was used as input image and then some important features were extracted to classify the image. To classify the image Fuzzy Interface System was used. The system used both neural network and fuzzy logic technique.

3. PROPOSED SYSTEM

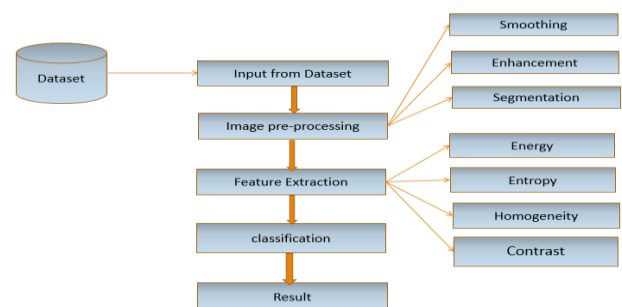


Fig -1: Proposed System

3.1 Image Database

The required database for system must contain the CT scan Images. The CT scan image is more reliable than other form of images. The images are in the DICOM format. It needs to be converted to use as input to the system.

3.2 Image Preprocessing

Smoothing, image enhancement are the techniques are used in pre-processing. In proposed model for median filter is used to remove the noise from the given image. Median filter give the best result because it removes the noise without blurring the image whereas other techniques remove the noise but give blur image as output. Image enhancement improves the visibility of features for better understanding. It also highlights the high frequency area and suppress the low frequency area.

3.3 Feature Extraction

The features like energy, entropy, homogeneity, contrast can be extracted after the segmentation process is done in the region of lung, for analyzing some diagnosis to identify the cancer in the lung correctly. In this study, the features are used by representing Gray level Co-occurrence matrix.

4. GLCM

The extraction of the texture features using GLCM has been considered as the significant technique and it has been used in several applications of remote sensing for analysis of the texture.

GLCM points to Gray level Co-occurrence matrix. It is of the second order statistics, so information with regards to pixels of pairs are collected by GLCM algorithm. GLCM exhibits how the pixel brightness occurs in an image. A matrix is built up at a distance of d=1 and at angles in degrees (0, 45, 90, 135).

GLCM texture picks up the relation between two pixels at a time, called the neighbour and the reference pixel. GLCM calculates the distance and angular spatial relationship over an image sub- region of specific size. GLCM is prepared by using gray scale values.

| | | | |
|---|---|---|---|
| 0 | 0 | 1 | 1 |
| 0 | 0 | 1 | 1 |
| 0 | 2 | 2 | 2 |
| 2 | 2 | 3 | 3 |

Fig -2: Test image figure

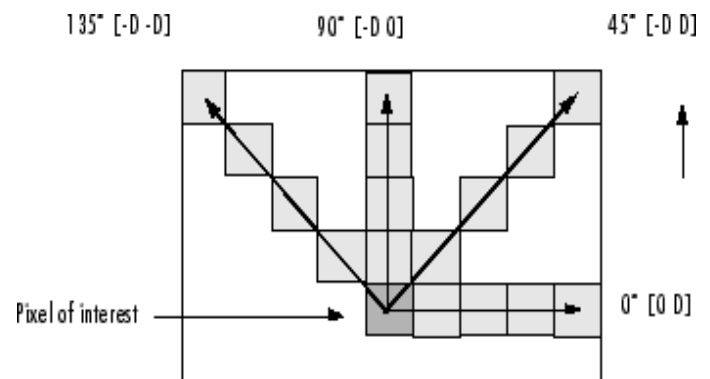
| | | | | |
|-------------------------|-------|-------|-------|-------|
| Neighbour pixel value-→ | 0 | 1 | 2 | 3 |
| Reference pixel value: | | | | |
| 0 | (0,0) | (0,1) | (0,2) | (0,3) |
| 1 | (1,0) | (1,1) | (1,2) | (1,3) |
| 2 | (2,0) | (2,1) | (2,2) | (2,3) |
| 3 | (3,0) | (3,1) | (3,2) | (3,3) |

Fig -3: GLCM Calculation

The extraction of features is done as the follows:

1. Divide the planes of the image into RGB
2. Calculate the matrices of GLCM as per given equation.
3. Calculate the geometric features for the obtained GLCM matrix:

Energy, Contrast, Entropy, Homogeneity.



Notations:

P_{ij} = Element ij of GLCM

N = No. of Gray levels in the image

4.1 Energy

$$Energy = \sum_{i,j=0}^{N-1} (P_{ij})^2$$

This statistic is also called as Angular second moment or Uniformity. It measures the textural uniformity that is the pixel pair repetitions. It detects disorders present in textures. Energy reaches a maximum value equal to the value one. High energy values occur when the distribution of gray level has a constant or periodic form.

4.2 Entropy

$$Entropy = \sum_{i,j=0}^{N-1} -\ln(P_{ij})P_{ij}$$

This statistic measures the complexity of an image. The entropy is large when the image is not texturally uniform and many of the GLCM elements have very small values. Complex textures tend to have a high entropy. Entropy is strongly, but inversely correlated to the energy.

4.3 Contrast

$$Contrast = \sum_{i,j=0}^{N-1} P_{ij} (i-j)^2$$

This statistic measures the spatial frequency of an image which is the difference moment of GLCM. It is the difference between the highest and the lowest values of a contiguous set of pixels present in an image.

4.4 Homogeneity

$$Homogeneity = \sum_{i,j=0}^{N-1} \frac{P_{ij}}{1+(i-j)^2}$$

This statistic is also called as the Inverse Difference Moment. It measures image homogeneity. It is sensitive to the presence of near diagonal elements in the GLCM. When all elements in the image are same it gives maximum value. GLCM homogeneity and contrast are strongly, but inversely correlated in terms of equivalent distribution in the pixel pair population. It means if contrast increases homogeneity decreases while energy is kept constant.

5. CONCLUSIONS

Lung Cancer is one of the largest occurred disease in the world. Only early detection of disease may help to decrease the death rate due to the lung cancer. Image processing techniques are widely used in some medical areas for image improvement in earlier detection and treatment stages, where the time factor is very important to discover the abnormality issues in target images.

This proposed system is able to detect the cancer in early stages more correctly. We proposed the methodology to detect the cancerous lung CT image based on GLCM.

REFERENCES

[1] Mokhled S. AL-TARAWNEH, "Lung Cancer Detection Using Image Processing Techniques". Leonardo Electronic Journal of Practices and Technologies, 2012

[2] Ritika Agarwal, Ankit Shankhadhar, Raj Kumar Sagar, "Detection of Lung Cancer using content based Medical Image Retrieval ,Fifth International Conference on Advanced Computing and Communication Technologies, 2015

[3] Khinn Mya Mya Tun ,Feature Extraction and Classification of Lung Cancer Nodule using Image Processing Techniques, International Journal of Advance Research in Engineering, Science & Technology, 2014

[4] Mr. Akshay Bhor, Mr. Aditya Likhar , Mr. Azhar Maner , Mr. Aditya Patil, Prof. Dipti Chaudhari , Classification of lung cancer stages using Quantitative Image and Genomic Biomarker, International Journal of Advance Research in Engineering, Science & Technology, 2017

[5] Tanushree Sinha Roy, Neeraj Sirohi, ArtiPatle , "Classification of Lung Image and Nodule Detection Using Fuzzy Inference System," International Conference on Computing, Communication and Automation, 2015