

**Detection of lung cancer tumor in its early stages using image processing techniques**¹Dongare Vijaykumar, ²Kadam Suraj, ³Bakhale Tushar, ⁴Dhalpe Somnath^{1,2,3,4}D. Y. Patil College of Engineering, Ambi, Pune, Maharashtra, India

Abstract — Lung Cancer is a Disease of uncontrolled cell growth in tissues of the lung. Discovery of Lung Cancer in its initial stage is the key of its treatment. Medical problems are often in each person. Cancer is one of the most unpredictable diseases a human can ever had. The Disease like lung cancer is very difficult to detect in its early stages. In numerous parts of the world far reaching screening by MRI is not yet realistic, so that midsection radiology stays in starting and most basic system. Important to discover the abnormality issues in target CT images, particularly in various cancer tumors such as lung cancer, Image quality and accuracy is the core factors of this research, image quality assessment as well as improvement are depending on the amplification stage where low pre-processing techniques, such as Lung Field Segmentation using watershed algorithm, Data Processing, Feature Extraction, Genetic Algorithm Classification utilizing neural system and SVMs are discussed in detail. This system provides more accurate results by using various images enhancement and segmentation techniques on CT images.

Keywords- Lung CT images, Pre-processing, Watershed, segmentation, SVM Classifier, Genetic Algorithm, Feature extraction.

I. INTRODUCTION

Lung Cancer is a noteworthy reason for Mortality in the western world as exhibited by the striking factual numbers distributed consistently by the American Lung Cancer Society. They demonstrate that the 5-year survival rate for patients with lung malignancy can be enhanced from a normal of 14 parentage up to 49 parentage if the ailment is analyzed and treated at its initial stage. Medicinal pictures as a vital piece of therapeutic determination and treatment were focusing on these pictures for good. These pictures incorporate success of concealed data that misused by doctors in settling on contemplated choices around a patient. Then again, removing this important shrouded data is a basic first stride to their utilization. This reason inspires to utilize information digging systems abilities for productive learning extraction find concealed lung.

Mining Medical pictures includes numerous procedures. Medicinal Data Mining is a hopeful region of computational insight connected to a consequently break down patients records going for the expose of new information valuable for restorative choice making. Affected information is expected not just to increment exact determination and effective infection treatment, additionally to improve security by diminishing blunders. The systems in arrange the progressive X-beam midsection movies in two classes: ordinary and strange. The ordinary ones are those portraying a solid patient. The irregular ones incorporate Type of lung tumor; we will utilize a typical arrangement technique specifically SVMs neural systems.

i. Problem Statement

Nowadays cancer is the most serious health problem for any human being in world and detecting lung cancer in its early stages is very difficult time consuming. So this system will take CT images as input and process on them using various technique like images enhancement, segmentation and detect stages of cancer.

ii. Proposed System

In this system use some techniques are used. Crucial to the task of medical image mining, Lung Arena Segmentation, Data Processing, Feature Extraction, Classification using neural network and SVMs. Different learning experiments were performed on two different data sets, created by means of feature selection and SVMs trained with different parameters; the results are compared and reported.

iii. Scope of Project

Given system describe what features are in the scope and what are not in the scope of the system to be developed. Our project is like a component which can be used on different ways in future. Component can be implemented in hospital management system to improve and support doctor's work. It can be also be used in as Android App for more generalized purpose.

iv. Notations and Preliminaries

$S = \{I, O, F, \text{Success}, \text{Failure}\}$
 $I = \{I \text{ is set of inputs}\}$
 $I = \{\text{CT images}\}$
 $O = \{O \text{ is set of outputs}\}$
 $O = \{O1, O2\}$
 $O1 = \{\text{Is size of Tumour}\}$
 $O2 = \{\text{Is level of lung cancer}\}$
 $F = \{\text{Set of Function}\}$
 $F = \{F1, F2, F3, F4, F5, F6, F7, F8, F9\}$
 $F1 = \{\text{Is grey scale conversion}\}$
 $F2 = \{\text{Is normalization}\}$
 $F3 = \{\text{Is Noise reduction}\}$
 $F4 = \{\text{Is SVM Classifier}\}$
 $F5 = \{\text{Is remove uncounted portion of image}\}$
 $F6 = \{\text{Binarization}\}$
 $F6 = \{g(x,y)=1, \text{ if } f(x,y) \geq T \text{ otherwise } 0\}$ Where T is the threshold value
 $F7 = \{\text{Is segmentation}\}$
 $F8 = \{\text{Is the thresholding method}\}$
 $F9 = \{F02, F03\}$
 $F9 = \{\text{Is feature extraction}\}$
 $F02 = \{(\text{center} - x = \text{weight}/2)\}$
 $F02 = \{(\text{center} - y = \text{height}/2)\}$
 Success = Display the size of tumour and level of the lung cancer.
 Failure = Does not show the size of tumour and level of the lung cancer.
 Success=Receive MMS and email on alternative mobile number and email address.
 Failure= Did not get any MMS or Email.

II. IMPLEMENTATION

Lung cancer is that the leading reason behind tumour-related deaths within the world. At identical time, it seems that the speed has been steady increasing. Carcinoma is caused by the uncontrolled growth of tissues within the respiratory organ. The yank cancer society estimates that 213, 380 new cases of carcinoma within the U.S are diagnosed and a hundred and sixty, 390 deaths as a result of carcinoma can occur in 2007. Tobacco smoking is the main behind all cases. Lung cancer is that the growth of a tumour, known as a nodule that arises from cells lining the airways of the system. The detection of carcinoma has been a tedious task in medical image analysis over the past few decades. Within the health trade, chest X-rays are thought-about to be the foremost wide used technique for the detection of carcinoma.

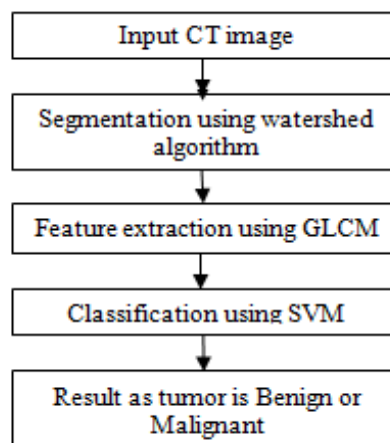


Fig. Procedure of image processing.

Methodology Used

1. SEGMENTATION USING MARKER BASED WATERSHED ALGORITHM

The segmentation of lungs from CT images is a critical step in any Computer Aided Design system which leads to the early diagnosis of lung cancer and also in other pulmonary diseases. The segmentation of lung CT image is a very challenging problem due to in-homogeneity in the lung region and pulmonary structures having similar densities such as veins, arteries, bronchioles, and different scanning protocols and scanners. The success of this technique can be measured in terms of accuracy, time complexity, processing time, and efficiency level. The tumor in lung form darker regions in CT images compared to other parts of the chest such as the heart and the liver.

Input:

Noise free Lung CT image from medical database.

output:

Segmented/ Partitioned image.

2. FEATURE EXTRACTION USING GREY LEVEL CO-OCCURRENCE MATRIX

Gray Level Co-occurrence Matrix (GLCM) is one of the most popular ways to describe the texture of an image. The extracted ROI can be distinguished as either cancerous or not using their texture properties. A GLCM denote the second order conditional joint probability densities of each of the pixels, which is the probability of occurrence of grey level i and gray level j among a given distance ' d ' and on the direction ' θ '. 7 options area unit thought-about for planned technique.

1. Area: It provides the particular variety of pixels within the ROI.
2. Convex Area: It provides the quantity of pixels in convex image of the ROI.
3. Mean: it's the proportion of the pixels within the convex hull that also are within the ROI.

$$\text{Mean} = \mu_i = \sum_{i,j=0}^{N-1} i (P_{i,j}) \quad \dots (1)$$

4. Energy: it's the summation of square parts within the GLCM and its price ranges between zero and one.

$$\text{Energy} = \sum_{K=0}^N P^2(i, j) \quad \dots (2)$$

5. Contrast: it's the live of distinction between AN intensity of constituent and its neighboring pixels over the total ROI. Where, N is that the variety of various grey levels.

$$\text{Contrast} = \sum_{i,j=0}^{N-1} P_{i,j} (i - j)^2 \quad \dots (3)$$

6. Homogeneity: it's the live of closeness of the distribution of parts within the GLCM to the GLCM of every ROI and its price ranges between zero and one.

$$\text{Homogeneity} = \sum_{i,j} \frac{P(i, j)}{1 + |i - j|} \quad \dots (4)$$

7. Correlation: it's the live correlation of constituent to its neighbor over the ROI.

3. IMAGE CLASSIFICATION USING SVM CLASSIFIER

SVM introduced by Cortes is mostly used for classification purpose. SVMs area unit economical learning approaches for training classifiers supported many functions like polynomial functions, radial basis functions, neural networks etc. it's thought-about as a supervised learning approach that produces input-output mapping functions from a labelled training dataset. SVM has vital mental capacity and thus is broadly speaking applied in pattern recognition. SVMs area unit universal approximators that rely upon the applied math and optimizing theory. The SVM is especially placing the biological analysis and capable to handle noise, massive dataset and enormous input areas.

The fundamental plan of SVM may be represented as follows:

- a. Initially, the inputs area unit developed as feature vectors.
- b. Then, by victimization the kernel perform, these feature vectors area unit mapped into a feature house.
- c. Finally, a division is computed within the feature house to separate the categories of training vectors.

Algorithm 1: Genetic Algorithm

1. Start with a randomly generated population of M chromosomes, where M is the size of population, l – length of chromosome x .
2. Calculate the fitness value of function $\phi(x)$ of each chromosome x in the population.
3. Repeat until M offspring's are created:
 - 3.1 Probabilistically select a pair of chromosomes from current population using the value of the fitness function.
 - 3.2 Produce an offspring y , using crossover and mutation operators, where $I = 1, 2, \dots, N$.
4. Replace current population with newly created one.
5. Go to step 2. In some cases of GA, the whole population is formed for strings having the same length. Thus GA is applied to find the optimum feature subset through the evolution process. The proposed fitness function of the GA is used to test the performance of an individual.

Algorithm 2: MD 5 algorithm

MD5 algorithm description MD5 algorithm takes input message of arbitrary length and generates 128-bit long output hash. MD5 hash algorithm consist of five steps:

Step 1. Append Padding Bits

Step 2. Append Length

Step 3. Initialize MD Buffer

Step 4. Process Message in 16-Word Blocks

Step 5. Output

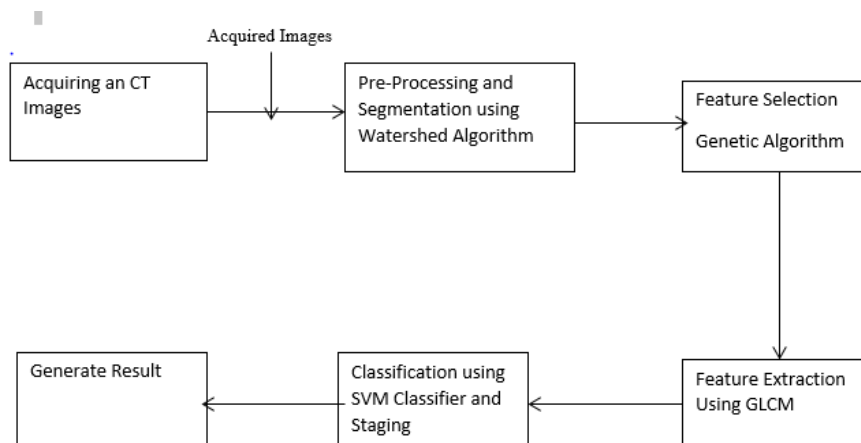


Fig. 1: System Architecture

III. RESULT ANALYSIS

IV.



Fig. 1(a).Select the image

Fig. 1(b). Segmented image

First user have to select the image and convert it into gray scale image. The tumor area is separated from original image shown in fig.(b).



Fig. 2(a) Dark image



Fig. 2(b) Final Result

After gray scale conversion, this image have to convert into the dark image and final result will be display as shown in fig. 2(b) as tumor size and the stage of cancer.

IV.CONCLUSIONS

The proposed model identifies and detects the stage of disease based on the features extracted. The approach starts by extract- ing the lung regions from the CT image using several image pro- cessing techniques, including binary image, image segmentation, binarization, watershed algorithm, svm classifier. we use the gray level co-occurrence in this system. Initially process is to read the image and need to reprocess because of high resolution and noise occur in the image, the noises are removed using median filter and the image is enhanced and segmented. In the future the reprocessed image will be the input for feature selection and extraction which are used to extract the particular region. The extracted features must be stored for classification. Based on the classification, stages will be identified which is used for a physi- cian to give some therapy suggestions. Correctness of this system is generate the result up to 65% to 70%.

ACKNOWLEDGEMENT

In performing our assignment, we had to take the help and guideline of some respected persons, who deserve our greatest gratitude. The completion of this assignment gives us much Pleasure. We would like to show our gratitude Prof. Vrushali Desale for giving us a good guideline for this research throughout numerous consultations. We would also like to expand our deepest gratitude to all those who have directly and indirectly guided us in writing this research.

REFERENCE

1. Zakaria Suliman Zubi and Rema Asheibani Saad, Using Some Data Mining Techniques for Early Diagnosis of Lung Cancer, Recent Researches in Artificial Intelligence, Knowledge Engineering and Data Bases, Libya, 2007.
2. Paola Campadelli, Elena Casiraghi, and Diana Artigiani, A Fully Automated Method for Lung Nodule Detection From Postero-Anterior Chest Radiographs, In Proc. of IEEE TRANSACTIONS ON MEDICAL IMAGING, VOL. 25, NO. 12, DECEMBER 2006.
3. JabaSheela L and Dr.V.Shanthi, An Approach for Discretization and Feature Selection Of Continuous-Valued Attributes in Medical Images for Classification Learning, International Journal of Computer Theory and Engineering, Vol. 1, No.2, June 2009.
4. V.Krishnaiah, Dr.G.Narsimha, Dr.N.Subhash Chandra. 2013, Diagnosis of Lung Cancer Prediction System Using Data Mining Classification Techniques, International Journal of Computer Science and Information Technologies, Vol. 4 (1), 2013, 39-45.
5. S. Swensen, et al., CT screening for lung cancer: five-year prospective experience, Radiology, vol. 235, no. 235, pp. 259-265, APR. 2005.
6. P. Bach, et al., Computed tomography screening and lung cancer outcomes, J. Amer. Med. Assoc., vol. 297, no. 9, pp. 953-961, Mar. 2007.
7. G. Bepler, et al., "RRM1 and PTEN as prognostic parameters for overall and disease-free survival in patients with 55 Detection of lung cancer tumour in its early stages using image processing techniques non-small-cell lung cancer," J. Clin. Oncol., vol 22, no. 10, pp. 1878-1885, May. 2004.
8. C. Poleri, et al., "Risk of recurrence in patients with surgically resected stage I non-small cell lung carcinoma: histopathologic and immunohistochemical analysis," Chest vol. 123, no. 6, pp. 1858-1867, Jun. 2003.
9. G. Loannidis, et al., "How close are we to customizing chemotherapy in early non-small-cell lung cancer," Ther. Adv. Med. Oncol., vol. 3, no. 4, pp. 185-205, Jul. 2011.
10. E. P. Diamandis, "Mass spectrometry as a diagnostic and a cancer biomarker discovery tool: opportunities and potential limitations," Mol Cell Proteomics, vol. 3, no. 4, pp. 367-378, Apr. 2004. H. Li, et al., "Cost-effectiveness of a novel molecular test for cytologically indeterminate thyroid nodules,"

BIOGRAPHIES



Kadam Suraj, D. Y. Patil College of Engineering, Ambi, Pune, Maharashtra, India.



Dongare Vijaykumar, D. Y. Patil College of Engineering, Ambi, Pune, Maharashtra, India.



Dhalpe Somnath, D. Y. Patil
College of Engineering, Ambi,
Pune, Maharashtra, India.



Bakhale Tushar, D. Y. Patil
College of Engineering, Ambi,
Pune, Maharashtra, India.