iJRASET

International Journal For Research in Applied Science and Engineering Technology

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 2 **Issue:** Issue- III **Month of publication:** November 2014

DOI:

International Journal for Research in Applied Science & Engineering Technology (IJRASET)

Classification of lung tumour Using Geometrical and Texture Features of Chest X-ray Images

Dr. P.V. Ramaraju[1], Satti Praveen[2]
Department of Electronics and Communication
SRKR Engineering College
Andhra Pradesh, INDIA

Abstract: Early detection is the most promising way to enhance a patient's chance for survival of lung cancer. One of the most important tasks in medical image analysis is to detect the absence or presence of disease in an image, without having precise delineations of pathology available for training. A computer algorithm for nodule detection in chest radiographs is presented. The algorithm consists of four main steps:

1. Image acquisition

2. Image pre-processing;

3. Nodule candidate detection;

4. Feature extraction.

Algorithm is applied on two main types of lung cancer images, like Small-Cell, Non-Small-Cell type and as well as on TB database. Total 75 images are used (25 from each category) during experiment to estimate geometrical and texture features. Active Shape Model (ASM) technique is used for lung field segmentation. Gary Level Co-occurrence Matrix (GLCM) technique is used to estimate texture features.
***Keywords*: Computer Aided Diagnosis (CAD), MATLAB, NCLC, SCLC DATABAES, benign, malignant.**

I. INTRODUCTION

Lung cancer is one of the most common and deadly diseases in the world. The prognosis and the cure of lung cancer depend highly on the early detection and treatment of small and localized tumours. The 5-year patient survival rate is approximately 40% when lung cancer is detected in the early stage. About 87% of lung cancers are thought to result from smoking or passive exposure to tobacco smoking. Physical characteristics of the nodules, such as rate of growth, pattern of calcification, type of margins are very important in the investigation of the solitary lung nodules. Every lung nodule grows in volume over time. However malignant nodules grow at an exponential rate, which is usually expressed as a tumour's doubling time. Malignant nodules have a doubling time of between 25 to 450 days, whereas the benign nodules are stable and have a doubling time more than 500 days. In addition to the rate of growth of the nodules, the pattern of the calcification is an important indicator of whether the nodule is benign or malignant. Nodules which are centrally or diffuse calcified are usually benign. Chest X-ray image has been used for detecting lung cancer for a long time. The early detection and diagnosis of pulmonary nodules in chest X-ray image are among the most challenging clinical tasks performed by radiologists. Some of these lesions may not be detected due to the fact that they may be camouflaged by the underlying anatomical structures, or the low quality of the images, or the subjective and variable decision criteria used by the radiologist. Computer-aided diagnosis (CAD) has been proven to be a very effective approach as assistant to radiologists for improving diagnostic accuracy. Numerous systems were reported for detecting lung nodules on chest X-ray images. However, the strong concern of almost all of them is that the false positives per image are too large. How to reduce the number of false positives while maintaining a high true positive detection rate is the most important work in realizing a chest CAD system. Most of the proposed computer-aided diagnosis systems (CAD systems) adopt a two-step pattern recognition approach, which is a combination of a feature extraction process and a classification process using neural network classifier or statistical classifier. The performance of the classifier depends directly on the ability of characterization of candidate regions by the adopted features. Many kinds of features have been proposed for discriminating between normal tissues and abnormal ones. However, there have been a few researches on comparing the effectiveness of those features. The purpose of our research is to find the optimal feature set from the available database for the classification.

II. METHODOLOGY

The lung cancers are classified as Small-Cell type of lung cancer (SCLC), and Non-Small-Cell type of lung cancer (NSCLC) (Fig.1). Usually Small-Cell type of lung cancer arises at alveolar level or at terminal bronchial level, and seen to be more scattered in nature on X-ray. Non- Small-Cell type of lung cancer arises in the larger, more central bronchi; tends to spread locally; and metastasizes somewhat larger than the other patterns, but its rate of growth in its site of origin is usually more rapidly than that of other types.

Special Issue-3, November 2014
ISSN: 2321-9653
International Journal for Research in Applied Science & Engineering Technology (IJRASET)

Fig. 1 SCLC and NSCLC images

Feature selection is a very important step in organizing a classifier. Theoretical approach cannot be applied to determine the optimal combination of features, and the only way to select the optimal feature subset is to evaluate all possible combinations of the features. Moreover, sufficient numbers of test materials are necessary to evaluate the performance of each feature combination. It means that the number of combinations and the total amount of computation time become impractically huge. Therefore Jun Wei, Yoshihiro Hagihara, Akinobu Shimizu, and Hidefumi Kobatake accepted heuristic algorithms such as a genetic algorithm, a forward stepwise and a backward stepwise selection technique to decide the optimal feature set.

A. Geometric Features

Spread-ness, circularity, area, equivalent radius, distance from the candidate point to the pulmonary hilum, and flatness are some of the geometric features. Such geometric features are calculated from the binary suspicious region (SR) using thresholding technique.

B. Texture or Contrast Features

Generally, tumor region is brighter than its background on X-ray image. So, the contrast information can be used as features. Contrast features are again classified under two categories, first order statistic and second order statistic. In this work, such ten kinds of features are calculated from SR regions.

1. *First-Order Statistics features*

First-order statistic features are calculated from histograms of the grey-scale values. The histograms are obtained from filtered images. Features calculated from each histogram include avg. gray level, standard deviation, contrast, skew ness, kurtosis, and entropy.

2. *Second-Order Statistics features*

Co-occurrence matrix method has been adopted to extract features of second-order statistics. They are obtained by using Haralick transformation. Co-occurrence matrices are obtained from the inner and the outer regions of each SR. Correlation, energy, homogeneity and contrast are the features computed by using co-occurrence matrix.

III. IMPLEMENTATION

The steps followed for analysis and feature extraction from lung cancer and TB X-ray images are described in this section. As an initial step, the images are obtained using image acquisition method and then the application of pre-processing algorithms, including size normalization and filtering of the image. The features, those are identified to be useful for diagnosis and analysis, require separation of the lung fields from the background. Lung field masks are prepared manually by segmenting the lung fields. As well as readily available masks developed by using Active Shape Model (ASM) technique are used to separate the lung fields. Thresholding along with region based segmentation techniques are used to segment the lung nodules (in case of NSCLC images) and cancerous portion (in case of SCLC images) from the separated lung field area. Next step deals with detail description and methodology adapted to estimate geometrical and texture features from the Postero-Anterior (PA) X-ray images.

The chest unit used for screening X-ray films is mobile KlinoskopH unit (Siemens, India make). Keeping the tube voltage equal to 150 kV, 500 mA at 2.2 mm Pb the images were printed on 14 X 17 centimeter film. Then these films were digitized with a high-resolution scanner (Scanjet 2400, HP India make).These films were collected from the private medical institutes. Lung cancer and TB images from the public database are also used in this study. Every image data is acquired with 256 gray levels (8 bits) and stored as JPEG (.jpg, .jpeg) data. Before extraction of the features from an image, it is necessary to pre-process the image to reduce irrelevant information or noise, and to enhance the image properties, which makes the feature measurement easier and more reliable. Scanned images are resized to a size of 512 X 512 pixels. Median filter is used to remove the noise or irrelevant information from the images. The purpose of the segmentation is to find suspicious regions within the lung fields. Segmentation of lung fields on PA chest radiographs has received considerable attention in the literature

Active Shape Model, a general technique for image segmentation has been developed by Cootes and Taylor and has been applied to various segmentation tasks in medical imaging Around 180 such lung field masks (separate for each right and left

Special Issue-3, November 2014
ISSN: 2321-9653
International Journal for Research in Applied Science & Engineering Technology (IJRASET)

lung field) are prepared using ASM technique. These masks (Fig. 2) are the part of JSRT (Japanese Society for Radiological Technology) public database and useful for lung field segmentation.

Fig. 2 Original X-ray image along with lung field masks prepared using ASM technique

Lung field masks (see Fig. 3) are also prepared manually by segmenting the lung fields. Manual segmentation is carried out by determining the peripheral lung field pixel co-ordinates with region based segmentation technique.

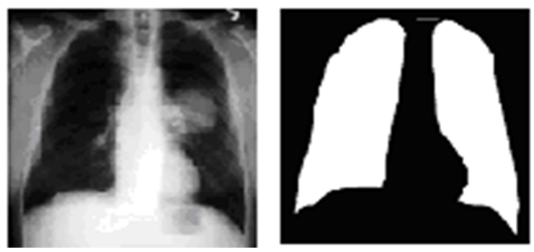


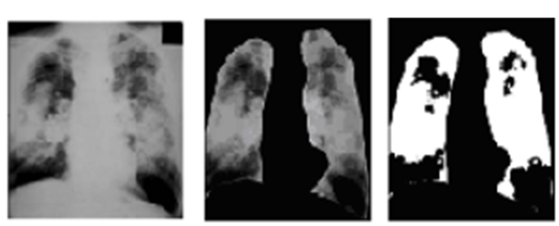
Fig. 3 Original X-ray image & manually segmented lung fields mask

Further the lung fields are separated from the background by multiplying the mask image with the filtered X-ray image. Above discussed technique is applied on SCLC, NSCLC and TB database images. Thresholding is applied on the separated lung fields' image to separate the nodule or infected portion. Valley point value between the two peaks of the histogram is selected as a threshold value. Region based segmentation techniques like region-growing (in case of NSCLC) and region-labeling (in case of SCLC) have been applied further to separate the nodules and infected portion.

(a) (b) (c)

Fig. 4 (a) segmented lung fields after multiplication, (b) Image after thresholding, (c) Separated nodule

Fig. 4 depicts the resultant images of the nodule segmentation (NSCLC type) technique. Resultant images for the SCLC type of image segmentation are included in Fig. 6



(a) (b) (c)

Fig. 5 (a) SCLC original image, (b) Separated lung fields, (c) Separated cancerous portion

Bit Quads technique devised by Gray is used to extract geometrical features like *area* and *perimeter*. Distance is a real valued function $d\ \{(j1, k1), (j2, k2)\}$ of two image points $(j1, k1)$ and $(j2, k2)$. The most common measures encountered in image analysis are the Euclidean distance, defined as:

$d_E = [(j1 - j2)2 + (k1 - k2)2]1/2$

In discrete images, the coordinate differences ($j1 - j2$) and ($k1 - k2$) are integers, but the Euclidean distance is usually not an integer. Here *diameter* is estimated using Euclidean distance. As per the literature survey it is seen that the growth of the malignant part (nodule here) is usually circular in nature, therefore roundness of the nodule is calculated using a simple equation:

$$I = 4 * pi * area / perimeter ^ 2$$

This metric value or roundness or circularity index or irregularity index (I) is equal to 1 only for circle and it is less than 1 for any other shapes. Here it has been assumed that, more the circularity of the object, the probability of that object being nodule is high.

Table –I

GEOMETRICAL FEATURES

Sr. No.	Features	value
1	Area	2815
2	Perimeter	226.85
3	Diameter	59.286
4	Irregularity Index (I)	0.69

The geometrical features estimated for the separated nodule, shown in Fig. 4 (c) are included in Table I.
An important approach for describing a region is to quantify its texture content. A frequently used approach used for texture analysis is based on statistical properties of the intensity histogram. One class of such measures is based on statistical moments. An expression for the nth moment about the mean is given by:

$$\mu n = \sum_{i=0}^{L-1} (zi - m)n\, p(zi)$$

where zi is a random variable indicating intensity levels in an image, $p(z)$ is the histogram of the intensity levels in a region, L is the possible intensity levels. A histogram component $p(zj)$, is an estimate of the probability of occurrence of intensity value zj , and the histogram may be viewed as an approximation of the probability density function (PDF). GLCM is the technique used to calculate PDF.

$$m = \sum_{i=0}^{L-1} zi\, p(zi)$$

Here m is the mean (average) intensity. These moments can be computed using MATLAB function *statmoments*, which is acting as a sub function in another MATLAB function known as *statxture*. This function is used to calculate first-order statistic texture features like mean, standard deviation, smoothness, third moment, uniformity and entropy.
A measure of average contrast or the standard deviation can be calculated by using following equation, where $\mu 2\ (z)$ is the second moment.

$$\sigma = \sqrt{\mu 2\ (z)} = \sqrt{\sigma 2}$$

Smoothness measures the relative smoothness of intensity in a region. R is 0 for a region of constant intensity and approaches 1 for region with large excursions in the values of its intensity levels. Smoothness is calculated by using the following equation.

$$R = 1 – 1/ (1 + \sigma 2)$$

Skew-ness of the histogram is also known as third moment. This measure is 0 for symmetric histograms, positive by histograms skewed to the right (about the mean) and negative for histograms skewed to the left. For smooth images this value comes to be negative. Following equation is used to calculate third moment.

$$\mu 3 = \sum_{i=0}^{L-1} (z\, i – m)3\, p(z\, i)$$

When all gray levels are equal, uniformity measures maximum and goes on decreasing from there for the inequality.

$$U = \sum_{i=0}^{L-1} p2\, (z\, i)$$

Entropy is nothing but the measure of randomness, given by the following equation.

$$e = - \sum_{i=0}^{L-1} p(z\, i) \log 2\, p(z\, i)$$

Special Issue-3, November 2014
ISSN: 2321-9653
International Journal for Research in Applied Science & Engineering Technology (IJRASET)

i = 0

The GLCM functions characterize the texture of an image by calculating how often pairs of pixel with specific values and in a specified spatial relationship occur in an image. However, a single GLCM might not be enough to describe the textural features of the input image. For example, a single horizontal offset might not be sensitive to texture with a vertical orientation. Therefore it is essential to generate multiple GLCMs with different offset values or at different angles. MATLAB function *graycomatrix* is used to generate such multiple GLCMs. Using multiple GLCMs, second-order statistic features like, Contrast Correlation, Energy, and Homogeneity are estimated.

IV. RESULTS

Results shown below are taken for 7 patients Shown below. There will be three types of classifications based on area of the tumour present in the resulted image. They are
1. Normal (i.e; no tumour present in the organ)
2. Benign (less dangerous)
3. Malignant (very dangerous)

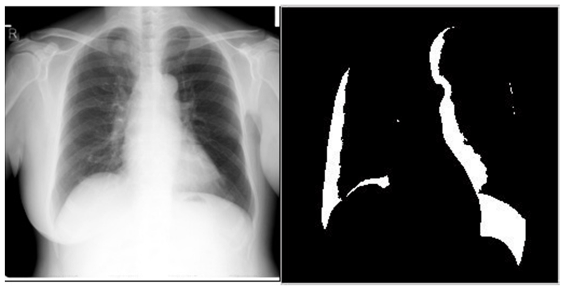


Fig. 6 showing x-ray of patient and detected benign tumour

Fig 7 showing x-ray and on tumour in it.

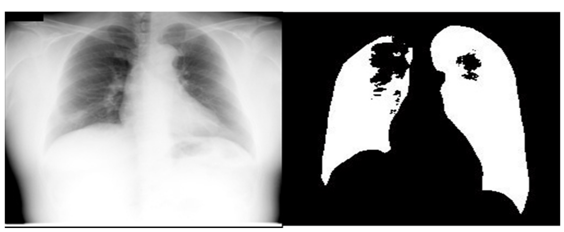


Fig 8 showing x-ray of patient and detected malignant tumour

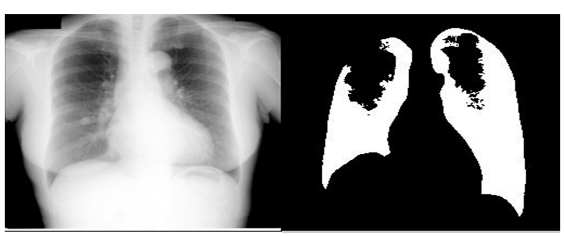


Fig 9 showing x-ray of patient and detected malignant tumour

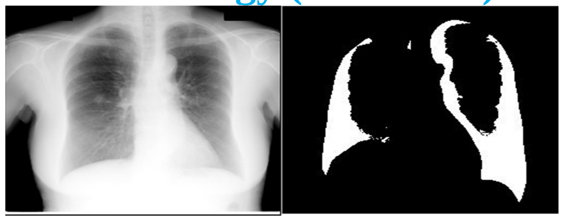


Fig 10 showing x-ray of patient and detected malignant tumour

Fig 11 showing x-ray of patient and detected benign tumour

Fig 12 showing x-ray of patient and detected benign tumour

V. CONCLUSION

Results of SCLC images show that, the area values are quite larger than the other types, like NSCLC. This is because of the scattered nature of the infected area. Irregularity index is always closer to ‘1’ for the circular objects, The segmented portion in case of NSCLC images are having irregularity index closer to ‘1’, which concludes that the segmented portion is a malignant portion or a lung nodule. Referring 1st and 2nd order statistic features, it is concluded that the uniformity and energy values are almost identical in all cases.

REFERENCES

[1] Lung Cancer Detection Using Image Processing Techniques by Mokhled S. AL-TARAWNEH, Issue 20, January-June 2012
[2] An efficient visualization and segmentation of lung ct scan images for early diagnosis of cancer by preeti aggarwal, h.k. sardana, renu vig
[3] Bram van Ginneken, Bart M. ter Haar Romeny, and Max A.Viergever, Member IEEE, “Computer-Aided Diagnosis in Chest Radiography: A Survey”, IEEE Trans., Med. Imag, vol. 20, pp. 1228 - 1230, Dec. 2001.
[4] H. Suzuki, N. Inaoka, H. Takabatake, M. Mori, H. Natori, and A.Suzuki, “An experiment system for detecting lung nodules by chest x-ray image processing”, SPIE. Biomedical Image Processing II, vol. 1450, pp. 99 - 107, 1991.
[5] J. Lin, S. B. Lo, A. Hasegawa, M. T. Freedman, and S. K. Mun, “Reduction of false positives in lung nodule detection using a twolevel neural classification”, IEEE Trans. on Med. Imag., vol. 15, pp. 206 - 217, 1996.
[6] Xin-Wei, Xu, K. Doi, T. Kobayashi, H. MacMahon, and M. L. Giger, “Development of an improved CAD scheme for automated detection of lung nodules in digital chest images”, Med. Phys., vol. 24, no. 9, pp. 1395 – 1403, 1997.
[7] H. Yoshida and K. Doi, “Computerized detection of pulmonary nodules in chest radiographs: reduction of false positives based onsymmetry between left and right lungs”, Proc. SPIE in Medical Imaging 2000, pp. 97 – 102, 2000.
[8] Y.Wu, M.L.Giger, K.Doi, C.J.Vyboorny, R.A.Schmidt, and C.E.Metz, “Artificial Neural Networks in Mammography: Application to Decision Making in the Diagnosis of Breast Cancer”, Radiology, vol. 187, pp.81- 87, 1993.
[9] M.Kupinski, M.L.Giger, P.Lu, and Z.Huo, “Computerized detection of Mammographic lesions: Performance of artificial neural network with enhanced feature extraction”, Proc. of SPIE, vol. 2434, pp. 598 – 605, 1995.

Dr. P.V. Rama Raju is a Professor at the Department of Electronics and Communication Engineering. SRKR engineering college AP, India. His research interests include Biomedical-signal Processing, VLSI Design And Microwave Anechoic Chambers Design. He is author ofseveral research studies published in national and international journals and conference proceedings

Satti Praveen is an M.tech student at the department of electronics and communication engineering. SRKR engineering college. AP. India. He received his B.tech degree from swarnandhra institute of engineering & technology in 2012