October 2013

SOFT TOOL DEVELOPEMENT FOR CHARACTERIZATION OF LUNG NODULE FROM CHEST X-RAY IMAGE

MANOJ R. TARAMBALE
Electrical department, M M's College of Engineering, Pune-52, University of Pune, Pune, India,
manoj_tarambale@yahoo.com

NITIN S. LINGAYAT
Electrical Department, BATU's Institute of Petrochemical Engineering, Lonere, Dist- Raigad, India,
nslingayat@yahoo.com

Follow this and additional works at: https://www.interscience.in/ijipvs

Part of the Robotics Commons, Signal Processing Commons, and the Systems and Communications Commons

SOFT TOOL DEVELOPEMENT FOR CHARACTERIZATION OF LUNG NODULE FROM CHEST X-RAY IMAGE

MANOJ R. TARAMBALE[1], NITIN S. LINGAYAT[2]

[1]Electrical department, M M's College of Engineering, Pune-52, University of Pune, Pune, India
[2]Electrical Department, BATU's Institute of Petrochemical Engineering, Lonere, Dist- Raigad, India
E-mail: manoj_tarambale@ yahoo.com, nslingayat@yahoo.com

Abstract- CAD (Computer Aided Diagnosis) is a concept established by taking into account equally the roles of physician and computer to comment on disease. With CAD system, the performance given by computer does not have to be comparable to or better than that by physician, but need to be complementary to that by physician. To reduce the false positive and false negative diagnosis in determining whether the tumor is malignant or benign, doctors are taking help of CAD system. CAD using image processing technique has become one of the major research subjects in medical imaging and diagnostic radiology. Radiologist uses the CAD system output as a 'second opinion' and make the final decision to conform the disease. In present paper, suspicious area is segmented from the chest X-ray image after doing pre-processing of the image. Characterization of lung nodule means describing distinctive essential features. By doing segmentation, the features are extracted from the segmented region. All features are calculated from their respective mathematical formulas. Calculated features values vary according to arrangement of pixels in the image. Further this information can be used as an input to the CAD system for determining whether the segmented suspicious tumor area is malignant or benign. The proposed system will not replace the doctor's role in detection of cancer but it will help doctor to take correct decision in short time with accuracy. It will act as second opinion before conformation of cancer.

Keywords- CAD; Cancer; Convex area; Malignant; Tumor;

I.   INTRODUCTION

On World Cancer Day, WHO survey indicates that more than half of all countries worldwide are still struggling to prevent cancer. These countries are failed to provide proper treatment and chronic care to cancer patients. This means, currently many of these countries do not have a functional cancer control plan that includes prevention, early detection, treatment and care. There is an urgent need to help countries to reduce cancer deaths and provide appropriate long term treatment and care to avoid human suffering and protect countries' social and economic development.

Today, cancer is one of the most formidable health problem faced by mankind. Cancer is one of the leading causes of death worldwide. 7.6 million people died from cancer worldwide in 2008 and every year almost 13 million cancer cases are newly diagnosed. More than two-thirds of the new cancer cases and deaths occur in developing countries where cancer incidences continue to increase at alarming rates. Research suggests that currently one third of all cancer deaths are due to modifiable risks including tobacco use, obesity, harmful use of alcohol and infections. If detected early many types of cancer can be successfully cured.

Lung cancer is an uncontrolled growth of abnormal cells in one or both of the lungs. Cancerous tumors grow out of control and can invade, replace, and destroy normal cells near the tumor. In some cases, cancer cells spread (or metastasize) to other areas of the body. Lung cancer generally takes many years to develop even though damage to the lungs may occur shortly after a person is exposed to cancer causing substances. Cancer is curable if it is detected in its early stage. Tumor may be benign when they are slow growing and localized without causing much difficulty to the host or malignant when they proliferate rapidly, spread throughout the body and may eventually cause death of the host. The common term used for all malignant tumors is cancer [1].

The noble cause for cancer will be because of changes in life style, changes in environment and increase in life expectancy. Research suggests that one-third of cancer deaths can be avoided through prevention, and another third through early detection and treatment. Despite having proven interventions for prevention, early diagnosis, treatment, and care for cancer and the medicines, technologies, and services are not widely available in low- and middle-income countries [2].

For early detection of cancer, radiologists are taking help of medical imaging (X-ray, CTs, MRIs and PETs). In medical imaging, the accurate diagnosis and / or assessment of a disease depends on both image acquisition and image interpretation. The image interpretation process, however, has only recently begun to benefit from computer technology. Most interpretations of medical images are performed by radiologists but image interpretation by humans is limited due to the presence of irrelevant information (noise etc.) in the image. Presence of dust particles, moisture, wrong setting of medical image machine lens, presence of hair or tattoos and some other type

of abnormality (burn skin, cuts etc.) present on the skin of the patients introduce irrelevant information in the medical image. Sometime, bad illumination, very low or very high contrast in the image, circuit noise, high temperature etc also affects the quality of the image and these factors are producing problem in interpretation of image by radiologist. Radiologist has to analysis large sets of medical images leads to a possibility of false positive or false negative errors in the interpretation.

To reduce irrelevant information from the medical image and to minimize human errors, radiologists / doctors are taking help of computer aided diagnosis system (CAD). Radiologist uses the output from a CAD system as a second opinion in detecting / analyzing lesions, assessing extent of disease, and improving the accuracy and consistency of radiological diagnosis to reduce the rate of false negative cases. The typical architecture of a CAD system shown in figure 1 includes image acquisition, enhancement, segmentation, features extraction and classification.
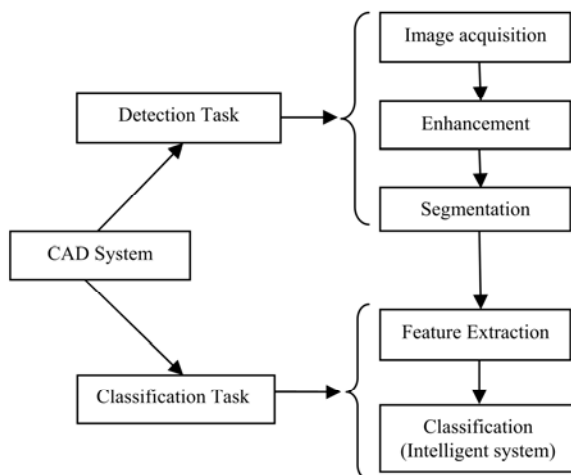
Fig.1. CAD system

The general approach of CAD system is to find the location of a suspicious region and also to determine / estimate the probability of a disease.

In first step, hard copy (X-ray) of images is converted into the softcopy with the help of high resolution scanner. These scanned images are stored in the storage drive for further analysis and processing. These scanned images are imported into the CAD system. CAD system performs two main tasks - detection task and classification task. Further detection task include three processes namely image acquisition, enhancement and segmentation. In image acquisition all the scanned images are resized to the standard size, as the CAD system algorithm are developed for specific size image. Enhancement process is required for the image before applying any algorithm on it. During the scanning and storing

process, lot of irrelevant information (noise) is added. For removing irrelevant information from the image and for enhancing the X-ray image quality, low pass filter, high pass filter and point processing techniques are employed. After enhancing the image, segmentation is done to extracted interested area from the rest of the image.

Second task of CAD system is classification task. In this task features / parameters / properties extraction and classification is done. In feature extraction mathematical, textural and geometrical properties are calculated from the segmented region. Further these properties are used for classification purpose. With the help of intelligent system and feature extracted properties, CAD system decide whether the given image contain some suspicious part or not. Present paper deals with the features / parameters / properties extraction, how they are calculated and how they are useful in analyzing of lung nodule.

In poor and developing countries, where people cannot afford expensive medical imaging (CTs, MRIs, and PETs) for the early detection of the cancer, in such countries, people's first preference is X-ray imaging only. As X-ray is less costly, it required less time to generate X-ray images and there are hardly any side effects to having an X-ray. For such peoples, it is required to develop CAD system which can operate on X-ray image efficiently and accurately which helps the radiologist / doctor in interpretation of X-ray image and detecting disease in its early stage.

Using feature extracted data, intelligent system will be trained so that for any new digital image is to be analyzed, intelligent system (CAD) will be capable to detect / classify lung nodule [3].

II. ENHANCEMENT

Enhancing of the image is done before doing segmentation. The quality of the input image plays a key role in image-based measurements. Image enhancement is a preprocessing task of paramount importance because it can significantly reduce the uncertainties in subsequent measurement procedures, such as parameter estimation [4].

Image enhancement includes noise removing and enhancing relevant detail in the image. This step is required because as image contain lot of irrelevant information. The possibility of false negative and false positive error increases if the irrelevant information is not removed or enhancement is not done.

Gaussian noise and salt and pepper noise are the type of noise that is commonly found in the X-ray image [4]. For removing noise, low pass filtering is used.

Low pass filtering removes noise but on the hand it produces blurring. High boost filtering is used to sharpening the image. Histogram equalization [5] and contrast streching [6] is the image enhancement technique that commonly used for improving the brightness and contrast problem in medical images.

III. SEGMENTATION

Segmentation is the process of partitioning a digital image into multiple segments i.e. sets of pixels (also known as super pixels) [19]. 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 [20]. 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 visual characteristics.

For segmentation, thresholding with labeling algorithm is used. Thresholding is useful in discriminating foreground from the background by selecting an adequate threshold value; the gray level image is converted to binary image [7], [16]. Labeling algorithm, label each white pixel area in the binary image. By selecting label one can extract interested area.

IV. FEATURE EXTRACTION

The goal in image analysis is to extract information useful for solving application based problems [8]. Feature extraction provides some important parameter which helps in diagnosing suspicious area from the chest X-ray image. All features are calculated from their respective formulas.

a) Area
Area of the segmented tumor is computed by counting the number of pixels which have the value '1' in the image array. The area (A) in the object is the just the count of the ones in the image array. For computing area, binary image is used. It is measure in terms of pixels [9], [10].

$$A = n[1] \tag{1}$$

Where, n [ ] represents the count of number of the patterns within the parenthesis.

b) Perimeter
Perimeter property is calculating the distance between each adjoining pair of pixels around the border of the region. If the image contains discontinuous regions, then it returns unexpected results. Perimeter is calculated by counting the pixels contained only in the boundary.

c) Irregularity Index
Lung cancer is characterized partially by the irregularity in its tumor border [11]. For this analysis, the irregularities in the tumor are computed by an index:

$$I = \frac{4\pi \times Area}{(Perimeter)^2} \tag{2}$$

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

d) Convex Area
Convex hull is the smallest convex polygon that can contain the region in binary image. The convex hull, with all pixels within the hull filled is known as convex image. Convex area is the number of pixels in convex Image [9].

e) Convex Perimeter
It is calculated by calculating the number of pixel at the boundary of the region in the convex image. This property is calculating the distance between each adjoining pair of pixels around the border of the region. If the image contains discontinuous regions, then it returns unexpected results [9].

f) Solidity
It gives the proportion of the pixels in the convex hull that are also in the region. It is defined as the ratio of the area of the original image to the area of the convex hull. For very irregular surface, solidity comes near to 0 values [9].

$$Solidity = Area / Convex\ Area \tag{3}$$

g) Convexity
It gives the proportion of the perimeter of the convex hull to the perimeter of the binary image. For very irregular surface convexity value comes nearer to zero [9].

$$Convexity = Convex\ Perimeter / Perimeter \tag{4}$$

h) Deficiency
It gives the sum of the difference between the convex hull image and the original image [6].

$$\sum_{x,y \in R} | convex\ hull\ (x, y) - object\ (x, y) | \tag{5}$$

Where x and y are the coordinates of the pixel in the image.

i) Equiv diameter
Scalar that specifies the diameter of a circle with the same area as the region [9], [10]. Computed by formula

$$E_{diameter} = \sqrt{\frac{4 \times area}{\pi}} \qquad (6)$$

j) Kurtosis

It is a measure of whether the data are peaked or flat relative to a normal distribution. That is data set with high kurtosis tend to have a distinct peak near the mean decline rapidly and have heavy tail. Data set with low kurtosis tend to have a flat top near the mean rather than a sharp peak. For standard normal distribution is three. Kurtosis of less than three indicates a low peak with a fat midrange on either side; this is referred to as platykurtic. Conversely, Kurtosis greater than three indicates a sharp/high peak with a thin midrange and fat tails; this is called leptokurtic. Therefore, put simply, Kurtosis describes how bunched around the center or spread at the endpoints a frequency distribution is.

$$Kurtosis = \frac{\sum_{i=1}^{N}(Y_i - \bar{Y})^4}{(N-1)s^4} \qquad (7)$$

Where Y_i is data, $\bar{Y}$ is the mean, 's' is the standard deviation and N is the number of data points.

k) Skewness

Skewness is a measure of symmetry, or more precisely, the lack of symmetry. A distribution, or data set, is symmetric if it looks the same to the left and right of the center point [14].

$$Skweness = \frac{\sum_{i=1}^{N}(Y_i - \bar{Y})^3}{(N-1)s^3} \qquad (8)$$

Where Y_i is data, $\bar{Y}$ is the mean, 's' is the standard deviation and N is the number of data points. The skewness for a normal distribution is zero, and any symmetric data should have skewness near zero. Negative values for the skewness indicate data that are skewed left and positive values for the skewness indicate data that are skewed right. By skewed left, we mean that the left tail is long relative to the right tail. Similarly, skewed right means that the right tail is long relative to the left tail. Some measurements have a lower bound and are skewed right. For example, in reliability studies, failure times cannot be negative.

l) Statistical Properties of an Image:

1. Entropy

Entropy measures the randomness of a gray-level distribution. The Entropy is expected to be high if the gray levels are distributed randomly throughout the image [11]. It is given by

$$E = -\sum_{x}^{m}\sum_{y}^{n} P[x, y] \log P[x, y] \qquad (9)$$

Where 'x' and 'y' are the coordinates of the pixel in the image. 'M' and 'N' are the total number of pixel in the row and column.

2. Mean

Mean provides the mean of the gray levels in the image. The Mean is expected to be large if the sum of the gray levels of the image is high [11]. It is given by –

$$\mu = \frac{1}{N \times M}\sum_{x=0}^{M}\sum_{y=0}^{N} P(x, y) \qquad (10)$$

3. Variance

Variance tells us, how spread out the distribution of gray levels is. The Variance is expected to be large if the gray levels of the image are spread out greatly [11]. It is given by

$$\sigma^2 = \frac{1}{M \times N}\sum_{x=0}^{M-1}\sum_{y=0}^{N-1}(p(x, y) - \mu)^2 \qquad (11)$$

4. Standard Deviation

Standard Deviation shows much 'variation' or 'dispersion' exists from the average [11].

$$Std = \sqrt{\sigma^2} \qquad (12)$$

V. RESULT AND DISCUSSION

Feature / parameter extraction is a very important part of CAD system. Before extracting features / parameter, image is enhanced and irrelevant information is removed. Images obtain from public JSRT database and the images collected from nearby local hospitals are used for testing purpose. All images are scanned and stored in the storage device for further processing. Pre-processing techniques are applied to remove any irrelevant information and to enhance relevant information. Suspicious area is segmented from the image by using thresholding and labeling algorithm. Thresholding with threshold is used to separate suspicious area from the image. Features are calculated from the segmented area. Calculated features, give relevant information about the suspicious region and helps doctor in diagnosing disease. Further these features can be given as an input to the CAD system for classification task.

For implementing the algorithm, Digital Image Processing and GUI toolboxes of MATLAB software are used. Following images Fig.2 (a) and Fig3 (a) are taken from lung cancer database which are certified for malignancy and benign by radiologist. Fig.2(b) and Fig.3(b) are the enlarge segmented lung tumor from the original X-ray image by the use of software, which is developed by using GUI and image processing toolbox in MATLAB.

Fig.2 (a) Malignant image   Fig.2 (b) Enlarge segmented image
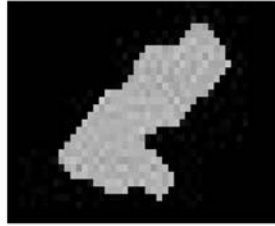
Fig. 3(a) Benign image   Fig.3 (b) Enlarge segmented image

The features which are calculated from the image fig.2 (a) and fig.3 (a) are listed in a table I and table II.

TABLE I.   PARAMETER COMPARISON TABLE FOR MALIGNANT AND BENIGN

PARAMETER	MALIGNANT	BENIGN
Area	857	330
Perimeter	191.4386	90.8112
Irregularity index	0.2939	0.503
Convex Area	1191	409
Convex perimeter	136.8528	81.7401
Solidity	0.7196	0.8068
Convexity	0.7149	0.9001
Deficiency	334	79
Equiv diameter	33.0328	20.498
Kurtosis	147.5456	229.8294
Skewness	11.8386	14.8822

TABLE II.   STATISTICAL PARAMETER COMPARISON TABLE FOR MALIGNANT AND BENIGN

PARAMETER	MALIGNANT	BENIGN
Entropy	0.0103	0.0045
Mean	0.2276	0.0877
Variance	5.57E+07	2.15E+07
Standard Deviatio	7.46E+03	4.63E+03

VI. CONCLUSION

Malignant tumor is more irregular as compare to benign. Because of its irregularity, its irregular index value is low. As compare to malignant tumor, benign tumor is less irregular. Convexity and deficiency values are high for benign tumor. Statistical properties are giving distinguishable values between benign and malignant nodule. Calculated values vary according to the arrangement and the values of the pixels in the image. CAD system will not replace the doctor role but it will help doctor, by providing second opinion regarding suspicious area. This method is less costly and this method will provide the second opinion to the doctor before conformation of cancer by clinical method (pathological and surgical).

ACKNOWLEDGMENT

REFERENCES

[1] Disha Sharma, and Gagandeep Jindal, "Computer aided diagnosis system for detection of lung cancer in CT scan images," International Journal of Computer and Electrical Engineering, vol. 3, no. 5, pp.714-718, Oct.2011.

[2] Dignam JJ, Huang L, Ries L, Reichman M, and Feuer E, "Estimating breast cancer statistic and other-cause mortality in clinical trial and population-based cancer registry cohorts, " vol. 115, pp. 5272–5283, Nov. 2009.

[3] S.N. Deepa, and B. Aruna Devi, "A survey on artificial intelligence approaches for medical image classification," Indian Journal of Science and Technology, vol. 4, no. 11, pp.1583-1595, Nov 2011,

[4] Fabrizio Russo, "An image-enhancement system based on noise estimation," IEEE Transactions on Instrumentation and Measurement, vol. 56, no. 4, pp. 1435 – 1442, AUG. 2007.

[5] Robert H. Sherrier and G. A. Johnson, "Regionally adaptive histogram equalization of the chest, " IEEE Transactions On Medical Imaging, vol. 6, no. 1, pp. 1 – 7, Nov. 2007.

[6] J. Damodhar, Dr. D. Sathyanarayana , "Implementation of contrast stretching algorithm on lung cancer effected images for analysis, " International Journal of Advances in Electronics Engineering, vol. 1, no.1, pp. 267-272,June 2012.

[7] Salem Saleh Al-amri1, N.V. Kalyankar and Khamitkar S.D,"Image segmentation by using thershod techniques," Journal of Computing, volu. 2, issue 5, pp. 83-86, MAY. 2010.

[8] Dr.M.Mohammed Sathik, "Feature extraction on colored x-ray images by bitplane slicing technique," International Journal of Engineering Science and Technology , vol. 2(7), pp. 2820-2824, 2010,.

[9] Rapheal C.Gonzalez and Richard E. Woods, Digital Image Processing, 3rd ed. Delhi, India: Pearson Education, 2009, ch. 10, pp. 711-783.

[10] A.K. Jain, Fundamental of Digital Image Processing, 1st ed. Prentice Hall, Englewood Cliffs, 1989, ch. 7, pp. 241-250.

[11] S.A. Patil and V.R.Udpi, "Chest x-ray features extraction for lung cancer classification", Journal of Scientific and Industrial Research, vol. 69, pp 271-277, April 2010.

[12] Rafael C. Gozalez, and Richard E. Woods, Digital Image Processing Using Matlab, 2nd ed. Gatesmark, USA, 2002, ch. 12, pp. 642-654.

[13] Pierre Jannin, J.Michael Fitzpatrick, Dave J. Hawkes, Xavier Pennec, Ramin Shahidi, and Michael W.Vannier, "Validation of medical image processing in image-guided therapy,"IEEE Transactions on Medical Imaging, vol. 21, pp. 1445-1449, 2002.

[14] K.A.G. Udeshani, R.G.N. Meegama and T.G.I. Fernando,"Statistical feature-based neural network approach for the detection of lung cancer in chest x-ray images," International Journal of Image Processing (IJIP), vol .5, issue. 4, pp. 425-437, 2011.

[15] Maryellen L. Giger, Samuel G. Armato and Nico Karssemeijer, "Guest editorial computer-aided diagnosis in medical imaging," IEEE Transactions on medical imaging, vol. 20, no. 12, Dec. 2001

[16] Anita Chaudhary and Sonit Sukhraj Singh, " Lung cancer detection using digital image processing," IJREAS, vol. 2, issue 2, pp. 1351-1359, Feb. 2012.

[17] P. Bountris, and N. Apostolou, "Advanced image analysis tools developmentfor the early stage bronchial cancer detection," World Academy of Science, Engineering and Technology, pp.151-156, 2007.

[18] Patil S. A., "Texture analysis of TB x-ray images using image processing technique," Journal of Bio-Medical and Bioengineering, vol. 3, issue 1, pp. 53-56, March 2012.

[19] PunamThakare, "A study of image segmentation and edge detection techniques", International Journal on Computer Science and Engineering (IJCSE), vol. 3, no. 2, pp. 899-904, Feb. 2011.

[20] Gautam Appasaheb Kudale and Mahesh D. Pawar (2010). Study and analysis of various edge detection methods for X-ray images. International Journal of Computer Science and Application [online]. pp.15-19.Available: http://www.sinhgad.edu/IJCSA-2012/pdfpapers/1_4.pdf

❖ ❖ ❖