Published by :
http://www.ijert.org

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 5 Issue 07, July-2016

Identification of Cancerous Nodules in Lung CT Scan Images

Prof. Mrs. Ratna Nitin Patil
Computer Science and Engineering
Vishwakarma Institute of Technology
Pune, India

Miss. Nayan Rajesh Yengul
Computer Science and Engineering
Vishwakarma Institute of Technology
Pune, India

***Abstract*— This paper aims to detect cancerous nodule in CT scan image of Lung by using image processing and machine learning algorithms. Lung cancer is the leading cause of death now a day in U.S. Early detection facilitates early treatment and increases patient's chances of survival. This work aims in identifying whether the detected nodule is malignant or benign, using MATLAB for pre-processing on CT images and WEKA/R-programming for classification purpose.**

Keywords—Lung cancer; Image processing; Machine learning; MATLAB; CT images

I. INTRODUCTION

Lung cancer is a serious health problem in human body. The mortality rate of lung cancer is much more as compared to other types of cancers. The growth of lung cancer nodule starts in trachea (windpipe) and often spreads towards the centre of chest. Doctors do biopsy in detection of cancer. Risk factor in biopsy involves swelling near puncture side, fever, pneumonia etc. Therefore earlier detection of lung cancer is necessary. The earlier detection increases the chances of successful treatment. To detect lung cancer in early stage requires a CAD (Computerized Aided Diagnosis) system. Low Dose Helical Computed Tomography (CT) scan images are used in diagnosis, which gives better clarity and less distortion than MRI or X-ray images.

There can be two types of nodules found in lung CT image benign and malignant. Benign is a non-cancerous and malignant is cancerous. Cancerous nodule spreads in other parts of the body. Malignant nodule differs from benign in size, shape as well as texture.

This sytstem acquires input image in the form of CT scan images. Before cropping and segmentation ,image is enhanced in this system by using power law transformations.Gamma correction is an image enhancement technique which is used to correct power law response,by varying different gamma values. Lung portion was extrated by applying cropping and segmentation techniques. Segmentation is done for means of removing unwanted part from image.The connected component labelling technique typically used to locate object in an image.These objects are then analyzed for further classification. Features of every object is calculated. The values of extracted features are useful for classification purpose.

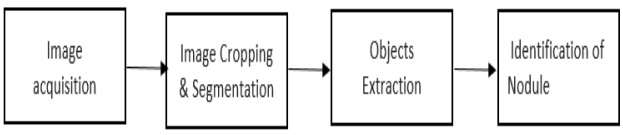

Fig.1.Stages of nodule detection

II. METHODOLOGY

A. *Image acquisition and pre-processing*

For detecting lung cancer CT images used as a input. CT scan images are in .dcm file format and can be viewed in DICOM (Digital Imaging and Communications in medicine) viewer. Spiral CT is advantageous as small nodules are not missed between slices as it may happen in older non spiral machines. It also increases the detection rate of nodule <5mm in diameter [4].

Image pre-processing involves two stages: image smoothing and image enhancement.

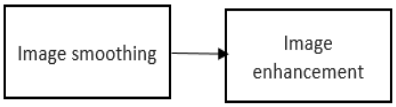

Fig. 2 Image Pre-processing

Quality of CT images is affected by the artifacts due to contrast variations and noise [2].To remove such noise median filtering was applied. It reduces blurring of edges. Generally median filtering is used to remove salt & pepper noise or impulsive noise. Median filter is advantageous as compared to other Denoising techniques, as it does not blur the edges while removing noise. Median filter replaces the value of a pixel by median of grey levels in the neighbourhood that pixel. In order to apply median filtering at a point in an image, we first sort the values of the pixel and its neighbours, determine median and assign this value to that pixel.

Image enhancement is a technique that improves the quality of the image for specific application. It was applied to adjust contrast of CT image. Gamma correction is one of the image enhancement techniques applicable when histogram of an image reveals that image is under-exposed.

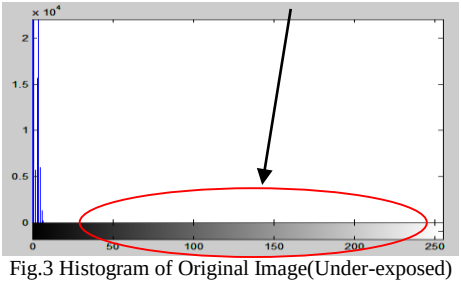

Fig.3 Histogram of Original Image(Under-exposed)

In gamma correction technique by varying different gamma values different transformation curves we get. For specific application specific gamma value is applied to get better enhanced image.

(This work is licensed under a Creative Commons Attribution 4.0 International License.)

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 5 Issue 07, July-2016

B. . Image Cropping and segmentation

Image cropping was done after proceeding segmentation. Segmented CT scan image was cropped into rectangle.

Image segmentation is process of partitioning connected set of pixels. Segmentation is a process which divides image into regions which contains more meaningful information. For any segmentation approach thresholding is first step. Thresholding algorithm depends on type of image and histogram of that image. There are different thresholding techniques available for different types of images. In this algorithm we have used simple global threshold algorithm. The threshold value is calculated by histogram of an image. To covert an image into binary image threshold value is applicable. Otsu's thresholding method was used to compute global threshold value using MATLAB commands. Otsu's thresholding technique gives threshold value between 0 and 1. Here we get threshold value (T) = 0.3588.

Pixel having intensity values above this threshold are assign to 1 and pixels having intensity value below this threshold value are assigned 0. And in this way we get binary image.

By applying morphological operations and cropping (rectangular cropping) we extract Region of Interest (ROI) i.e. Lung part from binary image. Connected component labelling technique was used to compute number of objects present in cropped image. Then features of every object computed in a single image and stored it in excel sheet.

C. Feature Extraction

This stage plays vital role in nodule identification process. Finite numbers of objects were extracted from segmented image. The purpose of feature extraction is creation of more usable form of raw images [3].The characters of features are area, perimeter, eccentricity, MajorAxisLength, MinorAxisLength, Orientation, ConvexArea, FilledArea, EulerNumber, EquivDiameter, Solidity, and Extent. These are measured in scalars. Features are defined as follows:

1. Area: It is scalar value that gives actual number of overall nodule pixels in the extracted ROI.
2. Perimeter: It is a scalar value which gives actual number of outline of the nodule pixels. It is length of extracted ROI boundary.
3. Eccentricity: This metric value is also called as roundness or circularity or irregularity complex (I) equal to 1 only for circular shape and is less than 1 for any other shape. It is defined as:
   Eccentricity = Length of major axis/Length of minor axis.
4. MajorAxisLength: This is a scalar value that specifies the length (in pixels) of the major axis of the ellipse (nodule) that has the same normalized second central moments as the region.
5. MinorAxisLength: This is a scalar value that specifies the length (in pixels) of the minor axis of the ellipse (nodule) that has the same normalized second central moments as the region.
6. Orientation: This is a scalar value that specifies the angle between the x-axis and the major axis of the ellipse that has the same second-moments as the region. The value is in degrees, ranging from -90 to 90 degrees.
7. ConvexArea: This is a scalar value that specifies the number of pixels in 'ConvexImage'.
8. FilledArea: This is a scalar value that specifies the number of on pixels in FilledImage.
9. EulerNumber: This is a scalar value that specifies the number of objects in the region minus the number of holes in those objects. This property is supported only for 2-D label matrices. Regionprops uses 8-connectivity to compute the Euler number measurement.
10. EquivDiameter: This is a scalar value that specifies the diameter of a circle with the same area as the region. Computed as sqrt (4*Area/pi).
11. Solidity: This is a scalar value that specifying the proportion of the pixels in the convex hull that are also in the region. Computed as Area/ConvexArea.
12. Extent: This is a scalar value that specifies the ratio of pixels in the region to pixels in the total bounding box. Computed as the Area divided by the area of the bounding box.

D. Classification

The general term classification refers to prediction of class labels using training dataset and values of class attributes .Classification of lung CT images is require identifying presence of cancer. Also we may detect which stage of cancer it is. There are different classification methods can applied on dataset of extracted features from objects, like Support vector machines (SVM) [1], Cellular Automata (CA) [3], Fuzzy C means (FCM) clustering [2] ,template matching and Genetic Algorithm(GA) [5].

Support Vector machines are supervised learning models with associated learning algorithms that analyse data and recognize patterns, used for classification. From given set of training examples, each marked as belongings to one of the two categories, an SVM training algorithm builds a model that assigns new examples into one category or the other. Best hyper plane is the one that represents the largest separation or margin between the two classes [1].

In this study we had applied three classifiers to identify object: Random Forest, SMO (Sequential Minimal Optimization), Naïve Bayes Classifier.

Random Forest is an ensemble learning method for classification. Random Forest works as large decorelation of decision trees. It randomly divides training dataset along with random attributes. Then constructs decision tree for every subset. Unknown data points pass through all decision trees. Each decision tree vote for unknown data point and use majority votes to make decision.

SMO classifier is used to train SVM in WEKA tool. It was implemented by using LibSVM library in WEKA.

The Naive Bayes algorithm is a simple probabilistic classifier that calculates a set of probabilities by counting the frequency and combinations of values in a given data set. The algorithm uses Bayes theorem and assumes all attributes to be independent given the value of the class variable. This conditional independence assumption rarely holds true in real world applications, hence the characterization as Naive yet the algorithm tends to perform well and learn rapidly in various supervised classification problems.

(This work is licensed under a Creative Commons Attribution 4.0 International License.)

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 5 Issue 07, July-2016

III. RESULTS

A. *Image enhancement*

The input CT image is enhanced by using gamma correction terminology. Gamma correction is special case of power law transformation. It can be defined as follows:

$s = c. r^{\gamma}$

Where c and γ are positive constants. r pixel value can be mapped into s pixel value using different combinations of c and γ (gamma). Figure 4 shows enhanced image with various gamma values.

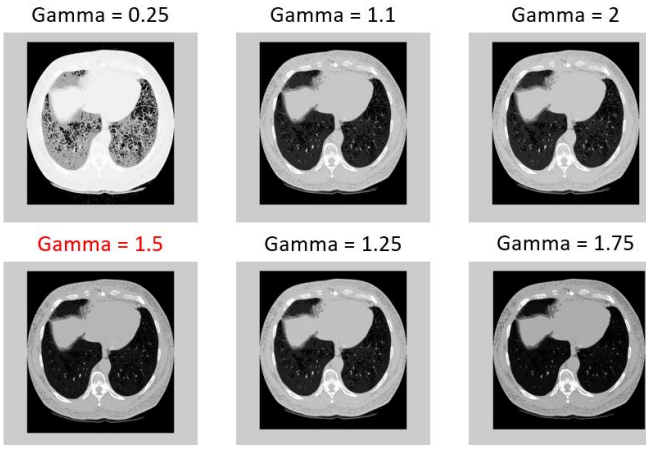

Fig.4 Enhanced Images with different gamma values

Figure 5 and 6 shows original CT scan image and its enhanced form respectively.

Fig.5 Original Image

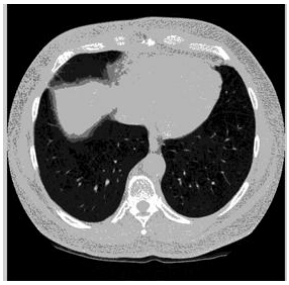

Fig.6 Enhanced Image

B. *Image cropping and segmentation*

By using MALTAB tool enhanced image was segmented, due to which further processing gets easier. Figure 6 shows result of image segmentation. This gives us objects present in an image. From this result we had calculated features of every object.

Fig.6 Binary Image

Fig.6 ROI (Region of Interest)

C. *Feature Extraction*

Area, perimeter, eccentricity, MajorAxisLength, MinorAxisLength, Orientation, ConvexArea, FilledArea, EulerNumber, EquivDiameter, Solidity, and Extent of each object in segmented image were extracted.

By using parameter values classification can be done by using three different classification technique. And accuracies of these classifiers was computed on WEKA which is mansion in below given table.

Classifier	Correctly Classified Instances	Incorrectly Classified Instances
NaiveBayes	64%(16)	36%(9)
SMO	56% (14)	44% (11)
Random Forest	52% (13)	48% (12)

IV. CONCLUSION

This methodology developed for identification of lung nodule in a spiral CT scan image. Image pre-processing techniques successfully enhanced CT scan image. Median filter along with gamma correction gives best result for pre-processing stage. The global thresholding method gives binary image with help of threshold value calculated by Otsu's algorithm. Segmentation of binary image was done by morphological operations. Number of objects obtained from segmented image gives features for further classification. Classification stage gives us type of lung nodule i.e. benign (non-cancerous) or malignant (cancerous). Accuracy tells us, which classifier we can use for further study. NaiveBayes classifier gives accuracy of 64% and other two classifiers gives accuracy more than 50%.Accuracy can be improved.

REFERENCES

[1] Anjali Kulkarni1 ,Anagha Panditrao2, "Classification of Lung Cancer Stages on CT Scan Images Using Image Processing" Phil. Trans. Roy. Soc. London, vol. A247, pp. 529-551, April 1955. (*references*)

[2] K.Punithavathy, M.M.Ramya, Sumathi Poobal "Analysis of Statistical Texture Features for Automatic lung Cancer Detection in PET/CT Images" ,International Conference on Robotics, Automation, Control and Embedded Systems- RACE 2015,18-20 February 2015, Hindustan University, Chennai, India .

[3] Nooshin Hadavi,Md.Jan Nordin, Ali Shojaeipour "Lung Cancer Diagnosis Using CT-Scan Images Based on Cellular Automata"

[4] N. Hollings, P. Shaw "Diagnosis imaging of lung cancer," ©ERS Journals Ltd 2002.

[5] Yongbum Lee, Takeshi Hara, Hiroshi Fujita, Shigeki Itoh and Takeo Ishigaki "Nodule Detection on Chest Helical CT Scans by Using a Genetic Algorithm" Intelligent Information System 1997.

[6] Yan Wang, Shuang Sun, Dian Qu, Anyu Chen,Zijian Cui, Yulu Yao "Preliminary Study on Early Detection Technology of Lung Cancer based on Surface-Enhanced Raman Spectroscopy" Biomedical Engineering College of Capital University of Medical Sciences Beijing, China, 100069.

[7] Sahil J Prajapati1, Kalpesh R Jadhav "Brain Tumour Detection By Various Image Segmentation Techniques with Introduction to Non Negative Matrix Factorization," International Journal of Advanced Research in Computer and Communication Engineering Vol. 4, Issue 3, March 2015.

[8] M. Gomathi , Dr. P. Thangaraj "Automated CAD for Lung Nodule Detection using CT Scans" 2010 International Conference on Data Storage and Data Engineering.

[9] Amal A. Farag, James Graham, Salwa Elshazly and Aly Farag "Data-Driven Lung Nodule Models for Robust Nodule Detection in Chest CT" 2010 International Conference on Pattern Recognition.

(This work is licensed under a Creative Commons Attribution 4.0 International License.)