International Journal of Applied Engineering Research ISSN 0973-4562 Volume 11, Number 4 (2016) pp 2697-2700

Feature Extraction Based Lung Nodule Detection in CT Images

Veerakumar Kuppusamy
Assistant Professor, Department of ECE, Bharath Niketan Engineering College, Tamil Nadu, India.

Ravichandran C Gopalakrishnan
Principal, SCAD Institute of Technology, Palladam, Tamil Nadu, India.

Abstract
Lung cancer diagnosis using pattern classification is the active research topics in medical image processing. Feature extraction is considered as an essential step in medical image analysis and classification. Computer aided segmentation for computed tomography (CT) and magnetic resonance imaging (MRI) are finding the application in computer aided diagnosis, clinical studies, and treatment planning. In medical images typically suffer from one or more imperfections such as low resolution (in the spatial and spectral domains) and low contrast images. This work proposed the black circular neighborhood algorithm for feature extraction and genetic algorithm (GA) based clustering using the extracted nodules. The performance of our algorithm is to reduce number of false positives (FP) and results are improves the accuracy.

Keywords: CT images, Feature extraction, Black circular neighborhood algorithm, GA algorithm

Introduction
This paper describes a fully automated feature extraction and clustering scheme, which is designed to recognize lung tissue and nodules from CT images. A new fully automatic method has been proposed based upon black circular neighborhood algorithm and genetic algorithm based image processing techniques to extract the nodules. Feature extraction can also be viewed as a problem of clustering, which is aimed to extract the nodules based on intensity values The early detection of cancer can be helpful in curing the disease completely. So the requirement of techniques is to detect the occurrence of cancer nodule in early stage is increasing. There are different technique exists but none of those provide better accuracy for detection.

Feature extraction is an important step in many computer vision algorithms. The separately obtained area is then analyzed for detection of nodules to diagnose the disease. Computed tomography (CT),

Magnetic resonance imaging, digital mammography, and other imaging modalities provide an effective means for noninvasively mapping the anatomy of a subject. These technologies have greatly increased knowledge of normal and diseased anatomy for medical research and are a critical component in diagnosis and treatment planning. In medical imaging, feature extraction is important for segmenting the lung nodules, image measurements, and image display.

Kanazawa et al (1996), described Computer aided diagnosis system for lung cancer based on helical CT images. The author describes a computer assisted automatic diagnosis system for lung cancer detection at early stage from CT images and to decreases the time complexity and increases the diagnosis confidence. Penedo *et al* (1998), consists of lung nodule by using a neural-network-based approach. The authors have provided a computer-aided diagnosis system based on two-level Artificial Neural Network (ANN) architecture for trained, tested and evaluated the test images. Yamamoto *et al* (2000), stated Computer aided diagnosis system with functions to assist comparative reading for lung cancer based on helical CT image. Yim *et al*. (2005) explained about Hybrid lung segmentation in chest CT images for computer-aided diagnosis. The Experimental result are automatically and accurately extracts the lung surfaces of the CT images from 10 patient datasets with lung cancer or pulmonary embolism. Jose *et al* (2003), presents a genetic embedded approach that performs the selection task for a SVM classifier. The main feature of the proposed approach concerns the highly specialized crossover and mutation operators that take into account gene ranking information provided by the SVM classifier. M. Gomati et al (2010), the lung region is segmented using Fuzzy Possibilistic C Mean (FPCM) clustering algorithm after that the features are extracted and the diagnosis rules are generated. These rules are then used for learning with the help of Support Vector Machine (SVM). Shahid Eqbal et al, (2015) proposed Neural networks and genetic algorithms based machine learning techniques for automated detection of pulmonary nodules in CT images. Their results could help the radiologists for a true diagnosis and decrease the number of the missing cancerous regions or unnecessary biopsies. Kishore et al, (2015) explained the performance of screens, ranks and selection methods. Also evaluate the performance, image analysis is being done for its recognition could predict the Classification accuracy for lung CT images. Jaffar et al (2008), has to considered lung segmentation based Genetic Algorithm (GA) and morphological image processing techniques for image analysis and classification

Lung nodule detection system
The Black circular neighborhood algorithm for identified the lung nodule pixels. The nodules identified from the CT image are clustered as benign and malignant based on shape, size

International Journal of Applied Engineering Research ISSN 0973-4562 Volume 11, Number 4 (2016) pp 2697-2700

and intensity of the nodule. The Genetic algorithm based clustering is used to cluster the lung nodules. GA based clustering is chosen to reduce the distances between the cluster center and other objects of the cluster. The overall block diagram of the proposed work is given in Figure 1.

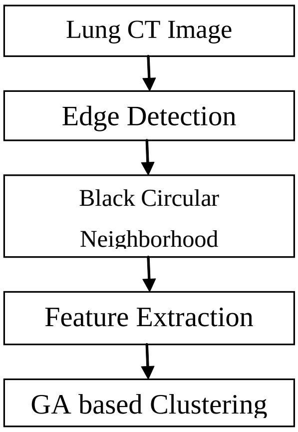

Figure 1: Block Diagram of proposed work

Results and Discussion

Ant colony optimization (ACO) algorithm based edge detected image is given as the input to black circular neighborhood algorithm. The assumption is that all the edges are black pixels. The edges here represent the edges of the lungs along with the nodule's edges. As the lung edges are fine lines and nodules are identified in the form of clustered black pixels, black circular neighborhood algorithm is applied. The work of black circular neighborhood algorithm is to find the center pixel of clustered black pixels. Initially, clustered black pixels are identified using the 4 – connected and 8 – connected properties of the pixels. The proposed algorithm detects nodules whose diameter is up to 5 units. It detects both spherical and elliptical nodules. From the black pixel clusters, the center pixel is identified to be the center of lung nodules. Black circular neighborhood algorithm is presented below. Using the spherical shape and size spotted the nodules in the edge detected image. These nodules have to be classified based on their intensity. The center of lung nodules detected is given to the intensity based feature extraction algorithm.

Algorithm 1. Black circular neighborhood algorithm
Input: ACO algorithm output image.
Output: Pixels of nodules center.
Assumption: ACO output is black and white image with black pixels representing edges.
begin
detect edge pixel positions
for each shape (spherical, elliptical)
for each edge pixel
check for 8-neighborhood
if accepted add to center_pixels
end for
end for
return center_pixels
end

Nodules are identified by the radiologists with the help of intensity change in the CT image. A whiter region on the image indicates the presence of nodule in lungs. Intensity feature of the nodule is used to identify them. In the proposed algorithm, an intensity based feature extraction algorithm is used for locating the nodules position over the CT image.

The center pixels of the lung nodules are retrieved from the black circular neighborhood algorithm. After identifying the nodules center pixel, the intensity is used to classify the nodule and to detect malignant nodules. A circular region is determined around the center pixel of the nodules. The intensity of the nodules within the radius of the circular region is retrieved from the input lung CT image. The sum of intensity of all the pixels inside the circular region is calculated and its average value is found. When the average intensity exceeds the threshold intensity, the specified pixel and the size of the surrounding region is entered in the feature matrix. The output feature matrix consists of average intensity of the region, size and the shape of the region. For spherical shape, 1 is given and for elliptical shape, 2 and 3 are given for x-axis as major axis and y-axis as major axis respectively. These features extracted from the CT images are used for classification.

Algorithm 2. Intensity based clustering algorithm
Input: Nodules center pixels, Input CT image.
Output: Feature matrix
begin
for each size_types
for each shape
for each center_pixel
find pixels within circular radius
find pixels intensity
determine circular region with figured pixels
calculate the average intensities within the circular region
if avg>threshold
assign center_pixel to nodule_pixel
end if
end for(center_pixel)
end for(shape)
end for(size_type)
end

Clustering

Hierarchical Clustering

The feature matrix consisting of size (i.e., area of the identified nodules) and intensity with the age of the patient is fed as input. Agglomerative approach is used to cluster the features in a bottom-up manner. The clustering is done by merging two groups having the smallest dissimilarity measure. Nearest-neighbor linkage is used to find the dissimilarity between cluster points. In figure 2, represented the accuracy of 302 lung CT images.

Genetic Algorithm Based Clustering

The feature matrix with shape, size and intensity features is used to cluster the lung nodules. Genetic algorithm based clustering is used for this purpose. The feature matrix is initially given to the κ-means algorithm to cluster the nodules and the cluster centers are identified. The cluster centers are refined based on the genetic algorithm. The searching capability of genetic algorithms is used in an effective way to find the appropriate cluster centers such that a similarity metric of the resulting clusters is optimized. The selection, crossover, and mutation processes of the genetic algorithm help to refine the cluster centers. It gives the lung nodules as benign and malignant based on the lung's size (small or large), shape (spherical or elliptical) and intensity.

The selection of fitness function is very essential step in genetic algorithm. It helps to accurately place the cluster centers. Similarity calculated by cross-correlation coefficient normalized covariance is used as the fitness function. Here m_a and m_b are the cross correlation co-efficient, a and b represent the image pixels.

$$\text{Similarity}_{a,b} = \frac{\sum_{i=0}^{n-1}(a_i - m_a)(b_i - m_b)}{\sqrt{\sum_{i=0}^{n-1}(a_i - m_a)^2}\sqrt{\sum_{i=0}^{n-1}(b_i - m_b)^2}} \quad (1)$$

where,

$M_a = \frac{1}{n}\sum_{i=0}^{n-1} a_i$, $m_b = \frac{1}{n}\sum_{i=0}^{n-1} b_i$.

In figure 3, represents the accuracy of 302 lung CT images of Genetic algorithm based clustering.

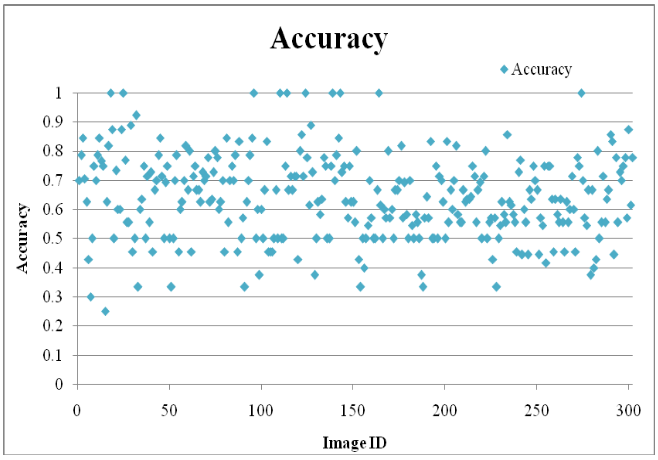

Figure 2: Accuracy for 302 images for hierarchical algorithm based clustering

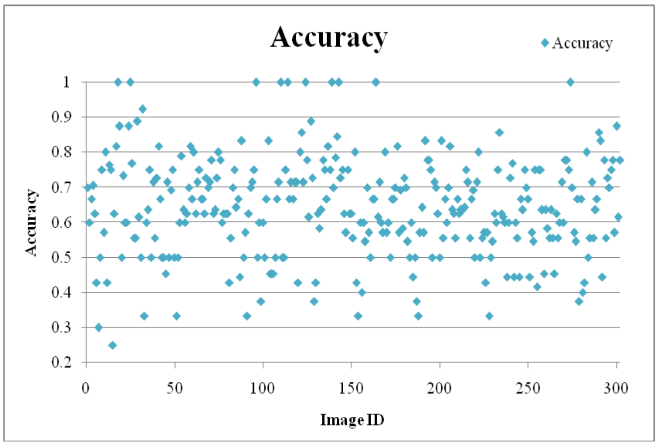

Figure 3: Accuracy for 302 images for genetic algorithm based clustering

Conclusion

In this paper, a black circular neighborhood algorithm has to extract lung nodules from edge detected CT images. The extracted nodules are clustered by using Genetic algorithm (GA). GA based algorithm has to reduced number of false positive (FPs) and give more accuracy. The overall accuracy of 302 lung CT images is 64%. This result is higher than the hierarchical based clustering algorithm.

References

[1] Shahid Eqbal Dr. M. A. Ansari, 2015, "Medical Image Feature Extraction for Computer Aided Diagnosis of Lung Cancer," International Journal of Advanced Research in Computer Science and Software Engineering, Volume 5, Issue 6.

[2] R. Kishore, 2015, "An Effective And Efficient Feature Selection Method For Lung Cancer Detection," International Journal of Computer Science & Information Technology (IJCSIT) Vol 7, No 4.

[3] Jaffar, M.A. Hussain, A. Nazir, M. Mirza, A.M, 2008, "GA and Morphology Based Automated Segmentation of Lungs from CT Scan Images," Computational Intelligence for Modelling Control & Automation, IEEE International Conference., pp. 265 – 270.

[4] M.Gomathi Dr. P.Thangaraj,2010, "A New Approach to Lung Image Segmentation using Fuzzy Possibilistic C-Means Algorithm", International Journal of Computer Science and Information Security, Vol. 7, No. 3.

[5] Kanazawa, K., M. Kubo and N. Niki, 1996, "Computer aided diagnosis system for lung cancer based on helical CT images," Proceedings of the 13th International Conference on Pattern Recognition, Newport Beach, CA, USA, pp: 381-385. ISBN: 0-8186-7282-X.

[6] Penedo, M.G, M.J. Carreira., A. Mosquera and D. Cabello, 1998, "Computer-aided diagnosis: a neural-networkbased approach to lung nodule detection," IEEE Trans. Med. Image., 17: 872-880. ISSN: 0278-0062.

[7] Yamamoto, T., Y. Ukai, M. Kubo, N. Niki and H. Satou, et al., 2000, "Computer aided diagnosis system with functions to assist comparative reading for lung cancer based on helical CT image," Image Process. Int. Conf. Proceed,1: 180-183. ISBN: 0- 7803-6297-7.

[8] Yim, Y., H. Hong and Y.G. Shin, 2005, "Hybrid lung segmentation in chest CT images for computeraided diagnosis," Proceedings of 7th International Workshop on Enterprise networking and Computing in Healthcare Industry, June 23-25, Seoul Nat. University, South Korea, pp: 378-383. ISBN: 0-7803-8940-9.

[9] Jose L. Marroquin, Edgar Arce Santana, and Salvador Botello, 2003, "Hidden Markov Measure Field Models for Image Segmentation," IEEE transactions on pattern analysis and machine intelligence, VOL. 25, NO. 11.

[10] S. Kumar Raja, ChiranjitMukhopadhyay, and T. V. Ramachandra, 2011, “Hybrid Bayesian Classifier for Improved Classification Accuracy,” IEEE geoscience and remote sensing letters, vol. 8, no. 3.

[11] M. Gomathi, P. Thangaraj, 2011, “A Computer Aided Diagnosis System for Lung Cancer Detection using Machine Learning Technique,” European Journal of Scientific Research ISSN 1450-216X Vol.51 No.2 (2011), pp.260-275.

[12] A.Bharathi, A.M.Natarajan, 2011, “Cancer Classification using Modified Extreme Learning Machine based on ANOVA Features,” European Journal of Scientific Research ISSN 1450-216X Vol.58 No.2 (2011), pp.156-165.

[13] SaleemIqbal, Khalid Iqbal, 2011, “Lungs Segmentation for Computer Aided Diagnosis,” International Journal of Academic Research Vol. 3. No. 5.

[14] Nisar Ahmed Memon, Anwar MajidMirza, and S.A.M. Gilani, 2006, “Segmentation of Lungs from CT Scan Images for Early Diagnosis of Lung Cancer”, World Academy of Science, Engineering and Technology 20.

[15] NingXua, NarendraAhujaaand Ravi, “Automated Lung Nodule Segmentation Using Dynamic Programming and EM Based Classification”, Urbana-Champaign, IL 61801 b Siemens Corporate Research, Inc. Princeton, NJ 08540.

[16] Murat Ceylan, Y¨Uksel ¨Ozbay, o. Nuriuc¸an, Erkanyildirim, 2010, “A novel method for lung segmentation on chest CT images: complex-valued artificial neural network with complex wavelet transform,” Turk J Elec Eng& Comp Sci, Vol.18, No.4