www.ijecs.in
International Journal Of Engineering And Computer Science ISSN:2319-7242
Volume - 3 Issue - 8 August, 2014 Page No. 7641-7645

A Novel Method for Classification of Lung Nodules as Benign and Malignant using Artificial Neural Network

Rohit B Kuravatti[1], Sasidhar B[2], Dr.Ramesh Babu D R[3]

[123]Computer Vision Lab, Dayananda Sagar College of Engineering, Bangalore, India
[1]rkrohitk76@gmail.com, [2]bolasasi@gmail.com, [3]bobrammysore@gmail.com

Abstract: ***Automated Segmentation and Classification of lung nodules into benign and malignant is a challenging task and is of vital interest for medical applications like diagnosis and surgical planning. It improves the accuracy and assist radiologist for better diagnosis. In this paper, a new method is proposed for the classification of lung nodules using Artificial Neural Networks based on Shape, Margin and Texture features. In order to reduce the complexity of the algorithm and the computational load, use of fewer features is particularly important, while maintaining an acceptable detection performance. The proposed algorithm was tested on LIDC (Lung Image Database Consortium) datasets and the results were satisfactory in terms of accuracy in classification.***

Keywords: Artificial Neural Network, Lung Nodules, Benign, Malignant, Classification.

1. Introduction

The death rate due to lung cancer is found to be more than any other cancer in both men and women with an estimated 159,260 deaths, accounting for about 27% of all cancer deaths are expected to occur in 2014[1]. Imaging technology already had lifesaving effects on ability to detect cancer early and more accurately diagnose the disease.

Computed Tomography (CT) is one of the best imaging techniques for finding soft tissues and it has high spatial resolution, minimizes artifacts. In early stage, the lung cancer is visible in CT as non-calcified lung nodule which are visible as low-contrast white with spherical in shape in the lung regions. The datasets were taken from LIDC database which are in Digital Imaging and Communications in Medicine (DICOM) format and measured in Hounsfield Units (HU).

The difficult tasks in CAD system are to improve the accuracy in classification of lung nodules as benign (harmless tumor) and malignant (harmful tumor) and to fully automate the classification process. Classification process is useful for the early diagnosis of lung cancer which in turn increases the survival rate of patient. Our proposed method solves the problem of early diagnosis of lung cancer with less number of false positives by extracting shape, margin and textural features and with the use artificial neural network for classification. Artificial neural network has several advantages such as the generalization and the capabilities of learning from training

data without knowing the rules in priori. The lung nodules are classified based on different features like shape, margin and texture (calcification pattern) as shown in the Figure 1.

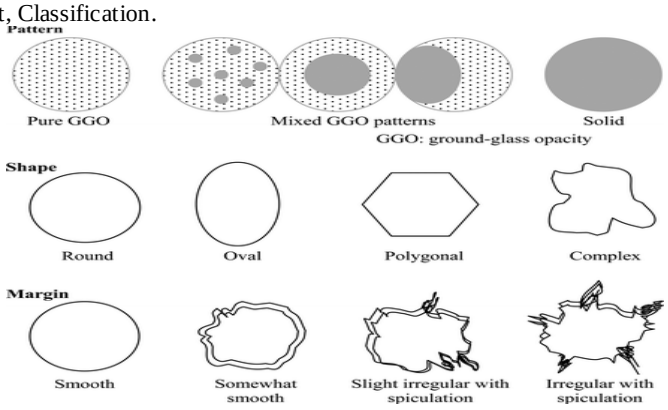


Figure 1: Morphological characterization of nodule [1]

2. Related work

The state-of-the-art on the classification of lung nodules as benign and malignant is presented in this section. Yongjun WU, Na Wang [2] proposed a method to classify the nodules based on 5 clinical parameters, using ANN and Logistic regression models; the accuracy was found out to be 84.6%, the results were compared by ROC curve analysis.

Kenji Suzuki, Feng Li [3] proposed a Scheme for Distinction between Benign and Malignant Nodules in Thoracic LDCT using multiple Massive Training Artificial Neural Network (MTANN) and an integration ANN.

S.L.A. Lee, A.Z. Kouzani, and G. Nasierding [4] came up with an unique architecture for classification-aided-by clustering which employs random forests as its base classifier.

Gurcan [5] used curvature analysis and candidate detection scheme using k-means clustering, rule-based followed by LDA which reported to reduce the false positive rate. Rule based

classifier produced 84% sensitivity with 5.48 FPs/slice. When LDA was applied after rule-based, the false positive rate falls to 1.74 per slice for the same sensitivity.

Weighted multi-scale convergence index (WMCI) and fisher linear discriminant (FLD) were combined by Hardie [6]. The system was evaluated on 154 chest radiographs from JRST database with 100 malignant and 54 benign nodules which result 78.1% sensitivity with 4 FPs/image. Also, a performance comparison between quadratic classifier and Gaussian Bayes linear classifier and FLD classifier was performed. The FLD's performance was superior to its studied counterparts.

S.K. vijay Anand [7] has used shape and textural features with ANN for classification purpose, accuracy of ANN was found out to be 81.8%.

Kanazawa [8] detected suspicious regions by using Gaussian and mean curvature, Rule-based method was used to eliminate non-nodules. This work resulted in sensitivity of 90% with 8.6 FPs per subject.

C Robert Falk [9] examines the effectiveness of geometric feature descriptors, for false positive reduction and for classification of lung nodules in low dose CT (LDCT) scans using active appearance models (AAM); which are then used to detect candidate nodules based on optimum similarity measured by the normalized cross-correlation (NCC).

As per the survey, we found out that various methods are being implemented to segment and classify the lung tumors. Scope is given to improve the accuracy of the classification process. In this direction an attempt is made to improve the accuracy of ANN classifier model by inputting more number of features and training the model with large datasets.

The rest of the paper is organized as follows: In section 3, Proposed Methodology is discussed. Section 4 gives the results and discussion of the proposed algorithm. Conclusion is given in section 5.

3. Proposed Methodology

The steps involved in segmentation and classification of lung nodule as benign and malignant are depicted in Figure 2. Chest CT image is taken as input and lung regions are segmented based on adaptive thresholding.

Lung nodules are extracted from lung region and features like shape, margin and texture features are extracted from lung nodules. Artificial Neural Network is used to classify lung nodules as benign and malignant based on the features extracted from the training data.

3.1 Segmentation of lung regions based on thresholding and morphological operations

The segmentation of lung regions plays an important role to speed up the process of detection and analysis of lung nodules.

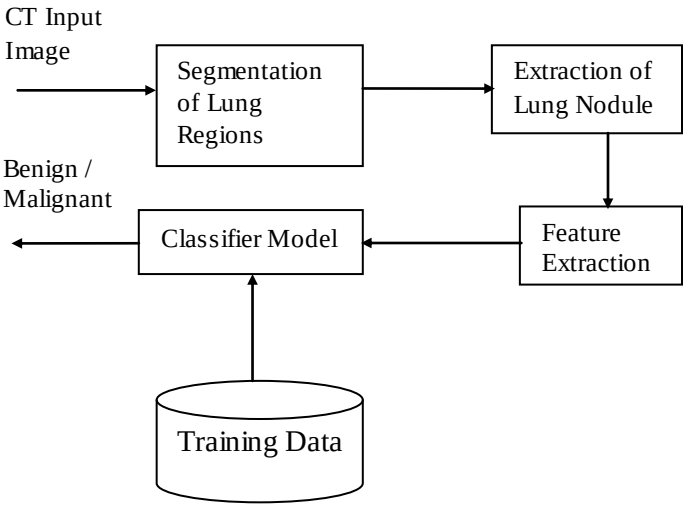


Figure 2: Block diagram of the proposed method

The steps involved in segmentation of lung regions are

1. Apply intensity threshold to extract lung parenchyma. Experimentally, threshold value was found at -420HU [10]

$$\text{binary}(i,j) = \begin{cases} 1, & \text{if } f(i,j) < T \\ 0, & \text{otherwise} \end{cases} \quad (1)$$

Here T is -420 HU, binary (i,j) is the binary image after applying thresholding. To convert Hounsfield Unit to gray level value, equation (2) is applied on binary (i,j)

$$\text{Gray Level value} = 1024 + T \quad (2)$$

2. Label various connected components and clear image borders to remove the regions which are connected to the border of the image.

3. Find the two largest connected components. The two largest connected components are two lung regions.

$$\text{Lungregion} = \begin{cases} \max(\text{area}(R)) & \text{lung1} \\ (\max - 1)(\text{area}(R)) & \text{lung2} \end{cases} \quad (3)$$

The first largest connected component is lung1 and second largest connected component is lung2. The connected components which are less than lung2 have been removed by using equation (3).

4. Superimpose the lung mask in binary image format with the input CT image to obtain the lung regions with original intensities.

3.2 Extraction of Lung Nodule based on intensity, shape and area features

After extracting lung regions, detection of lung nodules could be performed as explained below.

The steps followed in extraction of lung nodule are

1. Input image is the segmented lung region
2. Apply intensity threshold, shape and area features to get nodule mask in binary image format.
3. Superimpose the mask with the input image to extract lung nodules with original intensities.

3.3 Feature extraction of Lung Nodules

Lung nodules are classified into benign or malignant based on their appearance, texture and margin variation features.

1) Appearance features: Area, Perimeter, Convex Area, Equiv Diameter, Eccentricity, Extent, and Circularity.

2) Texture features: Energy, Contrast, Correlation, Solidity, and Homogeneity.

3) Margin Variation: It is a margin feature which specifies whether border is smooth or speculated. If the margin of the nodule is smooth then the nodule is benign otherwise it is malignant.

3.4 Classification of Lung Nodules using Artificial Neural Network

Before the actual classification process, neural network is first trained with training data consisting of the same 13 features extracted from different nodules (30 nodules are used for training purpose). The network is trained using log-sigmoid activation function (equation 10) with a learning rate of 0.1 to evaluate the feature vectors based on the current network state. The error threshold is set as 0.1 and maximum number of epochs (iterations) as 500(iterations will stop if error threshold is less than or equal to 0.1).

The classification process using neural network starts with the input feature values ($\boldsymbol{p}$) getting multiplied with random weights ($\boldsymbol{w}$). The weighted input is added to a scalar bias b to form net input. The net input is passed through the transfer function ($\boldsymbol{f}$), which produces an output ($\boldsymbol{O}$). The names given to these three processes are: the weight function, net input function and the transfer function, which is summarized in the following equation.

$$O = f(wp + b) \qquad (4)$$

Hyperbolic tangent sigmoid transfer function ($\boldsymbol{f}$) is used for the classification purpose. This is mathematically equivalent to tanh (n). But it differs in that it runs faster than the implementation of tanh.

$$o = tansig\,(n) = \frac{2}{1+e^{-2*n}} - 1 \qquad (5)$$

Where n is an input vector and $\boldsymbol{O}$ is the output of transfer function.

A multi-layer feed forward BPN for classification of tumor consists of an input layer, one hidden layer and an Output layer.

The total number of nodes in the input layer (m) is 13 representing features extracted from the ROI. The number of nodes in hidden layer (H) 10 was decided experimentally as the network produced satisfactory results. The output of BPN is a binary value; hence two nodes exist in the output layer. The nodes in one layer connect to the nodes in the next layer by means of directed communication links, each with an associated random weight.

Since the BPN for proposed method has one hidden layer and one output layer, it results in two weight matrices, w1 connecting the input layer to the hidden layer and the w2 connecting the hidden layer to the output layer.

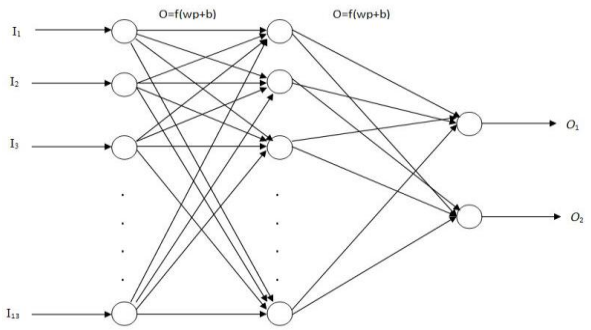


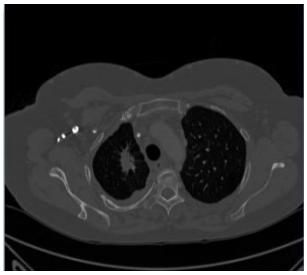
Figure 3: Multi-layer feed forward BPN for classification

The output values lie in the range 0 to 1. Since the output layer contain two neurons, the target values are defined as [1 0] or [0 1] representing 2 classes. The target values are assumed to be [1 0] for malignant nodules, and [0 1] for benign nodules.

The output values from neurons at the output layer are compared with target values, and error factor is computed. Again weights are adjusted at the hidden layer neurons to minimize the error factor.

4. Results and Discussion

The input image for the classification is as shown in the Figure 4(a).Lung region separation is done by using intensity threshold and morphological operations as shown in the Figure 4(b). The extraction of lung nodules are based on intensity threshold and appearance features which is depicted in the Figure 4(c).



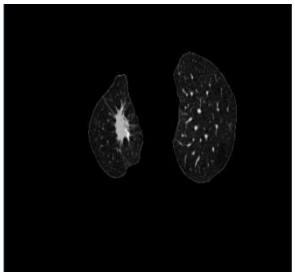
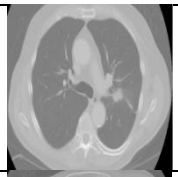
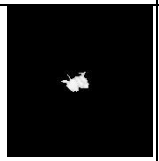
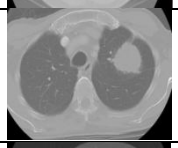
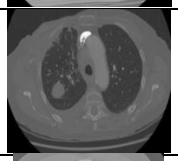
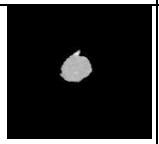
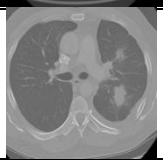
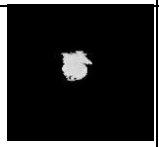


Figure 4: (a) Input image (b) Segmentation of lung Regions

(c) Extraction Lung Nodule

Once the Lung nodule is segmented, features are extracted from it for the classification purpose. Table I depicts segmented lung nodules and values of 3 major features extracted from these nodules.

Table I: Extracted Lung nodules and features

CT Image	Nodule	Area	Margin Variation	Eccentricity
		344	48	0.772
		4632	22	0.180
		1069	33	0.511
		967	53	0.773

From the segmented nodule, 13 major features mentioned in section 2 (C) were extracted from 22 test nodules and given as input to artificial neural network with 10 neurons in hidden layer. The output of artificial neural network is either 1(malignant) or 0(benign). The performance of proposed approach is evaluated based on four performance metrics such as accuracy, sensitivity, specificity and precision as shown in Table I.

Calculation performance metrics is based on four parameters viz, True positive (TP), False positive (FP), False negative (FN) and True negative (TN).

- **True positive (TP):** Lung nodules classified by the algorithm and the radiologist as malignant are known as True positive.
- **False positive (FP):** Lung nodules classified as malignant by the algorithm and benign by the radiologist are known as false positive.
- **False negative (FN):** Lung nodules classified as benign by the algorithm and malignant by the radiologist are known as false positive.
- **True negative (TN):** Lung nodules classified by the algorithm and the radiologist as benign are known as True positive.

Specificity is the number of correctly classified negative (Benign) nodules out of actual negative nodules

$$Specificity = (TN)/(TN + FP) \quad (6)$$

Accuracy is the number of correctly classified nodules (Benign and Malignant) out of all nodules.

$$Accuracy = (TP + TN)/(TP + FP + TN + FN) \quad (7)$$

Precision is the number of correctly classified positive nodules (Malignant) out of all positive nodules.

$$Precision = (TP)/(TP + FP) \quad (8)$$

Recall is the number of correctly classified nodules (Malignant) out of actual positive nodules.

$$Recall = (TP)/(TP + FN) \quad (9)$$

Table II: Performance Measures and comparative study

Performance Measures	Existing system[7]	Proposed Method
Specificity	77.7%	80%
Accuracy	81.8%	86.4%
Precision	84.6%	84.6%
Recall	84.6%	91.7%

The performance of the classifier model is determined with the help of Receiver Operating Characteristics (ROC) curve. ROC curve is a plot of the true positive rate versus the false positive rate as shown in the Figures 5 and 6. Area under thecurve for proposed method was found out to be 0.858.

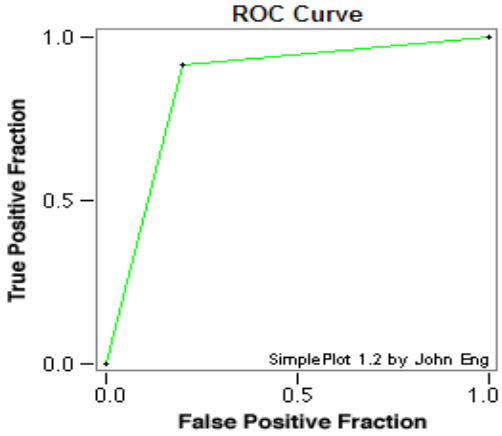


Figure 5: Receiver Operating Characteristic Curve for the proposed method.

Existing System [7] used the classifier model (Artificial neural network) with 9 features at the input layer and 5 neurons in the hidden layer. The performance of this approach is evaluated based on four performance metrics such as accuracy, sensitivity, specificity and precision as shown in Table II.
Area under the curve for existing method was found out to be 0.806 which indicates the false positive rate is high compared to proposed method.

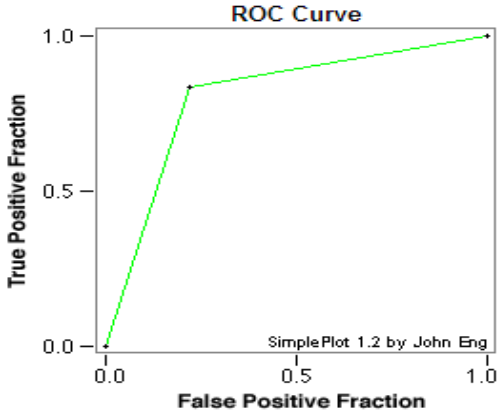


Figure 6: Receiver Operating Characteristic Curve in [7]

After comparing tables and ROC curves of both approaches, proposed method gives better results with respect to all four metrics.

5. Conclusion

The proposed method is able to predict whether the tumor is benign or malignant in nature with an accuracy of 86.4%. The plot of ROC curve depicts a significant reduction in false positive rate (Area under curve 0.858) compared to existing approach (Area under curve 0.806), hence reducing the misclassification probability. The proposed system would be effective in assisting the physician in identifying the lung tumor as cancerous (Malignant) or non-cancerous (Benign).

References

[1] American cancer society,cancer facts and figures, American cancer society, New York,NY,USA,2013.

[2] Yongjun WU, Na Wang, Hongsheng ZHANG, Lijuan Qin, Zhen YAN, Yiming WU," Application of Artificial Neural Networks in the Diagnosis of Lung Cancer by Computed Tomography", Sixth International Conference on Natural Computation,2010.

[3] Kenji Suzuki, Feng Li, Shusuke Sone, and Kunio Doi," Computer-Aided Diagnostic Scheme for Distinction Between Benign and Malignant Nodules in Thoracic Low-Dose CT by Use of Massive Training Artificial Neural Network", IEEE TRANSACTIONS ON MEDICAL IMAGING, VOL. 24, NO. 9, SEPTEMBER 2005.

[4] S.L.A. Lee, A.Z. Kouzani, and G. Nasierding,"Pulmonary Nodule Classification Aided by Clustering", IEEE International Conference on Systems, Man, and Cybernetics San Antonio, TX, USA - October 2009.

[5] M. N. Gurcan, B. Sahiner, N. Petrick, H. P. Chan, E. A. Kazerooni, P. N. Cascade, and L. Hadjiiski, "Lung nodule detection on thoracic computed tomography images: preliminary evaluation of a computer-aided diagnosis system" *Medical Physics*, vol. 29, no. 11, pp. 2552-2558, 2002.

[6] R. C. Hardie, S. K. Rogers, T. Wilson, and A. Rogers, "Performance analysis of a new computer aided detection system for identifying lung nodules on chest radiographs " *Medical Image Analysis* vol. 12, pp. 240- 258, 2008.

[7] S.K. Vijai Anand," Segmentation coupled textural feature classification for lung tumor prediction", ICCCCT, 2010.

[8] K. Kanazawa, Y. Kawata, and N.Niki, "Computer aided diagnosis for pulmonary nodules based on helical CT images," Computerized Medical Image and Graphics, Vol. 22,1998.

[9] Amal Farag, Asem Ali, James Graham, Aly Farag, Salwa Elshazly and Robert Falk," Evaluation of geometric feature descriptors for detection and classification of Lung Nodules in loe dose CT scans of the chest",IEEE transactions,2011.

[10] Eng. Michael Samir Labib habib,"A Computer aided diagnosis system (CAD) for the detection of pulmonary nodules on CT scans", systems and biomedical engineering department, faculty of engineering, cairo university, Giza, Egypt, 2009.

[11] Moulay A. Mexiane," Current Concepts in imaging and management of the solitary pulmonary nodule",J Med Liban 2009.

Author Profile

Rohit B Kuravatti currently perceiving M.Tech in Computer Networks from Dayananda Sagar College of Engineering, perceived B.E. in Computer Science Engineering from B.V.B College of Engineering in 2011. He along with the other two authors mentioned in this paper is working actively in Computer Vision Lab, DSCE Bangalore towards improving the accuracy of tumor classification.