IJCSE International Journal of Computer Sciences and Engineering Open Access
Research Paper Volume-4, Issue-12 E-ISSN: 2347-2693

Lung Cancer Classification

D.N. Sonar[1*] and U.V. Kulkarni[2]

[1] *Department of Computer Science, College of Computer Science & Information Technology, Latur-413512, India*
[2]*Department of Comp. Sci. & Engg., SGGS Institute of Engg., & Tech., Nanded-431606, India*

Available online at: www.ijcseonline.org

Received: 19/Nov/2016 Revised: 02/Dec/2016 Accepted: 20/Dec/2016 Published: 31/Dec/2016

***Abstract*—** Detection and diagnosis of lung cancer from chest radiographs is one of the most important and difficult task for the radiologists. In this paper, combination of statistical texture and moment invariant features are used to classify the lung cancer images. These features are extracted from JSRT raw chest X-ray images. The proposed approach is built on two-level architecture. In the first level architecture images are sharpened and segmented to extract the region of interest i.e. lung from the ribs using image processing techniques. In second level architecture, statistical texture and moment invariant based features are extracted depending on the shape characteristics of the region. These features are used as input pattern to the Fuzzy Hypersphere Neural Network (FHSNN) classifier. The experimental result shows that proposed approach is superior in comparison with only statistical texture features in terms of recognition rate, training and testing time.

***Keywords*—**Chest Radiography, Computer Tomography (CT), Fuzzy Hypersphere Neural Network (FHSNN), Lung Nodule, Gray level co-occurrence matrix (GLCM)

I. INTRODUCTION

Your body is made up of tiny building blocks called cells. When these cells change and grow out of control, it is called Cancer. Normal cells grow when your body needs them, and die when your body does not need them any longer. Cancer is made up of abnormal cells that grow even though your body doesn't need them. In most cancers, the abnormal cells grow to form a lump or mass called a tumor. These cells are often round in shape and are in bright contrast in chest X-rays. The detection and diagnosis of lung cancer has been a tedious task in medical image analysis over the past few decades. However, determination of cancerous tumors in its initial stage is essential for its early cure.

Various technologies have been raised to identify lung cancer at early stage. Computer Tomography (CT) is one of the more effective methods of detecting and diagnosing the lung cancer. In past research, it is observed that artificial neural networks, two level neural classifiers, hybrid lung nodule detection have been widely used for detection of lung cancer in medical images. Also the Computer Aided Diagnosis (CAD) system attracts more and more attention for early detection of lung cancer. Most common problem encountered during identification of lung nodules is the overlapping of ribs. Several rib suppression techniques have been proposed in recent research literature. However, by considering the cost limitation and safety precautions, X-ray diagnosing is more preferred by the radiologist.

In this paper, we used combination of statistical texture and moment invariant features to classify the lung cancer images. The proposed approach is built on two-level architecture. In first level architecture, image processing techniques such as median filtering, thresholding are used to sharpen the images to extract the region of interest. In second level architecture, combination of statistical texture and moment invariant features are extracted from the suspected area. These features are used as input pattern to the Fuzzy Hypersphere Neural Network which is built by using hypersphere fuzzy sets. The experimental results show that the proposed technique achieves a considerable rate of classification in terms of recognition rate, training and recall time.

This paper is organized as follows: Section 2 discusses the related work done in this area; Section 3 describes the lung cancer diagnosis system in brief. Section 4 presents experimental results and discussions. Finally, conclusions of the research are drawn in Section 5.

II. RELATED WORKS

M.N. Gurcan, et al.[1] has developed a computer-aided diagnosis (CAD) system for lung nodule detection on thoracic helical computed tomography (CT) images. The k-means clustering technique has been used by them to identify lung regions. Also rule-based classifiers are designed here to distinguish nodules and normal structures using 2D and 3D features. K. Kanazawa, et al.[2] presented an automatic diagnostic system to detect lung nodules from helical CT images of thorax. Kanazawa, Kubo, and Niki [3] used helical CT images for lung cancer and described Computer Aided Diagnosis system. This technique executes in two stages – analysis stage and diagnosis stage. In first stage, it extracts and analyses the features of lung and pulmonary blood vessel regions using image processing methods and in second stage, according to the features, rules are defined and those are then applied to identify the tumor

© 2016, IJCSE All Rights Reserved

region. Chiou, Lure, Ligomenides [4] designed a Hybrid Lung Nodule Detection (HLND) system based on artificial neural network architecture to improve diagnostic accuracy and speed of lung cancerous pulmonary radiology. This HLND system includes different phases of processing and the extracted features are classified using artificial neural networks. Yamomoto et al. [5] proposed image processing for computer-aided diagnosis of lung cancer by CT (LSCT). In this new LSCT technique, the author attempts to reduce the image information considerably to be displayed for the doctor with the help of image processing techniques. Zhou, Jiang, Yang and Chen [6] presented artificial neural network ensembles to detect cancer cells in specimen images of needle biopsies. According to the author, this technique achieves a high rate of nodule cell identification. Also it reduces the false negative identification rate. Yeny, Helen and Yeong [7] stated about Hybrid lung segmentation in chest CT images for computer aided diagnosis. The system first separates lungs and airways by an inverse seeded region growing and linked component labelling method. Secondly, three-dimensional region growing method is used to eliminate trachea and large airways from the lungs. In the final phase, exact lung region borders are obtained by subtracting the outcome of the second phase from that of the first phase. For pattern recognition and classification, artificial neural networks and fuzzy neural networks have been used widely. Penedo, Carreira, Mosquera, and Cabello [8] provided a computer-aided diagnosis scheme that depends on two-level artificial neural network (ANN) architecture. The outcome of this network is thresholded at a particular level of importance to provide a positive identification. Lin and Yan [9] stated a neural fuzzy model to formulate the diagnosis rules for identifying the pulmonary nodules. Initially, series of image processing methods to segment the lung area and obtain the region of interest are used. Next, features are obtained from region of interest and the nodules are detected with diagnosis rules that are formed with the help of neural fuzzy model. Patrick Simpson [10] has proposed supervised learning neural network classifier named fuzzy min-max neural network (FMN) that utilizes fuzzy sets as pattern classes. He has also proposed unsupervised fuzzy min-max clustering neural network [11]. Kulkarni and Sontakke [12] have proposed Fuzzy Hypersphere Neural Network (FHSNN) classifier that utilizes fuzzy sets as pattern classes and each fuzzy set is a union of fuzzy set hyperspheres. Kulkarni, Doye and Sontakke [13] have proposed a general fuzzy hypersphere neural network classifier which is an extension of fuzzy hypersphere neural network. It works in similar manner. Mundada, Murade, Vaidya and Swathi [14] used artificial neural network based software fault prediction technique to handle estimated solutions to optimization and search problems.

III. LUNG CANCER DIAGNOSIS SYSTEM

The diagnosis system for lung cancer detection is depicted in Fig. 1. Images are obtained from the public Standard Digital Image Database, JSRT. These images consist of 154 lung nodules (100 malignant cases, 54 benign cases), and 93 non-nodules. These images contain nodules within the range 5-60mm. In first level, pre-processing including normalization of size, filtering to enhance the contrast is done on the raw chest X-ray images. For diagnosis and analysis of lung nodules, separation of lung fields from the background is necessary.

For this, lung field masks as shown in Fig. 2.b. are prepared manually. Further, multiplication of lung field masks with the original X-ray image is carried out to separate lung fields from the background as shown in Fig. 2.c. On these separated lungs, Otsu's method [15] of thresholding is applied which chooses the threshold to minimize the intra-class variance of the black and white pixels as shown in Fig. 2.d. After that, the proposed approach performs a connected component analysis on the thresholded image which creates a connected group from a set of pixels. Shape characteristics or circularity index of the connected group is computed as (1) to see whether it is a tumor or not. The circularity index is 1 only for circles and is less than 1 for any other shape.

$$C_i = \frac{4\pi A_i}{R_i^2} \quad (1)$$

Where A_i is the area, R_i is the perimeter of the i^{th} connected component.

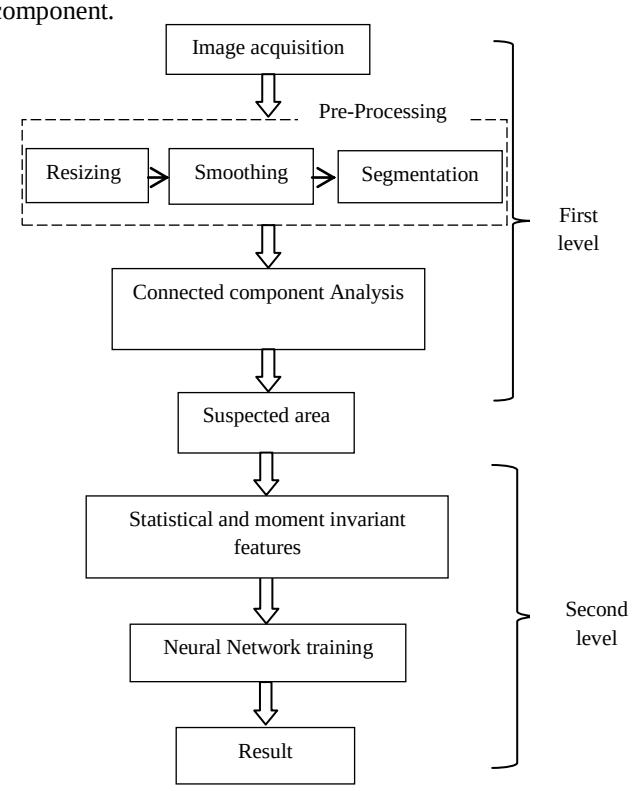

Fig.1. Lung Cancer Diagnosis System

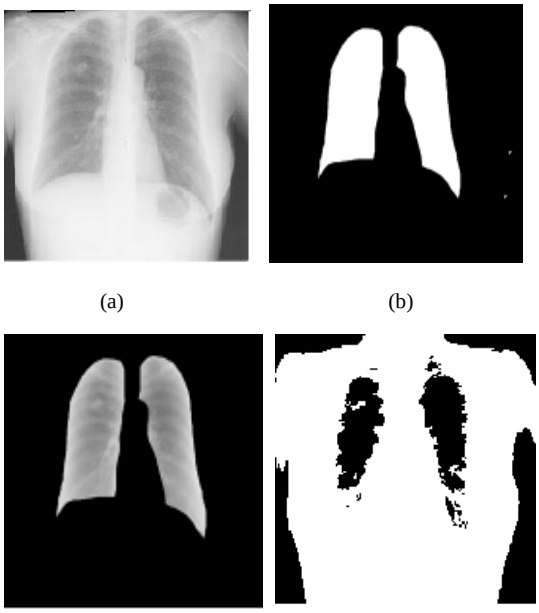

Fig. 2. Stages of suspected area detection. (a) Original image (b) Manually created lung masks (c) Segmented lung fields after multiplication (d) Image after thresholding.

After identifying the suspected area which shows a high probability of being a nodule, the second level of proposed approach starts.

Second level extracts the statistical texture and moment invariant features of the suspected area. These are categorized as first order statistical texture features and second order statistical texture features. The first order statistical texture features include average gray level, average contrast, Measure of smoothness, Third moment, Measure of uniformity and Entropy. While the second order statistical texture features are extracted using Gray-level Co-occurrence Matrix GLCM as in [15]. GLCM is a statistical method of examining texture of an image that considers the spatial relationship of pixels. It characterizes texture of an image by computing how often pairs of pixel with specific values occur in the image. For this, different angles are used which computes 1) the local variations in the GLCM, 2) correlation for the joint probability occurrence of the specified pixel pairs, 3) energy for the sum of squared elements and d) homogeneity for the closeness of the distribution of elements in the GLCM.

Similarly, geometrical moment invariants have been used as features for this system which helps in shape recognition and classification. Invariant shape recognition is performed by classification in the multidimensional moment invariant feature space. Several techniques have been developed that derive invariant features from moments for object recognition and representation. These moment invariant values are invariant with respect to translation, scale and rotation of the shape.

m_{pq} is the two dimensional moment of the function f(x,y). The order of the moment is (p+q), where p and q are both natural numbers, is implemented as,

$$m_{pq} = \sum_x \sum_y x^p y^q f(x, y) \tag{2}$$

To normalize for translation in the image plane, the image centroids are used to define the central moments. The co-ordinates of the center of gravity of the image are calculated using equation (2) and are given by:

$$\bar{x} = \frac{m_{10}}{m_{00}} \bar{y} = \frac{m_{01}}{m_{00}} \tag{3}$$

The central moments can then be defined in their discrete representation as:

$$\mu_{pq} = \sum_x \sum_y (x - \bar{x})^p (y - \bar{y})^q \tag{4}$$

The moments are further normalized for the effects of change of scale using the following formula:

$$\eta_{pq} = \frac{\mu_{pq}}{\mu_{00}^{\gamma}} \tag{5}$$

where the normalization factor $\gamma = (p+q/2)+1$. From the normalized central moments a set of seven values can be calculated and are defined by:

$$\phi_1 = \eta_{20} + \eta_{02} \tag{6}$$

$$\phi_2 = (\eta_{20} + \eta_{02})^2 + 4\eta_{11}^2 \tag{7}$$

$$\phi_3 = (\eta_{30} - 3\eta_{12})^2 + (\eta_{21} - \eta_{03})^2 \tag{8}$$

$$\phi_4 = (\eta_{30} - \eta_{12})^2 + (\eta_{21} - \eta_{03})^2 \tag{9}$$

$$\begin{aligned}\phi_5 = {} & (\eta_{30} - 3\eta_{12})(\eta_{30} + \eta_{12})[(\eta_{30} + \eta_{12})^2 \\ & - 3(\eta_{21} + \eta_{03})^2] + (3\eta_{21} - \eta_{03})(\eta_{21} + \eta_{03}) \\ & [(3\eta_{30} + \eta_{12})^2 - (\eta_{21} + \eta_{03})^2]\end{aligned} \tag{10}$$

$$\begin{aligned}\phi_6 = {} & (\eta_{20} - \eta_{02})[(\eta_{30} + \eta_{12})^2 - (\eta_{21} + \eta_{03})^2] \\ & + 4\eta_{11}(\eta_{30} + \eta_{12})(\eta_{21} + \eta_{03}) + (3\eta_{21} - \eta_{03}) \\ & (\eta_{21} + \eta_{03})[(3\eta_{30} + \eta_{12})^2 - (\eta_{21} + \eta_{03})^2]\end{aligned} \tag{11}$$

$$\begin{aligned}\phi_7 = {} & (3\eta_{21} - \eta_{03})(\eta_{30} + \eta_{12})[(\eta_{30} + \eta_{12})^2 \\ & - 3(\eta_{21} + \eta_{03})^2] + (3\eta_{12} - \eta_{30})(\eta_{21} + \eta_{03}) \\ & [3(\eta_{30} + \eta_{12})^2 - (\eta_{21} + \eta_{03})^2]\end{aligned} \tag{12}$$

These seven invariant moments, $\emptyset_I$, $1 \leq I \leq 7$, are computed over the shape boundary and its interior region.

IV. EXPERIMENTAL RESULTS

To evaluate the proposed approach, all the data samples are divided into five subsets according to the subtlety of nodule,

© 2016, IJCSE All Rights Reserved

which are subtle1 through suble5. The subtlety of the images are categorized as obvious, relatively obvious, subtle, very subtle and extremely subtle by expert radiologists. The main reason behind choosing these data sets is that they evaluate different capabilities of a pattern classifier. The experimental results are computed and the quantitative analysis is done.

Table 1 depicts the percentage recognition rate of lung cancer classification using Fuzzy Hypersphere Neural Network (FHSNN) classifier. The average recognition rate of proposed technique is 83.35, which is better as compared with statistical features, as depicted in last row of Table 1. Also, the recognition rate for subtle1, 4 and 5 are better, whereas for subtle 2 and 3 are less. The average number of hyperspheres created by proposed approach is more as compared with only statistical texture features.

Table 2 summarizes the Timing analysis of training and testing for FHSNN Classifier. The average training and testing time taken for classification is less as compared with only statistical texture features. The proposed approach takes 0.2206 and 0.2153 seconds respectively for training and testing as depicted in last row of Table 2. The average training and testing time difference is of 0.0034 and 0.0017 seconds respectively.

Table 1. Percentage Recognition Rate using FHSNN

Data Set	Statistical Texture features		Statistical and Moment Invariant features	
	Test Accuracy (%)	*No. of HPs*	*Test Accuracy (%)*	*No. of HPs*
Subtle1	66.66	16	93.75	18
Subtle2	100	14	89.68	15
Subtle3	79.16	19	58.33	17
Subtle4	50	08	75	11
Subtle5	100	08	100	08
Average	79.16	13	83.35	14

Table 2. Time Analysis

Data Set	Statistical Texture features		Statistical and Moment Invariant features	
	Training Time (Sec)	*Recall Time (Sec)*	*Training Time (Sec)*	*Recall Time (Sec)*
Subtle1	0.2220	0.2160	0.2375	0.2204
Subtle2	0.2263	0.2178	0.2103	0.2137
Subtle3	0.2311	0.2188	0.2251	0.2180
Subtle4	0.2296	0.2246	0.2151	0.2177
Subtle5	0.2110	0.2079	0.2153	0.2068
Average	0.2240	0.2170	0.2206	0.2153

Hence, this approach gives better recognition rate in terms of less training and recall time along with higher recognition rate and it can be effectively used for lung cancer or any medical image classification and detection.

V. CONCLUSION

This paper emphasis on using a combination of Statistical texture and Moment Invariant features to classify lung cancer image. The outcome of this approach concludes that the Statistical texture and Moment Invariant features in combination gives better result in terms of classification accuracy, training time and testing time.

The proposed work can be further expanded with other neural network classifiers and it will be interesting to observe the results. Also, other medical imaging data set can be collected and all the available classifiers can be compared for the optimum accuracy and performance time.

REFERENCES

[1] M.N. Gurcan, B. Sahiner, N. Petrick, H. 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.

[2] K. Kanazawa, Y. Kawata, N. Niki, H. Satoh, H. Ohmatsu, R. Kakinuma, M. Kaneko, N. Moriyama, and K. Eguchi, "Computer-Aided Diagnosis for Pulmonary Nodules Based on Helical CT Images", Compute. Med. Image Graph, Vol. 22, No. 2, pp. 157-167, 1998.

[3] K. Kanazawa, M. Kubo, and N. Niki, "Computer Aided Diagnosis System for Lung Cancer Based on Helical CT Images," International Conference on Pattern Recognition, pp. 381- 385, 1996.

[4] YSP Chiou, YMF Lure, PA Ligomenides, Neural networks image analysis and classification in hybrid lung nodule detection (HLND) system, in the Proceedings of the IEEE-SP workshop on Neural Networks for Signal Processing, 1993.

[5] Yamomoto, H. Jiang, M. Matsumoto, Y. Tateno, T. Iinuma, and T. Matsumoto, "Image Processing for Computer-Aided Diagnosis of Lung Cancer by CT (LSCT)", IEEE Workshop on Applications of Computer Vision, pp. 236 – 241, 1996.

[6] Z.H. Zhou, Y. Jiang, Y.B. Yang and S.F. Chen, "Lung cancer cell identification based on artificial neural network ensembles," Artificial Intelligence in Medicine, pp. 25 – 36, 2002.

[7] Y. Yeny, H. Helen and G.S. Yeong, "Hybrid Lung Segmentation in Chest CT Images for Computer-Aided Diagnosis", International Workshop on Enterprise networking and Computing in Healthcare Industry, pp. 378 – 383, 2005.

[8] M.G. Penedo, M.J. Carreira, A. Mosquera, and D. Cabello, "Computer-Aided Diagnosis: A Neural-Network-Based Approach to Lung Nodule Detection", IEEE Transactions on Medical Imaging, pp. 872 – 880, 1998.

© 2016, IJCSE All Rights Reserved

[9] D. Lin and C. Yan, "Lung Nodules Identification Rules Extraction with Neural Fuzzy Network", IEEE Processing on Neural Information, Vol. 4, pp. 2049- 2053, 2002.
[10] P.K. Simpson, "Fuzzy min-max neural networks-Part 1:Classification", IEEE Trans. neural networks, Vol. 3, No. 5, pp. 776 – 786, 1992.
[11] P.K. Simpson, "Fuzzy min-max neural networks-Part 2:Clustering", IEEE Trans. fuzzy systems, Vol. 1, No. 1, pp. 32 – 45, 1993.
[12] U.V. Kulkarni and T.R. Sontakke, "Fuzzy Hypersphere Neural Network Classifier", In 10th Int. IEEE Conference on fuzzy systems, University of Melbourne, Australia, 2001.
[13] U.V. Kulkarni, D.D. Doye and T.R. Sontakke, "General fuzzy hypersphere neural network", IEEE, IJCNN, pp. 2369 – 2374, 2002.
[14] D. Mundada,A. Murade, O. Vaidya and Swathi J.N., "Software fault prediction using artificial neural network and resilient back Propagation," International Journal on Computer Science and Engineering (IJCSE), Vol. 5, No.03, pp. 2319-7323, 2016.
[15] R.C. Gonzalez, R.E. Woods and S.L. Eddins. "Digital Image Processing", Addison-Wesley, M. Young, The Technical Writer's Handbook. Mill Valley, CA: University Science, 2002.
[16] Japanese Society of Radiological Technology, (JSRT), http://www.jsrt.or.jp

AUTHORS PROFILE

D.N. Sonar received the M.Sc.(CM) degree from S.R.T.M. University, Nanded in the year 2008. She is currently working as Asst. Prof. in the College of Computer Science and Information Technology, Latur, Maharashtra. Her current research area include Neural Networks and pattern recognition.

Dr.U.V. Kulkarni received the Ph.D. degree in Electronics and Computer Science Engineering from S.R.T.M. University, Nanded in year 2002. He is currently working as Professor in Department of Computer Science and Engineering, SGGS Institute of Engineering and Technology, Nanded, Maharashtra.

© 2016, IJCSE All Rights Reserved