A Review on Recent Methods for Identification of Lung Nodule

Sheetal Patel
P.G. Student
Department of Biomedical Engineering
Government Engineering College, Gandhinagar, Gujarat, India

Abstract— Lung cancer is the leading cancer among both men and women. Early detection of cancer is the most promising way to enhance a patient's chance for survival. Computer Aided Diagnosis (CAD) is one of the trusted methods in the field of medicine. CAD system assists the doctors for the diagnosis of diseases in higher degree of perfection within a short period of time. These days CAD system developed to a great extends, however it's not reached to 100% accuracy. In this article that discusses the different methodologies used for early identification of lung nodule.

Key words: Computer Aided Diagnosis, Image processing technique, ANN and SVM, Random Forest, Bagging and Adaboost classifiers

I. INTRODUCTION

According to the World Health Organization (WHO) the current Indian population is 1,270,272,105 (1.27 billion). The incidence of cancer in India is 70-90 per 100,000 population and cancer prevalence is established to be around 2,500,000 (2.5 million) with over 800,000 new cases and 5,50,000 deaths occurring each year. According to Indian Council of Medical Research (ICMR) cancer of breast with estimated 1.5 lakh (over 10 per cent of all cancers) new cases during 2016, is the number one cancer overall. Cancer of the lung is the next with estimated 1.14 lakh (83,000 in males and 31,000 in females) new cases during 2016 and 1.4 lakh cases in 2020.1

Cancer is medically knows as malignant neoplasm and simply it is an unregulated cell growth. Lung cancer is caused by uncontrollable irregular growth of cells in lung tissue. These unregulated cell growth causes tumor formation and those tumors are called as malignant tumors. The possibility of spreading lung cancer is much higher and this may spread to distant parts of the body. Cancer forms tumor, but all tumors are not cancerous. This makes more complexity for the treatment of lung cancer. The treatment methods for benign and malignant lung tumors are different and very high side effects for the treatment method of malignancy. So it is very important to diagnose whether the tumor is malignant or benign. Now a days one of the major research fields is the diagnosis of cancer with the help of medical image processing technique.

In this paper next section described the review of various methods for lung nodule identification.

II. METHODOLOGY

A CAD system for lung nodule identification using Image processing technique consists of 4 main steps including preprocessing, segmentation, feature extraction and classification.

Fig. 1: Original Lung CT-Scan

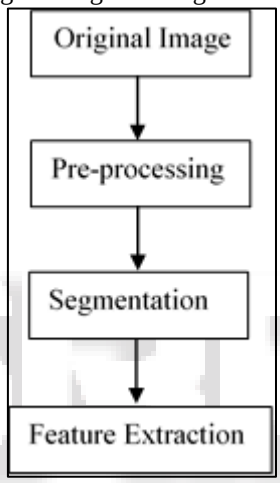

Fig. 2: CAD System

A. *Preprocessing:*

Image preprocessing is a process which eliminates primary noise and image distortion, and also enhances important features exists in CT images. By using some image enhancement methods including median filter and histogram equalization, a corrected image will be produced3.

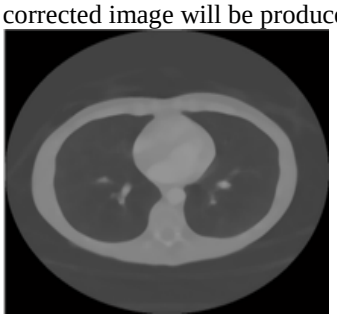

Fig. 3: Smoothing by Median Filter

B. *Segmentation:*

Segmentation the most important step in image processing. Segmentation is the division or separation of the image into meaningful regions that contents of each region have the same specifications such as pixels' intensity levels, color components, edges and texture. Segmentation is done by using region growing and thresholding methods. The segmentation based on Thresholding gives more accuracy3.

Fig. 4: Segmentation by Thresholding

C. Feature extraction:

To classify a two-dimensional (2D) pattern by using image processing algorithms, the features are obtained for each 2D pattern. The morphological features are computed by the regional descriptors of the 2D pulmonary nodule patterns based on the basic morphological shape information4. The geometric features of pulmonary nodules were obtained by using the regional descriptors of the 2D patterns based on the basic morphological shape information. The geometric features consist of the area, perimeter, diameter, solidity, eccentricity, aspect ratio, compactness, roundness, circularity, ellipticity of the patterns.

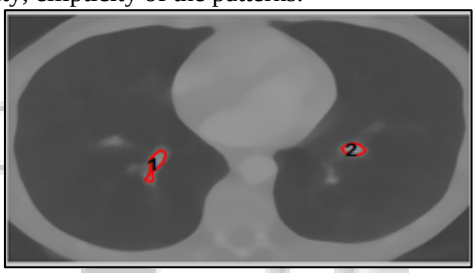

Fig. 5: Extract Features

D. Classification:

1) Artificial Neural Network (ANN):

Neural Networks are generally called as Artificial Neural Networks (ANN) is a mathematical model developed to mimic the functions of human brain. It consists of larger number of simpler units called neurons, which communicate with each other through weighted connections. The main benefit is that ANNs are non-linear classifiers, which makes them an excellent candidate for classification when underlying data relationship is not known. Also they possess is the ability to adapt and learn. They can correlate between input data and corresponding target vectors. L. narayanan et.al9 proposed a feed forward network is used for classification purpose which consists of three layers i.e. an input, one hidden and an output layer. The size of the hidden layer is chosen as 10. It uses tan- sigmoid transfer function for both hidden and output layers. The Networks are trained with 70% of the data along with a corresponding target vector. Then it is validated with 15% of the input data. Now the trained network is used to classify the tumors as either cancerous or non-cancerous. F. Taher et.al7 proposed a CAD system, compare two different classification methods, namely, ANN and SVM. The comparison demonstrates a clear superiority of the SVM classifier, where it gives the highest accuracy.

2) Support Vector Machines (SVM):

The classification is achieved through a SVM that aims to separate two data sets with maximum distance between them. The aim is to build a function f:IRK→{±1} using training data, consisting of K-dimensional patterns xi and class label yi. SVM works effectively in combination with kernel techniques. Thus the hyperplane defining the SVM corresponds to a non-linear decision boundary in the input space. If data utilizes a mapping φ(x) for Euclidean space, the training algorithm only depends on the data through dot products in such an Euclidean space, that is, on functions of the form φ(xi).φ(xj). If a Kernel function is defined as F(xi,xj)= φ(xi).φ(xj), knowing the φ function is not necessary during the training process. In the test process, a SVM is used by computing dot products of a given test point x and w,

$$f(x) = \sum_{i=1}^{Ns} \alpha i y i\, \varphi(si)\varphi(x) + w0$$
$$= \sum_{i=1}^{Ns} \alpha i y i\, F(si, x) + w0$$

Where si is the kernel function, the support vectors, respectively. The parameters αi are the solution of a quadratic optimization problem that are usually determined by the well-known Sequential Minimal Optimization (SMO) algorithm. A. Tarter5 proposed a system based on SVM and bagging SVM classifier. As a result of tests, bagging SVM is shown to be superior to a single SVM.

3) Bagging classifier:

Bagging is an unstable learning algorithm for small data set if small changes in the training data generate very diverse classifiers. L. Breiman10 proposed the bagging classifier to improve the performance by taking advantage of this effect. A single classifier could have a higher test error. Therefore the combination of classifiers may produce a lower test error than that of the single classifier because the diversity of classifiers usually compensates for errors of any single classifier. A learning algorithm combination in those small changes in the training set leads to relatively large changes in accuracy4.

4) Adaboost classifier:

Adaboost is one of the powerful methods for pattern recognition. Adaboost classifier is an ensemble classifier composed of many weak classifiers for the two-class classification problem. It generates a strong classifier by combination of weak classifiers by adaptively adjusting the weights at each loop. While the weights of the training patterns classified correctly by a weak classifier are decreased, the weights of the training patterns misclassified by the weak classifier are increased. Adaboost algorithm gives better performance effect because of its ability to generate expanding diversity4.

5) Random Sub-Space (RSS):

Random subspace method is composed of several classifiers. It utilizes a generalized algorithm of random forest classifier. While random forest algorithm is generated from decision trees, RSS may be consisted from any classifier according to the structure of problem. RSS may also be applied in classification problems with single class4.

6) Random Forest (RF):

Radom Forest was proposed by Leo Breimans in 19998. It is a new development in tree based classifiers and fast proven to be one of the most important algorithms in the machine learning systems. It is defined as a combination of tree predictors depend on the values of a random vector sampled independently and with the same distribution for all trees in the forest. It gives robust and improved results of classifications on standard data sets. It is providing very good

competition to neural networks and ensemble techniques on different classification problems. Random Forest is related to be special type of ensembles using bagging and random splitting methods to grow multiple trees. The Random Forest can predict what features are important in the classification. Also it can process efficiently on large data sets and can be utilized as an effective method to estimate missing data8.

III. Result & Discussion

The original chest CT images of patients used in the study are in size of 512x512 pixels and stored as DICOM (Digital Imaging and Communications in Medicine) format files. In preprocessing, median filter was used for smoothing the image. Thresholding based segmentation for nodule detection and features are extracted based on the shape of nodule.

Region No.	Area	Perimeter	Centroid	Equivalent Diameter	Eccentricity	Irregularity Index
1	242.0	84.8	118.0 154.2	17.6	1.0	0.4
2	181.0	49.7	274.7 139.0	15.2	0.7	0.9

Table 1: Output

References

[1] "WHO | Over 17 lakh new cancer cases in India by 2020: ICMR"
[2] P. Eskandarian, "Computer-Aided Detection of Pulmonary Nodules based on SVM in Thoracic CT Images," pp. 0–5, 2015.
[3] F. V. Farahani, A. Ahmadi, and M. H. F. Zarandi, "Lung Nodule Diagnosis from CT Images Based on Ensemble Learning," Comput. Intell. Bioinforma. Comput. Biol. (CIBCB), 2015 IEEE Conf., pp. 1–7, 2015.
[4] Tartar, A. Akan, S. Member, and N. Kilic, "A Novel Approach to Malignant - Benign Classification of Pulmonary Nodules by Using Ensemble Learning Classifiers," pp. 4651–4654, 2014.
[5] Ahmet TARTAR, Niyazi KILIÇ and Aydın AKAN, "Bagging Support Vector Machine Approaches for Pulmonary Nodule Detection *," pp. 47–50, 2013.
[6] H. Han, L. Li, F. Han, H. Zhang, W. Moore, and Z. Liang, "Vector Quantization-Based Automatic Detection of Pulmonary Nodules in Thoracic CT Images," pp. 0–3, 2013.
[7] F. Taher, N. Werghi, and H. Al-ahmad, "Computer Aided Diagnosis System for Early Lung Cancer Detection," pp. 5–8, 2015.
[8] Tarter, N. kilic, A. Akan, Senior Member, IEEE, "A New Method for Pulmonary Nodule Detection using Decission Trees," pp. 3–7, 2013.
[9] Lakshmi N arayanan A, Prof. Jeeva J.B, I. Conference, I. Systems, B. M. Engineering, and B. M. Engineering, "9. A Computer Aided Diagnosis for detection and classification of lung nodules," pp. 0–4, 2015.
[10] Kulkarni and A. Panditrao, "Classification of lung cancer stages on CT scan images using image processing," 2014 IEEE Int. Conf. Adv. Commun. Control Comput. Technol., no. 978, pp. 1384–1388, 2014.
[11] L. Breiman, "Bagging predictors," Mach. Learn., vol. 24, pp. 123–140, 1996.

