THE INTERNATIONAL JOURNAL OF SCIENCE & TECHNOLEDGE

Lung Nodule Detection Based on Classification Techniques-A Survey

S. Ramya Preethi
ME-CSE (Final Year), Muthayammal Engineering College, Rasipuram, Tamil Nadu, India
P. Deepa
Assistant Professor, Muthayammal Engineering College, Rasipuram, Tamil Nadu, India

Abstract:
Lung cancer is a major cause of cancer related deaths. Thus the identification of lung nodule is essential part for screening and diagnosis of lung cancer. The classification of four type of lung nodule in low dose computed tomography scans. i.e., Well-circumscribed, vascularised, juxta-pleural, and pleural-tail. So that classification of lung nodule is on three stages by combining the lung nodule with surrounding anatomical structures. First stage an adaptive graph patch based division is used to construct concentric multi level partition use of super pixel formulation. The second stage of the method is feature set designed to incorporate intensity, texture, and gradient information for image patch feature description use of scale-invariant feature transform, local binary pattern provides texture description of objects and Histogram of oriented gradients represents the local portions of object. And third stage is to classify the lung nodule based on SVM, ANN, k-NN, supervised, and semi supervised classifier with respect to feature descriptors. Classification of lung nodule is done using different classification and their performances are compared.

Keywords: *scale-invariant feature transform, local binary pattern, Histogram of oriented gradients, SVM, ANN, k-NN, supervised, semisupervised classifier.*

1. Introduction

Cancer is a disease characterized by uncontrolled growth of curious cells. The cancer cells types, which are classified by the cell they initially affect. The Lung Cancer remains a leading cause of humanity. Its cure rate is very low because it is usually detected at very last stages. Lung Cancer is characterized by uncontrolled cell growth originating in the lungs. The most common cause of Lung Cancer is smoking however in some cases it can also be attributed to genetic factors, gas, asbestos, smoking and air pollution. Lung cancers are detected by using chest radiograph and CT scan will be used to detect the lung nodule.

Lung nodules are small bags of tissue in the human lung and are quite common. They appear as around, white shade on a chest X-ray or computerized tomography (CT) scan. Most nodules (more than 60 percent) are not cancerous. The Lung nodule can be distorted by surrounding anatomical structures, such as vessels and the neighbouring pleura. Inter parenchymal lung nodules are more likely to be malignant than those connected with surrounding structures. If a lung nodule is new or has distorted in size, shape or appearance, your doctor may recommend further testing such as a CT scan, positron emission tomography (PET) scan, and bronchoscopy or tissue biopsy to determine if it is cancerous.

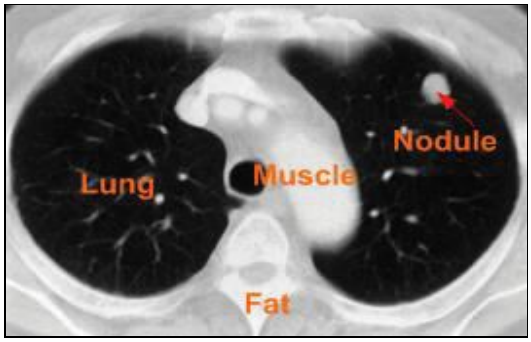

Figure 1: CT scans of lung nodule

The nodules are divided into a four types. Well-circumscribed (W) with the nodule located centrally in the lung without any connection to vasculature; vascularised (V) with the nodule located centrally in the lung but closely connected to neighbouring vessels; juxta-pleural (J) with a large portion of the nodule connected to the pleural surface; and pleural-tail (P) with the nodule near the pleural surface connected by a thin tail.

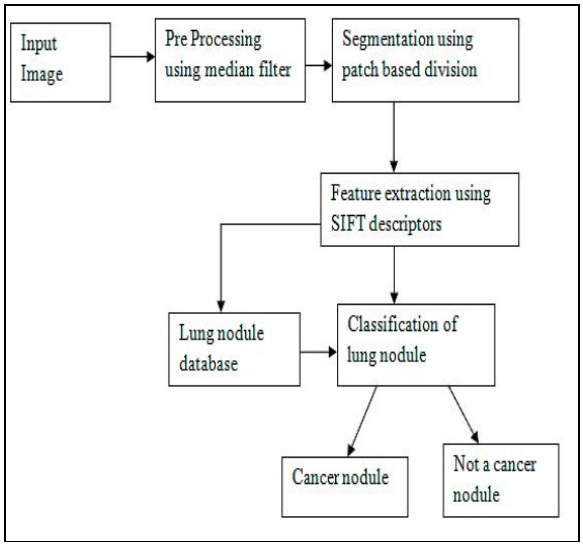

Figure 2: Steps to detection of lung nodule

2. Lung Nodule Detection

The detection and segmentation of lung nodule will provide a limited data for lung nodule classification. The performance will be improved by better feature design and a classifier. The filter based feature extraction techniques are also used to remove the noise in the image. Median Filter is a simple and powerful non-linear filter which is based order statistics. It is easy method of smoothing images. Median filter is for reduce the amount of intensity variation between one pixel and the other pixel.

3. Segmentation

The lung nodules with surrounding anatomical structures are only present in certain type of nodule. For example the nodules are similar in shape and location in W and V. So it is difficult to classify the nodule. In order to classify the adaptive graph patch based division is used which segment the nodule images into number of patches and are used to construct the multilevel partition. In that the single patch will represent the one type anatomical structures. The superpixel formulation based on quick shift algorithm is used. It will provide the lung nodule images with surrounding anatomical structures.

4. Feature Extraction

Next to describe the patch numerically, and translate into feature vector. The feature set will extract the intensity, texture, and gradient information of the image. The Scale Invariant Feature Transform (SIFT) is invariant to image translation, scaling, rotation, illumination changes, to local statistical data. For texture the description of object Local Binary Pattern (LBP) which provides the rotation invariant property. Histogram of Oriented Gradients will provide the gradient orientation in local portion of the image. It will show the effectiveness of individual image and get better the performance.

5. Classification

Next to feature extraction, the classifier will able to classify the nodule whether the nodule is cancer nodule or not. The number of classification techniques is used to classify the nodule with surrounding anatomical structures. Among that we propose the semi supervised classifier is to use the both of labelled and unlabeled images based on the contextual latent semantic analysis. It will calculate the probabilistic estimations for the rank based unlabeled relevant images.

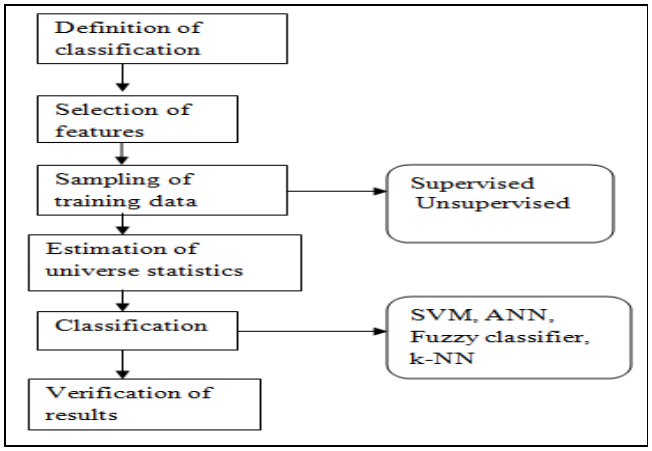

Figure 3: classification procedures

Nodule Type Classification is performed using the nearest-neighbour classifier which measure the Euclidean distance between the images and classify the lung nodule based on true positive rates. To classify the four type of lung nodule one of the data mining concept is used as SVM.

5.1 SVM

SVM is supervised learning models which use only the labelled data to classify the nodule. It takes a set of tested data against the training data and then it will classify the nodule. The hamomorphic filtering is used to enhance the edges in the image. The features can be obtaining by segmenting the nodule and perform edge detection and find the mean and variance of resultant edges of nodule. Now SVM classify the nodule the idea behind that is mapping input vectors X into a high dimensional feature space through non-linear mapping. The performance are calculated based on some criteria as

1. True Positive (TP): the percentage of nodules classified as nodules.
2. False Positive (FP): the percentage of non-nodules classified as nodules.
3. False Negative (FN): the percentage of nodules classified as non-nodules.
4. True Negative (TN): the percentage of non-nodules classified as non-nodules.

In addition to that evaluate their classifiers in terms of Sensitivity, Accuracy and Specificity. These criterions are defined to be:

$$\text{Sensitivity} = \frac{TP}{TP+FN}$$

$$\text{Accuracy} = \frac{TP+TN}{TP+TN+FP+FN}$$

$$\text{Specificity} = \frac{TN}{TN+FP}$$

The identification of nodule use of SVM classifier found be 87.5%

5.2 ANN (Artificial Neural Network)

Artificial Neural Network is developed for diagnosis and classification of candidate nodules obtained from application of diagnostic rules. ANN work by training and testing process applied to it. The ANN [10] system consists of three main layers input layer, hidden layer, output layer. The network is trained using Back propagation (BPA) algorithm. The idea of BPA is to reduce error produced by the difference between actual output and expected result. Initially the best optimized ANN is obtained by varying various parameters of network like hidden nodes, training percentage for ANN, number of epochs.

Here the input is converted into gray image and mean filter is to remove the Gaussian white noise and the segmentation of the nodule done by Otsu's threshold method and detect the edge nodule use of canny edge detector. Now the image features are extracted use of morphological processing and the image is ready for classification and we use the confusion matrix to calculate the performance of classifier. It is a table that consider the performance of a learning algorithm the column of a matrix defines the predicted class of the nodule. And row defines the actual class. The nodules are classified and it will improve the performance of 89.3%.

5.3. K-NN Classification

The k-Nearest Neighbours classification is a non-parametric method used for classification and regression. In both cases, the input consists of the k closest training examples in the feature space. The output depends on whether k-NN is used for classification or regression: In k-NN classification, the output is a class membership. A nodule is classified by a popular vote of its neighbours, with the nodule being assigned to the class most common among its k nearest neighbours (k is a positive integer, typically small). If k = 1, then the object is simply assigned to the class of that single nearest neighbour. In recent period, the median filter is to remove the noise in the image. They were used to extract the image of multiscale filter based feature extraction technique use of Gaussians functions and Laplacian transform. Now the classifier will classify the image. It will improve the performance as 85.2% based on image intensity, and gradient information.

The k-NN use stored database, distance between the nodules, value of k nearest neighbour are retrieved. Distance between k is calculated and classify the unknown structures nodule use of nearest neighbour to determine which class the nodule belongs to and are done by Euclidean distance. The KNN classifier works as follows:

- Calculate the distance between two points: In Cartesian coordinates, if a = (a1, a2,..., an) and b = (b1, b2,..., bn) are the two points in Euclidean n-space, then the distance from a to b, or from a to b is given by Euclidean distance, $d(a,b) = \sqrt{\sum_i (a_i - b_i)^2}$
- Find the class from nearest neighbour list
  - The k-nearest neighbours
  - Weigh the vote according to distance where weight factor, $w = 1/d^2$
- Choosing the k value of:
  - If k is too small, sensitive to noise points
  - If k is too large, neighbourhood may include points from other classes
- Kd-tree construction algorithm
  - Select the x or y dimension
  - Partition the space into two
  - Repeat recursively in the two partitions provided that there are enough points.

5.4. Supervised Classification

Supervised classifications use only the labelled information to classify the images. The training data consist of a set of training examples. In supervised classification, each example is a pair consisting of an input object (typically a vector) and a required output value (also called the supervisory signal). A supervised learning algorithm analyzes the training data and produces a secondary function, which can be used for mapping new examples.

Here the lung nodules are classified use of CT images. First the patch based division is to segment the image into number of patches use superpixel formulation based on quick shift algorithm.

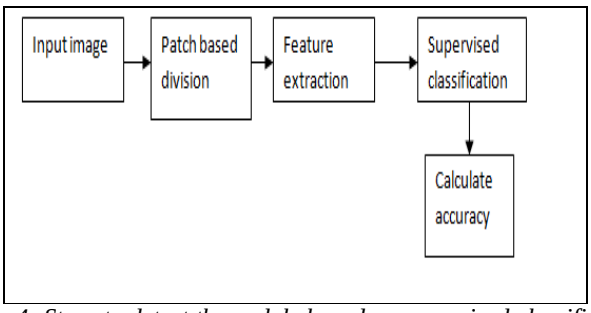

Figure 4: Steps to detect the nodule based on supervised classification

6. Conclusion

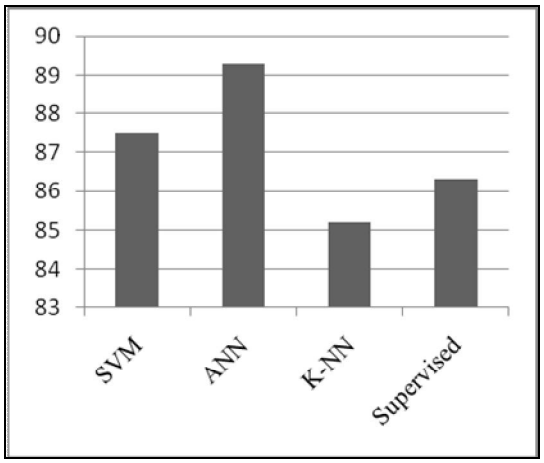

Figure 5: comparisons of types of classification

In this study a novel method for lung nodule image classification into four types, well-circumscribed, juxta-vascular, juxta-pleural and pleural-tail. In this survey discussed feature descriptor such as SIFT, LBP, and HOG is designed to describe the processed image quantitatively according to the characteristics of each type of lung nodules. For classification, SVM, ANN, K-NN, and supervised classifier achieve good classification result. And the classification rates obtained for SVM, ANN, K-NN, and supervised classifier were 87.5%, 89.3%, 85.2%, and 86.7%. The classifications are measured in terms of sensitivity, accuracy, specificity, precision and recall. In above classification techniques the ANN classifier shows better results. In future to improve classification result to deriving graph-based distances that emphasize labelled and unlabeled images classification propose a graph- based semi-supervised learning method exploiting the Unlabeled image classification.

7. References

1) prashant naresh, Dr. Rajashree shettar “early detection of lung cancer using neural network techniques”, international journal of engineering research and applications, ISSN:2248-9622,vol.4, issue 8, august 2014 pp 78-83.
2) Fan Zhang, “ lung nodule classification with multilevel patch based context analysis”, IEEE transaction on biomedical engineering, vol. 61, no-4, apirl 2014.
3) Ms. Swati P. Tidke, “classification of lung tumor using SVM” international journal of computational engineering research” vol 2, issue 5 September 2012.
4) Wail A.H Mousa and Mohammad A. U Khan “Lung Nodule Classification Utilizing Support Vector Machines” IEEE international geosciences and remote sensing symposium” June 2002.
5) [5] Anam Tariq , M. Usman Akram and M. Younus Javed ” Lung Nodule Detection in CT Images using Neuro Fuzzy Classifier” IEEE 2013 Fourth International Workshop on Computational Intelligence in Medical Imaging (CIMI) – Singapore 2013.
6) M.Gomathi, Dr.P.Thangaraj “An Effective Classification Of Benign And Malignant Nodules Using Support Vector Machine” Volume 3, No. 7, July 2012 Journal of Global Research in Computer Science.

7) M. Gomathi and Dr. P. Thangaraj "Lung Nodule Detection using a Neural Classifier" IACSIT International Journal of Engineering and Technology, Vol.2, No.3, June 2010, ISSN: 1793-8236.
8) Pooja Kamavisdar1, Sonam Saluja2, Sonu Agrawal "A Survey on Image Classification Approaches and Techniques" International Journal of Advanced Research in Computer and Communication Engineering Vol. 2, Issue 1, January 2013
9) S. Liu, L. Zhang, W. Cai, Y. Song, L. Wen, and D. Feng, "A supervised multiview spectral embedding method for neuroimaging classification," in Proc. Int. Conf. Image Process., 2013, pp. 602–605.
10) J. Yao, A. Dwyer, R. M. Summers, and D. J. Mollura, "Computer-aided diagnosis of pulmonary infections using texture analysis and support vector machine classification," Acad. Radiol., vol. 18, no. 3, pp. 306–314, 2011.
11) F. Zhang, Y. Song, W. Cai, Y. Zhou, M. Fulham, S. Eberl, S. Shan, and D. Feng, "A ranking-based lung nodule image classification method using unlabeled image knowledge," in Proc. Int. Symp. Biomed. Imag., 2014.
12) F. Zhang,W. Cai, Y. Song, M.-Z. Lee, S. Shan, and D. Feng, "Overlapping node discovery for improving classification of lung nodules,"
13) D. Wu, L. Lu, J. Bi, Y. Shinagawa, K. Boyer, A. Krishnan, and M. Salganicoff, "Stratified learning of local anatomical context for lung nodules in CT images," in Proc. Comput. Vis. Pattern Recog., 2010, pp. 2791–2798.
14) Y. Song, W. Cai, Y. Zhou, and D. Feng, "Feature-based image patch approximation for lung tissue classification," IEEE Trans. Med. Imag.,vol. 32, no. 4, pp. 797–808, Apr. 2013.