Classification of Lung Tumor from CT Images using Computer Aided Diagnosis Scheme

Amritha Vijayan
PG Student
Department of Electronics & Communication Engineering
Marian Engineering College, Trivandrum, India

Minnu Jayan, C
Assistant Professor
Department of Electronics & Communication Engineering
Marian Engineering College, Trivandrum, India

Abstract

Lung cancer is the second most prevalent type of cancer. Early detection and classification of lung tumors helps in better survival. The paper presents a computer aided diagnosis (CAD) system for chest computed tomography (CT) images that focus on a new algorithm to classify lung tumor into benign or malignant using advanced neural network based classifier. Statistical textural features are extracted which aids in better classification. The accuracy of the algorithm is also computed. Methods: Input image is pre-processed by median filtering and segmented to extract the portion of lungs. This is followed by identifying the tumor objects alone from the region of interest of the lungs. Statistical textural features are extracted only from the tumor objects. Classification of tumor into benign and malignant tumor is done using feed forward back propagation neural network classifier. The result shows proposed algorithm gives classification accuracy above 99% which suggests that the developed CAD system has greater potential for automatic classification of lung tumor.

Keywords: Artificial Neural Network (ANN), Benign Tumor, Computer Aided Diagnosis (CAD), Computerized Tomographic Images (CT), Levenberg Marquardt Training Algorithm, Malignant Tumor

I. INTRODUCTION

Lung cancer is the second most prevalent type of cancer in both men and women [1]. It was estimated by cancer society that 14% of all new cancers are lung cancer. Approximately one out of four cancer deaths are due to lung cancer. It results in higher mortality rate in recent years with rate higher rate than breast cancer, colon cancer and many other type of cancers. It accounts to 1.59 million deaths per year. Most people who develop lung cancer do not develop clinical symptoms until it become advanced as many people mistake them for other problems such as infections [2]. Being a relatively large body human organ that may involve many other chronic diseases, when tumor is found it had already resulted in 95% loss of pulmonary function. Early detection of lung cancer helps in early treatment and increases survival rate up to 80%. The common symptoms of lung cancer are cough, dyspnea, hemoptysis and systemic symptoms such as weight loss and anorexia [3]. It is mainly caused due to smoking, exposure to certain type of chemicals, family history. Abnormal rates of cell division of lung tissues results in lung tumor. There are mainly two type of lung tumor based on whether the cells invade other parts of body [4]. It is benign tumor if it does not invade its surrounding tissue and malignant if it spreads to other parts of the body. If the cells of the malignant tumor reaches other parts of the body through blood or lymphatics, that area will also be affected (metastasis). Hence it is important to find the malignancy of tumor before it spreads, resulting in multiple cancer and death. Based on histology lung cancer is classified into small cell and non-small cell lung cancer. The main imaging modalities to image lungs are chest radiography, computed tomography (CT), Magnetic resonance Imaging (MRI scan) [5]. Screening using CT is on the rise due to low cost, robustness and quality. A part of detected tumor can be cancerous. But for screening purposes CT with high radiation is harmful and hence recently developed low dose CT (LDCT) with less radiation. For detecting extra-encephalic malignant tumors and unexpected lesions, CT images shows high performance up to 30% [6]. It make use of the image processing techniques to assist radiologists to interpret the medical images and helps in diagnosis with high accuracy by avoiding chances of human error [7,8]. Based on the quantitative tumor related features even small lesions can be detected using CAD scheme. This paper proposes computer aided diagnosis system for classification of lung tumor with high accuracy using neural network based classifier.

II. RELATED WORK

To state the problem of automatic detection and classification of lung tumors, researchers had proposed several methods. The aim of all these researchers is to develop a system which can predict, detect and classify tumor in its early stage with high accuracy. However, all of their works involved these main steps. The steps are pre-processing, segmentation, feature extraction and classification. Some of the major research works are summarized below.

Tiwari [9] provided a review of the various image processing techniques and pattern recognition based approaches by the previous researchers in the prediction of lung cancer. A comparison of different classification techniques and their accuracies proposed by different researchers were also presented.
The CAD system developed by Jacobs et al [10] focus on the detection of sub-solid pulmonary nodules by using 128 image features. Besides using the existing features such as intensity, shape and texture features, context based features were also extracted. The combination of these set of features significantly improved the classifier performance. The developed system finally reaches a sensitivity of 80%.

Samuel H Hawkins et al [11] developed a CAD scheme using selected features from CT images for the prediction of lung cancer using different classifiers. Rule based, decision tree, Naive Bayes and support vector machine classifiers are compared for their accuracies in cancer prediction. Among them, decision tree classifier is found with best predictive accuracy.

Computer aided classification method by Frahani et al [12] is based on an ensemble of three classifiers. The classifiers includes multilayer perceptron, K-Nearest neighbor and support vector machines. Shape related features such as roundness, compactness, circularity, and eccentricity were extracted from the segmented part of lungs. In order to combine the result of these classifiers, majority voting method was used. The ensemble system showed high performance in the detection of pulmonary nodules from CT images.

Sheeraz Akram et al [13] had proposed an automatic pulmonary nodule detection system with the aid of artificial neural networks. The CAD system used a standard Lung Image Consortium Datasbase (LIDC). The lung volume is segmented out and the candidate nodules are extracted from the lungs based on the rules which are formed using the characteristics of nodules. The neural network is trained using geometric and intensity based statistical features. The developed CAD system has highest performance when compared to other classifier based CAD systems.

F. Paulin et al [14] studied the classification of breast cancer using different feedforward neural networks. The network performance is computed using various back propagation training algorithms. The highest accuracy of 99.28% is obtained when using the Levenberg Marquardt (LM) training algorithm. Nawaz Khan et al [15] provides a case study for the performance evaluation of LM algorithm in the error reduction of diabetes condition classification. A dynamically constructed neural network is trained and tested by different back propagation algorithms and a comparison of their output is done in terms of accuracy, sensitivity and specificity. From the implementation, experimentation and performance evaluation, it is found that LM algorithm outperforms all other training algorithms.

III. PROPOSED METHOD

The proposed CAD system is developed to classify the lung tumor into benign or malignant. The aim of the system is to minimize the computational time and to obtain the output with high algorithm accuracy. The CAD system is implemented in MATLAB using image processing and highly advanced neural network classifier. The classifier is validated using leave-one-out validation technique. The proposed CAD system consist of various stages which includes the image acquisition, pre-processing, segmentation, morphological filtering, region of interest selection, identifying tumor regions from the region of interest, feature extraction and classification. The block diagram of proposed system is shown in Fig.1.

A. Image Acquisition

The images are obtained from the public access research lung image database that was available to be used for performance evaluation of different CAD systems. The database consist of 35 low dose CT scans which were obtained in a single breath hold.

B. Pre-Processing

This stage aims to enhance the quality of image so as to increase the precision and accuracy in the later stages of image processing. Pre-processing removes the noise occurred during the image acquisition process. Using median filter, the unwanted noise components are removed while preserving the edges.

C. Segmentation

It is an important step in the image processing of medical images. It is done to extract only the portion of lungs by separating the objects and boundaries. It is done by initially converting the pre-processed grayscale image into binary image by thresholding. The binary image is then filled to obtain the boundary region. The filled image is subtracted from the binary image to extract only the portions of lungs. The complement of the extracted portion of lungs is then obtained.

D. Morphological Filtering

The contrast and quality of the extracted portions of lungs is enhanced by morphological filtering [16]. This is done by erosion followed by dilation. Erosion shrinks and removes objects of certain size and shape described by the structuring element. While dilation grows objects with certain size and shape. If the pixels have low intensity neighborhood, then their intensity values are decreased by erosion. While dilation does the opposite by increasing the intensity values of pixels with high intensity neighborhoods.
E. Region of Interest (ROI)

Region of interest is the area of the lungs where further focus of image analysis is required. It is graphically selected from the window displaying the image as a polygonal area of interest.

F. Identifying the tumor regions from the region of interest

The tumor regions from the region of interest are identified for further feature extraction and classification. This is done by extraction of connected objects from the displayed region of interest by making use of the fact that the tumor regions tend to be brighter than rest of the regions in the selected part of left or right lung.

G. Feature Extraction

Among the different type of image features available, histogram based statistical textural features are extracted. Textural features are selected as their orientation is different for different classes. The first order statistical features are extracted in this study [17]. The statistical features are extracted from the tumor regions alone. These features are used for training. The various first order statistical features used in the study are

1) Mean
   It is the measure of central tendency of probability distribution. The mean of all the pixel values of the tumor region is calculated. The mean of the tumor region is calculated by using (1)
   \[ \mu = \frac{1}{MN} \sum_{i=1}^{M} \sum_{j=1}^{N} P(i,j) \]  
   Where P(i,j) is the intensity value at the point (i,j) and M*N is the size of the image.

2) Variance
   It is comparatively more incredible feature than mean. It gives the amount of grey level fluctuations from the mean grey level values. The variance of the tumor region is calculated by using (2).
   \[ \sigma^2 = \frac{1}{MN} \sum_{i=1}^{M} \sum_{j=1}^{N} (P(i,j) - \mu)^2 \]  

3) Skewness
   It is the measure of asymmetry of probability distribution. It measures if there is a wider range of either lighter pixels or darker pixels. The skewness of the tumor region is calculated by using (3).
   \[ S = \frac{1}{MN} \sum_{i=1}^{M} \sum_{j=1}^{N} \left( \frac{P(i,j) - \mu}{\sigma} \right)^3 \]  

4) Kurtosis
   It is a measure of peakness of probability distribution. Also it is a measure of the shape descriptor of probability distribution. The kurtosis of the tumor region is calculated by using (4).
   \[ K = \frac{1}{MN} \sum_{i=1}^{M} \sum_{j=1}^{N} \left( \frac{P(i,j) - \mu}{\sigma} \right)^4 \]  

H. Classification of Tumor

Classification of tumor into benign or malignant type is done using feedforward backpropagation neural network. It outperforms other classifiers in terms of classification accuracy for binary classification [18]. Back propagation network consist of input layer, hidden layer and output layer. Back propagation is based on error correction rule and the weights are adjusted by back propagating the associated error. The input database is divided into training, testing and validation set. The classifier is trained using the statistical textural features of the tumor using the training dataset. Levenberg Marquardt backpropagation algorithm is used for training the classifier. It is the fastest training algorithm available in the toolbox and helps in obtaining highly optimized output [19]. Training the classifier is followed by testing the classifier with a new test input. Classification of tumor in the region of interest of the test input is done by recognizing trained patterns by the network. Validation dataset is used to measure the generalization capability of network to the new test inputs after training. Validation is done by leave-one-out validation technique. When the generalization stops improving, the training is stopped. The algorithm accuracy is computed and performance of the network is plotted with mean square error (mmse) against the number of epochs. The performance plot shows the rapid drop in
mean square error as the network learns. In the performance plot, blue line shows the decreasing error on the training data while the red line shows the error on test data. The green line shows the decreasing error on validation data as the network generalizes to new data.

**IV. Results**

The given image of lungs with tumor is converted initially into a binary image as given by Fig. 4a). The binary image is filtered using median filter. The filtered image is then filled to obtain the boundary region. The obtained boundary region is subtracted from the binary image to extract the lungs shown in Fig. 4b). The complement of Fig. 4b) is taken and complemented image is morphologically filtered and the finally obtained image is as shown in Fig. 4c). After morphological filtering the region of interest is selected given by Fig 4d). Then tumor regions are identified from the region of interest shown in Fig. 4e). It is followed by extraction of features from the tumor as shown in Fig. 4f). It is followed by the classification of tumor into benign or malignant as shown in Fig. 4g). The performance plot is shown in Fig. 4h) with training stopped at epoch 4 as generalization stops improving.
V. CONCLUSION

The result shows that the developed CAD system gives classification accuracy above 99% which suggests that the developed CAD system has greater potential for automatic classification of lung tumor.

REFERENCES


