Recent Trends in Computational Intelligence & Image processing - RICP 2017
23 February 2017, India
College of Engineering, Vadakara, Calicut, India

Lung Cancer Detection Using Neural Network

Athira PK[1], Rajin R[2], Sruthi[3]

[1]Assistant professor, Dept of computer science, college of engineering, vadakara, kerala
[2]Assistant professor, Dept of computer science, college of engineering, vadakara, kerala
[3]Assistant professor, Dept of computer science, ACKM ICA, Thrissur

Abstract: ***Image processing techniques are widely used in several medical areas for image improvement in earlier detection and treatment stages, where the time factor is very important to discover the abnormality issues in target images, especially in various cancer tumours such as lung cancer, breast cancer, etc. Lung cancer has been attracting the attention of medical and sciatic communities in the latest years because of its high prevalence allied with the difficult treatment. Statistics from 2008 indicate that lung cancer, throughout world, is the one that attacks the greatest number of people. Detection of lung cancer is very important for successful treatment. To cure the disease, its detection in its earlier phase is essential. Therefore, the need to develop new techniques for the detection of cancer nodules is on an all time increase. Image processing seems to be a supportive tool to solve this problem. lung cancer detection using image processing mainly carried out in two steps. Image processing and classification. In image processing image preprocessing, segmentation, feature extraction are carried out. And neural network is to classify the cancer nodule.***

Keywords: Cancer Detection, Image processing, Feature extraction, neural network

1. Introduction

Lung cancer is a cancer that starts in the lungs. When a person has lung cancer, they have abnormal cells that cluster together to form a tumor. Unlike normal cells, cancer cells grow without order or control and destroy the healthy lung tissue around them. Each person's lung cancer is unique which is why it is important to know the treatment options available for your particular cancer. Nearly 400, 000 people in the US are living with lung cancer. 81% of those living with lung cancer are over the age 60. The disease ails the elderly the most severely. The survival rate of lung cancer (17%) is drastically lower than breast cancer, with a survival rate of 90% and prostate with nearly 100%. Only 15% of lung cancers are caught at an early stage. When it has spread to other organs, the survival rate drops to 3.5%. More than half of those with lung cancer pass away within a year of diagnosis

A CT scan uses x-rays to make detailed cross-sectional images of your body. Instead of taking one picture, like a regular x-ray, a CT scanner takes many pictures as it rotates around you while you lie on a table. A computer then combines these pictures into images of slices of the part of your body being studied.

CT scans are more likely to show lung tumors than routine chest x-rays. They can also show the size, shape, and position of any lung tumors and can help find enlarged lymph nodes that might contain cancer that has spread from the lung. When a low-dose CT scan of the chest is done for lung cancer screening, it’s common to find small, abnormal areas (called nodules or masses) in the lungs, especially in current or former smokers. Most lung nodules seen on CT scans are not cancer. They are more often the result of old infections, scar tissue, or other causes.

If these defects are identify at early stage, then proper treatment and care can be given. But the precentage of detection of lung cancer at early stage is very less. Recently, image processing techniques are widely used in several medical areas for image improvement in earlier detection and treatment stages, where the time factor is very important to discover the abnormality issues in target images, especially in various cancer tumours such as lung cancer, breast cancer, etc.

2. Research Elaboration

Lung cancer is one of the major causes of deaths in children. A proper diagnosis can avoid such problem. Unfortunately, the percenatge of dtection is very less from ct scan. A system for automated medical diagnosis from CT image would enhance the accuracy of the diagnosis and reduce the cost effects. The proposed system is designed and developed by using MATLAB‟s GUI. Initially the lung region is extracted from the CT image by performing thresholding and morphological reconstruction. Then from the extracted lung region the nodules are segmented using global thresholding and morphological operation. The segmented nodules are used for feature extraction. Then the GLCM (Gray Level Co-occurrence Matrix) features are extracted. The features extracted are given for classification. A computer aided classification method in computed tomography (CT) images of lungs developed using artificial neural network for the detection of cancerous cells. In this method, two major steps are involved:

(1) Image Processing
(2) Classification Process

2.1 Image Processing

Image processing procedures play the most crucial part. An image obtained can be differ from its pixel, resolution, gray scale, and many else due to different sources where the image is obtained. Thus the interpretation of the chest x-ray needed to be done first as the essential skills for everyone involved in clinical medicine, and also as a systemic method to examine the image required.

2.1.1 Image Pre-Processing

The original image format is in PNG and in various sizes and contrast. Therefore, all images have been undergoing several pre-processing process in order to standardize all the image characteristics. Image pre-processing process involved are cropping, resizing, and contrast adjustment are done based on the histogram analysis made. Before any procedures made, all the images have been converted into gray scale images as to facilitate intensity analysis.

2.1.2 Segmentation

Lung field segmentation procedures have been proposed due to the approximate intensity value between the lung cancer nodule and the unwanted background region. The approximation of this intensity value makes it difficult to perform perfect nodule detection. The most important step for the process is thresholding, followed by morphological operation to remove unwanted noise. Subtraction of image is done for obtaining the segmented lung. After extracting the lung region from the original image, the lung nodules are segmented using thresholding operation.

2.1.3 Feature Extraction

The features act as the basis for classification process. Regarding on the type of the image processed, which is binary image, the only colour presented is black and white.The features are extracted from the region of interest. The region of interest is the segmented single slices containing 2 lungs. This extracted features are given for classification process.

2.2 Classification Process

ANN classification work by the training and the testing process is applied to it. The network of the ANN consists of three main layers that are input, hidden node and output layer. In training the ANN network, back propagation (BP) procedures will be used. BP algorithm is based on the error correction rule. Error propagates via forward and backward pass where weight is fixed and adjusted. Finally, a set of outputs is produced as the actual response of the network.
The input data for the network is obtained from the feature extracted from the images, while the output for the network will be the classification of the CT images. The data have been normalized with respect to their maximum value. This is because it is proved that the ANN works better with input and output of the network in the range of 0 to 1.

3. System Architecture

The system presents four basic steps. The first step starts with taking a gathering of CT images from the available database. The second step is image segmentation which uses Otsu's thresholding method and the third step contains the calculation of feature extraction. The final step is the classification of disease stages using neural network. Figure 3.1 illustrates a block diagram of lung cancer nodules feature extraction and classification of this system.

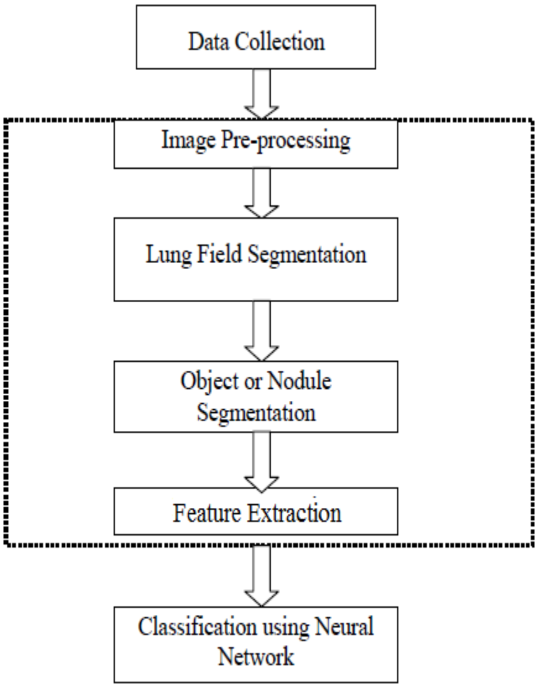

Figure 1: Architecture of proposed method

3.1 Data Collection And Image Preprocessing

The images are collected from a database of Lung Image Database Consortium (LIDC). Computed Tomography (CT) images have better clarity, low noise and distortion for lung diagnosis. So, CT scan of lung images are given as input for this system. Dimensions of images are 512x512 pixels in size. CT images are collected including both men and women. The average age of the patients considered is 64.2 years (age of the youngest patient is 18 years and the oldest patient is 85 years). The low dose CT scan images are obtained at kilo voltage peak distribution of 120–140 KVp with a current varying from 25 to 40 mAs depending upon the age of the patient. Figure 3.2 shows the original CT lung image with nodule.

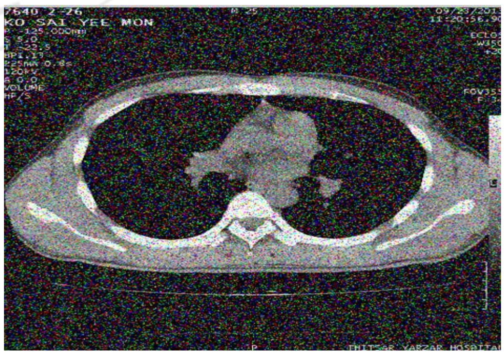

Figure 2: Original CT lung image with nodule

The original image format is PNG and varies in size and contrast. Therefore, all the images have to undergo several pre-processing process in order to standardize all the image characteristics. Image pre-processing processes involved are cropping, resizing, and contrast adjustment are done based on the histogram analysis made. Noise such as white noise, salt and pepper noises etc are also eliminated in this stage. Before any procedures made, all the images have been converted into gray scale images as to facilitate intensity analysis. The gray scale image is first converted to binary image. All the pixels in the input image with a intensity greater than a threshold level is replaced with value '1' and all pixel values with a intensity less than threshold level is replaced with value '0'. The threshold level is calculated by Otsu method. The Otsu method chooses the threshold level to minimize the intraclass variance of the black and white pixels.

The images are collected from a database of Lung Image Database Consortium (LIDC). Computed tomography (CT) images have better clarity, low noise and distortion for lung diagnosis. So, CT scan of lung images are given as input for this system. Dimensions of images are 512x512 pixels in size. The input of CT image contains noises such as white noise, salt and pepper noises etc. Therefore, image preprocessing stage is needed to eliminate noises.

3.2 Lung Field and Nodule Segmentation

The morphological opening operation is performed to the binary image with a structuring element. The structuring element is a shape, used to probe or interact with a given image, with the purpose of drawing conclusions on how this shape fits or misses the shapes in the image. The image is then inverted and clear border operation is performed. The clear border operation suppresses structures that are lighter than their surroundings and that are connected to the border of the image. The edge of the binary image is detected and which is subtracted from the filled image then we get the lung field. Main advantage of morphological operation is their speed and simplicity of implication.

Once the lung have been segmented from the image, nodule present is next to be detected. As for lung nodule detection procedures, several operations will again be applied such as the thresholding, and also the morphological operation of removing unwanted pixel in an image. Other morphological operation such as erosion and dilation operation will also be implemented. Erosion operator makes a region smaller while dilation operator enlarges a region.

3.3 Feature Extraction

These features act as the basis for classification process. Regarding on the type of the image processed, which is binary image, the only colour presented is black and white. The GLCM features are extracted from the segmented image. A statistical method of examining texture that considers the spatial relationship of pixels is the gray-level co-occurrence matrix (GLCM), also known as the gray-level spatial dependence matrix. The GLCM functions characterize the texture of an image by calculating how often pairs of pixel with specific values and in a specified spatial relationship occur in an image, creating a GLCM, and then extracting statistical measures from this matrix.

3.4 Classification Using Neural Network

ANN classification work by the training and the testing process is applied to it. The network of the ANN consists of three main layers that are input, hidden node and output layer. In training the ANN network, back propagation (BP) procedures will be used. BP algorithm is based on the error correction rule. Error propagates via forward and backward pass where weight is fixed and adjusted. Finally, a set of outputs is produced as the actual response of the network. Figure 3.3 shows

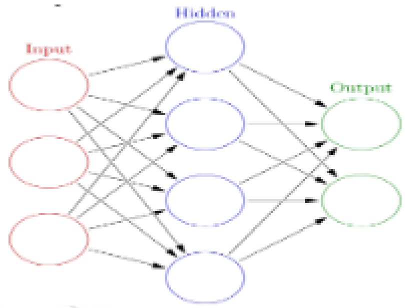

Figure 3: Neural network

The input data for the network is obtained from the feature extracted from the images, while the output for the network will be the classification of the CT images. The data have been normalized with respect to their maximum value. This is because it is proved that the ANN works better with input and output of the network in the range of 0 to 1.

The back propagation is a systematic method of training multilayer neural networks in a supervised manner. The back propagation method, also known as the error back propagation algorithm, is based on the error-correction learning rule. The back propagation network consists of at least three layers of units: an input layer, at least one intermediate hidden layer and one output layer. The units are connected in feed forward fashion with inputs units connected to the hidden layer units and the hidden layer units are connected to the output layer units. An input pattern is forwarded to the output through input to hidden and hidden to output weights. The output of the network is the classification decision.

4. Conclusion

Lung cancer detection technique is proposed here using different image processing techniques. It primarily uses CT images. Images are preprocessed first. The image segmentation is done by Otsu method. Then features are calculated. Finally A three layer neural networks is implemented for classification of images. The proposed technique gives very promising results comparing with other used techniques.

References

[1] http://www.ncbi.nlm.nih.gov
[2] http://www.cancer.org
[3] Lung Cancer Database, Available at: https://eddie.via.cornell.edu/cgibin/ datac/signon.cgi, (accessed July 2011).

[4] Gonzalez R.C., Woods R.E., Digital Image Processing, Upper Saddle River, NJ Prentice Hall, 2008.
[5] Cristobal G., Navarro. R., Space and frequency variant image enhancement based in Gabor representation, Pattern Recognition Letters, Elsevier, 1994, 15, p. 273-277.
[6] Krishan A., Evaluation of Gabor filter parameters for image enhancement and segmentation, in Electronic Instrumentation and Control Engineering, Master. Punjab: Thapar University, 2009, p. 126.
[7] Nunes É.D.O., Pérez M.G., Medical Image Segmentation by Multilevel Thresholding Based on Histogram Difference, presented at 17th International Conference on Systems, Signals and Image Processing, 2010.
[8] Venkateshwarlu K., Image Enhancement using Fuzzy Inference System, in Computer Science & Engineering, Master thesis, 2010.
[9] Disha Sharma, Gagandeep Jindal, Identifying Lung Cancer Using Image Processing Techniques, International Conference on Computational Techniques and Artificial Intelligence (ICCTAI'2011), 17(1), 2011, 872-880.
[10] Anita chaudhary, Sonit Sukhraj Singh, International Transaction on Computing Sciences, 4, 2012.
[11] Disha Sharma, Gaga deep Jindal, International Conference on Computational Techniques and Artificial Intelligence (ICCTAI'2011), 17, 2011, 872-880.
[12] American Cancer Society, Cancer Statistics, CA, A Cancer Journal for Clinicians, 55, 2005, 10-30

Author Profile

Rajin R received B-tech degree in computer science and engineering from CUSAT in 2012 and M-tech from Kannur University in 2015. He is currently working as Asst. Professor at College of Engineering Vadakara, Kerala

Athira P K received B-tech degree in computer science and engineering from Calicut University in 2013 and M-tech from MG University in 2015. She is currently working as Asst. Professor at College of Engineering Vadakara, Kerala