www.currentsciences.info

Research

International Journal of Current Trends in Science and Technology

Online ISSN: 0976-9730

Print ISSN: 0976-9498 **Section: Information Technology**

A Survey on Lung Cancer Detection in CT scans Images Using Image Processing Techniques

Jayshree Talukdar[1], Dr. Parismita Sarma[2]

[1]Department of Information Technology, Gauhati University

[2]Assistant Professor, Gauhati University Department of Information Technology, Gauhati University

Received 2018-02-10; Accepted 2018-03-13

DOI: http://currentsciences.info/10.15520/ctst.v8i03.388.pdf

Abstract:

Lung cancer is one of the most common form of cancers worldwide, and is responsible for a large number of deaths. Human radiologists typically uses low dose CT scans to assess individual's risk of lung cancer. The presence of tissue growths called "nodules" that are a common precursor to cancer. However, even for highly trained radiologists, detecting nodules and predicting their relationship to cancer are challenging tasks, leading to both false positive and false negative results that can adversely affect patient health. Also a constant pressure on them to analyses a huge amount of data and making a decision as quickly as possible based on the analysis. So a possible way to decrease this burden on radiologist is by developing a Computer aided diagnosis system that can learn features quickly. In this paper we are highlighting use of image processing techniques, Deep learning algorithms and convolutional networks for analyzing medical images of lung cancer disease.

Keywords: Image Preprocessing, Segmentation, Otsu's Thresholding,

1. Introduction:

Lung cancer is the most deadly disease causing large scale deaths in human. For a healthy person, cells in the lungs divide and reproduce at a controlled rate to repair worn-out or injured tissues. Lung cancer develops when cells inside the lungs multiply at an uncontrolled rate. These abnormal tissue masses are called tumors. Tumors are either non-cancerous (benign) or cancerous (malignant). The major proportion of deaths in cancer is due to improper primary diagnosis that is raising the need for Computer Aided Diagnosis (CAD).

Surgery, radiation therapy and chemotherapy are used in the treatment of lung carcinoma. In spite of that, the five-year survival rate for all stages combined is only 14%. However, early detection helps significantly—it is reported that the survival rate for early-stage localized cancer (stage I) is 49%.

The computed tomography (CT)scan is considered to be the most accurate imaging modality available for early detection and diagnosis of lung cancer. It allows detecting pathological deposits as small as 1mm in diameter. These deposits are called lung nodules. However, the examination of large amount of data makes the interpretation tedious and difficult, leading to a high false-negative rate for detecting small nodules. A simulation study demonstrated [1] the overall detection rate to be only 63% for nodules of 1–7 mm in diameter. As the size of the nodule decreased, the sensitivity fell to 48% for nodules smaller than 3mm, and only 1% of nodules smaller than 1.5 mm in diameter were detected.

The knowledge for diagnosis of lung cancer involves many different procedures and techniques. It involves steps like preprocessing, enhancement of images, segmentation etc. Till now many computers aided diagnosis system are developed for its detection but still it is a vast area of research for the researches. The main step in the diagnosis of lung cancer is its detection which can be done from the nodule formed. But in most cases mistakes are made in proper detection of these nodule. So, emphasis should be made for proper detection of these nodules. Once the nodule is detected, the next process become easy to determine whether it is benign or malignant. And this classification of whether the nodule is benign or malignant are determined from the previous knowledge about the features of different nodules stated by the various radiologist

2. Reviewed Paper:

In this paper Chon, Albert Balachandra, N and Lu, Peter (2017) [1], the main goal of the researchers were to develop a Computer-aided diagnosis (CAD) system to detect the presence of lung cancer in patient's CT scans of lungs with and without early state lung cancer. They have used methods from computer vision and deep learning, mainly 2D and 3d convolution neural networks, to build an accurate classifier. There classification model mainly comprised of image processing, followed by nodule candidate detection and lastly malignancy classification. Out of various image segmentation techniques like Watershed, clustering (Kmeans and Mean shift), and thresholding was used as initial segmentation approach to segment lung tissues from CT scan images. Then normalization of the 3D image was done using linear scaling, followed by interpolation to downsample the data (scale 0.5). Zero centering of data was done by subtracting the mean of all images in training set. The datasets used was LUNA16 data, which was used to train a U-Net for nodule detection and Kaggle's Data Science Bowl 2017, which contain labeled data of 2101 images was used to predict the accuracy of the model. The researchers have used linear classifiers as baseline, a vanilla 3D CNN for 3D CNN and 3D Gogglenet model on 2D model for image classifications. The classifiers used a weighted softmax entropy loss, Adam optimizer, ReLU activation and dropout after each convolution during training. The result of linear model was 66% of accuracy, sensitivity 65%, Specificity 67%, AUC of 66%, the Vanilla 3D CNN gave accuracy of 70%, sensitivity of 59%, specificity of 76% and AUC of 69%. 3D Gogglnet gave an accuracy of 75%, sensitivity 77%, specificity of 74% , and AUC of 75%.

The work published by Gruetzemacher, R and Gupta, a (2016) [2], proposed a methology which mainly focused on image recognition algorithm and deep learning for detection of malignant pulmonary nodules. The Caffe, an open source deep neural network solver was used to train the DNN model. This paper performed 4 architecture of convolution layer with convolution layer of 3,4,5,6 and the architecture with five (5) convolution layers gave the best performance with an accuracy of 82%. Sensitivity of this method is 78.2% and specificity of 86.3%.

GLF da Silva, AC Silva, AC de Paiva, M Gattass[3] proposed three different architecture of convolution neural networks(CNN), for classification of malignancy of lung nodule without computing morphology and texture features.Their methodoly was tested onto Lung Image Database Consortium and Image Database Resource Initiative(LIDC-IDRI). The nodules with smaller dimension of the CT scan slice was resized to28*28 and nodules with larger dimensions were rescaled to dimensions of 28*24 and put to background of 28*28 and was later submitted to CNN for classification purpose. The researches had proposed 3 architecture CNN of which 3 architecture gave best accuracy with 82.3% sensitivity of 79.4% and specificity of 83.2%. This architecture consisted of more convolution layer, thus generating more features maps, and the presence of hidden layer in the fully connected layer. Regarding sensitivity, success rate of malignant nodules was better given by architecture 2 with rate of 82.9%, in which each convolution layer was followed by pooling layer showing the importance of pooling layer.

In [4] Alakwaa, W., Nassef, M. and Badr, A (2017), proposed a CAD system for lung cancer classification in CT images. The image processing step included segmentation (thresholding) followed by normalization down sampling and lastly zero centering. The 3D CNN produces an accuracy of 86.6 % which was tested in Kaggle's CT scans. The future work of the researchers aims to detect the exact location of the cancerous nodules with the help of the current model. Also, they have thought of use of watershed algorithm for segmentation and making the network deeper.

In [5] Kumar, Devinder, Wong, Alexander and Clausi, A. David, had proposed a CAD system, in which deep features was extracted by using auto encoder along with binary decision tree as a classifier. The features from layer four of auto encoder of layer five was used to create the feature vector and which was later feed as an input to binary decision tree to perform binary classification. The system gave an accuracy of 75.01% with a sensitivity of 83.25%. In this paper, Kuruvilla, Jinsa, Gunavathi, K. (2014) [6], presented a computer aided classification method where the entire lung was segmented from the CT images and the parameters were calculated from the segmented image. The statistical parameters like mean, standard deviation, skewness are used for classification. The classification process comprised of feed forward and feed forward back propagation neural networks. The better classification was given by feed forward back propagation network. The parameter skewness gives the maximum classification accuracy. Among the already available thirteen training functions of back propagation neural network, the Traingdx function gives the maximum classification accuracy of 91.1%. Two new training functions was proposed out of which training function 1 gave an accuracy of 93.3%, specificity of 100% and sensitivity of 91.4% and a mean square error of 0.998. The training function 1 included the momentum factor and the learning rate is proposed. In the training function 1 each variable was adjusted according to the gradient descent with momentum.

3. Proposed Methodology:

In this paper we have proposed one model which may be used to detect lung cancer using Artificial neural network. Fig 1 shows the block diagram for the same

.

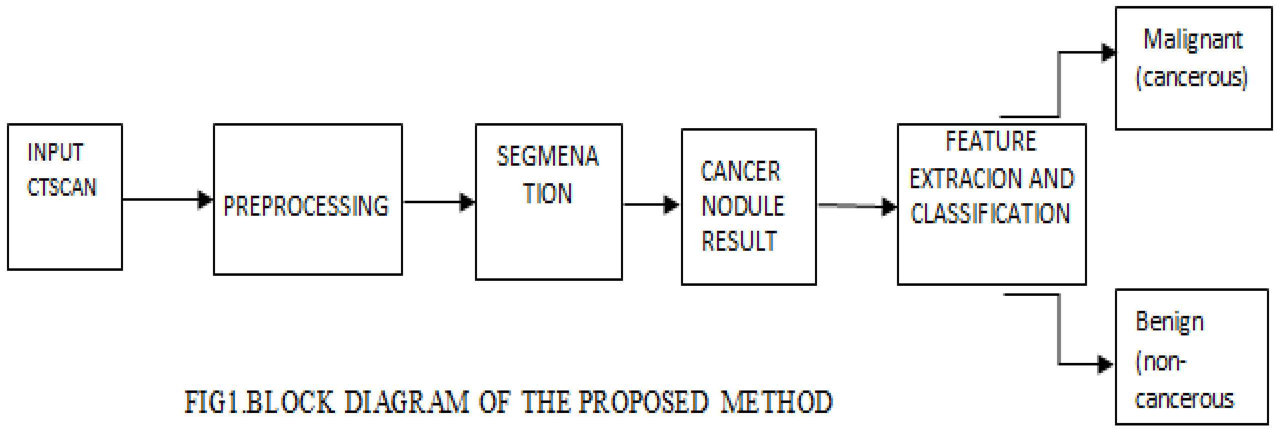


FIG1.BLOCK DIAGRAM OF THE PROPOSED METHOD

3.1 Preprocessing: The basic definition of image processing refers to processing of digital image, i.e. removing the noise and any kind of irregularities present in an image using the digital computer. The noise or irregularity may penetrate into the image either during its formation or during transformation etc. As the medical data consist of noise that is mainly Gaussian noise, so the first step is to remove the noise from the image.

3.2 Image segmentation: The is the process of partitioning a digital image into multiple segments (sets of pixels, also known as super-pixels). The goal of segmentation is to simplify and/or change the representation of an image into something that is more meaningful and easier to analyze. It is typically used to locate objects and boundaries in images. For the segmentation of lung Thresholding is found to be best method.

Thresholding is the simplest method of image segmentation. From a grayscale image, thresholding can be used to create binary images. It replaces each pixel in an image with a black pixel if the image intensity is less than some fixed constant T, or a white pixel if the image intensity is greater than that constant. An example of thresholding technique is shown in snapshot 1.

Otsu's thresholding method is used to automatically perform clustering-based image thresholding. It is the reduction of a gray level image to a binary image. The algorithm assumes that the image contains two classes of pixels following bi-modal histogram (foreground pixels and background pixels), it then calculates the optimum threshold separating the two classes so that their combined spread (intra-class variance) is minimal, or equivalent(because the sum of pairwise squared distances is constant).

Snapshot 1: Simple thresholding.

The dataset used is QIN LUNG CT from The Cancer Imaging Archive (TCIA) collections [12]. The Computed tomography (CT) Image data was obtained on patients diagnosed with Non-Small Cell Lung Cancer (NSCLC) with mixed stage. The dataset comprises of 67 images.

Pre -processing: Digital images are prone to various types of noise. Noise is the result of errors in the image acquisition process that result in pixel values that do not reflect the true intensities of the real scene. Most of the medical image data consist of Gaussian noise, so in order to make our image noise free at the first step of preprocessing, we implemented Gaussian filter in order to remove noise [10]. The type of noise present in the image was determined from the histogram analysis.

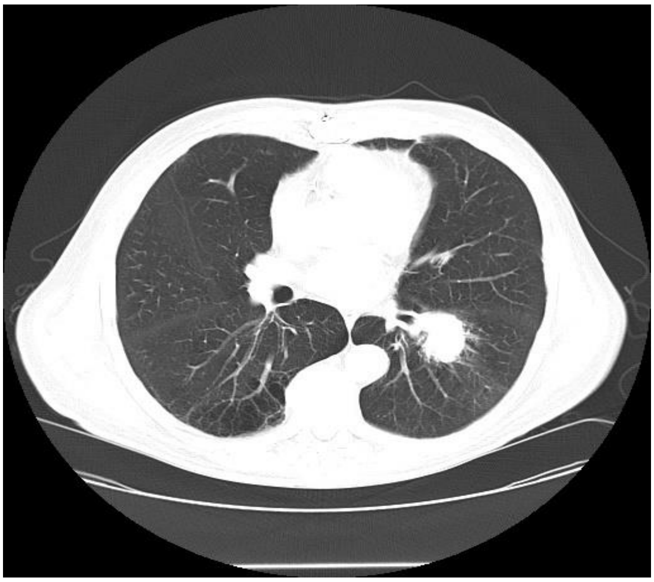
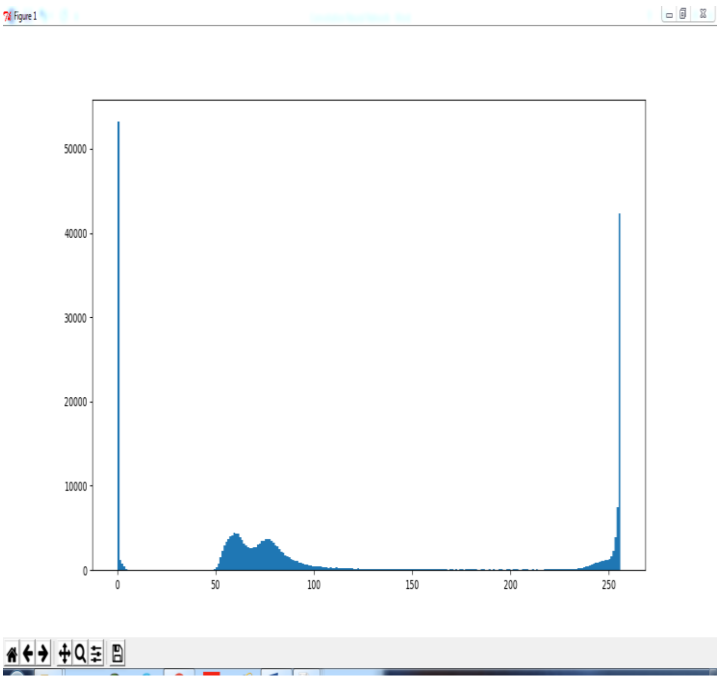


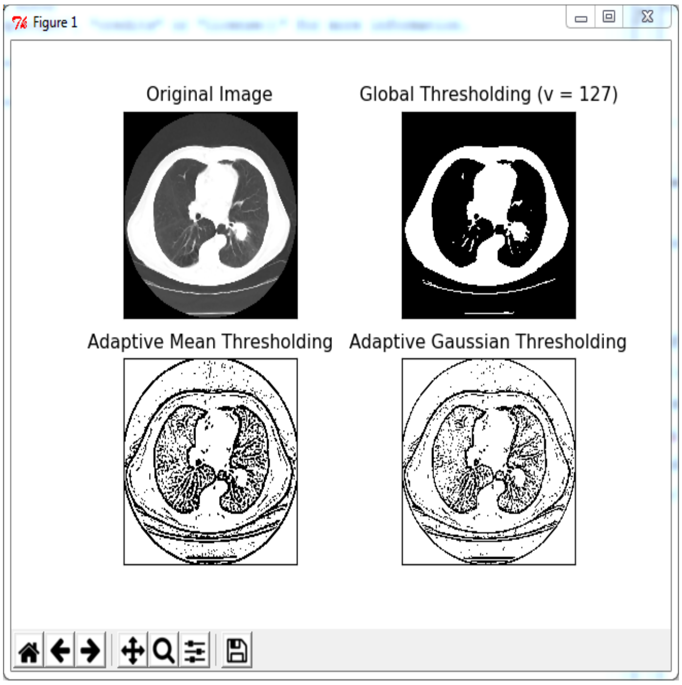
Fig 3: image1.jpg



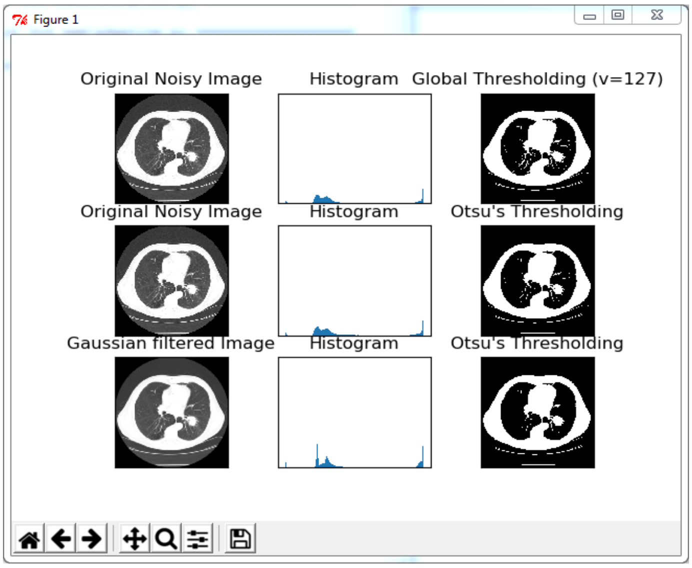
Snapshot 2. Histogram Analysis for Gaussian Noise of Image1

The snapshot 2 gives the histogram of the image in fig 7. As we know that images contain noise which get added during acquisition only. In order to know the type of noise present in the image we used the simplest way of analyzing the histogram of the image. We have found that the images consisted for Gaussian noise. So, we used Gaussian filtering in removing Gaussian noise from the image.

Segmentation: The goal of segmentation is to simplify and/or change the representation of an image into something that is more meaningful and easier to analyze. In our work it's typically used to locate boundaries in images. For locating the boundaries of the image we have used different thresholding techniques like Adaptive thresholding, Otsu's thresholding and simple thresholding. As Otsu's thresholding automatically determine the optimal threshold for segmentation, it is found to be the better method for segmentation of lung boundaries. We tried with different thresholding techniques to find out the best one. The snapshot 3 depicts the adaptive thresholding techniques in which we are showing the output of the original images with a threshold point 127 and found that the output of the adaptive Gaussian thresholding is not that clear as compared to Otsu's technique shown in snapshot 4. We tried with different thresholding point below 127, and above 127 but the result at the point 127 is found to be the best.



Snapshot 3: Adaptive Threshold (V=127)



Snapshot 4: Otsu Thresholding

The literature survey has been carried out thoroughly to make an in depth analysis. From all the papers it is seen that detection of lung cancer in early stage is very difficult. Also the large amount of data per examination makes the interpretation tedious and difficult, leading to a high false-negative rate for detecting small nodule. Unfortunately, even for the highly trained radiologists, potentially the detection of identification of cancerous nodules is very difficult. The reasons being Firstly, nodules are typically small, particularly in the pre-cancer stage. Secondly, there appearance is not always distinct from that of other than benign tissue formation in lungs. And thirdly, the resolution of CT images can vary in ways making precise identification challenging. The above mentioned points can lead to false positive and false negative results in diagnosis which later can lead to misdiagnosis affect the health of patients.

5. Conclusion and Future Work:

In this paper we have discussed the preprocessing of the CT scan image of lung. Firstly, by removing the noise i.e. the Gaussian noise present in the image and then implemented different thresholding techniques to find the good one for the segmentation process.

The main aim with our project is to determine the lung cancer in its earlier state. As we all know the major problem is with the availibity of data. Medical data are not easily available as they are kept confidential between the patients and the doctors itself. So, there is still a lack of proper availibity of data to carry out with our work. Although we know that at present deep learning is rising but one major problem with working with this technique is availability of data. And with low data we will not get proper result. Another problem is that this method is extremely computationally expensive to train. The most complex models take weeks to train using hundreds of machines equipped with expensive GPUs. We mainly need a system that takes less time with effective result. So, before moving with this technique our project will focus on effective method of segmentation of lung for proper diagnosis of nodule for cancer detection. Although there are many work already done in this field but till date no proper method is found for its early detection. So according to our survey it is found that since real data are not easily available, we should focus on processing of image with better techniques in the limited set of data and then only move to our next phase of model.

References:

1. Chon, A., Balachandra, N. and Lu, P., 2017. Deep convolutional neural networks for lung cancer detection. tech. rep., Stanford University.
2. Gruetzemacher, R. and Gupta, A., 2016. Using deep learning for pulmonary nodule detection & diagnosis
3. da Silva, G.L., Silva, A.C., de Paiva, A.C. and Gattass, M., Classification of Malignancy of Lung Nodules in CT Images Using Convolutional Neural Network.
4. Alakwaa, W., Nassef, M. and Badr, A., 2017. Lung Cancer Detection and Classification with 3D Convolutional Neural Network (3D-CNN). *Lung Cancer*,
5. Kumar, D., Wong, A. and Clausi, D.A., 2015, June. Lung nodule classification using deep features in CT images. In *Computer and Robot Vision (CRV), 2015 12th Conference on (pp. 133-138). IEEE.*
6. Kuruvilla, J. and Gunavathi, K., 2014. Lung cancer classification using neural networks for CT images. *Computer methods and programs in biomedicine*, ELSEVIER.
7. D. P. Naidich, H. Rusinek, G. McGuinness, B. Leitman, D. I. McCauley. I. Henschke. Variables affecting pulmonary nodule detection with computed tomography: evaluation with three-dimensional computer simulation.
8. https://en.wikipedia.org/wiki/Feedforward_neural_network
9. https://en.wikipedia.org/wiki/Softmax_function
10. https://en.wikipedia.org/wiki/Image_noise
11. https://en.wikipedia.org/wiki/Thresholding_(image_processing)
12. http://www.cancerimagingarchive.net/
13. Liu, T., Fang, S., Zhao, Y., Wang, P. and Zhang, J., 2015. Implementation of training convolutional neural networks. *ArXiv preprint arXiv: 1506.01195.*
14. Baker, D., Kilpatrick, J. and Chaudhry, A., Predicting Lung Cancer Incidence from *CT Imagery*. *2017*, SPRING
15. https://www.verywell.com/lung-nodules-symptoms-causes-and-diagnosis-2249304
16. https://en.wikipedia.org/wiki/Cancer
17. Paul, R., Hawkins, S.H., Hall, L.O., Goldgof, D.B. and Gillies, R.J., 2016, October. Combining deep neural network and traditional image features to improve survival prediction accuracy for lung cancer patients from diagnostic CT. *In Systems, Man, and Cybernetics (SMC*), 2016 IEEE
18. Makaju, S., Prasad, P.W.C., Alsadoon, A., Singh, A.K. and Elchouemi, A., 2018. Lung Cancer Detection using CT Scan Images. *Procedia Computer Science*, *125*,
19. Sharma, D. and Jindal, G., 2011. Computer Aided Diagnosis System for Detection of Lung Cancer in CT Scan Images. *International Journal of Computer and Electrical Engineering*,
20. Tiwari, A.K., 2016. PREDICTION OF LUNG CANCER USING IMAGE PROCESSING TECHNIQUES: A Review. *Advanced Computational Intelligence: An International Journal (ACII)*,