Research on Different Classifiers for Early Detection of Lung Nodules

Madan.K, Bhanu Anusha K, Pavan Kalyan. P, Neelima. N

Abstract: Lung cancer is one of the most common deadliest disease which has highest death rate as per the recent medical research. However, research indicates that early detection of lung cancer improves chances of survival. The disease is identified using nodules attached to lung walls and lung parenchyma. Nodules plays significant role in identifying cancer in lungs. The proposed approach to determine lung nodules has three stages preprocessing, feature extraction and classification. Segmentation is the preprocessing technique involves two phases namely lung parenchyma segmentation and lung nodules segmentation. Then, texture features and geometric features are extracted using feature extraction algorithms. Lastly, using classification techniques the nodules are classified as benign or malign. TCIA dataset was used for validation of the proposed approach. Form the dataset, CT images were used which have high density resolution and adequate information which helps to find every small detail easily. The proposed method helps in improving accuracy to find number of the lung nodules in lung region and also helps is differentiating benign and malign nodules using CNN architecture. Different classifiers such as SVM, MLP and CNN classifiers are used in comparison analysis. As the result, we conclude that the approach of feature extraction with CNN decreases the false positive rate significantly compared to the existing classification approaches.

Index Terms—lung nodules, thresholding, morphological operators, segmentation, CNN classifier

I. INTRODUCTION

Cancer is defined as the spread and growth of the abnormal cells in various areas inside the body. The exact reason for causing this disease is not yet identified. It may be due to some of the habits like consuming tobacco, alcohol etc. or it may cause due to inheritance and also due to immune conditions. In the present growing world the technology is developing much faster along with that cancer cases are also developing in the world. Major cases of cancer can be cured in the early stages of detection. Cancer cells will grow faster than normal cells in the body and they form clusters to form nodules or tumors. If the cells form clusters primarily in lungs then the nodules are called primary lung nodules. However, if lung cancer cells breaks down and travel to different areas of body through blood vessels and forms cancer at various other fast for example near bones. This type of cancer is called metasis or the secondary lung cancer.

According to the global lung cancer coalition, lung cancer is the major killer among the various cancers like breast cancer, oral cancer, blood cancer etc. In every 30 seconds, someone, somewhere in the world due to lung cancer. The survival rate is about 20 percent after the diagnosis for the lung cancer. Recent researches indicate growth of lung cancer cases in women are twice that of the men among the smokers. Even among the non-smokers also the lung cancer cases are high in women only. Majorly there are two types of lung cancers and they are non-small cell lung cancer (NSCLC) and small cell lung cancer (SCLC). NSCLC is the most common lung cancer among the 75% cases of the lung cancer. All the nodules doesn't be a cancerous cells. Nodules which are less than 3mm are benign tumors which does not leads to cancer and nodules which are greater than 3mm are malign tumors which leads to cancer. Our aim is to find the both malign and benign tumors along with the nodules count which are growing inside and on the walls of lungs.

In present growing world, computer aided diagnosis (CAD) systems helps in identifying various diseases in their early stages in medical field. Radiologists and doctors uses CADs for finding various diseases. In general skin cancer, lung cancer, prostate cancer symptoms are not identified in early stages so the use CADs plays major role here in identifying them. CAD systems success is measured in terms of accuracy, speed and automation level. When CT scans are compared to x-ray images CT scanned images are better to use for identifying lung nodules when compared to x-ray images. During x-ray waves are passed through body to identify the denser parts of body like bones. It is less likely used for identifying tumors. During CT scanning they takes the images of the lung in in 360 degrees so it helps us to find the nodules which are attached to bones, trachea etc. It helps to find the small abnormalities in the lung which helps radiologists to carry their diagnosis easily.

Manual detection is possible only when the accurate information is acquired from the CT scanned images. However if the cancer is in benign stage, machine learning algorithms are required to determine the lung cancer. The detection of lung nodules is carried out in two steps. First, lung parenchyma is separated from CT scanned images and then lung nodules are detected from the lung parenchyma. In order to achieve better results we used morphological operators like hit, miss, close, open, erosion and dilation. Thresholding is used for segmenting both lung parenchyma

Revised Manuscript Received on July 10, 2019.

Madan.K, Department of Electronics & Communication Engineering, Amrita School of Engineering, Bengaluru Amrita Vishwa Vidyapeetham, Tamil Nadu India.

Bhanu Anusha K, Department of Electronics & Communication Engineering, Amrita School of Engineering, Bengaluru Amrita Vishwa Vidyapeetham, Tamil Nadu India. (Email: bhanuanu126@gmail.com)

Pavan Kalyan. P, Department of Electronics & Communication Engineering, Amrita School of Engineering, Bengaluru Amrita Vishwa Vidyapeetham, Tamil Nadu India. (Email: pavankalyan6543@gmail.com)

Neelima. N, Department of Electronics & Communication Engineering, Amrita School of Engineering, Bengaluru Amrita Vishwa Vidyapeetham, Tamil Nadu India. (Email: n_neelima@blr.amrita.edu)

and lung nodules. In the recent years, Machine learning has made a remarkable presence in medical field applications. Using machine learning concepts the efficiency of classification and differentiation of nodules is improved. But the bottleneck is it needs a large amount of data. There are many ways for the classification of benign and malign nodules like SVM classifier, KNN classifier, EM algorithm and many others. The proposed model use 3D CNN architecture to differentiate them. Most of the medical images are in 3D format but CNN architecture uses 2D slices for classification. Existing machine learning techniques in medical analysis converts 3D data into 2D multi-channel representations i.e triplanar representation [1], and then feed them into 2D CNN. The proposed model uses CNN architecture to differentiate between malign and benign nodules.

II. RELATED WORK

Segmentation of lung region from CT scanned images can be done using various methods adaptive thresholding, gray level thresholding, water shedding, SNAKES, otsu's thresholding. Some of the factors like subject tissue volume, physical atria properties, image acquisition protocol and degrees of inspiration makes selection of single gray scale value difficult [2]. Even adaptive thresholding method also have a limitation on it's segmentation results like missing boundaries when there is niot much discontinuity is there[3]. Limitation of a water shedding method is t over segmentation of the images [2]. Yoshinori itai et al [4] uses active contour model i.e SNAKES for segmenting the lung region. Some of its limitation of this model is that to construct initial contours requires some human interactions and iterations to be specified for segmenting desired object. Helen r et al [5] proposed a method for segmenting the lung nodules using otsu's algorithm. Using this the computational for segmenting the lung region from CT scanned images requires 90 seconds/slice. In general we will have 60-120 slices of images in lung cancer cases so it takes more time to compute and give the results. So to overcome the above said limitations, the proposed method uses thresholding using HU values. The medical image data is in digital imaging and communications in medicine (DICOM) format which is measured in HU units. In our body each part has different Hu values. CT scanned images of lungs have water, fat, bones, tissues etc.so the table 1 shows different attenuation values of different tissues in our body[6].

Table 1: Attenuation values of different tissues [6]

Substance	HU range
Air	-1000
Lung	-400 to -600
Nodule	-150
Fat	-60 to -100
Water	0
Soft tissue	+40 to +80
Bone	+400 to +1000

For classification of the nodules there are many methods and each method has it's limitations and advantages. According to the reference paper [7] limitations of that are GGO and juxeta-pleural nodules are not addressed and it

was limited to cross validation. By using EM algorithm there is a limitation of missing nodules which are less than 5mm but it's applicability in some regions is very affective [8]. Using SVM classifier the accuracy of 84% is achieved but false positive rate is high. So the proposed method uses HU value for segmenting lung region then applying morphological operation false reduction and finally 3D CNN architecture is used to improve the efficiency.

III. PROPOSED METHOD

The work flow of the proposed method is described in Fig.1. The input images are CT scanned images, each image is of digital imaging and communication in medicine i.e in DICOM format with an extension of 'dcm'. Each slice has a size of 512*512. In the proposed method DICOM images are used directly without converting. The input images are segmented, in two stages for both lung parenchyma and nodule segmentation. And then false reduction is applied to determine lung nodules which is shown in Fig.2
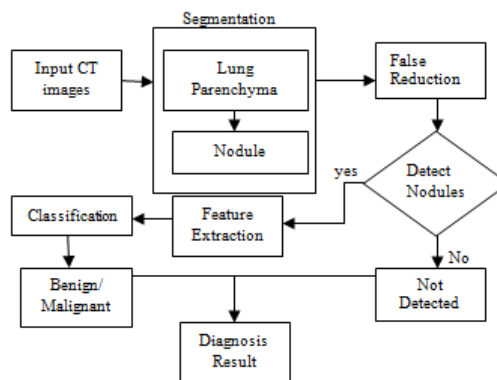

Fig.1: Block diagram of the proposed approach

The lung parenchyma segmentation is one of the important challenge problem because of dissimilarities in lung region and segmentation of lung regions will help in doing early diagnosis of lung cancer.

The lung parenchyma from the given DICOM image is segmented by setting a threshold of -420 HU [6]. Apply closure operation with a disk radius 2 to close small area which is at the border of lungs. Then find the connected areas in the image to acquire the 2 largest portions of image i.e lungs. To fill holes inside the connected components we used closure operation with a structuring element which is disk of radius 12. Superimpose the resultant image with given image to get the lung regions, after this lung walls are removed by subtracting eroded image from the input image.
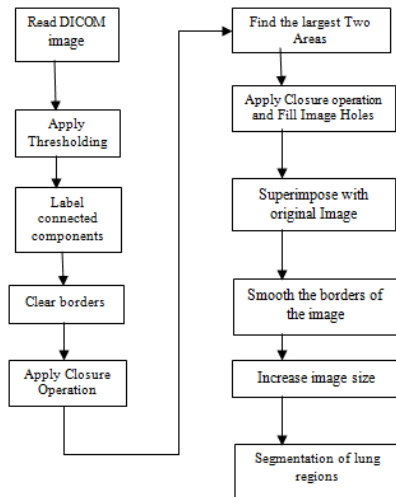
Fig.2 Flow chart for segmentation and false reduction
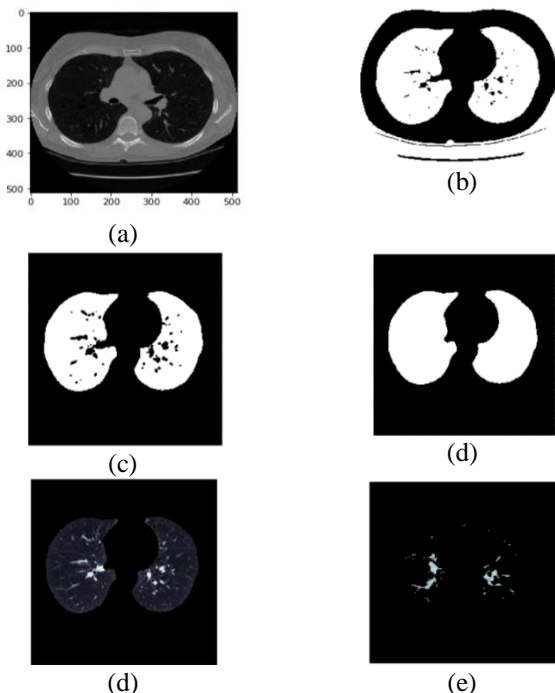

(a)

(b)

(c)

(d)

(d)

(e)

Fig.3. a) Input image b) Thresholded image c) Eroded image d) Binary mask e) Segmented lung image f) Nodule detected 2D-Image of a single slice.

Erosion is performed using the structured element disk with radius of 2, to increase the visual clarity of small nodules. To classify whether the nodules are benign or malign we need to extract their features. In this paper geometric and texture features are extracted. Nodules are 3D structural components which have both geometric and textural features. For geometric features binary mask is used and gray scale values are used for the textural features [7].

A. Geometric features:

The following geometric features are calculated and used as features.

(i) Volume(v): volume of the voxel that belongs to each object.

(ii) Radius(r):

$r = \sqrt{\frac{3v}{4\pi}}$

(iii) Spherical disproportion(d):

$d = \frac{a}{4\pi r^2}$

(iv) Spherical density(s): Compress values of each 3D object

$s = \frac{100n}{v}$

(v) Elongation:

$E = \frac{AY_{min}}{AY_{max}}$

Where AY_{max} is the measure of the bigger corner and the AY_{min} in the measure of the smaller corner.

B. Texture features:

(i) Skewness: It is the measure of lopsideness and also the measure of symmetricity.

$\sum_{i=0}^{n}(f(x,y,z) - \mu)^2$

Skewness = $N\sigma^2$

Where N is no voxels and f(x,y,z) is the gray level at the location at the (x,y,z) where as μ and σ are the average intensity and standard deviation of the object.

(ii) Kurtosis: It is the measure of peakedness of the histogram of the object.

$\sum_{i=0}^{n}(f(x,y,z) - \mu)^4$

Kurtosis =

$N\sigma^4$

(iii) Contrast:

$\text{Contrast} = \sum_{i=0}^{\sigma-1} \sum_{i=0}^{\sigma-1} Co(i,j)(i-j)^2$

(iv) Energy:

$\text{Energy} = \sum_{i=0}^{\sigma-1} \sum_{i=0}^{\sigma-1} Co(i-j)$

(v) Entropy:

$\text{Entropy} = \sum_{i=0}^{\sigma-1} \sum_{i=0}^{\sigma-1} -Co(i-j)log(Co(i-j)))$

(vi) Homogeneity:

$\text{Homogeneity} = \sum_{i=0}^{\sigma-1} \sum_{i=0}^{\sigma-1} \frac{1}{1+(i-j)^2}Co(i-j)$

(vii) Moment:

$\text{Moment} = \sum_{i=0}^{\sigma-1} \sum_{j=0}^{\sigma-1} \frac{(Co(i-j)^2)}{1+|i-j|^2}$

IV. CLASSIFIERS

CNN or conv NET is a class of deep learning which is most commonly used in imagery. To require minimal pre-processing CNN uses multilayer perceptron. CNN is inspired from the biological process of neurons. CNN uses little pre-processing than the other image classification algorithm. CNN architecture takes advantage to constrain the architecture in more sensible way takes the images as the input. Unlike the other deep learning architectures, CNN takes the consideration weight, height and depth, where depth means the number of layers. CNN is made up of many hidden layers. Each hidden layer is made up of neurons where each neuron of present layer is connected to all the previous layer neurons. The last layer of the series hidden layers is output layer which gives the class scores. CNN layers has simple API which converts 3D volume input data into the 3D volume output data by using some differential function with may or may not have a parameters.

A simple CNN architecture has mainly 3 layers. They are convolution layers, pooling layers, fully-connected layer. They are arranged in the following order like Input-Conv-Relu-Pool-Fc. The function of the each layer is explained briefly below.

(i) Input layer: it takes raw pixel values of original image as the input.

(ii) Convolution layer: This layer applies the convolution operation to the input image and passes the information to the next layer. This operation emulates the response of the individual neuron to visual stimuli.

(iii) Relu layer: In this layer element wise activation function is applied. The size remains as unchanged as in convolution layer.

(iv) Pooling layer: Down sampling operation is applied along the spatial dimensions. Pooling layers in the convolution networks may be local or global. Pooling layer reduces the dimension by connecting neuron clusters of the one layer to the single neuron of the other layer. Small clusters are combined in local pooling whereas global pooling acts on all the neurons in the layer.

(v) FC layer: it means fully connected layer in which every neuron in one layer will be connected to other neuron in other later. It will follow the principle same as of multi-layer perceptron (MLP). It helps to classify or identify the images.

Also for comparison analysis SVM (Support Vector Machine) and MLP (Multi-Layer Perceptron) classifiers are used.

V. RESULTS

The proposed model used the dataset "The Cancer Imaging Archive" (TCIA) dataset from the LIDC-IDRI database. Each dataset has 60 to hundreds of images from different angles. The proposed algorithm is compared with the existing model in [1]. The existing approach considered only the nodules present in lung walls. The approach does not considered juxta-pleural nodules which are presented behind the bones. But the proposed approach considered these nodules into consideration which makes complex for the classification. Fig.4 shows the nodule detection in both 2dimensional and 3dimensional view. The classification accuracy with MLP classifier is 87.75% and with CNN classifier is 91.28%. The comparison table is given in Table 2.
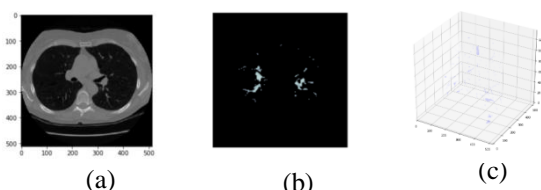

(a) (b) (c)

Fig.4. a) Input image b) Nodule detected 2D-image c) Nodule detected 3D-image.

Table II. Comparison analysis of different classifiers for lung nodule detection

Type of Classifier	Accuracy (%)
SVM	79.15
MLP	87.75
CNN	91.28

Some approaches claim more accuracy than 91% but the proposed approach considered the juxta-pleural nodules into consideration for classification. So, we conclude that the proposed approach provides better accuracy along with false reduction. The proposed approach is compared with existing classifiers such as SVM and MLP.

REFERENCES

1. Qi Dou, Hao Chen, Lequan Yu, Jing Qin, and Pheng-Ann Heng, "Multilevel Contextual 3-D CNNs for False Positive Reduction in Pulmonary Nodule Detection", IEEE Transactions On Biomedical Engineering, VOL. 64, NO. 7, JULY 2017.
2. Omid Talakoub,Javad Alirezaie,Paul Babyn,"Lung Segmentation in Pulmonary CT Images Using Wavelet Transform", ICASSP, IEEE, 2007
3. Xiaojie Huang, Junjie Shan, and Vivek Vaidya, "Lung Nodule Detection In CT using 3D Convolutional Neural Networks", 2017 IEEE.
4. Omid Talakoub,Javad Alirezaie,Paul Babyn,"Lung Segmentation in Pulmonary CT Images Using Wavelet Transform", ICASSP, IEEE,2007
5. R.Helen, Dr. NKamaraj, Dr.K.Selvi, V.Raja man," Segmentation of Pulmonary Parenchyma in CT Lung Images based on 2D Otsu optimized by PSO", Proceedings of ICETECT, IEEE.2011
6. Eng. Michael Samir Labib Habib," A computer aided diagnosis system (CAD) for the detection of pulmonary nodules on CT scans", systems and biomedical engineering, cairo university, Giza, Egypt, 2009.
7. Bhavana.V, Krishnappa H.K, " A Survey on multi modality medical image fusion", IEEE International conference on Wireless Communications, Signal Processing & Networking, (WiSPNET), 2016, pp: 1326-1329.
8. Bhavana.V, Krishnappa H.K, "Multi-modality medical image fusion using Discrete Wavelet Transform", 4th International conference on Eco-friendly Computing and Communication System, ICECCS 2015, Procedia Computer Science, pp: 625-631, 2015.
9. Nima Tajbakhsh, Kenji Suzuki, "Comparing two classes of end-to-end machine-learning models in lung nodule detection and classification: MTANNs vs. CNNs", Pattern Recognition 63 (2017) 476–486
10. Qian Yiming, Weng Guirong, "Lung Nodule Segmentation using EM algorithm", 2014 Sixth International Conference on Intelligent Human-Machine Systems and Cybernetics.
11. S. L. A. Lee, A. Z. Kouzani, E. J. Hu, "Automated detection of lung nodules in computed tomography images: a review", Machine Vision and Applications (2012) 23:151–163 DOI 10.1007/s00138-010-0271-2.
12. M. Gomathi, Dr. p. Thangaraj, "Automated CAD for Lung Nodule Detection using CT Scans", 2010 International Conference on Data Storage and Data Engineering.
13. Lakshmi N arayanan A, Prof. Jeeva J.B, "A Computer Aided Diagnosis for detection and classification of lung nodules", IEEE Sponsored 9th International Conference on Intelligent Systems and Control (lSCO)2015.
14. A.R. Talebpour, H.R. Hemmati, M. Zarif Hosseinian, "Automatic Lung Nodules Detection In Computed Tomography Images Using Nodule Filtering And Neural Networks", The 22nd Iranian Conference on Electrical Engineering (ICEE 2014), May 20-22, 2014, Shahid Beheshti University.
15. Anindya Gupta, Olev Märtens,Yannick Le Moullec and Tõnis Saar Thomas Johann Seebeck, "A Tool for Lung Nodules Analysis based on Segmentation and Morphological Operation", 2015 IEEE.