TEST Engineering & Management

January - February 2020
ISSN: 0193 - 4120 Page No. 455 - 460

Lung Cancer Detection using Neural Network and Content Based Image Retrieval

Rejiram .R[1], Kanniga.E[2*,] Sundararajan.M[3*]
[1]*Research Scholar, Bharath Institute of Higher Education and Research,
[2]*Professor Department of Electronics &Communication/ Instrumentation Engineering
CEDSE -Bharath Institute of Higher Education and Research, Chennai, Tamil Nadu,India
[3]*Professor Department of Electronics &Communication/ Instrumentation Engineering
CEDSE -Bharath Institute of Higher Education and Research, Chennai, Tamil Nadu,India
Corresponding email:kanniga.etc@bharathuniv.ac.in, msrajan69@gmail.com,rejiramnair@gmail.com

Article Info
Volume 82
Page Number: 455 - 460
Publication Issue:
January-February 2020

Article History
Article Received: *14 March 2019*
Revised: *27 May 2019*
Accepted: *16 October 2019*
Publication: *02 January 2020*

Abstract:
Computer aided detection (CAD) systems that automatically detect and localize lung nodules in CT scans. A major problem in this system is the large number of false positives because of no provision for comparison of the predicted output. This paper commends anewsystemwiththecombination of CBIR and neural network t ofull fill the gap inthe areaof early detection of lung cancer. From the preprocessed CT scan image the system identifies whether it contains nodules using Circular Hough Transform and classifies into benign or malignant nodule using Probabilistic Neural Network. Then, it search for the most similar images and retrieved it from the database. From the retrieved image it is easy to identify the present cancer stage of the patient. Experiments done based on both LIDC database and the locally collected database. The performance evaluation of the system is done by using both. The experimental results showsthat the proposed system can easily classify benign and malignant nodules with an efficiency of 97 % accuracy on LIDC dataset, 95 % accuracy on Local dataset and similar images are retrieved with its present stage from the available database with a higher precision and recallrate.

Keywords: CT, CAD, Circular Hough Transform, Benign Malignant, Probabilistic Neural Network, LIDC.

I. INTRODUCTION

Lung cancer is one of the malignancies spread in the world. Radiologists nowadays use computed tomography (CT) scans of chest to detect lung tumors as it has a high sensitivity and low error rate. However, even with these CT scans, it requires

They have many benefits such as reducing the error rate of nodule detection, reducing the operation time and detecting tumors that are overlooked by the radiologists. Several studies have shown that CAD systems offer a useful second opinion. But, one of the problems in this system is large number of false positive which may occur during detection. To solve these, content based medical image retrieval system is used. Firstly database prepared as benign and malignant using neural network detection.Apoptosis are classified as the benign and malignant query images are used to the different malignancies stages are described in database wise classified in the biomedical field.

Nodule classification and nodule detections are classified in the mainly SOM method using and segmentation. This method is based on the size shape and nodules balance and imbalance [1,2].Region growing algorithms are implemented to detect nodules inlungs from a CT Scan image of Lungs. Obtain accurate and

effective result of pulmonary parenchyma misdetection based on CT images. Algorithmic and proposed techniques are used to the structural elements are detected and modified [4]

The datasets are analyzed in the computer aided diagnosis methods used to lung nodules are evaluvated.Images systems are used to the lung mask are separated in different type of algorithm based on LUVEM.Then ,three groups off feature sare extracted from he database based on these nodules are classified in to benign and malignant using probabilistic neural network. Performance assessment index are analyzed in the different datasets are summarized from the database finally and described a system of CBIR approach. The results in the performance evaluation stage shows that accuracy of the proposed system is more than 97% and 95% in the standard database and the locally collected database respectively. The images are retrieved with higher precision and recallrate.

I. PROPOSEDWORK

The proposed system contains the data summaries and lungs volume segmentation and nodule collection classification and image processing.

A. Image Acquisition

First stage of the image analyzing method contain in the CT images are demonstrated and the whole process was obtained and saved in the Digital imaging and communication medicine.[5] files from LIDC dataset [6] and locally collected database.

The CT scanning images are changed in the JPEG format and the input pixel is 512x512.

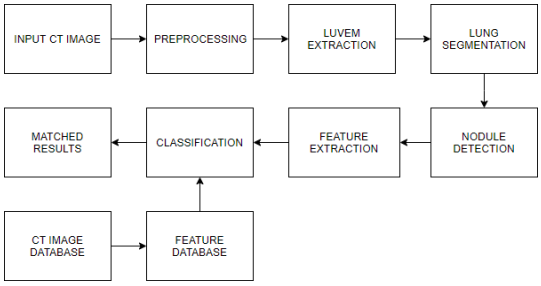

Fig.4. 1 Block Diagram of Proposed System

B. Preprocessing

The main concept of image proceeding is the prevent the unwanted results and outputs. Its mainly based on the images colour and sharpen.

Sharpens are denoted in different colors occurred in the results. The transition images are denoted as gradual transition of gray and white. The images corners are contrast and increased them denote the image was sharpened. The true color input image was denoted as version of the gray scale and input. Some of the image processing software's also used to analyze.

Lung Volume Extraction

The morphological operations used to the lung lobes are partially extracted in the CT image. The LUVEM algorithm based on the CT image side and edge then segments are easily removed.

STEPS OF LUNG VOLUME EXTRACTION

1. Creates an initial lung mask from the sharpened image
   - Find RGB values used for thresholding of marker/colorimage
   - Userselectmultiplepixelsonthemarkerimage
2. Creating output image based on selected pixels on marker
   - Morphological operation Filling and dilation done

TEST Engineering & Management

January - February 2020
ISSN: 0193 - 4120 Page No. 455 - 460

C. Lung Volume Segmentation

The second type of nodule selection segmentation based on the algorithm of K means clustering.

The algorithm contains various steps of K point represent the objects are clustered in the centroid of the group. all objects positions are recalculated and measured the K value.

$$V_i = \frac{1}{a}\sum_{j=1}^{a} x_i$$

The step 2 and 3 cenroids are separated in the object the group of metric minimization calculated.

D. Nodule Detection

Nodules are the main concept of detected the segmentation. First of all pulmonary nodules are identified and nodules are separated. Mostly nodules are helical or circular shapes. The circulatory determination round images are probably used. Round images are identified in the circular Hough transform method. it is the one of the basic techniques performed for image processing. In this method nodules are selected for different size and radius. Different extraction techniques are used to detect the circular images. It is identified the images inputs are find circle and imperfect. It is produced the different matrix are called accumulator matrix [7].

The described equation given below,

$$r^2=(x-a)^2+(y-b)^2 \quad (2)$$

The Hough transform can be used to determine the parameters of a circle when a number of points that fall on the perimeter are known. A circle with radius R and center (a, b) can be described with the parametric equations,

$$y=b+RSin\theta \quad (3)$$

E. Feature Extraction

Feature is the one of the technique used to the nodules are extracted. its is based on the division of malignant and beningn.the first step of feature is the nodules gray levels are calculated by histogram values based on the images. the basics studies of feature contains variance, kurtosis ,skewness, means and entropy are analyzed and calculated[10].In this study contain various angles and direction the total features are 88 calculated GLCM method and in different region and energy distribution. The CT images are classified in 2D wavelet decomposition the three images are denoted by high frequencies remaing decomposed in transform of difrent images.

Table I. Number of Extracted 2d Features from Ct Image Of Lungs

Feature Extraction Method	Number of Features
Statistical Features	6
GLCM Features	22*4 = 88
Wavelet	13

F. Nodule Classification

Probabilistic neural network(PNN) model is the one of the best network contain input and output layer. They contain 107 input layer and they classified.[10].

G. Content Based Medical ImageRetrieval

The most similar images are retrieved from the database using cbir technique. The images are stored as malignant and benign as separately along with its medical description. When an input is fed to the system , the system classifies into benign and malignant and secondly it retrieve most similar images from the database using Euclidean distance measurement. In this work, Euclidean distance is used for image retrieval. It is also called l2 distance. If u= (a ,b) and v = (c , d) are two points,thentheEuclideandistancebetweenuandvis givenby:

TEST Engineering & Management

January - February 2020
ISSN: 0193 - 4120 Page No. 455 - 460

$$(u,v) = \sqrt{(a-c)^2 + (b-d)^2} \quad (5)$$

Based on these measurements, images are retrieved in the order of similarity.

RESULTS AND ANALYSIS

The proposed system was implemented using MATLAB. The dataset for the project was collected from LIDC dataset and from local hospital for validating the results. A total of 100 images were taken from LIDC database out of that 50 were benign and 50 images were malignant. For testing the performance of the system, 10 images that belongs two different classes were used. After that validation was done using 60 images from local database.

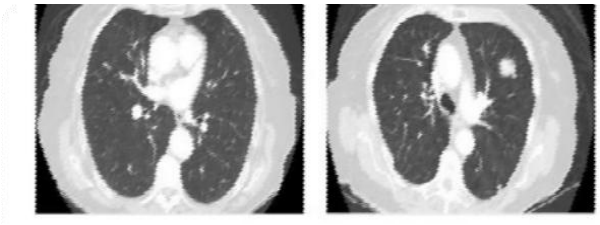

Figure 1.Input CT images of lungs

The sample images are then sharpened for removing the noises occurred during imaging of lungs using CT scan as shown in Figure 2.

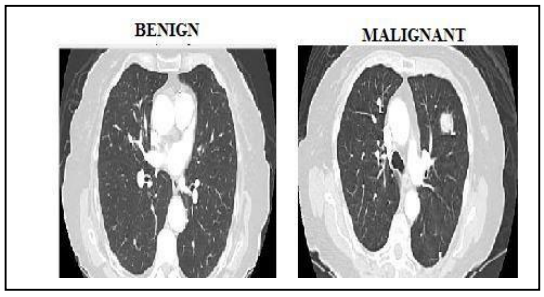

Figure 2 Sharpened images

The lung volume is extracted after the morphological operation is given below

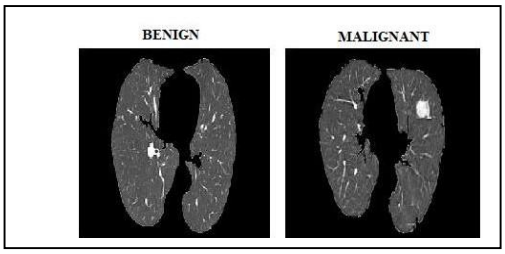

Figure 3. Lung Volume Extracted

The lung Segmentation is done through K-mean clustering is given below.

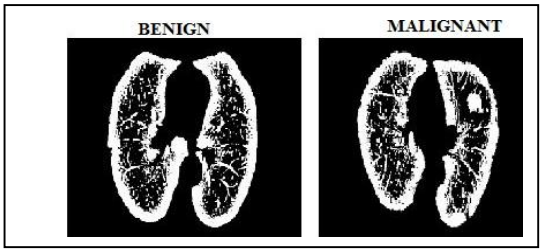

Figure 4 . Segmentation using K-means Algorithm

The lung nodules are detected using Circular Hough Transform with smaller and larger radii.

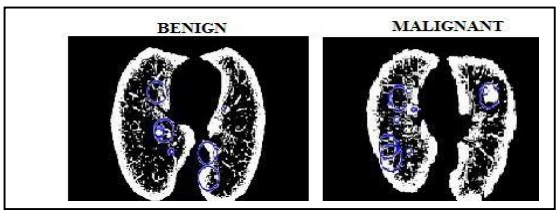

Fig 5 . Nodule detection

The lung nodules with smaller and larger radii are extracted.

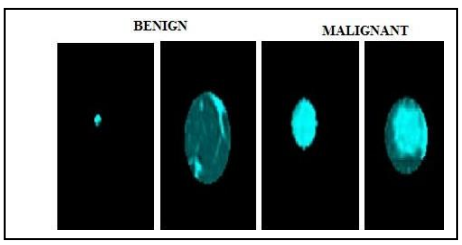

Fig 6. Nodule Extracted with small and large radii

From the extracted nodules each of benign and malignan trespectively, six statistical features, 22 GLCM features and wavelet features are extracted.Out of that more accurate results

January - February 2020
ISSN: 0193 - 4120 Page No. 455 - 460

obtained from GLCM features and more weightage is given to that. Finally, features fed to the neural network and classification is done. In Table (i), GLCM features of benign and malignant nodules aregiven.

GLCM FEATURES	BENIGN	MALIGNANT
Contrast	1.04997	1.22824
Correlation	0.00602	0.013074
Cluster Prominence	0.87967	0.961336
Cluster Shade	0.87967	0.961336
Dissimilarity	1.51303	33.59047
Energy	0.35612	0.693887
Entropy	0.00387	0.004838
Homogeneity	0.98170	0.981364
Homogeneity	0.06901	0.076555
Contrast	0.99833	0.998292

TABLE 1.GLCM FEATURESCOMPARISON

Most of the GLCM features have high values for malignant nodules. For getting accurate results, these features are fed to neural network. In Table (ii), shows that LIDC database have higher accuracy on neural network detection of cancer.

Features Used	Local Data base	LIDC database
SSF	78%	81%
GLCM	92%	93%
Wavelet	90%	91%
Combined	95%	97%

TABLE 2. DIFFERENT FEATURES ACCURACY ON LIDC DATABASE AND LOCAL DATABASE

N=100	*BENIGN*	*MALIGNANT*	Total
BENIGN	TN = 47	FP = 3	50
MALIGNANT	FN = 0	TP = 50	50

The Performance evaluation of PNN neural network is done using Confusion matrix is shown in table 3

TABLE 3. PERFORMANCE EVALUATION OF PNN NEURAL NETWORK USING CONFUSIONMATRIX

The content based image retrieval of benign and malignant tumor are shown In Fig 8. and Fig 9.

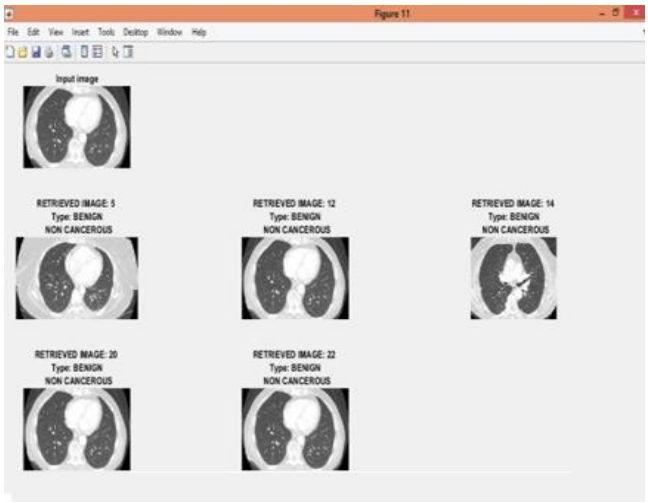

Fig 6. Content based image retrieval in LIDC Database(Benign)

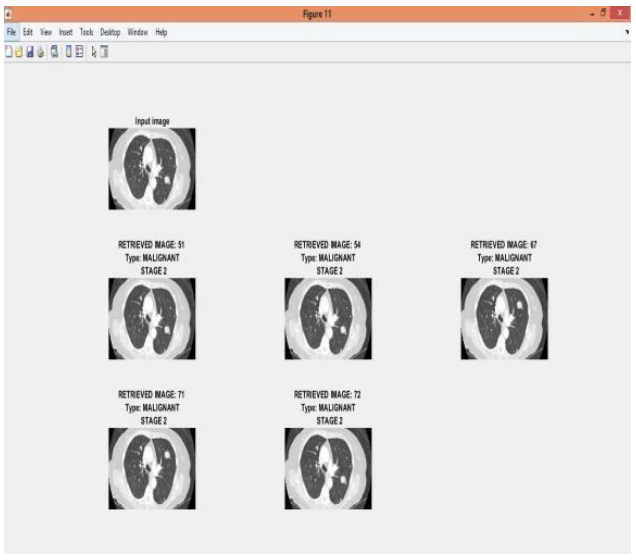

Fig 7. Content based image retrieval in LIDC Database (Malignant)

The performance evaluation of CBIR is shown in Table 4 and 5

January - February 2020
ISSN: 0193 - 4120 Page No. 455 - 460

Number Of Images	Number of True Retrieval	Precision Rate
5	5	100 %
10	10	100 %
15	14	93 %
20	18	90 %

TABLE 4. PERFORMANCE EVALUATION OFCBIR(BENIGN)

Number Of Images	Number of True Retrieval	Precision Rate
5	5	100 %
10	10	100 %
15	15	100 %
20	19	95 %

TABLE 5. PERFORMANCE EVALUATION OF CBIR(MALIGNANT)

I. CONCLUSIONS AND FEATURE SCOPES

The proposed Lung cancer detection system implemented successfully with 97 % accuracy using PNN neural network. Lung nodule database prepared as benign and malignant. From the set of large unsorted LIDC database, similar images are retrievedusingcontentbasedimageretrieval.Thep erformance of the proposed system compared with local database also for validation and obtained an accuracy of 95% detection. This systemhelpsthedoctortoimprovethetreatmentlev elbycorrect diagnosis and also help to identify the present stage of the patient with the help of retrieval from large database. In future, successful treatment method of each patient can be attached with the image which can act asreference.

REFERENCES

[1] EmreDandil, A Computer-Aided Pipeline for Automatic Lung Cancer Classification on Computed Tomography Scans, in Journal of Healthcare Engineering , vol. 2018, Article ID 9409267, 12 pages,2018.

[2] Sori, WorkuJifara, Jiang Feng, and Shaohui Liu., "Multi-path convolutional neural network for lung cancer detection.", *Multidimensional Systems and Signal Processing* , pp. 1-20,2018

[3] Chaudhary, Anita, and SonitSukhraj Singh. ,"Lung cancer detection on CT images by using image processing", *International Conference on Computing Sciences*. IEEE,2012.

[4] John, Jibi, and M. G. Mini. "Multilevel thresholding based segmentation andfeatureextractionforpulmonarynoduledetection",*ProcediaTechnology*, vol. 24 , pp. 957-963,2016

[5] DICOM (Digital Imaging and Communications in Medicine),http:// medical.nema.org.

[6] M. F. Mcnitt-Gray, S. G. Armato, C. R. Meyer et al., The Lung Image Database Consortium (LIDC) data collection process for nodule detection and annotation, in Academic Radiology, vol. 14, no. 12, pp. 1464-1474,2007.

[7] R.O.Duda and P.E.Hart, Use of Hough transformation to detectlines and curves in pictures, in Communications of the ACM, vol. 15, no. 1, pp. 11-15, 1972.

[8] G. Van, P. Wouver, D. Scheunders, and D. Van, Textural features for image classification, in IEEE Transactions on Image Processing , vol. 8, no. 4,pp. 592-598,1998.

[9] Hollitt, Review on feature extraction methods in pattern classification, in European Journal of Scientific Research, vol. 71, no. 2, pp. 265-272,2012.

[10] G. Van, P. Wouver, D. Scheunders, and D. Van, Textural features for image classification, in IEEE Transactions on Image Processing, vol. 8, no. 4, pp. 592-598,1998.

[11] K. Z. Mao, K. C. Tan, and W. Ser, Probabilistic neural network structure determination for pattern classification, in IEEE Transactions on Neural Networks, vol. 11, no. 4, pp. 1009-1016,2000.