International Journal of Engineering and Advanced Technology (IJEAT)
ISSN: 2249-8958 (Online), Volume-8 Issue-5, June 2019

Deep Lung Cancer Prediction and Segmentation on CT Scan

Anuja J, Smitha Vas P

Abstract: ***Lung cancers are one of the world's lethal ailments and early prognosis of cancer is a complex mission in the detection of lung cancer. Analysis and treatment of lung malignancy has been one of the greatest problem faced by humans in the last few years. Early identification of the tumour would consistently make it easier to save a large number of lives across the globe. This paper presents an approach to classify tumour found in the lung as malignant or benign using a Convolutional Neural Network. Here, an Inception V3 model is used to predict if the lung is malignant or benign. The accuracy obtained through CNN is 97 percent, which is more efficient than traditional neural network system.***

Index Terms**: *Chest CT image, Computed Tomography, Convolutional Neural Network, Deep Learning, Lung cancer.***

I. INTRODUCTION

Tumors can be benign or malignant; when we speak of "cancer" we refer to those tumors that are considered malignant. Benign tumors are not spreading to other parts of the body. Malignant tumors which are fast growing and tending to spread other body parts, enabling tumor cells to enter the bloodstream or lymph system and other sites in the body. This method of spreading is called metastasis; metastasis is called the zones of tumor development at these remote locations. Lung cancer tends to spread at an early stage; it is one of most challenging task to treat the cancer. Some organs are the most prevalent places for lung cancer metastasis, especially the adrenal glands, liver, brain, and bone. The lung is also a very common site for tumor metastasis in other areas of the body. Tumor metastasis consist of the same cell type as the tumor originally or primarily. Lung cancer is the most common cancer and the most common cause of cancer deaths in U.S. males. Lung cancer is caused by the lung tissue cells' uncontrollable uneven development. These pulmonary tissue defects are often referred to as pulmonary nodules. They are about 5 millimeters to 30 millimeters in size, tiny and approximately spherical tissue masses. The early phase of pulmonary cancer is characterized by pulmonary nodules. Research has shown that the curability of this lethal disease is almost 75%, if acknowledged soon enough because it is simpler to treat and has less risks. In order to decrease mortality, early diagnosis of malignant nodules is a vital problem. Many imaging methods, including Computed Tomography (CT), Chest X-ray, Magnetic Resonance Imaging (MRI), PET and Blood testing are used to identify early lung tumor. Detection means the classification of two classes of tumor, benign and malignant. The chance of survival at the advanced stage is lower when compared to treatment and lifestyle to survive cancer therapy when diagnosed at the early stage of the cancer. With the implementation of image processing techniques, manual analysis and diagnosis system can be greatly improved. A number of researches on image processing techniques to detect early stage cancer detection are available in the literature. But there is not much improvement in the hit ratio of early stage cancer detection. With the advancement in machine learning techniques, many researchers are trying to diagnose cancer early. Neural network plays a major part in recognizing cancerous area(malignant), which provides an effective tool for cancer detection by using convolutional neural networks(CNNs). The cancer treatment will only be effective if the tumor cells are precisely separated from the normal cells Classification of the tumor cells and training of the neural network on the basis of machine learning based cancer diagnosis. This paper presents a technique based on the Convolutional Neural Network (CNN) to detect whether the lung cancer as malignant or benign.

II. ANATOMY OF LUNGS AND RESPIRATORY SYSTEM

The respiratory system supplies the oxygen needed by body cells and carries off their carbon dioxide waste. Inhaled air passes via the trachea (windpipe) through two narrower tubes, the bronchi, to the lungs. Each lung comprises many fine, branching tubes called bronchioles that end in tiny clustered chambers called alveoli. Gases cross the thin alveolar walls to and from a network of tiny blood vessels. Intercostal(rib) muscles and the muscular diaphragm below the lungs operate the lungs like bellows, drawing air in and forcing it out at regular intervals.

III. LUNG CANCER TYPES

Small cell lung cancer and non-small cell lung cancer are the two types of lung cancer. The type of cancer under the microscope is based on how it looks. These are the following types of NSCLC: squamous cell carcinoma, large cell carcinoma and adenocarcinoma. SCLC tends to grow faster than cancer of non-small cell. Because it grows faster, when it has spread outside of the lung, SCLC is often found.

Manuscript published on 30 June 2019.
* Correspondence Author (s)
Anuja J*, MTech Scholar, Computer Science and Engineering, LBS Institute of Technology for Women, Thiruvananthapuram, India.
Smitha Vas P, Assistant Professor, Computer Science and Engineering, LBS Institute of Technology for Women, Thiruvananthapuram, India.

IV. CAUSES OF LUNG CANCER

Most cases of lung cancer are caused by smoking, although people who have never smoked can also develop the condition. Other causes of lung cancer are passive smoking, asbestos fibers and radon gas**.**

V. GENERALIZED SIGNS AND SYMPTOMS OF LUNG CANCER

Lung cancer symptoms may vary from person to person. You may feel like you're having bronchitis or a bad cold that isn't getting better or You may have no symptoms at all. Generalized signs and symptoms of lung cancer are: a cough that gets worse or doesn't go away, more breathing problems (shortness of breath) than usual, blood coughing, chest pain, heavy voice, frequent lung pain, drop in weight, headaches, mental status changes or neurologic findings, enlarged liver, gastrointestinal disturbance and abdominal pain. Lung diseases affecting the airways, air sacs, blood vessels, pleura and chest wall are the certain group division for the common lung diseases.

VI. LUNG IMAGING

Imaging plays an important role for the investigation and detection of lung cancer types, with the most common methods including CT, CAD, Chest X-Ray, PET, magnetic resonance imaging (MRI) and blood test.

A. *CT Scan*

Using special X-ray tube to acquire picture information from distinct angles around the body, and then using information processing to show a cross section of body tissues and organs. Some of the fundamental concepts of CT are projection reconstruction, meaning that patient information is measured at various location and angles. CT modalities can demonstrate distinct kinds of tissues, lung, soft tissue and bones, and using specialized devices and knowledge to generate and interpret body CT scans, radiologists can more readily diagnose tumour, cancer or other lesion, and assess their size, accurate place, and the magnitude of tumour participation with other neighbouring tissue. The pictures taken from a CT scanner may show some soft tissue and other structures that in standard X-rays are not even noticeable.

B. *CAD*

Computer assisted diagnosis of pulmonary CT picture was a notable and revolutionary move in the early and premature identification of pulmonary defects. The CAD systems include automatic detection systems for lung nodules and restoration of 3D lung systems that help radiologists in their final choices. in order to clarify and improve the picture and separate the region of concern from the whole picture, advanced image processing algorithms are introduced to the picture. The region acquired individually is then analysed for identification of nodule, tumour or cancer to diagnose the disease. Efficient pulmonary segmentation method enables precision and greater confidence value.

C. *Chest X-Ray*

The X-ray is the most prevalent first diagnostic phase when any pulmonary cancer symptoms occur. A view from the back to the front of the chest as well as a view from the side is often involved in the chest X-ray operation. Like any X-ray operation, X-rays expose the patient briefly to a minimum quantity of radiation for a short time. Chest X-rays may reveal region of suspicion in the lungs but cannot determine whether these areas are cancerous.

D. *PET*

Tomography of Positron emission *(PET)* scanning is an imaging technique that create colored three-dimensional picture of the pulmonary tissues. PET scans can determine the tumor types and growth of the cell tissues.

E. *MRI*

Magnetic resonance imaging *(MRI)* scans may be suitable location of a tumor requires accurate detail. The MRI method utilizes magnetism, radio waves, and a computer to create body structure pictures. The patient is put, as with CT scanning, on a mobile bed that is inserted into the MRI scanner. MRI scanning has no known side impacts, and radiation is not exposed. MRI's picture and resolution are very comprehensive and can detect small structural modifications within the body.

F. *Blood Test*

Blood tests may reveal biochemical or metabolic defect in the body that involve cancer whereas regular blood tests itself cannot assess lung cancer.

VII. LITERATURE SURVEY

Prof. Samir Kumar Bandyopadhyay [8] to detect the corners of CT pulmonary image for cancer identification, CAD method is used. In [5] Nikita Pandey, Sayani Nandy Proposed method to identify the cancer cell from lungs CT scan. Here, the suggested technique efficiently detects the cancer cell from lung CT scan based on the identification of the sobel edge detection and label matrix. Sobel operator helps to finding the edges and image gradient of an image. Image gradient changes the intensity of an image. Fatm Taher, Naoufel Werghi and Hussain Al-Ahmad [9] the suggested technique filtering thresholding algorithm is used to identify lung cancer sputum cells from the raw sputum image. Qinghua Ji, Ronggang Shi [7] to segment the image, conduct watershed conversion. Then morphological opening and closing operations to done the image gradient and eliminate the area over segmentation. Reconstruction can maintain the shape of the image gradient. Maintain the contours of the precise place of the dividing line and eliminate the root causes of the phenomenon.

VIII. PROPOSED METHODOLOGY

This paper uses CNN to detect lung cancer based on chest CT images. Lung regions are extracted from the CT image in the first stage and each slice are segmented to get tumours in that region. The tumour regions segmented are used to train the architecture of CNN. CNN is used to test the images of the patient. The study's main objective o is to detect whether the tumour in the lung is malignant or benign. Figure1 shows the proposed system block diagram. As shown in the figure, the trained system can detect the cancer presence in the CT image of the lung.

Published By:
Blue Eyes Intelligence Engineering
& Sciences Publication (BEIESP)

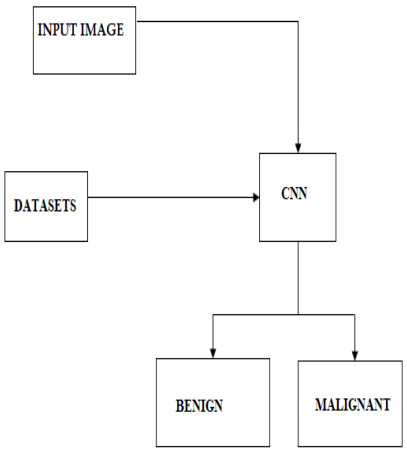


Fig 1: Block Diagram of the Proposed System

IX. DATASET

Lung Image Database Consortium (LIDC) and Image Database Resource Initiative (IDRI) provide training dataset. LIDC and IDRI consist of 1000 CT scans of large and small tumours saved in the format of Digital Imaging and Communications in Medicine (DICOM).

X. EDGE DETECTION

Edge detection is a mechanism of distinguishing edges in an image which could be used in image recognition as a key asset. It is used to locate regions with powerful contrasts of intensity. Sobel, Prewitt, Kirsch and Robinson were also classical gradient edge detection. 2-D spatial gradient measurement on an image is carried out by sobel operator and displaying locations of high spatial frequency that corresponds to edges. It is typically used in an input grayscale image to discover the approximate absolute gradient magnitude at each point in a stage. The operator comprises of a couple of kernels with 3x3 convolution as shown in Table I. one kernel is merely a 90 degrees rotation of the other.

Table I: 3x3 convolution kernels

-1	-2	-1
0	0	0
+1	+2	+1

hx

-1	0	+1
-2	0	+2
-1	0	+1

hy

XI. SEGMENTATION

Segmentation is the method where a digital image is partitioned into various segments/regions. Segmentation's primary objectives is to obtain the ROI (area of interest) for the evaluation of images. A well-known technique is the watershed algorithm that introduces several techniques and approaches are introduced into the segmentation region.

A. *Marker Controlled Watershed Segmentation*

We used markers to decrease regional minimum numbers. The markers idea is a nice approach to segmentation control. The markers are an image element that is linked. These are internal markers and external markers in which inner markers are linked to the object of concern and external markers are linked to the context. Marker watershed based on morphological rebuilding.

B. *Morphological Operation*

The use of marker is the immediate implementation of the watershed segmentation algorithm, usually results in over segmentation owing to noise and other local gradient irregularities. A practical alternative is to restrict the amount of areas permitted by integrating a pre-processing phase intended to incorporate extra expertise into the segmentation process. The idea of markers is based on a strategy used to regulate segmentation.

C. *Otsu Algorithm*

Otsu thresholding involves iterating technique which involves all possible threshold values. These iterated values calculate the pixel levels on each side of the limit. That is; pixels falling either in the foreground or in the background.

XII. DEEP LEARNING

Deep learning composed of multiple layers of nonlinear nodes combines input data with a set of weights so that the algorithm attempts to learn in supervised and/or unsupervised behavior by assigning inputs for the corresponding task. The product sum of these input and weights is passed through node activation function. The output of each layer's is fed simultaneously as input to the subsequent layer. Learning can correspond to different levels of abstraction in multiple levels of representations. Clearly indicate the units in an equation for each quantity.

XIII. CONVOLUTIONAL NEURAL NETWORKS(CNNs)

A CNN is a DNN type consisting of multiple hidden layers like convolution layer, layer of RELU, Pooling layer and a fully linked standardized layer. CNN shares weights in the convolutional layer that reduced memory footprint and enhances network performance. CNN's key features include 3D neurons volumes, local connectivity and shared weights. By converting different sub-regions of the input image with a learned kernel, a feature map is produced by convolution layer. Non-linear activation function is then applied via ReLu layer to improve the convergence properties when the error is low. A region of the image/feature map is selected in the pooling layer and the pixel with maximum value between them or average values is selected as the representative pixel s in order to reduce a 2x2 or 3x3 grid to a single scalar value. This results in a large sample size reduction. Sometimes, in conjunction with the convolutionary layers towards the output stage, traditional Fully-Connected (FC) layer will be used. Convolution layer and pool layer are usually used in some combination in CNN architecture. Usually two types of operations are max pooling and means pooling.

Published By:
Blue Eyes Intelligence Engineering
& Sciences Publication (BEIESP)

In mean pooling, the average neighborhood is calculated within the feature points and the maximum number of feature points is calculated in max pooling. Mean pooling reduces the error resulting from the limitation of neighborhood size and retains background information. Max pooling reduces the estimated error of the convolution layer parameter caused by the mean deviation and thus retains more information about the texture. Fig 2. illustrates the architecture of CNN.

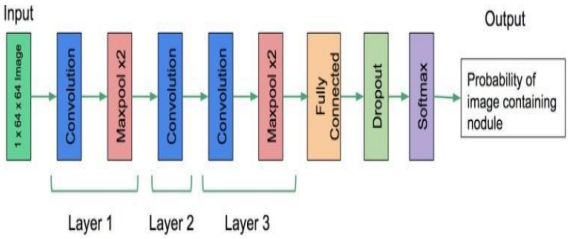


Fig 2: Architecture of CNN

XIV. FEATURE EXTRACTION

Extraction of features is an important phase which reflects the final outcomes to determine an image as normal or abnormal. These characteristics form the foundation of the classification method. Only these elements were usually considered to be extracted: average intensity, perimeter, area and roundness. The features are defined as follows:

1) Perimeter: It is a scalar valuation giving the real number of the pixel of the nodule outline that is acquired by the summing up the registered pixel's interconnected overview in the binary image.

2) Area: it is a scalar value that gives the actual overall pixel number of nodule. It is obtained by the summing pixel area in the image recorded as 1 in the obtained binary image.

3) Roundness: This value is metric or index of irregularity (I) is only 1 in circular form and it is <1 for any other form or shape.

XV. CLASSIFICATION

Lung nodule is smallest lung growth measuring in size from 5mm to 25mm. Malignant nodules in size >25 mm tend to be bigger and have a faster rate of growth. The size of the nodule in the normal image is less than 25mm. And its size is greater than 25mm in an abnormal image. In the segmentation that nodule is identified and then we use extraction function to obtain the primary characteristics from the segmented image by which we can define the stage of lung cancer. Lung nodule on chest x-rays and computed tomography scans appears as round, white opacity. Previous scanning of x-ray and the present x rays and CT-scan are used to determine whether the shape, size, or appearance of the nodules is changed. If after 2 years monitoring the nodule doesn't grow larger, no further treatment is needed.

XVI. IMAGE CLASSIFICATION TRANSFER LEARNING WITH INCEPTION V3

Transfer learning is a machine learning method that uses a neural network that is pre-trained. For example, there are two parts of the image recognition model called Inception-v3:

- Feature extraction part with a convolutional neural network.
- Classification part with fully-connected and SoftMax layers.

The pre-trained model Inception-v3 achieves the latest technology accuracy for recognizing 1000 classes of general objects such as "Zebra", "Dalmatian", and "Dishwasher". The model extracts general features in the first part from input images and classifies them in the second part based on those features.

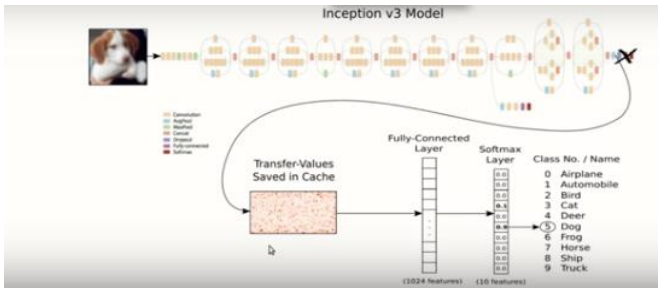


Fig 3: Architecture of Inception V3 Model

In transfer learning, you reuse the feature extraction part and re-train the classification part with your dataset when you build a new model to classify your original dataset. Since you don't have to train the feature extraction part (which is the model's most complex part), you can train the model with fewer computational resources and less training time.

XVII. TRAINING

Inception V3 model is used to train the Deep CNN to detect lung tumors in size 5×20×20 CT image. It made up of two phases. A CNN is the first phase. The classifier is the second phase. It has several layers of FC and threshold, followed by a layer of SoftMax to perform the neural network's high-level reasoning. No scaling was applied to the dataset's CT images to preserve as much as possible the initial DICOM image values. The random sub-volumes extracted from the training set's CT images and during training, normal distribution of voxel values of dataset is estimated. RELU layer will apply an activation function and will not be changed the volume size. POOL layer conducts a spatial dimension down sampling. FC layer computes the class results are calculated by FC layer.

XVIII. RESULTS

The neural network based on convolutional has been implemented in PYTHON.

A. *Marker Extraction and Edge Detection*

Marker extraction is done by internal, external marker and watershed marker. Sobel gradient method is used for edge detection.

Published By:
Blue Eyes Intelligence Engineering
& Sciences Publication (BEIESP)

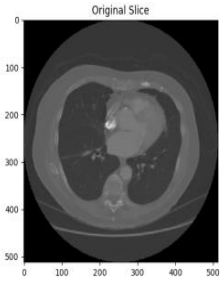


Fig 4: Original slice

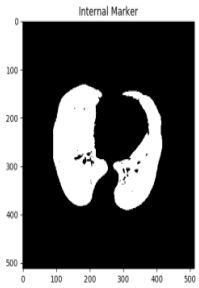


Fig 5: Internal Marker

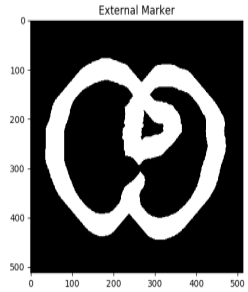


Fig 6: External Marker

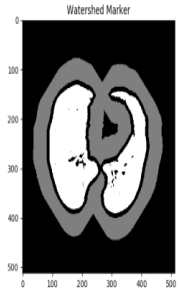


Fig 7: Watershed Marker

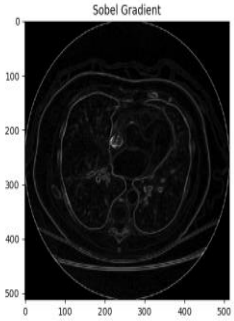


Fig 8: Sobel Gradient image

B. *Segmentation*

Watershed segmentation and morphological operation is used for segmenting the lung. Otsu thresholding is used for extracting the lung nodule.

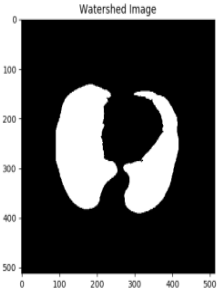
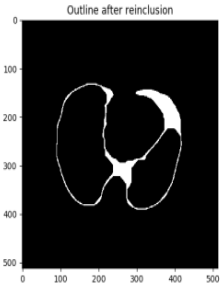


Fig 9: Watershed image



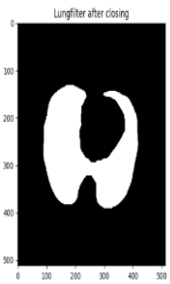
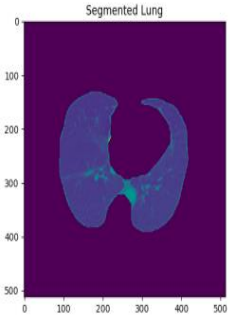


Fig 10: Lung filter after opening & closing



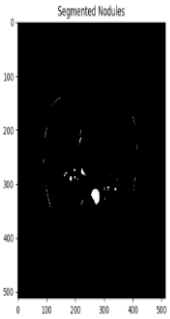


Fig 11: Image of segmented Lung & segmented lung nodule

C. Prediction Using Inception V3 Model

The system is trained with sample data sets to understand and familiarize the Inception V3 model with lung cancer. A sample image was fed to the trained model and the model can tell the presence of cancer at this stage and locate the cancer spot in the sample image of a lung cancer.

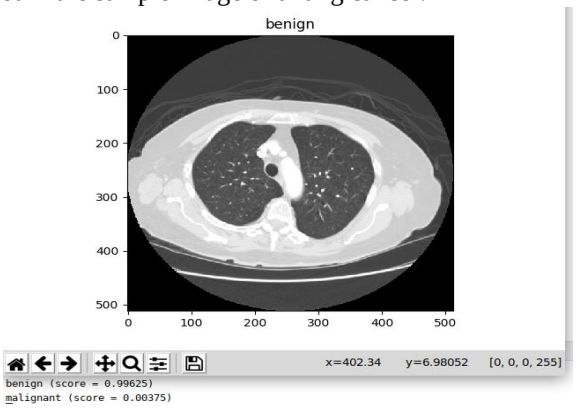


Fig 12: Output for benign image of lung

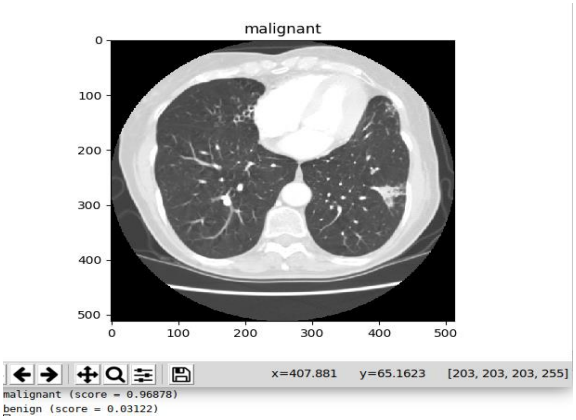


Fig 13: Output for malignant image of lung

Table II. Represents the scores of each output image.

Figures	Benign	Malignant
Fig 12	0.99626	0.00374
Fig 13	0.03147	0.96853

Published By:
Blue Eyes Intelligence Engineering
& Sciences Publication (BEIESP)

Table III. Parameters derived from CNN output.

SL NO	PARAMETERS	VALUES
1	Training steps	4000epoch
2	Training images	200
3	Test images	40
4	Training accuracy	100%
5	Cross entropy	0.01
6	Validation accuracy	70.0%
7	Precision	77.8%
8	False positive	0
9	Final test accuracy	97.0%

Prediction will be done through the trained model and predicting that given test images are either benign or malignant tumors and displaying the image with the corresponding label in the window as shown in the Fig. 12 and Fig. 13 each figure will be showing the scores of benign and malignant. The result is as benign or malignant according to the score of benign and malignant of each figure (see Table II).

The value of false positive is 0 (see Table III), which means that there is no misclassification of image in our work.

XIX. CONCLUSION

To detect the malignancy tissues, present in the input lung CT image, a convolutional neural network-based system was implemented. At the input for training the system was fed lung image with different shape, size of the cancerous tissues. The proposed system can detect the presence and absence of cancer cells with approximately 97% accuracy. Our proposed work obtained 100% training accuracy, which shows that there is no false positive detection. Other parameters such as cross entropy and validation accuracy are also high in our work.

REFERENCES

1. Lin, D. T., & Yan, C. R, "*Lung nodules identification rules extraction with neural fuzzy network*," *In Proceedings of the ICONIP* (pp. 2049-2053), November 2002.
2. Bhargavi, K., & Jyothi, "*S. A Survey on Threshold Based Segmentation Technique in Image Processing," IJIRD*, *3*(12), 76-85, 2014.
3. Chaudhary, A., & Singh, S. S, "*Lung cancer detection on CT images by using image processing*," *In Proceedings of the ICCSE* (pp. 142-146), 2012, Phagwara: India.
4. Dai, S., Lu, K., Dong, J., Zhang, Y., & Chen, Y, "*A novel approach of lung segmentation on chest CT images using graph cuts*," *Neurocomputing* journal, Elsevier, *168*(2), 799-807, 2015.
5. Nikita Pandey, *"A Novel Approach of Cancerous Cells Detection from Lungs CT Scan Images,'' IJARCSSE* Volume 2, Issue 8, August 2012.
6. Al Mahmud, M. A., Karim, A. Z., & Rahman, M. M, "*Brain Tumor Detection using MR Images through Pixel based Methodology*," *GJCST, 15*(4), 302-315, 2015.
7. QinghuaJi, Ronggang Shi "*A Novel Method of Image Segmentation Using Watershed Transformation*," *ICCSNT* , 2011.
8. Prof. Samir Kumar Bandyopadhyay, "*Edge Detection from CT Image of Lung," IJESAT* Volume - 2, Issue - 1, 34 – 37, Jan-Feb 2012.
9. FatmTaher, Naoufel Werghi, "*Extraction of Sputum Cells using Thresholding Techniques for Lung Cancer Detection*," *IIT*, March 2012.
10. Kumar, V., & Saini, "*A. Detection system for lung cancer based on neural network: X-Ray validation performance*," *IJERMCA, 2*(9), 40-47, 2013.

AUTHORS PROFILE

Anuja J, currently doing Master of Technology, department of CSE at Lal Bahadur Shastri institute of technology for women, Poojappura, Thiruvananthapuram. BTech graduated from Kerala University.

Smitha Vas P, currently working as an Assistant professor, department of CSE at Lal Bahadur Shastri institute of technology for women, Thiruvananthapuram. MTech graduated from Kerala University & BTech received from Cochin University. Area of interest are Image Processing & ML.