LUNG CANCER DETECTION AND CLASSIFICATION USING SVM

[1]Dasari Swetha, [2]V. Siva Macharla, [2]P. Abhishek, [2]S. Sivaiah,[2]R. Gopi Krishna,[2]P. Pratyusha

[1]Assistant Professor, [2]UG Student,
[1&2]Department of ECE,
[1&2]Bapatla Engineering College, Bapatla, India

Abstract: Lung cancer is a common cause of death for people all over the world. The most prevalent type of lung cancer has a significant fatality rate. Lung cancer screening can help identify cancer at an early stage if the sickness is recognized and treated while it is still in its early stages. As a result, early detection of lung cancer increases patients' chances of survival. Computed tomography (CT) scans of the lungs may be more accurate than X-rays and MRIs in detecting the presence of cancer. The lungs are scanned, and images are obtained using DICOM (Digital Imaging and Communications in Medicine). We apply a range of preprocessing techniques to enhance and smooth images. In this study, the region of interest for the lung tumor is segmented using edge detection and thresholding (ROI). Finally, we compute several geometrical features of the retrieved ROI using Support Vector Machine and classify them into benign and malignant severity levels (SVM). supervised machine learning methods called SVM classifiers are used to categorize data. They have the advantage of handling tiny samples of high dimensional data. Performance of the SVM is assessed using each feature as an input. As a result, a system that employs image processing techniques is used to diagnose lung cancer in CT scans.

Index Terms - Lung Cancer Detection, Lung Malignant Growth, Lung Diseases.

I. INTRODUCTION

Malignant cells develop in the lung tissue as a result of lung cancer. Tumors can develop as a result of the improper division of body cells, which is known as cancer. Malignant cancerous tumors have the ability to metastasize and spread to nearby tissues. Lung cancer is the most common cause of cancer death in both men and women among all forms of malignancies. Small cell lung cancer (SCLC) and non-small cell lung cancer (NSCLC) are the two main types of lung cancer [1]. Cancer mortality can be decreased by early identification and treatment. is terminally ill, making lung cancer early diagnosis challenging. The majority of lung cancers are brought on by smoking, however lung cancer can also strike non-smokers. Early cancer detection is aided by lung cancer screening. When a person has no symptoms of cancer, the test is used to check for several cancer types. Chest x-rays and computed tomography (CT) scans can both detect lung illness. The most dependable and efficient screening method available is the CT scan. The results of a biopsy support this choice. Often, bronchoscopy or CT guidance is used for this.

Lung cancer at Stage I, Stage II, Stage III, and Stage IV are the basic divisions of lung cancer in terms of severity and dissemination. Visual interpretation might result in manual mistakes and incorrect staging of cancer. A computerized and automated system is therefore necessary for accurate diagnosis. With his various image processing strategies and machine learning techniques, he has published several papers on the detection of lung cancer. a CAD system that uses a convolutional neural network as a classifier to find lung cancer. To extract a region of interest, they employ a circular filter (ROI).[4] Although the cost of implementation is lower, the precision is still poor. For her segmentation or grouping, the author utilizes the K-means unsupervised learning technique, and she gets an accuracy of roughly 90.7%. The author of suggests a system for dividing lung cancer into benign and malignant forms. To determine an area of interest

(ROI), extract various shape and texture features, train, and classify a support vector machine, the system uses a priori knowledge and Hounsfield Units (HU). benign or cancerous. However, it has the drawback of requiring prior knowledge of the area of interest. For preprocessing, the author employs a median filter and a Gaussian filter, and for ROI determination, he used watershed segmentation. Further features were retrieved and identified using the SVM classifier for her malignant nodules that were found.[4] In, the authors use morphological operations for image preprocessing and segmentation of tumor regions and her SVM classifier for classifying the extracted features as normal or abnormal. The authors of use deep learning and machine learning to detect and classify tumors. However, did not implement a classification system for benign or malignant tumors.

In this study, a novel method for classifying and detecting lung tumours using CT imaging is proposed. Employ a variety of image smoothing and enhancement preprocessing techniques. The lung tumour regions are then separated from the lung mask using thresholding and edge detection.[4] Lastly, we calculate the size, eccentricity, perimeter, compactness, and circularity of the retrieved ROI. Based on common training features for benign and malignant tumours, support vector machines are utilised to categorise test features as benign or malignant. Using our suggested strategy, we discovered that.

II. PROPOSED SYSTEM MODEL

We propose a system that can detect lung tumor regions using CT imaging and classify lung disease as benign or malignant using machine learning. Figure 1 shows the system flow diagram of the proposed system model. The system works in main steps:image preprocessing, image segmentation, training feature extraction, test feature extraction and classification.
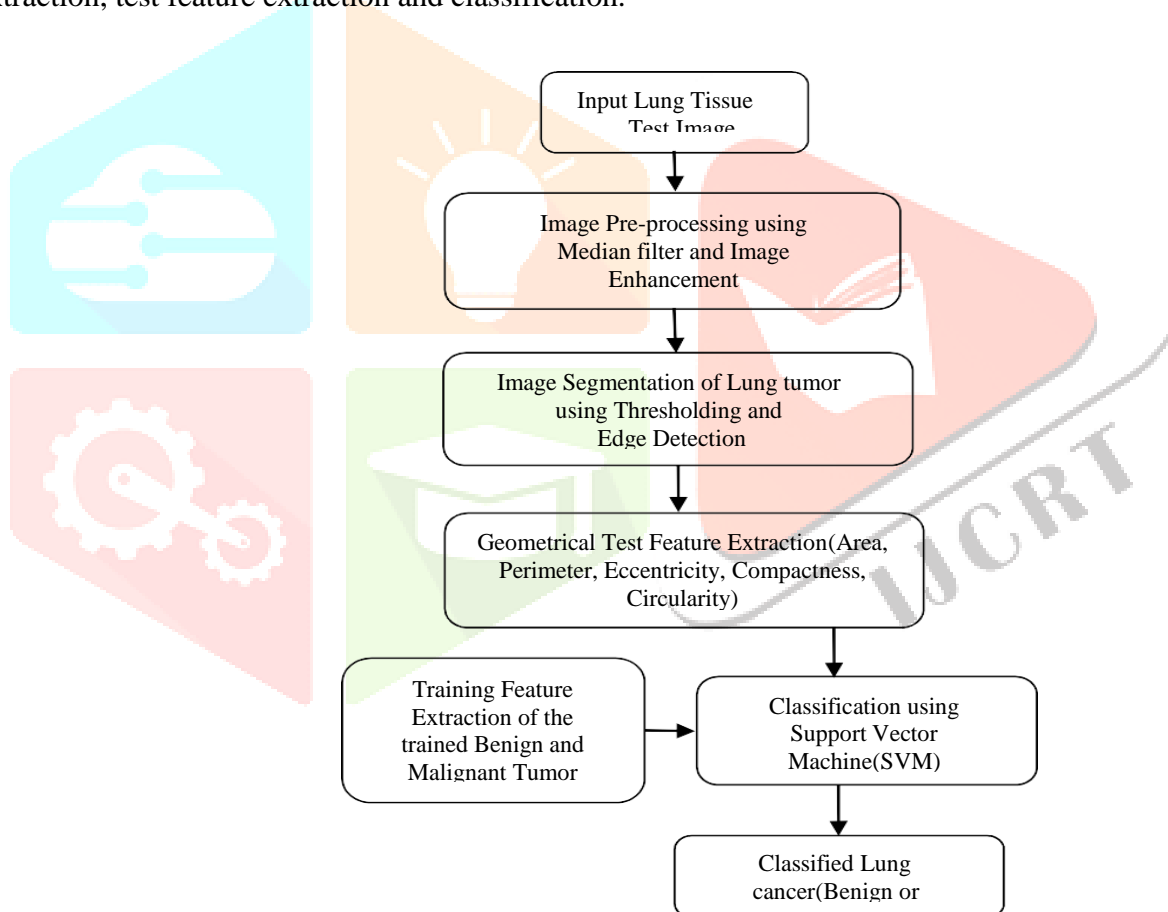
Fig. 1. System flow diagram of the Proposed system model [4]

The system first uses the input image as a test image. The first phase then begins with the pre-processing of the photographs, which uses a few techniques for image scaling, smoothing, and getting the best level of value and clarity through sharpening.[4] The next phase employs edge detection and thresholding techniques to segment the lung tumour. Afterwards, the trained benign and malignant tumours' characteristics are extracted. The segmented tumour region of the test image is then used to extract five characteristics in the following phase. The characteristics include Circularity, Eccentricity, Compactness, Area, and Perimeter. The test features are classified as benign or malignant in the last stage using support vector machines (SVM), and the output of the classification is provided as a marker.

III. **METHODOLOGIES**

We initially gather information from the accessible Cancer Imaging Archive to build the system (TCIA). The collection includes a variety of Computed Tomography (CT) pictures of both healthy people and people with lung cancer. We extract DICOM-formatted CT images from the database of patients with lung illness, both benign and malignant.[4] The system is being developed using the photos as training and test images. Here is an explanation of the methodologies:

A. *Image pre-processing:*

Pre-processing an image tries to reduce undesirable noise and highlight important aspects of the image. Image resizing, picture smoothing, and image enhancement are the three steps that are included.

1) *Image Resizing:*

The first step is to scale all CT images to 256 by 256 pixels. The images are then cropped to obtain a particular position and dimension.

2) *Image smoothing:*

We do image averaging using median filtering, smoothing. While de-noising and maintaining the edges' integrity, it is a non-linear operation that is more effective than convolutional operation. The filter employs a zero-padding approach on the edges in order to preserve the median value of the edges.

3) *Image Enhancement:*

Eventually, sharpening visual features like boundaries and edges as well as reducing artefacts brought on by contrast changes are the fundamental goals of image improvement operations. We adjust the input image's contrast level for our suggested system. After the image pre-processing operation, we can check the contrast by finding and displaying the histogram of the image data. Histogram plot utilises 256 equally separated bins where each of the bin represents different range of data value. We have to smooth the image to produce a less pixelated image if the contrast level is not satisfactory.

B. Image Segmentation:

The division of a picture into various parts is known as image segmentation. Segmentation aims to represent the image in a way that makes it simpler to analyse. It is typically used to identify any borders or objects in a picture. The specific issues or application requirements will determine how well the strategies work. Finding and segmenting the lung cancer nodules or the region of interest (ROI) of the disease in the lung tissues is the aim of our image segmentation technology.[4] To segment the lung tumours, we first use thresholding and edge detection, then contrast adjustment and clear border operation.

1) *Thresholding:*

To convert a grayscale image into a binary image, the thresholding method uses a clip level (or threshold value). To determine a threshold or level, we employ the thresholding method. To convert grayscale photos to binary images, the threshold level's value is used. The level in this case denotes the image's intensity level. The maximum entropy approach, balanced histogram thresholding, Otsu's method (maximum variance), and k-means clustering are some of the techniques utilised in the industry for thresholding. Using Otsu's approach, one may calculate global. Threshold based on the grayscale picture. To reduce the intraclass variance of the black pixels and white pixels, this method employs a threshold [6]. Any imaginary portion of the image that is not zero is ignored. The pre-processed grayscale images are then converted to binary images based on the calculated threshold [4].

2) *Edge detection:*

The edge detection process finds the edges of the background binary image. There are different edge detection methods as Sobel, Prewitt, Roberts, canny edge etc. We detect the edges of the binary images using 'Roberts' method as it performs best for binary images. The detected edge is then subtracted from the binary image.

C. Feature Extraction:

An important step in the processing of images is feature extraction. The process of gathering informative features or measured data that are obtained from measured data is known as feature extraction. By taking the mathematical characteristics of the tumour ROI and analysing them, the lung tumor's condition is evaluated. We measure each region property of the tumor's labelled ROI using the regionprops operation. Then, using the measured properties, we compute five features of the tumour region that are suggested for the classification of benign and malignant tumours. The extracted features are Area, Perimeter, Eccentricity, Compactness and Circularity. The attributes are described as follows:

1) Area:

The area is determined by the regionprops operations as the actual number of pixels in the region. Our CT scan images have a 300-dpi resolution. Given that 1 dpi = 1 ppi = 25.8 mm, we can use the following formula to determine the area in mm:

$$300\text{ppi} = \frac{300}{25.8} \, pixels \, per \, mm = 11.8 \, pixels \, per \, mm$$

Hence, we can determine the area in mm by multiplying the actual number of pixels in the area by 11.8.

2) Perimeter:

The circumference of the region's boundaries. The number of pixels in the region's boundaries is used by region props to calculate the perimeter. The perimeter is computed as follows:

$$\text{Perimeter} = \sum_{i=1}^{N-1} d_i \tag{1}$$

To determine the perimeter in millimeters, divide the measured number of pixels by 11.8.

3) Eccentricity:

The ellipse that indicates the eccentricity has an equivalent second-moment as the region. It is the ratio of the ellipse's principal axis' length to the space between its foci. This metric value is sometimes referred to as the irregularity index of circularity or roundness. The value ranges from 0 to 1. If the eccentricity is 1, an ellipse can be defined as a line segment, and if it is 0, it can be defined as a circle.

4) *Compactness***:**

The ratio of an object's area to the area of a circle with the same perimeter, when the circle is employed with the most compact shape, is referred to as compactness. In the case of lung tumours, benign tumours have a more rounded, smooth contour. As a result, it is expected that a malignant tumour will have a higher value of compactness than a smooth, spherical benign tumour. The alternative compactness formula is applied in this situation as follows:

$$\text{Compactness} = \frac{Perimeter^2}{2.Area} \tag{2}$$

A circle in this situation has a compactness value of zero, and it rises with shape elongation or roughness.

5) *Circularity***:**

It is a measurement of circularity or roundness that eliminates local imperfections. You can get it in the following ways:

$$\text{Circularity} = \frac{4\pi.Area}{12(Perimeter)^2} \tag{3}$$

It is equal to 1 for circular objects and less than 1 for non-circular objects.

D. Classification:

The identified tumour is now officially classed as malignant or benign. Support vector machine (SVM) is used as a classifier. The classification method is depicted [4]. It uses supervised machine learning to classify data based on patterns after analysing the input data. SVM divides the training datasets into two classes or groups before defining the group or classifier function. The test datasets are then split up into one of these two categories. The defined predicted group function is where wT is an m-dimensional vector, b is a bias term, and xi are training inputs. Since i=1 here, M. G (xvalue)'s is dependent on ||w||.
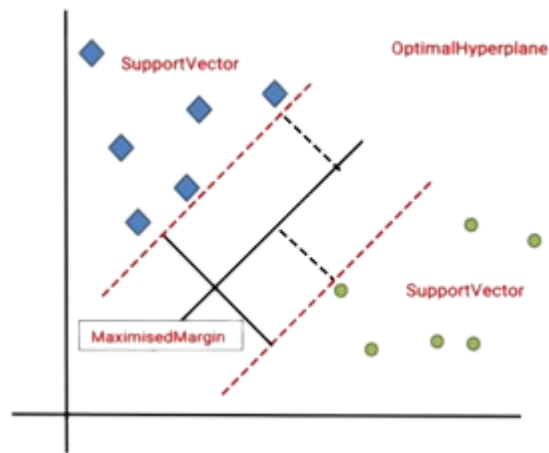
Fig.2: SVM for regression

IV. SIMULATION RESULTS

Using MATLAB, we create the proposed system model for classifying and detecting lung cancer. From the Cancer Imaging Archive, we extract the CT scan pictures for the training and test dataset. The stages of image preprocessing, image segmentation, and feature extraction are applied to the input test images. Figure 3 displays two input CT scan pictures from two patients with lung disease together with their corresponding histograms. A histogram image is a graphic representation of a photograph that shows pixel distribution throughout the image's various dark levels.

Later in the preprocessing stage, salt and pepper noise was eliminated using the Median filter [3x3]. Following median filter image resizing and smoothing operations, we obtain the cropped and filtered images depicted in with contrast adjustment, the image is further improved to improve it. shows the histograms for the improved CT scans. The histograms are significantly better following the image pre-processing processes, as can be seen.

Edge detection and thresholding are used to segment lung tumours. The Otsu's method of thresholding is used to extract the threshold values, which are then used to transform the image to a binary format.[4]

The ROI of the lung tumour for the two CT scan images is then obtained following contrast correction and the clear border operation, as illustrated in Figure 5. We compute five geometrical features from the training and test CT images using feature extraction, including area, perimeter, eccentricity, compactness, and circularity.[4] Using MATLAB's regionprops method, features are extracted. Using equations (2) and (3), compactness and circularity are then calculated (3). Training aspects are employed as the defining characteristics for the classification of diseases into benign and malignant tumours.
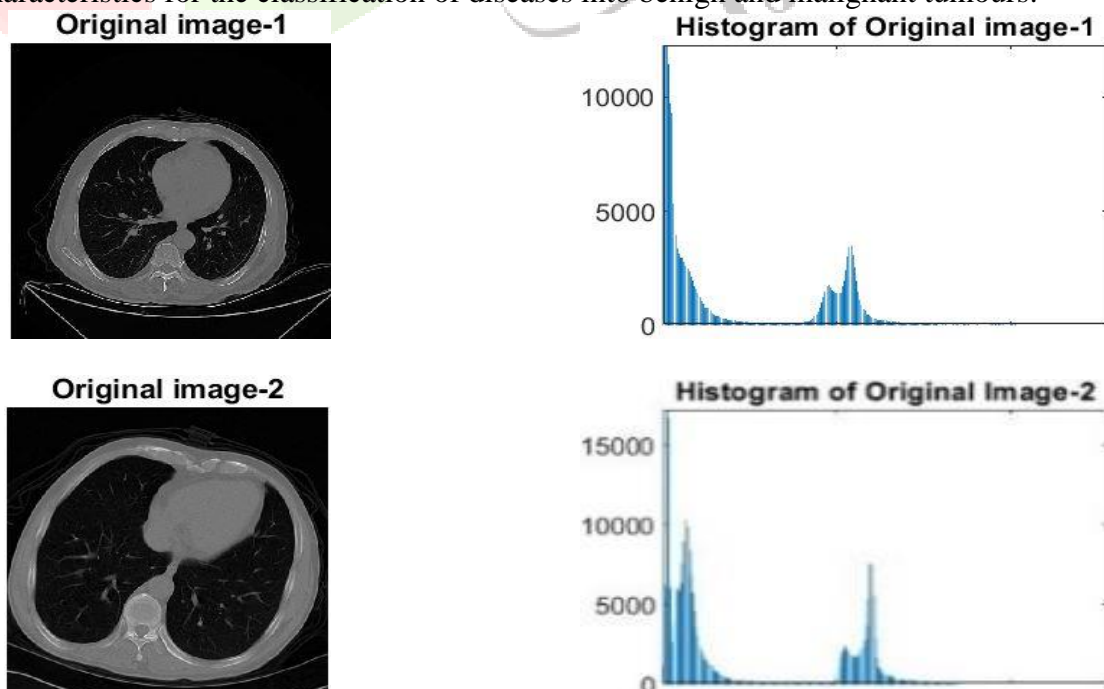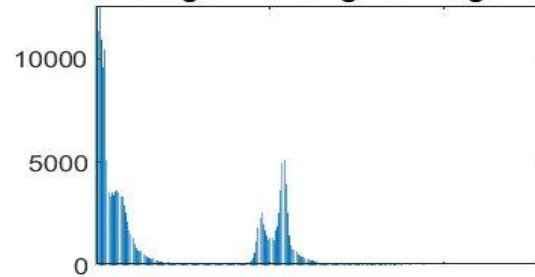

Fig. 3. Two of the Lung disease CT images and their Histogram
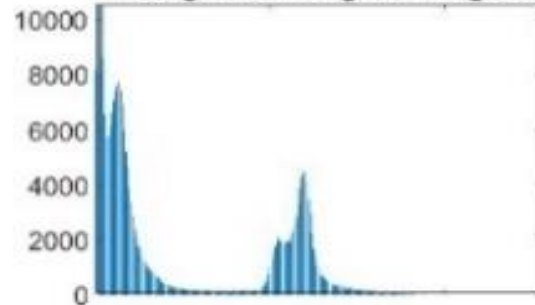
Fig. 4. The Enhanced Images and their Histograms

Fig. 5. ROI of the Lung Tumor and their Classification

Figure 3 shows two CT scans for lung disease along with their histograms. shows the median-filtered image of the CT images and the cropped image. To train the training features as benign and malignant tumours, we employ SVM. The SVM classifier is used to compare the extracted test features from the identified tumour ROI with the training features, and based on how closely the values match, either group is designated as benign or malignant [4]. For test CT image 1 and 2, Figure 5 classifies the lung tumour that was found. Using our suggested methods for lung cancer detection and classification, we generated 50 of these CT images in simulation. Table I displays the training characteristics of benign and malignant tumours as well as the categorization of the test features that were collected from two CT scan images. Table II contains the simulation results and our suggested method's classification of a few CT images.

TABLE I. CLASSIFICATION BASED ON THE EXTRACTED TEST FEATURES AND THE TRAINING FEATURE

ROI Properties	Training Feature		Test Features and Classification of the detected ROI			
	Benign	Malignant	Image 1	Classification	Image 2	Classification
Area(mm)	>6	>6	18.5		7.9	
Perimeter(mm)	>6	>6	18.88		13.8	
Eccentricity	0.11	0.8	0.7905		0.32	
Compactness	0.3	1	2.6	Malignant	0.4	Benign
Circularity	0.89	0.11	0.071		0.71	

TABLE II. CLASSIFICATION OF NODULES BY PROPOSED SYSTEM METHOD

Image	Nodule	Classification	Final Class
Image 15	Detected ROI is Malignant Nodule 1	Malignant	Malignant
Image 23	Detected ROI is Benign Nodule 1	Benign	Benign

V. CONCLUSION

In this project, we Compute geometrical Features of the Extracted ROI and Classify them into Benign and malignant Using SVM Model. We find Significant accuracy in the detection of lung cancer. In this technique, we have used CT images Which gave better image quality as compared with the other images. By using the Proposed model, we can detect the tumor region in the CT scan images accurately compared other techniques.

Using image processing and machine learning techniques, we have suggested a method for detecting and classifying lung cancer in this work. Using an SVM classifier, we were able to determine whether a tumour was benign or malignant based on its characteristics. We tested our proposed method on 50 CT scan simulations taken from the Cancer Imaging Archive and found that it can accurately identify lung tumours and classify them.

VI. ACKNOWLEDGMENT

REFERENCES

[1] Makaju, S., Prasad, P. W. C., Alsadoon, A., Singh, A. K. and Elchouemi, A., "Lung cancer detection using CT scan images", Procedia Computer Science, 125, pp.107-114, 2018.

[2] Lung Cancer- Patient Version", National Cancer Institute, https://www.cancer.gov/types/lung, 2020.

[3] "Non-Small Cell Lung Cancer", NCCN clinical practice guidelines in oncology, version 6.2020, June 15, 2020.

[4] Nusaat Nawreen,Umma Hany, Tahmina Islam, " Lung Cancer Detection and Classification using CT Scan Image Processing",July 2021.

[5] https://www.mathworks.com

[6] Rangayyan, R. M., El-Faramawy, N. M., Desautels, J. L. and Alim, O. A., "Measures of acutance and shape for classification of breast tumors", IEEE Transactions on medical imaging, 16(6), pp.799-810, 1997.

[7] Otsu, N., "A Threshold Selection Method from Gray-Level Histograms." IEEE Transactions on Systems, Man, and Cybernetics. Vol. 9, No. 1, 1979, pp. 62–66.

[8] Khobragade, S., Tiwari, A., Patil, C., and Narke, V., "Automatic detection of major lung diseases using Chest Radiographs and classification by feed-forward artificial neural network." IEEE International Conference on Power Electronics, Intelligent Control and Energy Systems, 1-5, 2016.

[9] Mutiullah, Bari, M., Ahmed, A., Sabir, M., and Neveed, S., "Lung Cancer Detection Using Digital Image Processing Techniques: A Review", Mehran University Research Journal of Engineering & Technology, Vol. 38, No. 2, 351-360, April 2019.

[10] Pradeebha, R. and Karpagavalli, S., "Prediction of lung disease using HOG features and machine learning algorithms", International Journal of Innovative Research in Computer and Communication Engineering, 4(1), pp.3158-3169, 2016.

[11] Rajbongshi, N., Bora, K., Nath, D. C., Das, A. K. and Mahanta, L. B., "Analysis of morphological features of benign and malignant breast cell extracted from fnac microscopic image using the pearsonian system of curves", Journal of cytology, 35(2), p.99, 2018.

[12] Nadkarni, N. S. and Borkar, S., "Detection of Lung Cancer in CT Images using Image Processing", International Conference on Trends in Electronics and Informatics, pp. 863-866, IEEE, April, 2019.

[13] Rendon-Gonzalez, E., & Ponomaryov, V., "Automatic Lung nodule segmentation and classification in CT images based on SVM", International Kharkiv Symposium on Physics and Engineering of Microwaves, Millimeter and Submillimeter Waves, 2016.

[14] Asuntha, A. and Srinivasan, A., "Deep learning for lung Cancer detection and classification", Multimedia Tools and Applications, pp.1-32, 2020.

[15] Jin, X., Zhang, Y., & Jin, Q., "Pulmonary Nodule Detection Based on CT Images Using Convolution Neural Network", International Symposium on Computational Intelligence and Design, 2016.

[16] Sangamithraa, P., and Govindaraju, S., "Lung tumour detection and classification using EK-Mean clustering", International Conference on Wireless Communications, Signal Processing and Networking, 2016.