Classification and Segmentation Techniques for Detection of Lung Cancer from CT Images

Prenitha Lobo
Dept. of Computer Science and Engineering
St Joseph Engineering College
Mangalore, India
prenithalobo94@gmail.com

Sunitha Guruprasad
Dept. of Computer Science and Engineering
St Joseph Engineering College
Mangalore, India
sunithag@sjec.ac.in

Abstract— **Information Technology(IT) has played an important role in all aspects of human life. In the recent years, it has influenced the medical and healthcare field. Image processing together with machine learning and other technologies are used to study medical images for earlier disease detection and treatment. Diseases like cancer should be discovered as fast as possible where time factor is vital. Medical images have factors such as noise, uncertain tumor boundaries and large variation in tumor appearance which makes it difficult to find exact tumor region. To overcome these problems, several methods are available for helping the radiologist in accurately detecting the cancer. This paper has reviewed classification and segmentation techniques used in detecting the lung cancer tumor and evaluated the performance of each approach. It has proposed a method for effective identification of lung cancer. The proposed method gave 79.166% accuracy.**

*Keywords— **Image processing, classification, segmentation, lung cancer, medical imaging.***

I. Introduction

Lung cancer is the multiplication of abnormal cells which is uncontrolled, that starts in either or both the lungs. When cancer starts in the lung cell, it is called as primary cancer. Abnormal cells split quickly to form tumors. As the tumors grow in size, they restrict the ability of the lungs to provide bloodstream with oxygen. There are two types of tumors: Benign tumors and Malignant tumors. Benign tumors stay in one particular place and do not spread [1]. Malignant tumors are the hazardous ones and extend to the other part of the body through bloodstream [1]. According to the World Health Organization (WHO), each year 7.6 million deaths are caused by the cancer and represents 13% of all global deaths. Lung cancer is considered as the number one cancer killer [1].

Lung cancer is diagnosed using Computed Tomography (CT) scans, chest X ray etc. Lung cancer forms the tumor but all tumors are not cancerous which makes it hard to detect the lung cancer. Exploring the medical image is a challenge and involves obtaining the insight value, analysis and diagnosis of the specific disease. In this modern world of computerized living, several Computer aided Diagnosis (CAD) systems are developed for detecting cancer at initial stage. It helps the radiologists to reduce the errors and the misinterpretations. But the CAD systems developed for detecting the cancerous tumor are not so accurate. There is a need of the system that fully automates the cancer detection showing the tumor region.

Literature review in this paper provides several segmentation and classification techniques used in lung cancer detection. The purpose of this paper is to review the related work and summarize different techniques used and propose a method that improve the performance of the system. The paper is organized as follows: Section II gives an overview of the CAD system, section III gives Literature review, section 1V gives the summary of the related work, section V gives proposed method and section VI provides results and discussion and VII gives the conclusion.

II. Overview of the CAD System

The system consists of Pre-processing stage, Segmentation stage, Feature Extraction stage and Classification stage. In the pre-processing stage, quality of the image is enhanced and noise from the image is removed. This is important since CT images have noise and makes it difficult to identify the tumor region correctly. In the Segmentation phase, the image is divided into several segments. The ultimate aim is to represent the image into something valid for easier analysis [2]. In feature extraction, the large data set is reduced into set of features. Features are distinctive properties of the input and helps in differentiating the categories of the input pattern. Features represent the raw image and are given to the classification process. In classification, the images are classified according to the features given to it. Classification forms classes and differentiates the images. The aim of classification is not only to reach the accuracy but to identify which part of the body is infected by cancer. Lung cancer images can be classified as cancerous or noncancerous. Several machine learning and data mining techniques can be used for segmentation and classification purpose for improving the accuracy of the cancer detection. Figure 1 shows the main stages in the lung cancer detection system.

III. Literature Review

Dr.T.Arumuga Maria Devi et.al [3] implemented a technique for automatic detection of lung cancer from CT images. Image was enhanced using the contrast limited adaptive histogram. Segmentation was performed using the thresholding based on entropy. Finally, SVM was used for classification where it gave accuracy of 93%.

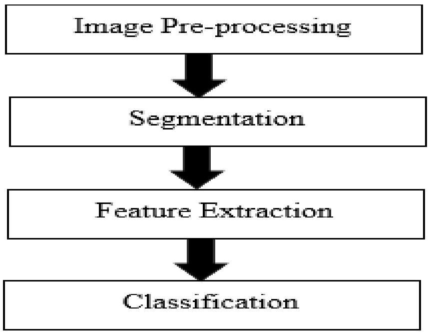

Figure 1: Main stages in the Lung Cancer Detection System

P.B.Sangamithraa et.al [4] presented a method to classify the lung cancer CT image as cancerous or noncancerous. Noise removal of the image was done by applying median and wiener filter. Fuzzy k means was carried out for segmentation and the result was made more accurate using EK means clustering. Features were exacted and given to back propagation network(BPN). System gave 90.87% accuracy. For provide better accuracy, SVM is suggested.

Yosefina Finsensia Riti et.al [5] developed a system for lung cancer detection. Initially, image was pre-processed to obtain the region of interest. Otsu thesholding was performed and false areas were eliminated using connected component labelling. Classification of the lungs lesion margin as regular or not were done by using Multi-Layer Perception (MLP). Results showed 85% accuracy.

Suren Makaju et.al [6] implemented a model to identifying the type of cancer. Median filter and Gaussian filter were used for noise removal. Watershed technique was applied for segmentation. SVM was used for classification to distinguish the cancerous region as benign or malignant. The system gave 92% accuracy. Different stages of cancer can be found in the future.

Eman Madagy et.al [7] executed an approach to differentiate the normal and cancer CT images. Weiner filter was applied for pre-processing. For segmentation, histogram analysis, morphological operations and thresholding were combined. Amplitude-Modulation Frequency-Modulation were used to exact the features and further given to Partial Least Squares Regression. Several classifiers were compared. The CAD system achieved 95% accuracy with linear classifier.

Emre Dandil et.al [8] created a CAD model to classify the types of lung cancer. The image was enhanced using median filter, histogram equalization was used for increasing constrast by altering the image intensities. Lung lobes were extracted using the morphological operations and remaining pieces on the edges were removed using double thresholding. Self-organizing map was used for segmentation of the lungs. GLCM was used for feature exaction and was further given for Principle Component Analysis for feature reduction. Artificial neural network (ANN) was applied for classification. The CAD system gave 90.63% accuracy.

Anam Tariq et.al [9] implemented a technique for detecting the lung cancer. Image processing was done after the segmentation phase. Segmentation of the image was done using elimination of background and threshold segmentation. Several morphological operations were used for image post processing. The features were extracted and given to hybrid classified, i.e. neural network and fuzzy. It was called Neuro-fuzzy. The method gave 95% accuracy.

Salsabil Amin et.al [10] implemented a method for detecting the lung cancer from CT images. Pre-processing was done using contrast stretching and enhancing. Combination of morphological operations, thresholding and region growing were used for lung image segmentation. Rule based classifier was used which reduced the false positives. The system gave 70.53% accuracy. In the future, CNN or SVM for classification is suggested.

Jaspinder Kaur et.al [11] presented the CAD system to find the lung cancer tumor. Median filter was used to eliminate noise. Gradient mean was used to remove the background of the image. Optimal thresholding was used for segmentation. After the segmentation, post processing was carried out which included morphological opening, edge detection, morphological closing. Region growing method was used to find the interested region. Features were extracted and given to back propagation network(BPN). The system gave 98% accuracy.

S.Kalaivani et.al [12] presented a method for automating the lung cancer identification. The image was pre-processed using the histogram equalization to increase the contrast. Thesholding was performed to segment the image and remove the unwanted areas. Features were extracted from the images and given to a feed forward BPN. The system gave 78% accuracy. Results can be improved by using large datasets.

Md. Badrul Alam Miah et.al [13] presented a technique for identifying the lung cancer tumor. Image was pre-processed using the binarization method. Thresholding was used for segmentation. Neural network was trained for the classification. Results showed 96.67% accuracy.

Omnia Elsayed et,al [14] presented an approach to automatically detect the lung cancer nodules. Image was pre-processed by converting the pixels into Hounsfield units (HU). Region growing was applied, chest filling was done in order to remove the interfering objects. Segmentation was done using gray level thresholding and binary lung filling was done to extract vascular tree and nodule. The features were extracted and classification was carried for the combination of four classifiers: quadratic, linear, parzen and neural network. The combination of parzen and neural network classifier worked best. The results showed 98% accuracy.

Jinsa Kuruvilla et.al [15] proposed an approach for detecting the lung cancer. Morphological operations are performed to segment the images. Features were selected and given to the feed forward back propagation for the classification. Their proposed training function gave maximum accuracy. The results showed that system gave 93.3% accuracy.

Niranjan Shukla et.al [16] implemented a method for identifying the lung cancer tumor as benign or malignant. Fuzzy filter was used for removing noise. Watershed was used for segmentation. Features were extracted using GLCM and given to SVM classifier. The system gave 92.5% accuracy.

Manasee Kurkure et.al [17] proposed the Naive Bayes and Genetic algorithm for the identification of lung cancer nodule. Initially histogram was applied to the image. Identification of strong edges was done by canny edge detection. Naïve Bayes was used for classification and performance of the classification was enhanced using genetic algorithm. Classification results showed 80% accuracy. The drawback was that GA is slow and requires complex computation.

Allison M Rossetto et.al [18] implemented ensemble of Convolution neural networks (CNN) with the voting system for differentiating the lung CT images. Two CNN with and without matlab neural network toolbox were used. The two CNN acquired the features and made predictions. Voting system was used at the. It gave the accuracy of 97% with toolbox. In the future, complex weighted voting is used.

Bhagyarekha U. Dhaware et.al [19] developed a method for determining the lung cancer. Contrast Limited Adaptive Histogram Equalization was applied for enhancing the contrast of the image. Gray level co-occurrence matrix was employed for exacting the features. Sequential forward selection for selecting the features. Bayesian classifier was trained to classify the CT image as cancerous or not. The cancerous image was segmented using Fuzzy c means. It gave high accuracy. For better accuracy, SVM is suggested.

Avinash. S et.al [20] presented a solution for locating the lung cancer nodule. Image was pre-processed using Gabor filter. Marker watershed technique was used for the segmentation which separated the touching objects in the image. Classification of the lung CT images as cancerous or noncancerous was done by calculating the area of the tumor to the standard tumor value from hospital.

IV. SUMMARY OF THE LITERATURE REVIEW

A plethora of work has been carried out in developing a system that automatically detects the lung cancer tumor. Table 1 presents the summary of the related work. It gives different segmentation and classification methods used in the existing work, accuracy of the work and the future work.

V. PROPOSED METHODOLOGY

The proposed system detects whether a given lung CT image is cancerous or noncancerous. It helps in early diagnosis of the disease and increases the survival rate. In the proposed method, Support vector machine (SVM) is considered for classifying the image and Fuzzy c means (FCM) for image segmentation. SVM is chosen as a classifier because it proves to be more efficient machine learning algorithm for pattern recognition in comparison with other techniques like neural network and Bayesian classifier. It has better performance in case of large datasets to simplify the problem. In the literature review, SVM was widely used for classification. Most of the researchers have suggested SVM for better accuracy. Its major advantage is that data training is relatively simple. SVM is a supervised learning model. First, the SVM is trained with the given examples that belong to one of the two categories.

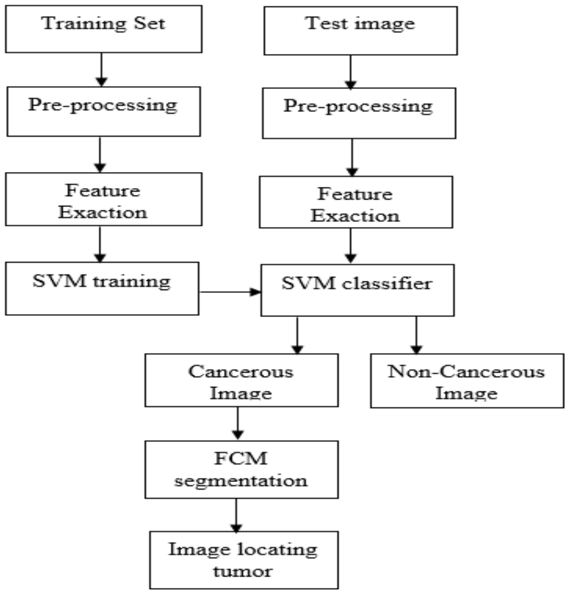

Figure 2: Workflow of the proposed system

The SVM creates a training model that then assigns the new data to the category it belongs. This makes it a probabilistic binary linear classifier. FCM is an unsupervised clustering algorithm and one of the suitable method for medical image segmentation. The technique is based on finding the cluster centers and grouping the similar data values in the same cluster. Usually, in the regular CAD system, the segmentation of the image is carried out in the earlier stage. But in the proposed system, segmentation is carried out at the end and only on cancerous images to find the exact lung cancer tumor region. It is not carried out on all images. Figure 2 gives the workflow of the proposed system.

VI. RESULTS AND DISCUSSION

Lung CT scan dataset was taken from an oncology center in Mangalore, Karnataka. The CT images were in the dicom format and had a slice thickness of 2.5mm. The experiment was carried out in MATLAB R2016b. The images of both men and women of all age groups were taken for experiment. The steps carried out and their results are listed below.

A. Image Pre-processing

CT image contains noise and other artefacts and disturbance. Contrast Limited Adaptive Histogram Equalization was applied for enhancing the contrast of the image. CLAHE was used to improve the image quality and restrict amplification of noise. Figure 3(a) and 3(b) shows the normal and cancerous image. Figure 4 shows the input lung CT image. Figure 5 and 6 shows graph before and after CLAHE.

B. Feature Extraction

Gray level co-occurrence matrix was employed for exacting the features. GLCM matrix was created and features were extracted from it. Texture features like contrast, correlation, energy, sum variance, maximum probability, dissimilarity were extracted from the image. Features generated were real value spatial data. These features are vital in classification.

Table 1: Summary of the different approaches

Reference	Year	Method Used		Accuracy	Limitation /Future work
		Segmentation	Classification		
[3]	2016	Thresholding	SVM	93%	Threshold method depends on peaks rather than special data
[4]	2016	Fuzzy k means	BPN	90.87%	Obtain better results with SVM
[5]	2016	Otsu thresholding	Artificial neural network	85%	Otsu methods assumes the image histogram is bimodal.
[6]	2017	Watershed	SVM	92%	Determine the stages of cancer.
[7]	2015	Histogram+ Threshold+ Morphological operations	Linear classifier	95%	Not suitable on nonlinear data.
[8]	2014	Self-Organizing Map (SOM)	Artificial neural network	90.63%	In SOM, lack of data adds randomness to grouping clusters.
[9]	2013	Background elimination+ Threshold	Neuro-fuzzy	95%	Computationally intensive.
[10]	2017	Morphological operations+ threshold + region growing	Rule based classifier	70.53%	Reduction in false positives using SVM or convolutional neural networks
[11]	2014	Optimal thresholding	Feed forward back propagation network	98%	Optimal thresholding needs some prior knowledge about object and background distribution.
[12]	2017	Thresholding	Feed forward and BPN	78%	Increase in the performance using large databases.
[13]	2015	Thresholding	Neural network	96.67%	Technique can be used for brain cancer detection.
[14]	2015	Gray level threshold+ binary lung filling	Parsen+ neural network	98%	Improve the results for larger dataset.
[15]	2014	Morphological operations	Feed forward and BPN	93.8%	Time taken for the neural network training is too high.
[16]	2015	Watershed	SVM	92.5%	Watershed algorithm over segments the images.
[17]	2016	-	Naïve Bayes+ Genetic Algorithm	80%	Genetic Algorithm is time consuming for large data.
[18]	2017	-	CNN with voting system	97%	Complex system voting system for better accuracy.
[19]	2016	Fuzzy c means	Bayesian classifier	-	Better accuracy using SVM.
[20]	2016	Watershed	Calculate tumor area	-	Watershed is affected by noise

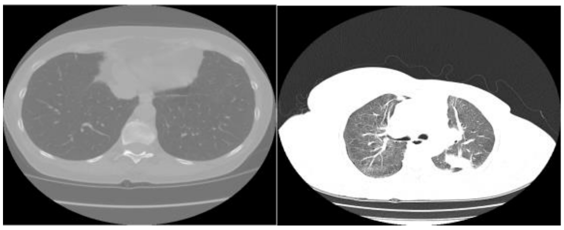

Figure 3(a): Normal Image Figure 3(b): Cancerous Image

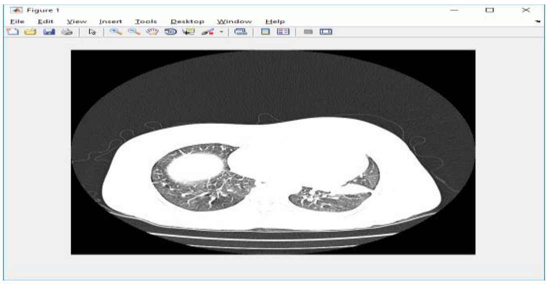

Figure 4: Input image

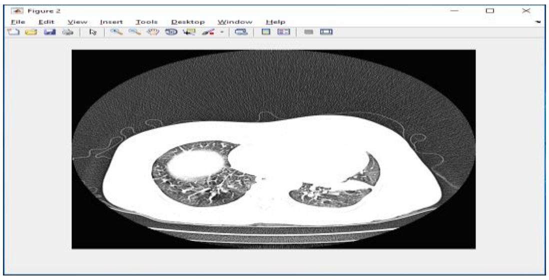

Figure 5: Pre-processed image

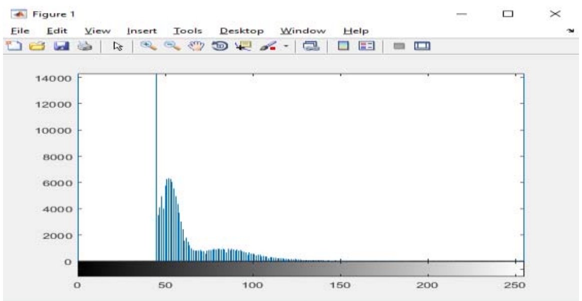

Figure 6: Histogram before CLAHE

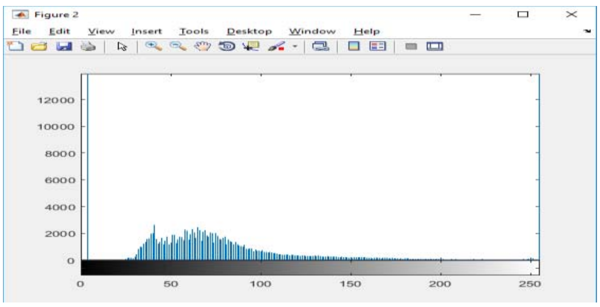

Figure 7: Histogram after CLAHE

C. Classification

Support vector machine was used for identifying the image as normal or cancer. SVM is a widely used machine learning supervised classifier. First, SVM was trained using the train data and model was validated using the test data. Features of train and test images were extracted. The extracted features were given to SVM which efficiently classified the images. SVM first labelled the train images. It internally creates a hyperplane and divides the new data depending on the previously encountered data. Figure 8 shows a dialog box for the classification of cancerous image.

D. Segmentation

Segmentation was carried out only on cancerous images. If a particular image was classified as containing cancer, then Fuzzy C Means clustering was applied to highlight the tumor area in the image. FCM groups the data point to one or more cluster. Cluster centers were initialized and membership of each data was given between 0 to 1. Then cluster centers were updated based on membership. Procedure was iterated and the output image was a clustered image. Figure 6 shows the image highlighting the cancerous tumor. It shows where the cancerous tumor is located.

The SVM used for classification gave good results. It could accurately detect the given image as normal or abnormal. It also produced less false positives. FCM on the other hand could highlight the tumor region in the image but it could not extract the cancerous region separately. Table 2 shows the performance of the proposed approach. Proposed method gave less false positives and achieved 79.166% accuracy, 83.33% recall and 76.92% precision.

Figure 8: Image classified as cancerous

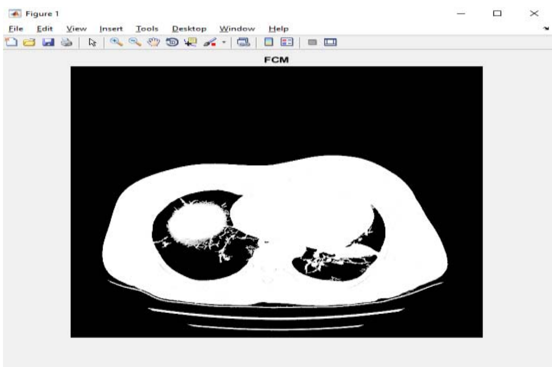

Figure 9: Image locating the tumor

Table 2: Performance of Proposed Approach

Proposed Method	
Accuracy	79.166%
Recall	83.33%
Precision	76.92%

VII. CONCLUSION

This paper gives an overview of current segmentation and classification techniques used in the identification of lung cancer tumor from CT images. It may help researchers to select an appropriate method. The paper also proposed a method for efficient segmentation and classification. SVM gave good results with 79.166% accuracy. FCM highlighted the tumor in the image but it could not extract the tumor. In the future, FCM can be combined with other method to extract the tumor area. Over the last decade, there has been lot of improvement in lung analysis techniques. However, there are still issues that need to be solved.

REFERENCES

[1] https://www.medicalnewstoday.com/info/lung-cancer

[2] Mokhled Altarawneh," Lung Cancer Detection Using Image Processing Techniques", Leonardo Electronic Journal of Practices and Technologies, 2012, Issue 20, ISSN 1583-1078.

[3] Dr. T. Arumuga Maria Devi, Mr. Mebin Jose V.I, Dr. P. Kumar Parasuraman," A novel approach for automatic detection of non-small cell lung carcinoma in ct images", International Conference on Control, Instrumentation, Communication and Computational Technologies (ICCICCT), 2016, ISBN: 978-1-5090-5240-0, DOI: 10.1109/ICCICCT.2016.7988059.

[4] P.B.Sangamithraa, S.Govindaraju," Lung tumour detection and classification using EK-mean clustering", International Conference on Wireless Communications, Signal Processing and Networking (WiSPNET), 2016, ISBN: 978-1-4673-9338-6, DOI: 10.1109/WiSPNET.2016.7566533.

[5] Yosefina Finsensia Riti, Hanung Adi Nugroho, Sunu Wibirama, Budi Windarta, Lina Choridah," Feature extraction for lesion margin characteristic classification from ct scan lungs image", International Conference on Information Technology, Information Systems and Electrical Engineering (ICITISEE), 2016, ISBN: 978-1-5090-1567-2, DOI: 10.1109/ICITISEE.2016.7803047.

[6] Suren Makaju, P.W.C. Prasad, Abeer Alsadoon, A.K.Singh, A. Elchouemi, "Lung cancer detection using ct scan images", 6th International Conference on Smart Computing and Communications, ICSCC 2017, https://doi.org/10.1016/j.procs.2017.12.016.

[7] Eman Magdy, Nourhan Zayed, and Mahmoud Fakhr, "Automatic classification of normal and cancer lung ct images using multiscale am-fm features", International Journal of Biomedical Imaging Volume 2015, http://dx.doi.org/10.1155/2015/230830.

[8] Emre Dandil, Murat Cakrolu, Ziya Eksi, Murat Ozkan, Ozlem Kar Kurt, Arzu Canan," Artificial neural network-based classification system for lung nodules on computed tomography scans", 6th International Conference of Soft Computing and Pattern Recognition (SoCPaR), 2014, ISBN: 978-1-4799-5934-1, DOI: 10.1109/SOCPAR.2014.7008037.

[9] Anam Tariq, M. Usman Akram, M. Younus Javed, "Lung nodule detection in ct images using neuro fuzzy classifier", IEEE Fourth International Workshop on Computational Intelligence in Medical Imaging (CIMI), 2013, ISSN: 2326-9928, ISBN: 978-1-4673-5919-1, DOI: 10.1109/CIMI.2013.6583857.

[10] Salsabil Amin El-Regaily, Mohammed Adbel Megeed Salem, Mohamed Hassan Abdel Aziz, Mohamed Ismail Roushdy," Lung nodule segmentation and detection in computed tomography", Eighth International Conference on Intelligent Computing and Information Systems (ICICIS), 2017, ISBN: 978-1-5386-0821-0, DOI: 10.1109/INTELCIS.2017.8260029.

[11] Jaspinder Kaur, Nidhi Garg, Daljeet kaur," An automatic CAD system for early detection of lung tumor using back propagation network", International Conference on Medical Imaging, m-Health and Emerging Communication Systems (MedCom), 2014, ISBN: 978-1-4799-5097-3, DOI: 10.1109/MedCom.2014.7006014.

[12] S.Kalaivani, Pramit Chatterjee, Shikhar Juyal, Rishi Gupta, "Lung cancer detection using digital image processing and artificial neural networks", International conference of Electronics, Communication and Aerospace Technology (ICECA), 2017, ISBN: 978-1-5090-5686-6, DOI: 10.1109/ICECA.2017.8212773.

[13] Md. Badrul Alam Miah, Mohammad Abu Yousuf," Detection of lung cancer from ct image using image processing and neural network", International Conference on Electrical Engineering and Information Communication Technology (ICEEICT), 2015, ISBN: 978-1-4673-6676-2, DOI: 10.1109/ICEEICT.2015.7307530.

[14] Omnia Elsayed, Khaled Mahar, Mohamed Kholief, Hatem A. Khater, "Automatic detection of the pulmonary nodules from ct images", SAI Intelligent Systems Conference (IntelliSys), 2015, ISBN: 978-1-4673-7606-8, DOI: 10.1109/IntelliSys.2015.7361223.

[15] Jinsa Kuruvilla, K. Gunavathi, "Lung cancer classification using neural networks for ct images", Computer Methods and Programs in Biomedicine Volume 113, Issue 1, January 2014, https://doi.org/10.1016/j.cmpb.2013.10.011.

[16] Niranjan Shukla, Aakash Narayane, Aniket Nigade, Krishnakumar Yadav, Harshada Mhaske, "Lung cancer detection and classification using support vector machine", International Journal of Advanced Trends in Computer Science and Engineering, 2016, DOI: 10.18535/Ijecs/v4i11.20.

[17] Manasee Kurkure, Anuradha Thakare, "Lung cancer detection using genetic approach", International Conference on Computing Communication Control and automation (ICCUBEA), 2016, ISBN: 978-1-5090-3291-4, DOI: 10.1109/ICCUBEA.2016.7860007.

[18] Allison M Rossetto, Wenjin Zhou," Deep Learning for Categorization of Lung Cancer CT Images", IEEE/ACM International Conference on Connected Health: Applications, Systems and Engineering Technologies (CHASE), 2017, ISBN: 978-1-5090-4722-2, DOI: 10.1109/CHASE.2017.98.

[19] Bhagyarekha U. Dhaware, Anjali C. Pise, "Lung cancer detection using bayasein classifier and FCM segmentation", International Conference on Automatic Control and Dynamic Optimization Techniques (ICACDOT), 2016, ISBN: 978-1-5090-2080-5, DOI: 10.1109/ICACDOT.2016.7877572

[20] Avinash. S, Dr. K. Manjunath, Dr.S.Senthil kumar," An improved image processing analysis for the detection of lung cancer using gabor filters and watershed segmentation technique", International Conference on Inventive Computation Technologies (ICICT), 2016, ISBN: 978-1-5090-1285-5, DOI: 10.1109/INVENTIVE.2016.7830084.