International Research Journal of Engineering and Technology (IRJET) e-ISSN: 2395-0056
Volume: 08 Issue: 06 | June 2021 www.irjet.net p-ISSN: 2395-0072

Lung Cancer Detection Extraction and Classification Using Digital Image Processing and Machine Learning

Lakhvir Kaur[1], Kushagra Sharma[2], G. N. Verma[3], Kapil Kumar[4]

[1]M. Tech Student, Department of Computer Science & Engineering, SSIET, IKGPTU, Punjab, India
[2,4]Assistant Professor, SSIET, IKGPTU, Punjab, India
[3]Professor, SSIET, IKGPTU, Punjab, India

Abstract - *There are various types of diseases that have created various challenges for human society. Out of these diseases cancer has shown the most devastating results. Various researchers are trying to analyze and also to find out the growth pattern of this disease so that cancer can be finding out in its early stages and can be cured easily. In this research there is a hybrid way of detecting, extracting as well as classifying the various lung cancers using image processing and machine learning is proposed. Lung cancer actually growths so fast that in some weeks it becomes untreatable and ultimately can cause the death of that person. Lung cancer has four stages. So it is advisable to identify and treat it in its early stages. In this present methodology first of all databases of various lung cancer is gathered and then depending upon the feature extraction the database images are labeled into various classes of cancers. Then this machine learning approached with the combined effect of digital image processing segmentation approach is used to check the lung cancer present in various images. So this automated system can help the radiologists to collect and analyze the data. All of the above work will be implemented in Matlab. Various parameters like Precision, Recall with True Positives, False Positives, True Negatives as well as False Negatives will be used for checking the authenticity of the proposed system.*

Key Words: Image Processing, Lung cancer, machine learning, Matlab, Segmentation, disease.

1. INTRODUCTION

Cancer is a serious general heath disease around the globe which is usually due to unbalanced life culture of humans where the mortality rates increasing year by year. Lung cancer is one of the most common type of cancer out of all other cancers but it is deadliest of all of these and it also does not care the gender of the person. One name of lung cancer is carcinoma where there is formation of malignant lung tumors which are nothing but the cancerous nodules which appeared in the human body due to uncontrolled growth of cells in the interior of lung tissues.

There could be many reasons behind this but main of these could be the eating tobacco as well as smoking which could be the leading risk factors for causing cancer. From the literature survey it is found that he survival rate of lung cancer patient after combining all four stages is nearly to 14% with a life time span of about 5 to 6 years only. [1] This is due to the fact that most of the time these diseases are not recognized in its earlier stages. If the patient reaches to third or fourth stage then it becomes very difficult for doctors to survive the patient. So this type of disease should be diagnosed as early as possible and should be diagnosed at first level if possible. It is because in the earlier stages the recovery rate of survival rate is quite good which is near to 70%. [1-2]

There are two types of cancer classifications relied on the cell characteristics.

1. Small cell lung cancer and
2. Non-small cell lung cancer

Out of these two types from the literature survey it is found that the usually the non small lung cancer occurs in approximately 85% cases and the small lung cancer is just got 15% cases. As lung cancer has many stages which further depend on the spread as well as size of tumor in the lungs. As already stated lung cancer has four stages which are described here. [3]

- Stage I- At this stage cancer is just confined to the lung,
- Stage II and III- At this stage cancer is moved to the chest
- Stage IV-At this stage the lung cancer has spread from the chest to various sections of the body

Traditionally the physicians or radiologists use the visual interpretation method to diagnose the disease and its local as well s global spread in the body. But this visual interpretation is not a well defined solution as this visual interpretation efficiency is dependent on the individual to individual. So it will lead to human error and could lead to misclassification other disease as cancer or vice versa. So there is urgent need of some automation diagnosis system which is capable of diagnoses the disease and extract its periphery and also help the physicians to save time for diagnosis the disease. [3-5]

This process of automatic diagnosis of lung cancer can be firmly done by using the digital image processing, pattern recognition methodologies, feature extraction and reduction to optimize the database and various classification methods so that we can label the data as types of cancer and its stages wise. Also due to machine learning it becomes possible to

learn the model so that this model can be further used to diagnose the disease. There are various algorithms which could help in this regard. So all of these theory has give us the insight to use this technology to diagnose the disease. [6]

2. LITERATURE SURVEY

N. S. Nadkarni et al. [1] proposed a methodology for finding the lung cancer in the Computed Tomography scanned images. Authors utilized various methods for the methodology. Authors used the median filtering algorithm to preprocess the digital CT scanned image to remove noise. Further authors used the segmentation method so that region of interest can be extracted from the image and it is done by with the help of morphological operators. Authors used the three geometrical features which were area, perimeter and eccentricity that were used to extract the tumor region from the scanned image. Various geometrical features were extracted by authors and these features were utilized for classification of input scanned images into two categories which are normal and abnormal. Here the authors had used the support vector machine for classification. Results of the proposed methodology showed the accuracy and early stage detection with ease.

W. Rahane et al. [2] proposed the system in which prime target was to detect lung cancer in the computed tomography scanned digital images. Authors had created a web application in which patient could upload his diagnosis medical report and then that report was analyzed by the server side script that captured the data in the report and finally outputs the result whether the user had cancer or not. Authors here had used image processing and machine learning algorithms for finding the outcomes. Authors had also used filtering and segmentation techniques to get the results. Authors had used some objective parameters to find the region of interest of the tumor. Authors had used the support vector machine (SVM) for classification purposed to get the positive and negative samples of the tumor.

K. Gopi et al. [3] had proposed method that had series of operations like preprocessing of scanned digital image, count of mask that could be generated by use of thresholding approach, also use of various noise removal filtering techniques to remove unwanted information from input scanned digital image. Further authors had used the region of interest (ROI) as tumor extraction using gray-level co-occurrence matrix (GLCM) method. After then authors had used the Support vector machine (SVM) classifier for classification purposes where authors had used the entropy, correlation and convexity parameters. Authors had used the famous LIDC-IDRI datasets in the proposed system which had 850 lesions of lungs. Authors found that the proposed system showed the accuracy approximately upto 92% and also showed good efficiency in comparison to other segmentation algorithms which were used for comparison in against to proposed algorithm.

S. M. Salaken et al. [4] proposed a deep autoencoder classification mechanism. This process first of all started learning deep features and then it trained the artificial neural network with the given learned features. Actually the process was divided into the two stages. In the first stage various classifiers were examined with the use of holdout validation. Outcome of the study showed that the deep learned classifier worked better in comparison to other classifiers It is also demonstrated that the performance improvement is statistically significant. The prime target of this stage was to check for one classifier that show good results on the taken dataset to get the best performance. In the second stage by using the above classifier and used the cross validation methodology for make a comparison among various classifiers. Here the authors used the four fold strategy so that training and testing could be done efficiently. The above proposed classifier showed good result in comparison to ANN, KNN, SVM and Adaboost classifiers.

A. M. Rossetto et al. [5] presented an group of Convolution Neural Networks(CNN) which had various preprocessing techniques so that maximum accuracy of the automated labeling of the scans could be generated. Authors used the voting system to get benefits of two networks. Authors had used the lung cancer images database which was delivered from Kaggle Data Science 2017 contest. This dataset had a collection of data of more than 1000 patients which had more than 150,000 individual slices of computed tomography scanned images. The database classified each dataset object into positive or negative. Here the attributes of each slice had different values of thickness, quality of image and other attributes. Results of the proposed algorithm showed that the significance over the other techniques by providing very high accuracy of nearly 97% and quite low value of false positive rate.

3. RESEARCH METHODOLOGY

In this proposed research machine learning support vector machine is utilized for the purpose of training the dataset. Proposed system extracts various features of the lung dataset images and records various classes of cancer. Then the input lung image is tested with the trained system. Matlab 2016a is utilized for the implementation of proposed research work.

3.1 Proposed Technique

Various steps which are used in the proposed research work are given below

(1) Read the scanned digital image
(2) Apply image processing methods to remove noise and other limitations
(3) Use classifier to extract the various features of images from the labelled database
(4) Use machine learning to train 80% of sample database. Test the network with the remaining 20% of samples

(5) Apply the trained network to on the scanned images and verify the results
(6) Compute the efficiency of the proposed algorithm in comparison to the other method for lung cancer detection
(7) Visualize the results

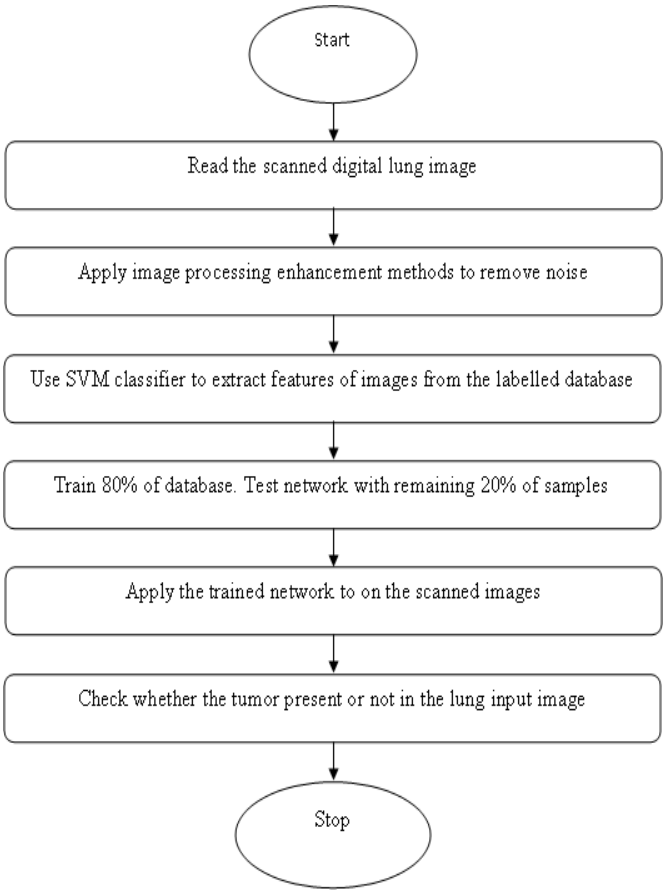

Fig -1: Flowchart of proposed system

3.2 Objective Performance Evaluation

For evaluating the performance of the proposed algorithm various objective parameters like precision, recall are used.

(i) **Precision**

$$\text{Precision} = \frac{\text{True Positive}}{\text{True Positive} + \text{False Positive}}$$

(ii) Recall

$$\text{Recall} = \frac{\text{True Positive}}{\text{True Positive} + \text{False Negative}}$$

4. RESULTS

In the given experiment analysis the supposed algorithm is compared with the standard genetic algorithm.

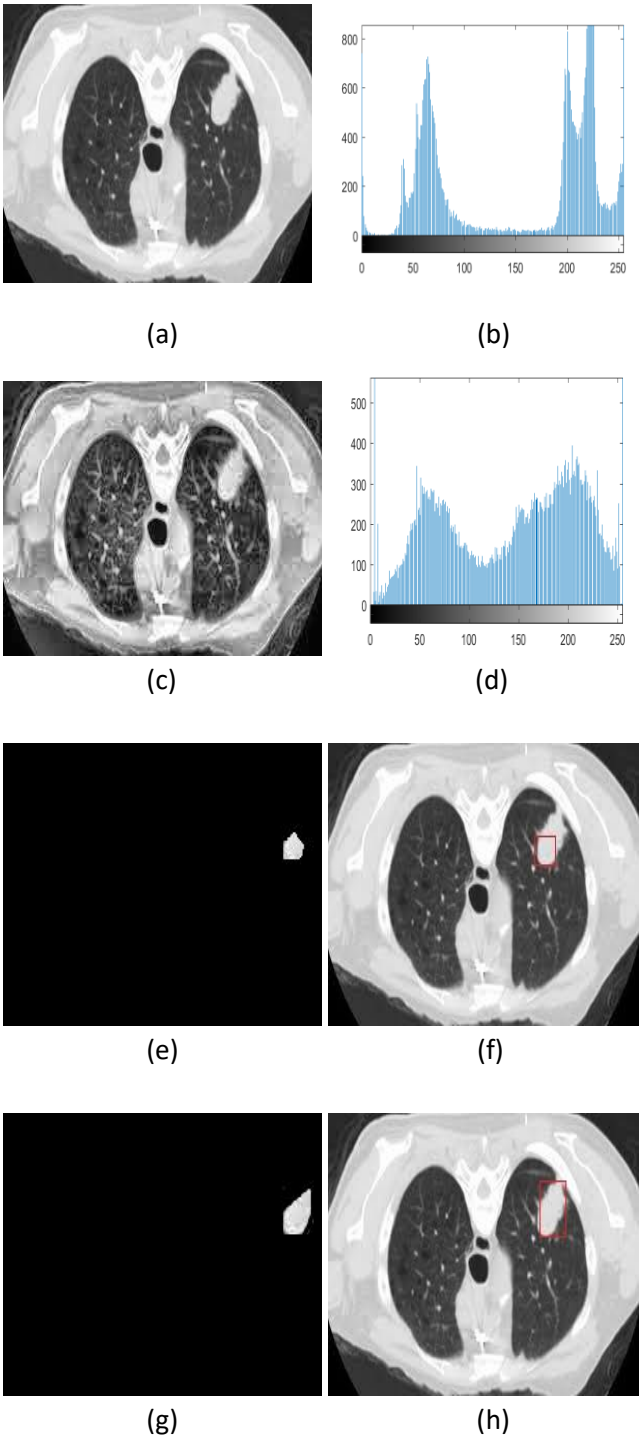

Fig-2(a-h): **Original image, histogram of original, enhanced, histogram of enhanced, segmented tumor by GA approach, segmented tumor by GA in original, segmented tumor by proposed approach and segmented tumor by proposed approach in original image 1**

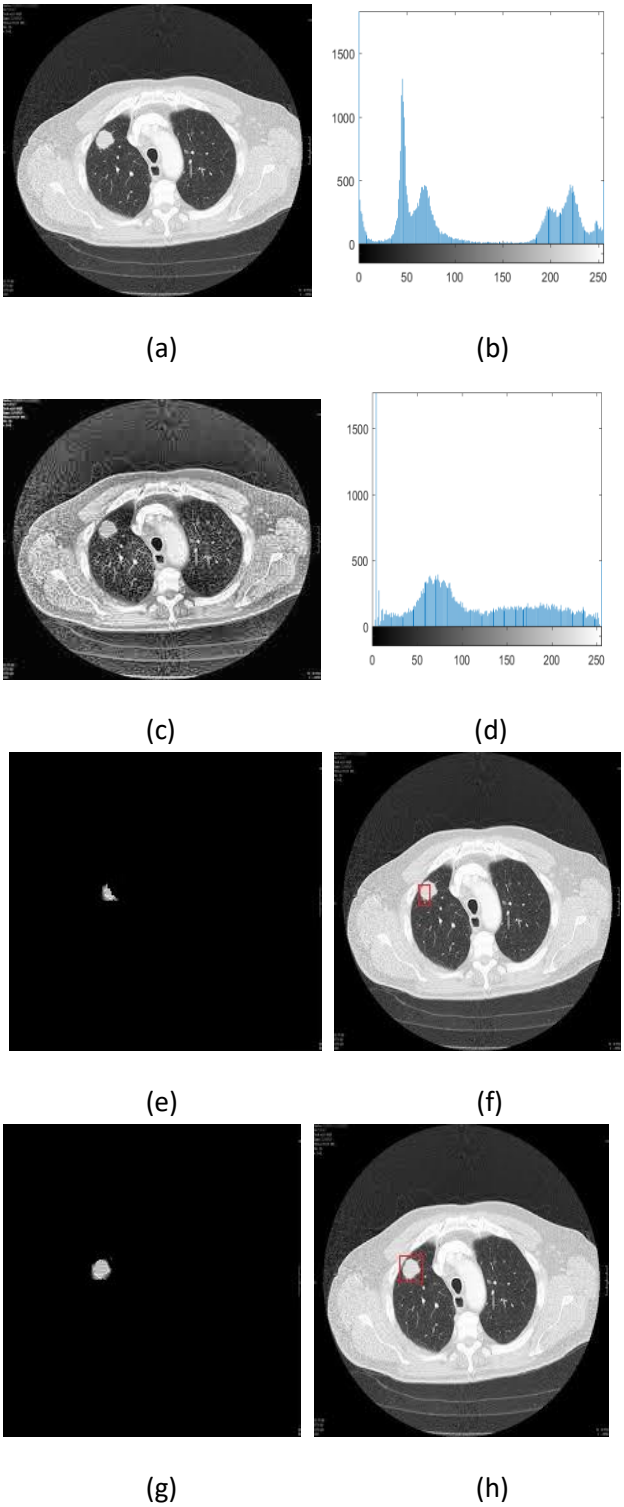

Fig-3(a-h): **Original image, histogram of original, enhanced, histogram of enhanced, segmented tumor by GA approach, segmented tumor by GA in original, segmented tumor by proposed approach and segmented tumor by proposed approach in original image 2**

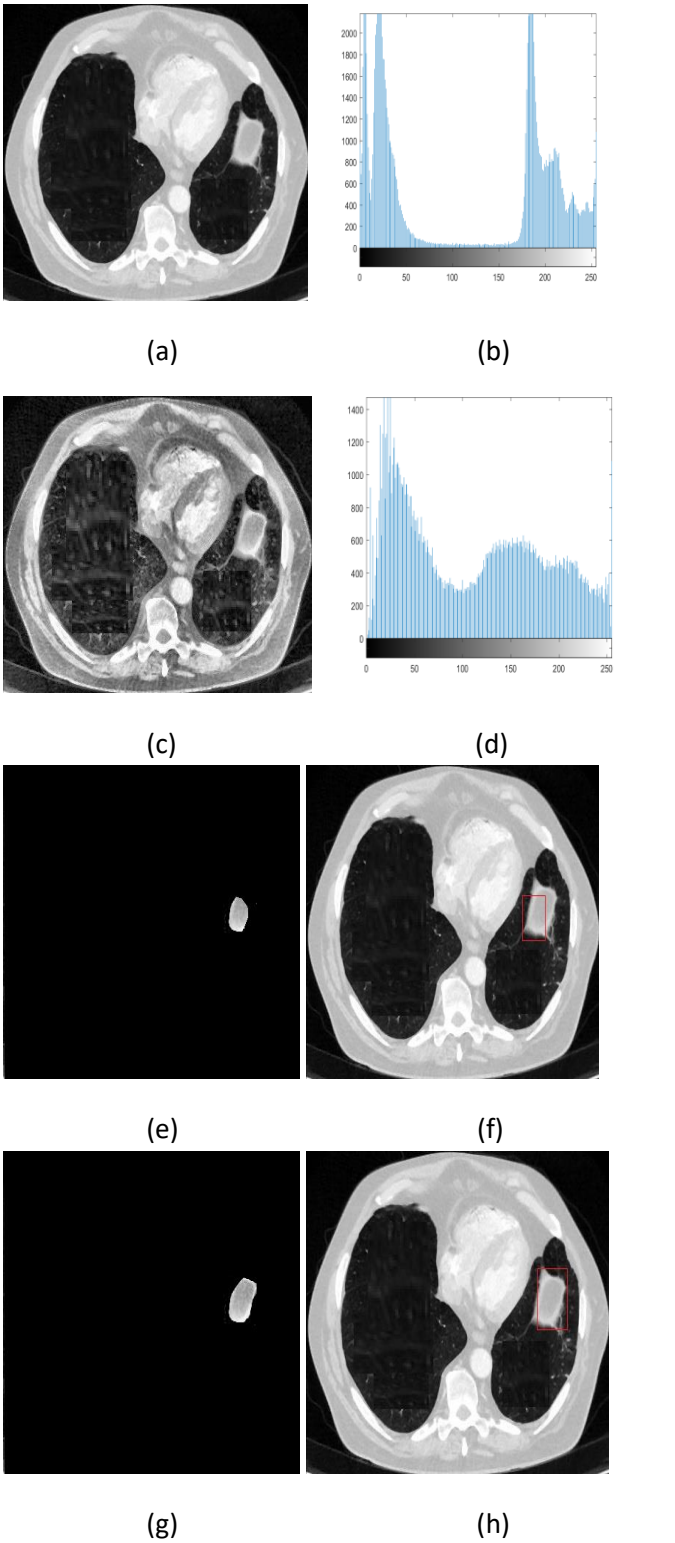

Fig-4(a-h): **Original image, histogram of original, enhanced, histogram of enhanced, segmented tumor by GA approach, segmented tumor by GA in original, segmented tumor by proposed approach and segmented tumor by proposed approach in original image 3**

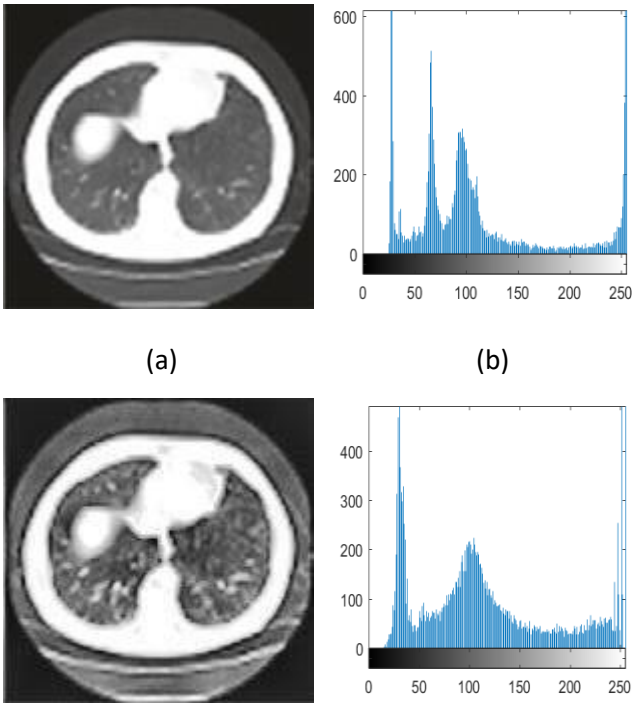

Fig-5(a-h): **Original image, histogram of original, enhanced, histogram of enhanced, segmented tumor by GA approach, segmented tumor by GA in original, segmented tumor by proposed approach and segmented tumor by proposed approach in original image 4**

Table-1: Characteristics of the labeled training set and test set

Database	Classes	Total Images	Lung Cancer Images	Rest Images
1	3	1000	500	500
2	4	2000	1200	800

Table-2: Result obtained on the test set by genetic algorithm

Database	TP	FN	TN	FP	Precision	Recall
1	720	110	10	160	0.81	0.86
2	1570	120	30	280	0.84	0.92

Table-3: Result obtained on the test set by proposed model

Database	TP	FN	TN	FP	Precision	Recall
1	950	10	30	10	0.99	0.99
2	1800	60	60	80	0.97	0.98

5. CONCLUSIONS

From the above results it is cleared that the proposed algorithm has achieved high efficiency in comparison to genetic algorithm for detection of lung cancer in the digital CT scan images. Value of objective parameters recall and precision has high values in comparison to the standard genetic algorithm. True positive rate of proposed algorithm is more than the genetic algorithm. So the proposed model for lung cancer detection has got high performance.

In the future work the performance of the proposed algorithm can be evaluated with other well known algorithms for lung cancer detection. Also author can use various other known objective parameters like eccentricity, entropy for checking the performance of the proposed algorithm.

REFERENCES

[1] N. S. Nadkarni and S. Borkar, "Detection of Lung Cancer in CT Images using Image Processing," IEEE Proceedings of the Third International Conference on Trends in Electronics and Informatics (ICOEI 2019), pp. 863-866, 2019.

[2] W. Rahane, H. Dalvi, Y. Magar, A. Kalane and S. Jondhale, "Lung Cancer Detection Using Image Processing and Machine Learning HealthCare," Proceeding of 2018 IEEE International Conference on Current Trends toward Converging Technologies, pp. 1-5, 2018.

[3] K. Gopi and J. Selvakumar, "Lung Tumor Area Recognition and Classification using EK-Mean Clusstering and SVM", IEEE International Conference on Nextgen Electronic Technologies, 2017

[4] S. M. Salaken, A. Khosravi, A. Khatami, S. Nahavandi and M.A. Hosen, "Lung Cancer Classification Using Deep Learned Features on Low Population Dataset", IEEE 30th Canadian Conference on Electrical and Computer Engineering (CCECE), 2017.

[5] A. M. Rossetto and W. Zhou, "Deep Learning for Categorization of Lung Cancer CT Images", IEEE/ACM International Conference on Connected Health: Applications, Systems and Engineering Technologies (CHASE), 2017.

[6] P. R. Katre and A. Thakare, "Detection of Lung Cancer Stages using Image Processing and Data Classification Techniques", 2nd International Conference for Convergence in Technology (I2CT) 2017.

[7] M. C. Vas and A. Dessai, "Classification of Cancerous and Non-cancerous Lung Cancer Nodule using Image Processing Technique" International Conference on Academic Research in Engineering and Management (ICAREM-17), 2017

[8] G. P. Pratap, R. P Chauhan, "Detection of Lung Cancer Cells using Image Processing Techniques", 1st IEEE International Conference on Power Electronics, Intelligent control and energy systems, IEEE, 2016.

[9] A. S. Deshpande, D. D. Lokhande, R. P. Mundhe and J. M. Ghatole, "Lung cancer detection with fusion of CT and MRI Images using Image Processing and Machine Learning", International Journal of Advanced Research in Computer Engineering & Technology (IJARCET), Vol.4, pp.763-767, 2015.

[10] K. Punithavathy, M. M. Ramya, S. Poobal, "Analysis of Statistical Texture Features for Automatic Lung Cancer Detection in PET/CT Images", International Conference on Robotics, Automation, Control and Embedded systems (RACE), IEEE, pp. 18-20, 2015.

[11] A. Kulkarni and A. Panditrao, "Classification of Lung Cancer Stages on CT Scan Images Using Image Processing", IEEE International Conference on Advanced Communication Control and Computing Technologies (lCACCCT), 2014.