Lung Cancer Detection and Multi Level Classification Using Discrete Wavelet Transform Approach

V. Veeraprathap, G. S. Harish, G. Narendra Kumar

Abstract—Uncontrolled growth of abnormal cells in the lung in the form of tumor can be either benign (non-cancerous) or malignant (cancerous). Patients with Lung Cancer (LC) have an average of five years life span expectancy provided diagnosis, detection and prediction, which reduces many treatment options to risk of invasive surgery increasing survival rate. Computed Tomography (CT), Positron Emission Tomography (PET), and Magnetic Resonance Imaging (MRI) for earlier detection of cancer are common. Gaussian filter along with median filter used for smoothing and noise removal, Histogram Equalization (HE) for image enhancement gives the best results without inviting further opinions. Lung cavities are extracted and the background portion other than two lung cavities is completely removed with right and left lungs segmented separately. Region properties measurements area, perimeter, diameter, centroid and eccentricity measured for the tumor segmented image, while texture is characterized by Gray-Level Co-occurrence Matrix (GLCM) functions, feature extraction provides Region of Interest (ROI) given as input to classifier. Two levels of classifications, K-Nearest Neighbor (KNN) is used for determining patient condition as normal or abnormal, while Artificial Neural Networks (ANN) is used for identifying the cancer stage is employed. Discrete Wavelet Transform (DWT) algorithm is used for the main feature extraction leading to best efficiency. The developed technology finds encouraging results for real time information and on line detection for future research.

Keywords— ANN, DWT, GLCM, KNN, ROI, artificial neural networks, discrete wavelet transform, gray-level co-occurrence matrix, k-nearest neighbor, region of interest.

I. INTRODUCTION

CANCER is due to abnormal cells developing into a tumor, mutation and excessive reproduction of improper cells are two important reasons [1]. LC is the leading cause of cancer death globally in high, low and middle income countries [2]. The process cancer cells spread to other parts of human body is called as metastasis; early detection of cancer plays an important role to avoid cancer cells leading to metastasis [3].

Cigarette smoking is chief cause containing copious amount of chemicals. Air pollution, excessive alcohol consumption being other factors, about 90% of cases arises due to tobacco use [4]. Small cell LC and non-small cell LC are the two types. The latter which is the most common type can lead to metastasis [5]-[7]. The techniques to diagnose are chest radiograph (x-ray), CT scan, PET scan, MRI [8]-[10]. These procedures detect disease in its advanced stages, reducing

Veeraprathap V., Harish G. S., and Narendra Kumar G. are with the Electronics and Communication Engineering Department, University Visvesvaraya College of Engineering, Bangalore University, Bangalore, India (e-mail: veeraprathap2001@gmail.com, gsharish.mce@gmail.com, gnarenk@yahoo.com).

chances of survival, and a new technique to identify in its early stages is foremost important.

The CT images assist in detecting the severity of the lung diseases since the visibility of soft tissue is better. Lung disease affects breathing, common forms are acute bronchitis, asthma, Chronic Obstructive Pulmonary Disease (COPD), Acute Respiratory Distress Syndrome (ARDS) and LC [11], [12]. Image processing techniques provide a good quality tool for improving the manual analysis, used in various research areas such as military, space, medical and many more. Radiologists classify manually with their own medical knowledge consuming more time and accuracy obtained may tend to decrease while dealing with wide number of CT images [1]-[3].

In the proposed method, an unsupervised non-invasive automatic system for lung tumor detection and classification on CT images is presented. A database containing different cases along with Diagnosis information is provided by the Early Lung Cancer Action Program (ELCAP) related to cancer patients are assessed [4]. Simulation results show the high precision and dependability of the proposed algorithm. The results are highly helpful for online doctors and radiologists to easily estimate the size and position of a tumor.

The paper is organized as follows. Section II presents a brief literature review and their implications with proposed system in Section III, while Section IV presents System Implementation including flow chart and block diagram. Section V shows the advantages of the proposed system. Section VI presents the test results obtained from the simulations performed concluding in Section VII.

II. LITERATURE REVIEW

Chunran et al. [1] proposed a fully convolutional network (FCN) based on level set method. FCN used for segmentation nodules detected inside lung area by threshold method and image processing techniques, while segmentation by threshold method is based on the coordinate system transformation.

Shaffie et al. [2] proposed a system fused texture and shape features to get an accurate diagnosis for the extracted lung nodules. 3D Local Binary Pattern (LBP) and higher-order Markov Gibbs random field (MGRF) models were utilized. Modelled features are used to differentiate between the malignant and benign nodules.

Zhao et al. [3] proposed a patch-based 3D U-Net and contextual Convolutional Neural Networks (CNN) to automatically segment and classify lung nodule. 3D U-Net employed for segmentation, while Generative Adversarial Network (GAN) is used to enhance model performance, online

sampling strategy determined malignant nodule.

Abdillah et al. [4] proposed a detection method based on image segmentation using marker control watershed and region growing approach. Enhancement using Gabor filter, feature extraction was performed providing good results with high accuracy and robustness.

Cristian et al. [5] proposed a system using transthoracic ultrasonography (TUS) and US-guided biopsy technique.

Chapaliuk et al. [6] proposed an automated diagnose system using 3D convolution and neural networks. Accuracy of the networks was evaluated.

Bhattacharya et al. [7] proposed a MEMS based sensor technique. Based on the volatile organic compound concentration exhaled from breath of a person, system identified the cancer patient.

Katre et al. [8] proposed a system with Median filter for noise removal, High boost operator for enhancement, and marker controlled watershed for segmentation. Various classification techniques for detecting LC based on suspicious ROI obtained by feature extraction was discussed.

Prathamesh et al. [9] proposed a system using median filter, Watershed Segmentation and morphological operations.

Wong et al. [10] proposed a LC identification method by exhaled breath using KNN and SVM classifiers with results at accuracy of 84.4%.

III. PROPOSED WORK

The proposed Lung Cancer Detection System (LCDS) examines CT scan images and identifies tumour affected lungs further subdividing into stages as shown in Fig. 1.
1) Image pre-processing
2) Image segmentation
3) Feature extraction

The three phases of image processing listed employed with best techniques along with feed forward neural networks and KNN algorithm creating an efficient and accurate system than existing methods. More significance is given to the feature extraction of tumour region properties given as input to classifiers. During diagnosis suspected patient's CT image given as input to system checks the patient condition using trained classifiers. Patient lungs found abnormal further examined to identify the cancer stage and appropriate treatment recommended. Tumour classification performed as per criteria's listed in 8th edition of Tumour Node Metastasis (TNM) staging of LC. 360 CT scan images of Abnormal and normal lungs are collected from ELCAP image database. System is developed using MATLAB.

IV. SYSTEM IMPLEMENTATION

Medical images have factors such as noise, uncertain tumor boundaries and more dissimilarity in tumor appearance which makes difficult to find exact tumor region [1]-[4]. To overcome problems, methods which are helpful for radiologists in accurately locating tumor are essential [2]-[10]. The doctor suggests CT scan based on the positive results obtained in pre-operative evolution, prostate malignancy and other clinical tests performed in laboratory as shown in Fig. 2.
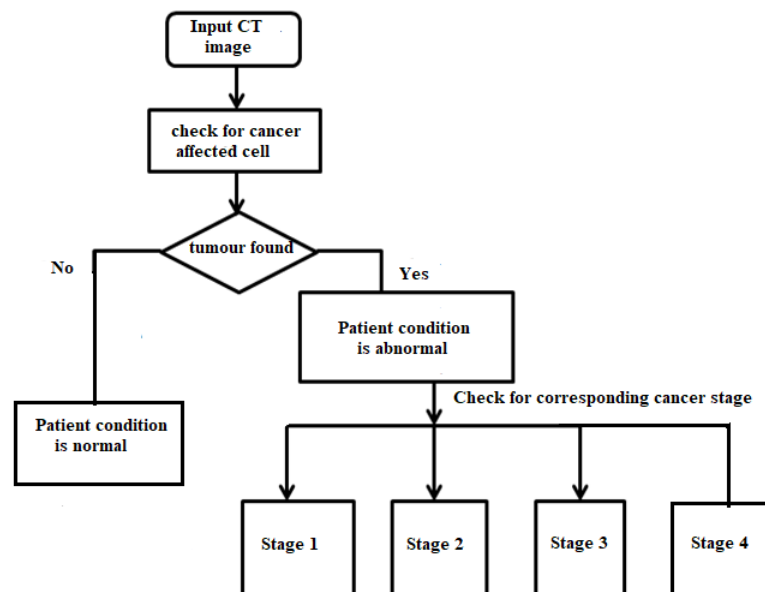

Fig. 1 Flowchart of the proposed system

An image read initially by system refers to image acquisition. During process, very crucial to eliminate the effect of noise content and improve image quality before an examination. Noise removal and contrast enhancement are primary steps of significance.

Pre-processing performed in stages as follows:
1) Image resizing
2) Image conversion to gray scale image

3) Gaussian filter applied
4) Histogram equalization
5) Media filter applied

 Image pre-processing is improvement of the image data suppressing unwanted distortions. Image scaling, color space transformation, and contrast enhancement is employed in image enhancement. Masking approach used to identify

brighter pixels and darker pixels to separate them, value greater than threshold value implies '1' otherwise '0' [3]. Basic morphological operations such as dilation, erosion, opening, closing and bwareaopen are employed [4]. Dilation adds pixels to the boundaries of objects in an image, whereas erosion removes pixels on object boundaries.
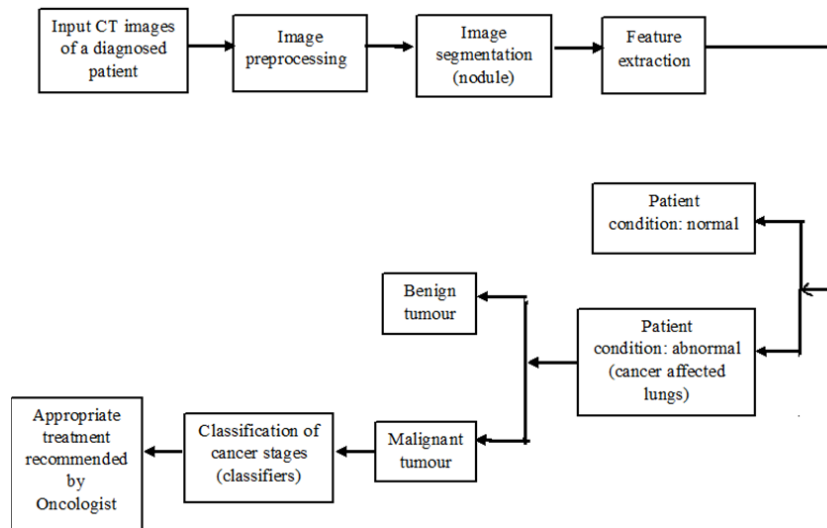
Fig. 2 The proposed system block diagram

A. LCDS Stages

1) Acquire CT scan image of suspected person.
2) Pre-processing of image.
3) Segmentation using ROI and Otsu's threshoding approach.
4) Feature extraction using DWT algorithm.
5) KNN and Neural network classification.
6) Diagnosed result.
7) Treatment method.

B. Proposed System Techniques

1) HE
2) ROI extraction
3) Morphological Operations
4) Region properties measurement
5) GLCM
6) ANN
7) KNN

 Based on size and shape of structuring element involved, pixels are manipulated from the objects. In medical images, segmentation accuracy is utmost important accounting impact on human lives. Region growing method and Otsu's thresholding approach are used for segmentation. The former technique performs pixel-based image segmentation involves the selection of initial seed points by examining neighboring pixels of initial seed points and determines to add pixel neighbors to the region [4]. The process is repeatedly performed as shown in Fig. 3.

 Feature extraction performed on data set of normal, tumor

affected lung CT images. Region properties Area, perimeter, centroid, diameter and eccentricity of Region of Interest (ROI) are included in feature extraction all being scalar quality. The extracted features shape and size given as input to neural networks. GLCM is texture analysis technique in medical imaging categorizes different combinations of pixel intensity value occurred in an image [4]. Homogeneity, contrast, correlation, energy, entropy features extracted are stored in matrix used in training of ANN.

 DWT algorithm applied on image, single level wavelet decomposition is computed. Advantage in training the large data as size of features extracted and stored in the form of 2-D matrix is reduced. Eccentricity value indicates deviation of tumor from being circular in shape, differentiates benign and malignant tumor [5]. Centroid gives the exact location of tumor in the form of x and y co-ordinates, while diameter helps in analysis of rough size of tumor. Perimeter indicates the number of pixels existing on ROI boundary, whereas area gives the overall number of pixels element acquired by the extracted ROI. Conversion relation 37.79 pixels equivalent to 1 cm is used, area value pixels estimated after normalization, size of a tumor is assessed accordingly as shown in Fig. 4.

 KNN, used to get level data, classifies images based on the similarity with the stored data, ANN trained by different samples produces better accuracy results even for non-trained images compared to other classifiers [6]-[10]. K= 1, 2 and 3 is used to vote for a class of testing data. At the first level of classification KNN checks for normal lungs, abnormal lungs and tumors very large in size which cannot be segmented
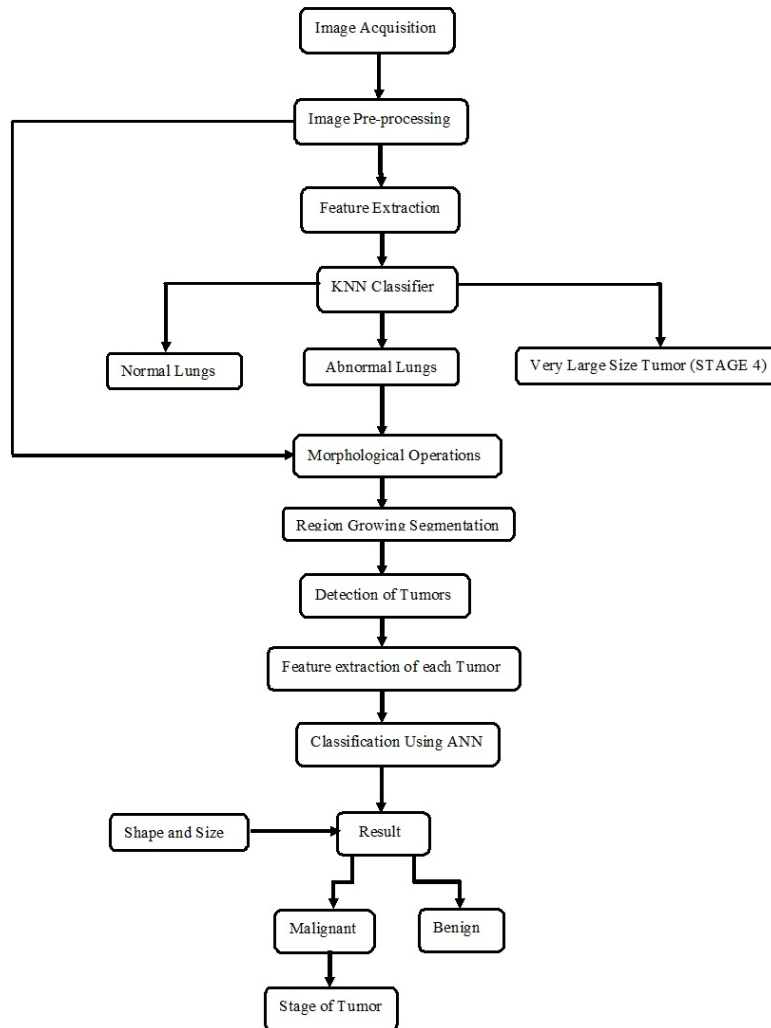
properly.

Fig. 3 Work-flow diagram of the proposed system

Stages of Lung Cancer	Tumor area assessed in pixels
Stage 1	0-114
Stage 2	115-190
Stage 3	191-266
Stage 4	>267

Fig. 4 Classification as per TNM staging

Second level of classification involves feed forward neural network type of ANN identifying tumor detected as Malignant or Benign, further evaluating the stages based on features extracted, i.e. tumor size (area) and node location (centroid).

V. ADVANTAGES OF THE PROPOSED SYSTEM

1) System assists medical consultants, cost effective, least invasive and time saving [11].
2) Early detection and prediction of LC play vital role in the diagnosis process and increase survival rate [12].

3) Identification along with accurate classification of
4) LC stages considering several cases are proposed.

VI. RESULT EVALUATION
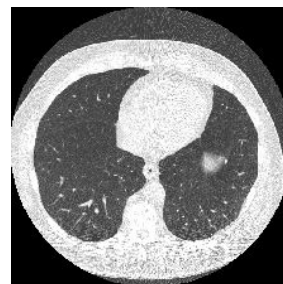
Fig. 5 Input CT image; CT scan image of suspected patient lungs is given as input to LCDS by consultant to examine the condition, and he/she give his/her opinion

Fig. 6 Resized image; First step is resizing, for further effective processing without reducing quality of an original image
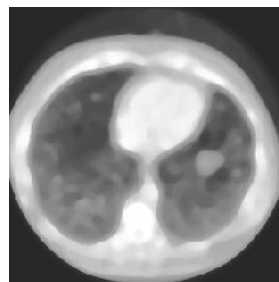

Fig. 7 Gray scale conversion; Gray scale conversion is performed to the original CT image to convert the color (24 bit) into a grayscale (8 bit), if any
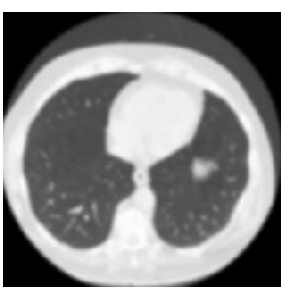

Fig. 8 Gaussian filtered image; Gaussian filter is a linear filter used in blurring an image, performs smoothing and reduces noise
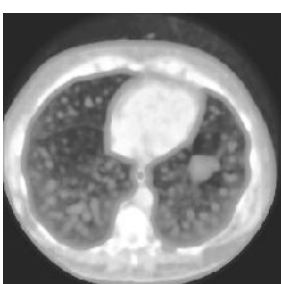

Fig. 9 Histogram equalized image; HE technique enhances the contrast of image, while Uniform distribution of intensities exist in the image

Fig. 10 Median filtered image; Median is a nonlinear filter removes salt and pepper noise, reduces impulsive noise and performs edge detection

Fig. 11 Region growing segmented image; Segmented suspected lungs boundary is seen. Pixels representing information in image are extracted and grouped together

Fig. 12 Binarization achieved; Binarization technique is achieved which helps further in thresholding operation

Fig. 13 Left lung; Left lung is segmented separately using row and column pixel values to obtain better accuracy in detection

Fig. 14 Right lung; Right lung is segmented separately using row and column pixel values
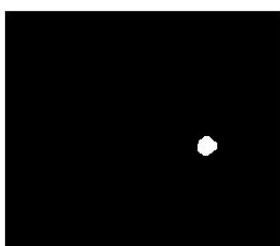

Fig. 15 Tumour part; Tumor part is extracted from right lung after segmentation
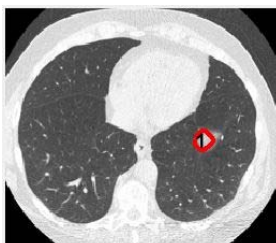

Fig. 16 Abnormal lungs; Tumor is detected in right lung finally, after 10 steps mentioned above and declared as abnormal
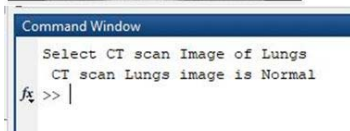

Fig. 17 Normal lungs; CT image of normal patient lung is given as input to LCDS, examined and identified as normal. Output is seen in command window. Different cases as such are verified
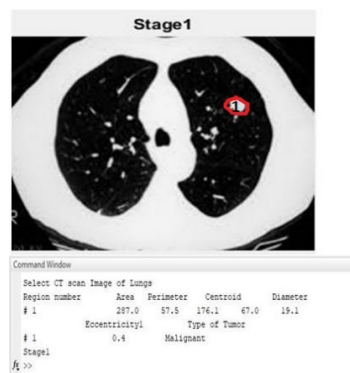

Fig. 18 Stage 1 cancer; Nodule area is assessed, declared as stage 1 and malignant type of tumor, computations seen in command window
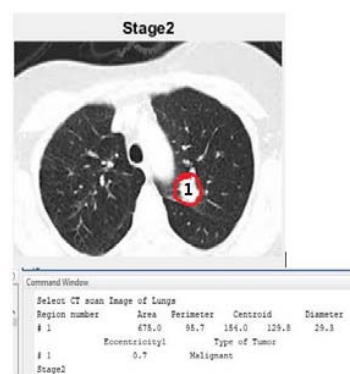

Fig. 19 Stage 2 cancer; CT image of a different patient is being examined and evaluated as stage 2, as seen in command window
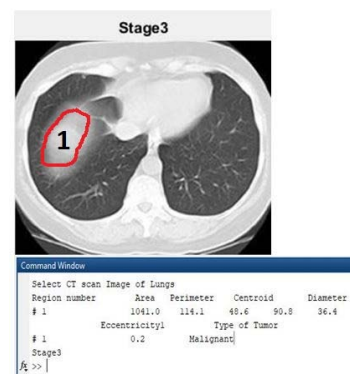

Fig. 20 Stage 3 cancer

Based on ROI, area size in pixels is evaluated and declared as stage 3. Output is seen in command window.
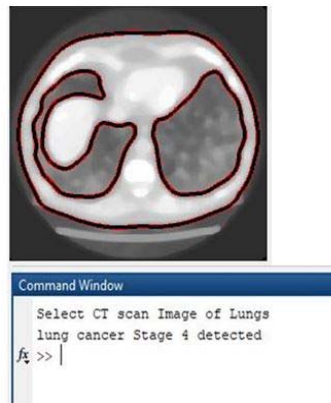
Fig. 21 Very large tumour found; KNN identifies very large sized nodule attached to left lung boundary, segmentation is not possible for this kind of case henceforth declared as stage 4 cancer, as seen in command window
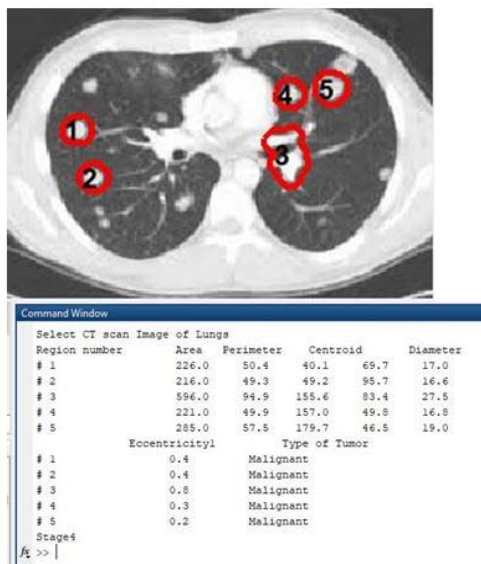

Fig. 22 Multiple tumours located; CT image of a case consists of multiple tumors identified in both the lungs, each tumor is being assessed and declared as stage 4 cancer as per TNM staging seen in command window

VII. CONCLUSION

System successfully detects and evaluates the LC stage from CT scan images. The results show that best approach for main features extraction is DWT algorithm, dataset is trained using ANN and KNN along with binarization and GLCM techniques lead to high efficiency of the system at 98.67%. The consultants suggest appropriate treatment based on the severity of the LC stage. Work can be extended to locate tumours in other parts of body.

REFERENCES

[1] Yang Chunran, Wang Yuanyuan and Guo Yi, "*Automatic Detection and Segmentation of Lung Nodule on CT Images*"- 11th International Congress on Image and Signal Processing, BioMedical Engineering and Informatics (CISP-BMEI 2018) –2018, IEEE.

[2] Ahmed Shaffie, Ahmed Soliman, Hadil Abu Khalifeh, Mohammed Ghazal, Fatma Taher, Robert Keynton, Adel Elmaghraby and Ayman El-baz, "*On The Integration of CT-Derived Features for Accurate Detection of Lung Cancer*" – 978-1-5386-7568-7/18/$31.00 – 2018, IEEE.

[3] Chen Zhao, Jungang Han, Yang Jia and Fan Gou, "*Lung Nodule Detection via 3D U-Net and Contextual Convolutional Neural Network*"–2018 International Conference on Networking and Network Applications, IEEE.

[4] Bariqi Abdillah, Alhadi Bustamam, and Devvi Sarwinda, "*Image processing based detection of lung cancer on CT scan images*" – The Asian Mathematical Conference 2016 (AMC 2016) - IOP Conf. Series: Journal of Physics: Conf. Series 893 (2017) 012063.

[5] Achim Cristian, Rusu-Both Roxana, Dulf Eva-Henrietta and Chira Romeo Ioan, "*Lung Cancer Diagnosis based on Ultrasound image processing*" – 22nd International Conference on System Theory, Control and Computing (ICSTCC)- IEEE, 2018.

[6] Bohdan Chapaliuk and Yuriy Zaychenko, "*Deep Learning Approach in Computer-Aided Detection System for Lung Cancer*" – 978-1-5386-7196-2/18/$31.00, 2018, IEEE.

[7] Diptarup Bhattacharya, Pradeep Kumar Rathore and Shubhankar Majumdar, "*Non-Invasive Piezoelectric Sensor for Detection of Lung Cancer*" - International Conference on Communication and Electronics Systems (ICCES 2018) - IEEE 2018.

[8] Pooja R. Katre and Dr. Anuradha Thakare, "*Detection of Lung Cancer Stages using Image Processing and Data Classification Techniques*", 2nd International Conference for Convergence in Technology (I2CT) – IEEE 2017.

[9] Gawade Prathamesh Pratap and R.P. Chauhan, "*Detection of Lung Cancer Cells using Image Processing Techniques*"- 1st IEEE International Conference on Power Electronics. Intelligent Control and Energy Systems (ICPEICES-2016) - IEEE 2016.

[10] De-Ming Wong, Chen-Yu Fang, Li-Ying Chen, Chen-I Chiu, Ting-I Chou, Cheng-Chun Wu, Shih-Wen Chiu and Kea-Tiong Tang, "*Development of a Breath Detection Method Based E-nose System for Lung Cancer Identification*" – IEEE International Conference on Applied System Innovation 2018 – IEEE ICASI 2018.

[11] Kelly M. Latimer, Md, Mph and Timothy F. Mott, Md, Naval Hospital Pensacola, Pensacola, Florida -"Lung *Cancer: Diagnosis, Treatment Principles, and Screening*"- Medical Journal–American family Physician website - www.aafp.org/afp -2015.

[12] Ajay Aggarwal, Grant Lewison, Saliha Idir, Matthew Peters, Carolyn Aldige, Win Boerckel, Peter Boyle, Edward L. Trimble, Philip Roe,Tariq Sethi, Jesme Fox, Richard Sullivan, - "*The State of Lung Cancer Research: A Global Analysis*"- 2016, International Association for the Study of Lung Cancer - Journal of Thoracic Oncology Vol. 11 No. 7: 1040-1050 – 2016, http://creativecommons.org/licenses/by-nc-nd/4.0/

Veeraprathap V., Research Scholar, Electronics and communication Engineering Department, University Visvesvaraya College of Engineering, Bangalore University, Bangalore, India (e-mail: veeraprathap2001@gmail.com).

Harish G. S., Graduate scholar, Electronics and communication Engineering Department, University Visvesvaraya College of Engineering, Bangalore University, Bangalore, India (phone: 9844310083; e-mail: gsharish. mce@gmail.com).

Dr. Narendra Kumar G, born in Bangalore on 5th February, 1959. Obtained Master's Degree in Electrical Communication Engineering, (Computer Science & Communication) from Indian Institute of Science, Bangalore, Karnataka, India in 1987. Was awarded PhD in Electrical Engineering (Computer Network) from Bangalore University, Karnataka, India in 2006. Currently Professor in the Department of Electronics & Communication Engineering, University Visvesvaraya College of Engineering, Bangalore University, Bangalore, Karnataka, India. He has published more than 70 papers in referred journals and conferences. Research interests include Mobile Communication, Image Processing, Wireless Communication, M-Commerce, Robotics, WSNs and Computer Networks. (Phone: 9886076757; e-mail: gnarenk@yahoo.com).