Detection and Characterization of Lung Tumor by Using Convolution Neural Network

M Sangeetha¹, S Mythili²

¹Assistant Professor, Department of Electronics and Communication Engineering, IFET College of Engineering, Villupuram.
²UG Scholar, Department of Electronics and Communication Engineering, IFET College of Engineering, Villupuram.

geethu.sparkles@gmail.com¹, mythili.sivaperumal@gmail.com²

Abstract. In today’s modern world, lung cancer is the most common cause of death amongst humans. Most of the lung cancers are curable at early stages. So an early and fast detection of lung cancer can save the patient’s life. With the new technology, early detection of lung cancer is possible at the initial stage. Formal method for diagnosis lung cancer detection is biopsy method. We have proposed lung cancer detection system using convolution neural network for early detection of lung cancer disease. It is more advantageous to patients. The diagnosing methodology uses image processing methods algorithm. The DICOM (Digital imaging and communications in medicine) image of lungs is taken as input and it goes under various pre-processing technique for noise removal and image segmentation. These features are given as the input to classifier. Convolution neural network is used for classification purpose. It classifies the given input images into cancerous or non-cancerous.

1. Introduction
Cancer is the most vicious disease, the cure of which must be the prime target through scientific investigation. Cancer causes cells to divide uncontrollably. This can results in tumors, damage to the immune system, and other impairment that can be fatal.

Cancer is the uncontrolled growth of abnormal cells anywhere in the body. These abnormal cells are termed as cancer cells, malignant cells, or tumor cells. Many cancers and the abnormal cells that compose the cancer tissue are further identified by the name of the tissue that the abnormal cells originated from (for example breast cancer, lung cancer, and colorectal cancer). Lung cancer is the most common cancer in both men and women. According to the survey of national cancer institute (2016), estimated new cases and estimated deaths are 2, 24, 390 and 1, 58, 080 due to lung cancer.

In the United States, an estimated 15.5 million people with a history of cancer were living as of January 1, 2016, according to a 2018 report from the American cancer society [1]. The tumor is the important causes for cancer. Usually the lung tumor is classified as benign lung tumor and malignant lung tumor. The benign lung tumors are non-cancerous one and it does not spread to neighbor tissue or other parts of the body. The growth rate of benign tumor is very slow when compare to the malignant tumor. The benign tumor is further classified as hamartomas and papillomas [2]. Malignant tumor often causes cancer in human body. It is cancerous. It may leads to death in human race.

In this paper, the early detection of different stages of cancer such as cancerous (benign) and non-cancerous (malignant) are done by using different advanced technologies of digital image processing. The early detection of cancer will reduce the death rate in the world due to lung cancer by providing appropriate treatment for different stages of tumor.

The early detection is based on different stages of medical image processing such as preprocessing and feature engineering steps (together with bring out extraction and selection) and succeeded by classification steps [3], [4], [5], [6].
The process flow is categorized as two stages such as training stage and testing stage. The training stage is used to train the image for future reference and the trained data is stored in database. Section 3 provides a brief explanation about the above process.

This paper is systemized as follows: Section 2 describes related works pertaining to neural networks for early diagnosis of lung tumor and different stages of lung tumor. The existing system and its limitations are discussed in the section 3. We present the proposed methodology is based on GLCM feature extraction and CNN classification approach in section 4. In section 5, the experiments and results are discussed. In the same section, we can also study the various steps of image processing thoroughly. At last, section 6 provides conclusions about our works.

2. Related Works
This section summarizes the advances in machine learning applied to the medical imaging and different segmentation and methodology used for lung cancer diagnosis.

Automatic detection and classification of different stages of tumor in the lung is important for early detection of lung cancer. The automatic and early detection is possible by various techniques of image processing. In recent years, the various techniques and methodology for medical image processing has been proposed and analyzed by a no of scholars. In 2017, the authors in [7] conducted a study on early detection of lung cancer by analyzing the exhaled breath of the patients. They considered as potential use of exhaled breath analysis is easy, non-invasive method for early detection of lung cancer. They used G C- MS method (Gas chromatography, mass spectrometry) in that, the exhaled breath is collected and treated with helium stream.

In 2013, the authors in [8] proposed a complete study on GLCM (Gray level co-occurrence matrix) approach for image texture feature extraction. In image processing, feature extractions are more important step. The features may be color, texture and shape of the objects. The author aims to extract the second order statistical texture features for motion images by using GLCM. They mainly focused on angular second moment, correlation, inverse difference moment and entropy. The main proposal is GLCM approach greatly reduces a compression time. In literature [9], a method is based on CAD system. The methodology is classified as supervised learning and unsupervised learning. The final result is based on comparative analysis of the above methods. The author aims to propose the deep learning specifically CNN and clustering the pixels and SVM for classification purpose. The author present a advanced new 3D CNN architecture based on the graph regularized sparse multi-task learning and characterization of lung nodule from CT scan for clustering purpose proportional SVM ($\propto$SVM) algorithm and used for categorization. They proposed lung nodule classification with accuracy of 77.97% for the feature of VGG-fc 7. They also proposed largest evaluation of CAD system for IPMN classification. Literature [10] uses a method based on watershed technique for segmentation purpose. They also propose watershed segmentation is more accurate than other approach and it gives 84.55% of accuracy. They detect the lung cancer on CT scan based on the extracted features such as area, perimeter, eccentricity and roundness. The roundness is a term used to predict the shape of tumor in lungs.

3. Proposed System
Lung cancer is the number one cause of deaths in both men and women in the universe. The 5-year relative survival rate for lung cancer is 15% for men and 21% for women. However, the survival rate is 55% for early detection of lung cancer. Early detection of lung cancer is challenging, due to the various shapes, size, density, and location of nodules. Studies have been adopted to tackle the problems.

In the proposed model, early detection is possible by GLCM (gray level co-occurrence matrix) and further classification by CNN (Convolution neural network).
The above figure shows the block diagram of the proposed methodology. In proposed method, the early detection is achievable by CNN and GLCM. After pre-processing, the image is used for segmentation. In segmentation, the image data in the form of pixel is segmented and grouped as data which having same characteristics.

Feature extraction is the process of transforming the raw pixel values from an image, to a more meaningful and useful information that can be used in other techniques, such as point matching or machine learning. Mainly GLCM used to measure the second order statistical texture such as angular second moment, correlation, inverse difference moment and entropy.

The GLCM also include contrast, correlation, homogeneity and energy.
The flowchart is divided into two stages, which are training and testing stage. The training stage is a process of training the framework to classify the image sample based on the training data. The testing stage is used to classify the test sample by applying the trained framework. In the training stage, the important features are extracted from the training sample images. The extracted features are used to grow a neural architecture. At the end of training stage, the architecture is built and the performance of the trained architecture is evaluated.

ADVANTAGES:
- Accuracy is improved and also used for large data set
- It is very easy for implementation

4. Result and Discussion
In this area underneath the consequences of our examination are given.

4.1 To load an input image in MATLAB
We pick the images from database. The input image is in the form of DICOM format.
4.2 Pre-processing
The aim of pre-processing is an improvement of image data that having unwanted distortion or enhances some important feature for further steps.

A. BW Thresholding
The process of converting a raw image into binary image, and it is based on threshold value. The main aim of BW thresholding is to convert a raw image into understandable image by computer as 0’s and 1’s. In the image, high luminance area is replaced as 1 (white) and others as 0 (black).

![BW Thresholding image](image1.png)

**Fig.4 BW Thresholding image**

4.3 Segmentation
Image segmentation is the process of partition a digital image into multiple segments (sets of pixels, also known as image objects). The medical image segmentation is the process of automatic or semiautomatic detection of boundaries within a 2D or 3D image.

![Segmented image](image2.png)

**Fig. 5 Segmented Image**

4.4 Output
After further steps, the output will be displayed and it is use full for early detection of cancer in lungs based on different stages of tumor. The output is finally displaced as normal, benign, and malignant based on size and character of the tumor in the lungs.
5. Conclusion
An image improvement technique is developing for earlier disease detection and treatment stages; the time factor was taken in account to discover the abnormality issues in target images. Pre-processing techniques based on median filter. The proposed technique is efficient for segmentation principles to be a region of interest foundation for feature extraction obtaining using GLCM. The proposed technique gives very promising results comparing with other used techniques. The CNN classifies detect the cancer accurately.

References

[1] Society, A.C.: Cancer Facts & Figures. American Cancer Society (2016).
[2] Basturk, O., et al.: Tumor-associated neutrophils and malignant progression in intraductal papillary mucinous neoplasms: an opportunity for identification of high risk disease. Annals of surgery (2015).
[3] El-Baz, A., El-Ghar, M.A., Elnakib, Falk, R., A., Gimelfarb, G., Khalifa, F., Nitzken, M.: 3D shape analysis for early diagnosis of malignant lung nodules. In: IPMI. pp. 772–783. Springer (2011).
[4] Han, F., Han, H., et al.: Texture feature analysis for computer aided diagnosis on pulmonary Nodules. (2015).
[5] Way, T.W., et al.: cad diagnosis of pulmonary Nodule on CT scans: segmentation and classification using 3D contours. (2006).
[6] Lee, M., Powell, c., et al.: computer aided diagnosis of pulmonary nodules using a two steps approach for feature selection and classifier enabler. (2010).
[7] Inbar Nardi-Agmon, Nir peled: Exhaled Breath analysis for the early detection of lung Cancer: recent developments and future Prospects. (2017).
[8] Gurukumar, L., mohanaiah, P., Sathyanarayana., P.: Image texture feature extraction by using GLCM (2013).
[9] Candice W. Bolan, Pujan kandel, et al.: lung and pancreatic tumor characterization in the deep Learning era: novel supervised and unsupervised learning approaches. IEEE (2018).
[10] Aniket Gaikwad, Azharuddin Inamdar, and Vikas Behera: Lung cancer detection using digital Image processing On CT scan Images. IRJET (2016).
[11] Fei-Fei, L., Karpathy, T., Leung, A., Shetty, S., Sukthankar, R., Toderici, G.: Large-scale video classification with convolutional neural networks. In: IEEE CVPR. pp. 1725–1732 (2014).