

Lung Cancer Diagnosis from Ct Images Based on Pre-Processing and Ensemble Learning



Denny Dominic, K. Balachandran

Abstract: Lung cancer is one of the dreaded diseases. Early detection is highly recommended to save the lives of the people. If the early detection is focused, we can reduce the mortality rate and increase the life expectancy accuracy 13% of the all new cancer diagnosis and 24% of all cancer deaths. There are various methods to detect the lung cancer from x-rays and CT images but the CT images are preferred. The medical images are always preferred to get better results on the same disease. The proposed method here will discuss about how the Watershed segmentation can be applied to the pre-processing of CT scans. The results are used for the deep learning methods so as to get the more accuracy from more image scans. The results are then used for the verification by the medical examiner for the validity of the results. The major pre-processing is done by using Median Filter, Gabor Filter and Watershed segmentation. In this research work we will discuss how the image manipulation can be done to achieve better results from the CT images through various image processing methods. The construction of the proposed method will include smoothing of the images with median filters, enhancement of the image and finally segmentation of the images.

Keywords: Metastasis, IMS, Watershed Segmentation, ROI, Threshold, CT morphologic

I. INTRODUCTION

The early prediction of Lung cancer disease seems to be difficult due to poor prognostic methods that are available now. There is better survival from cancer if it is diagnosed early and proper treatment is given [1]. Cancer cells are spread through the blood vessels, or lymph fluid that surrounds lung tissue. Lymph flows through lymphatic vessels, which drain into lymph nodes located in the lungs and in the Centre of the chest. The Cancer in the lungs is spread towards the chest as it is carried by the blood vessels. Non-small cell lung cancer and small cell lung cancer are the major two categories of lung cancer. These cancer types are depending on their cellular characteristics. There are four stages of lung cancer. The stages vary according to the size and shape of the cancerous cells. The CT scan is better than X-ray for the diagnosis of the lung cancer. The main cause of lung cancer is said to be smoking. Hence, a frame work has to be implemented to build an intelligent computer aided diagnosis system. This system has to be fast and accurate to develop grading of CT images thus to become a guiding feature to predict the lung cancer in the early stage.

The challenge here is that the images that are collected from various sources may vary in the size and format. Therefore, it is necessary to pre-process all of them before processing to get desired result. The purpose of this paper is to implement various pre-processing methods such as Median filter, Gabor filter and water segmentation. The medical images in normal case have more inner structures with different pixel density and textures. The traditional feature would not be sufficient enough to characterize all the classes efficiently[1].The pre-processing has been done for the purpose of getting the desired values from the images in the form of the binary numbers, metrics values, pre-processed images and charts. The chest computed tomography (CT) scan, has high resolution and is widely accepted for the diagnosis of the lung tumours. Small cells are frequently malignant and difficult to find them out in the early stages. The accuracy in the pre-processing is necessary to develop more informative tool for clinicians to differentiate the nature of pulmonary nodules noted on CT scan remains an urgent task. The major areas where the pulmonary nodules appear are lung, chest wall, airway, pulmonary fissure, vessel and this becomes complicated for the technician to find out the exact location of the malignancy. The pulmonary nodules can be diagnosed based on the shape that is, its sphericity, speculation and composition of interior structure. Thus, it is highly important to have a proper and accurate pre-processing of the image collected from various sources.

The main purpose of this study is to get insight about lung cancer cells and image processing on the CT images so as to predict the cancer cells early.

II. LITERATURE SURVEY

Noise filtering is the first step in detecting the lung cancer. The major two steps in pre-processing are Denoising and Weiner filtering. The white noising is one of the problems in image processing [3].Ilya Levner[2] introduced watershed segmentation for topographical function and object marker. The segmentation and pixel grouping are two operations in computer vision. Pixel grouping is necessary to cluster the classified pixels into objects, when objects of the same predefined class are in close proximity to one another. The unsupervised setting can use watershed algorithm for a better result into a set of non-overlapping regions. The description of quantification with respect to semantic nodule characteristics of sphericity, speculation and clarification have become major concern [7,8,9]. Therefore, it has become an urgent need that the image segmentation needs to acquire quantitative features. A pulmonary nodule could occur anywhere within the lung anywhere.

Manuscript published on November 30, 2019.

* Correspondence Author

Denny Dominic*, Department of Computer Science and Engineering, Christ (Deemed to be University), Bengaluru, India

K. Balachandran, H.O.D. Department of Computer Science and Engineering, Christ (Deemed to be University), Bengaluru, India

© The Authors. Published by Blue Eyes Intelligence Engineering and Sciences Publication (BEIESP). This is an [open access](https://creativecommons.org/licenses/by-nc-nd/4.0/) article under the CC-BY-NC-ND license <http://creativecommons.org/licenses/by-nc-nd/4.0/>

Lung Cancer Diagnosis from Ct Images Based on Pre-Processing and Ensemble Learning

In the Feature extraction a final result is explained as if there is any normality or abnormality in the image processed [4] and this extracted image is the basis for image classification. Ginneken[4] has divided the classification of lung images extract into either rule-based or pixel classification based category. Rule-based approach is preferred by the most of the category [5-6], with different steps, rule is introduced for the extraction of a particular result.

Thresholding, region growing, edge detection, and ridge detection, morphological operations, fitting of geometrical models or functions and dynamic programming are the major operations in image manipulation on CT images.

III. METHODOLOGY

The image manipulation for pre-processing and feature extraction is carried out with different steps involved in it. The pre-processing is done with smoothening of images, enhancement of images and finally segmentation as well. The feature extraction is carried out using morphological operation and colorimetric operations.

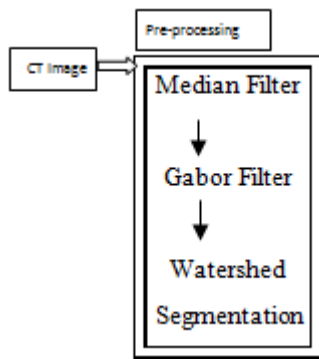


Fig. 3.1 Image Manipulation System

The image manipulation system (IMS) has used Gaussian pulse with convolution filters to smooth the given input CT images, hence, the enhancement of colour and contrast are made possible. Thresholding is the method used for nucleuses in the images to segment. Once these steps are achieved the IMS utilizes colorimetric and morphological to extract feature from image of the nucleuses [3]. The average intensity, area, perimeter and eccentricity of the nucleuses include from the extracted morphologic features. In the image processing the image acquisition is the step and we have selected CT images because it has less noise. CT images are with more accuracy and less distortion. The images are collected from various hospitals. The standard of CT image is DICOM (Digital Imaging and Communications in Medicine) has become a standard for medical Imaging. The image that are collected are raw data, hence, to improve the contrast, clarity and separate the background noise it is necessary to have a pre-processing to done on the images. Therefore, smoothing and enhancement are to be used to get an appropriate image required.

3.2 Image Pre-processing

3.2.1 Smoothing

Smoothing can suppress all kinds of fluctuations in the input image and it is almost similar to the suppression the frequency in the frequency domain. It is also used to blur the sharpened edges to smoothen the images. Median filter is used to remove the noise from the images and this is a

non-linear operation to reduce the salt and pepper appearance in the images. In general, a great deal of high spatial frequency detail is being allowed by the median filter and it is very effective at removing noise on images. But less than half of the pixels in a smoothing neighbourhood have been affected by the smoothing process. $\text{storage}=\text{medfilt2}(I,[a,b])$ is the equation of the median filter to perform the matrix operation. Each output pixel contains the median value in the $x \times b$ neighbourhood around the corresponding pixel in the image. Medfilt2 pads the image with 0's on the edges, so the median values for points within one-half the width of the neighbourhood $([a,b]/2)$ of the edges might appear distorted as shown in fig.3.2.1

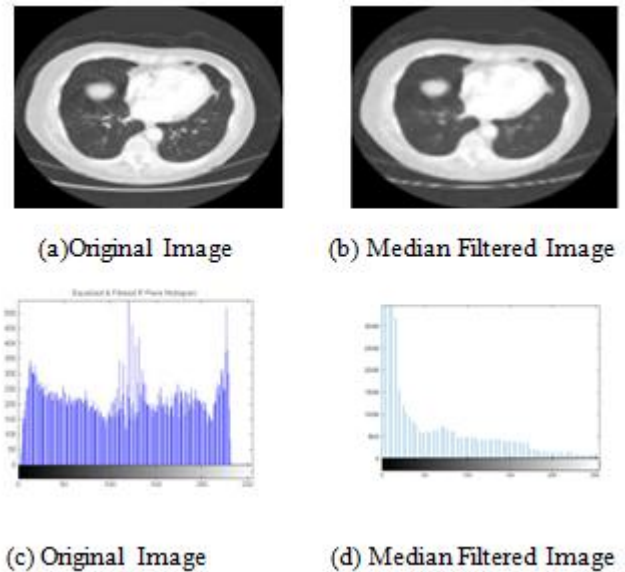


Fig.3.2.1

3.2.2 Enhancement

It a method used for the human viewer to have better interpretability and wider perception about the image, thus to facilitate better input for other automated image processing techniques. Spatial domain and frequency domain are the two main image enhancement classification categories. In this research we have used Gabor filter for enhancement purpose. It has outperformed compared to Fast Fourier and auto enhancement[2]. The mathematical equation of a 2D Gabor kernel can be explained as:

$$G(x, y) = \exp\left(-\frac{x'^2 + \gamma^2 y'^2}{2\sigma^2}\right) \cos\left(2\pi \frac{x'}{\lambda}\right)$$

Where, $x' = x \cos\theta + y \sin\theta$; $y' = -x \sin\theta + y \cos\theta$ the parameters involved in the construction of a 2D Gabor filter are: 1. The variance σ of the Gaussian function 2. The wavelength λ of the sinusoidal function 3. The orientation θ of the normal to the parallel stripes of the Gabor function 4. The spatial aspect ratio γ specifies the ellipticity of the support of the Gabor function.

For $\gamma = 1$, the support is circular. For $\gamma < 1$, the support is elongated in the orientation of the parallel stripes of the function as shown in fig.3.2.2

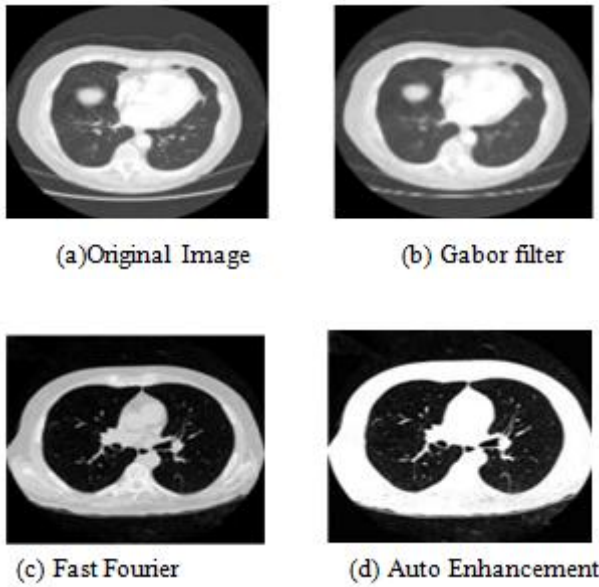


Fig. 3.2.2

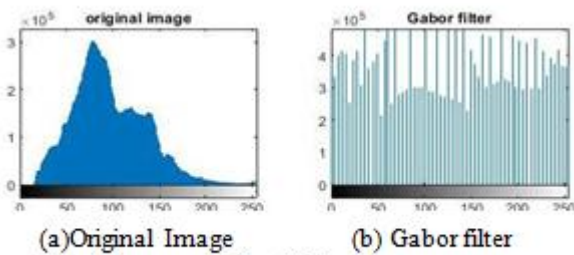


Fig. 3.2.3

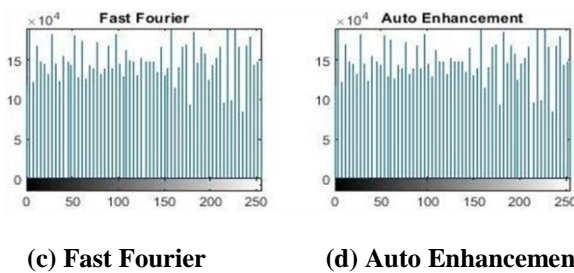


Fig. 3.2.3

3.2.3 Image Segmentation

The image getting into multiple segments is usually known as image segmentation. The reason behind segmentation is to analyze any image easily and to get better understanding. It segments image into its constituent regions or objects. The final outcome of segmentation is that it covers entire images into different segments or a set of contours extracted from the image [5]. Watershed segmentation of Marker-controlled uses this method as shown in fig.3.2.3. Its main use is to segment gray matter, white matter etc. The watershed approach is very powerful because it combines frontier approach and region approach. The working of the segmentation is given in Fig.3.2.4.

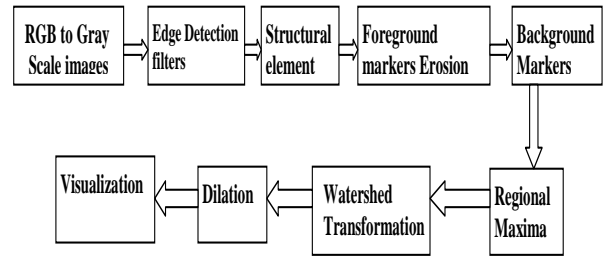
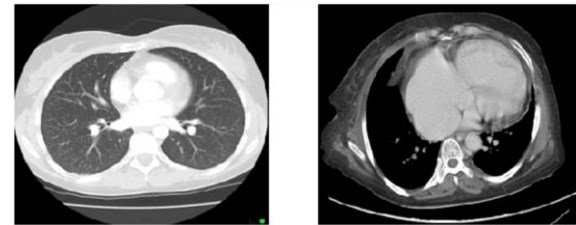
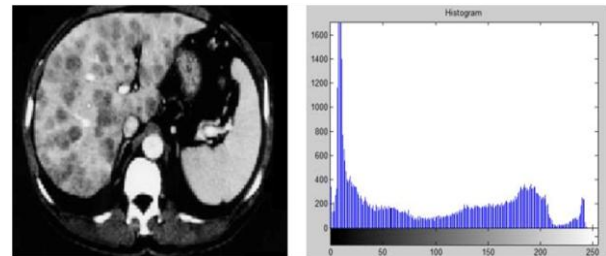


Fig.3.2.4

Steps of Marker- Controlled Watershed Segmentation



(a)Original Image (b)Segmented Image



(c)Marker-controlled (d) Histogram
Fig.3.2.5

IV. CONCLUSION AND FUTURE SCOPE

The paper has discussed with wide range of pre-processing methods used for image like CT scans. Major discussion of the paper has been how the pre-processing can be improved using different pre-processing methods and brought out with filtering methods with median filter and Gabor filters with watershed segmentation. Here in the smoothening method the image is smoothened on the edges where the sharp edges are chosen for it. Enhancement is done on the image with respect to better interpretability and wider perception about the image. The enhancement is done with spatial domain and frequency domain. Finally, the segmentation has been done with watershed segmentation and the result covers entire images into different segments. The watershed segmentation of Marker controlled uses this method with the support of regional maxima. The abnormality tissue has to take into account of the cells in the lung to find out the stages of the cancer. The region of interest is considered to find out the cancer in the cell. Gabor filter and watershed segmentation gives best results for pre-processing stage. For the further work the deep learning method can be implemented as the more number of images can give better result thus increasing the accuracy and detection of cancer nodules with high clarity.

REFERENCES

1. A.Cruzroa, J.C.Caicedo, and F.A.Gonzalez, "Visual pattern mining in histology image collections using bag of features," *Artificial Intelligence in Medicine*, vol. 52, no. 2, pp. 91–106, 2011.
2. Ilya Levner and Hong Zhang, "Classification driven Watershed segmentation", *IEEE TRANSACTIONS ON IMAGE PROCESSING* VOL.16, NO.5, MAY 2007
3. Anita chaudhary and Sonit Sukhraj Singh, "Lung Cancer Detection on CT Images Using Image Processing", *International transaction On Computing Sciences*, VOL4, 2012
4. B.V. Ginneken, B.M. Romenyand M.A. Viergever, "Computer- aided diagnosis in chest radiography: a survey", *IEEE, Transactions on medical imaging*, vol.20, NO.12, DEC-2001
5. Disha Sharma and Gagandeep Jindal, "Identifying Lung Cancer Using Image Processing Techniques", *International Conference on Computational Techniques and Artificial Intelligence (ICCTAI'2011)*, vol:17, pp:872-880, 2011
6. Nguyen, H.T., et al, "Water snakes: Energy-Driven Watershed Segmentation", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Volume 25, Number 3, pp.330-342, March 2003.
7. Farag A, Ali A, Graham J, Elshazly S and Falk R. "Evaluation of geometric feature descriptors for detection and classification of lung nodules in low dose CT scans of the chest. Paper presented at the *Biomedical Imaging: From Nano to Macro*", 2011 IEEE International Symposium, Chicago, IL, USA, March 30 to April 2, 2011.
8. Lin P-L, Huang P-W, Lee C-H and Wu M-T. "Automatic classification for solitary pulmonary nodule in CT image by fractal analysis based on fractional Brownian motion model". *Pattern Recognition*. 2013; 46 (12): 3279–3287.
9. Farag A, Elhabian S, Graham J, Farag A and Falk R. "Toward precise pulmonary nodule descriptors for nodule type classification. *Medical Image Computing and Computer-Assisted Intervention*", MICCAI 2010: Berlin, Germany: Springer; 2010.