

Machine Learning Approach to Predict Lung Cancer using CT scan Images

Dr.A.Nagarajan¹, J. Vasanth Wason²

¹Assistant Professor, ²Research Scholar,

^{1,2}Department of Computer Applications,

^{1,2}Alagappa University, Karaikudi – India.

¹nagarajana@alagappauniversity.ac.in

²jvasanthwason@gmail.com



ABSTRACT

Machine learning methods have become a popular tool for medical researchers to find accurate prediction. Machine learning techniques can identify the patterns in complex data set and it can effectively predict cancer. In the recent medical era, medical imaging plays an essential role to perform effective diagnosis and analysis of the complete human body. Medical experts completely depend on the medical imaging results which are obtained from the image sensors. Cancer is a disease of human cells. As the World Health Organization (WHO) reported, 9 million people died cancer all around the world. Today medical diagnosis process needs computational methods to find the results with improved accuracy. Machine learning techniques are widely used to support medical practitioners to find the diseases in the earlier stage. The CT scan images are used along with computational methods gives better results than normal tests. Early cancer detection plays a vital role in human health in the healthcare industry. In the human body respiration system, lungs are the fundamental element. In all over the world, Lung cancer is a severe disease of people died. The accurate prediction of lung cancer is a very challenging task for medical practitioners. In this research work, the patient's real CT scan images are gathered from the Lung Image Database Consortium (LIDC) archive. There are five main phases involved in this proposed research work such as image preprocessing, lung segmentation, nodules detection, feature extraction and image classification. The image segmentation is performed by using the OTSU method. OTSU is an image thresholding based segmentation method. This segmentation algorithm is used to find accurate segmentation of the lung region from the CT scan images. Machine learning-based Classification method is used to detect whether the nodule is either benign or malignant.

Key words: Machine Learning, Cancer Prediction, Classification, Cancer Susceptibility.

1. INTRODUCTION

Lung Cancer is one of the major reasons for death for both men and women. It has been characterized as a heterogeneous

disease and it has consists of various subtypes. When a group of human body cells are uncontrollable and irregular growth it causes cancer tumors [1]. The early prediction of lung cancer plays a vital role in choose the proper treatment to the doctors. The death rate of cancer may be reduced if the cancer is early predicted, so the medical practitioners start the treatment within the specified time. In the biomedical research field, Machine Learning (ML) based classification techniques are used to find out cancer by its symptoms in the earlier stages. Because the classification of cancer patients into low risk or high risk has been lead the accurate treatment [2].

Lung Cancer occurs due to the abnormal growth of cells. When lung cancer occurs, the cells in the lung grow abnormally [3]. Early diagnosis is the only way to the survival of the patient. The diagnosis of lung cancer is done by physical examination, imaging and lab tests. Early determination of lung disease is troublesome because a lung tumor huge enough to cause indications of ailment is normally advanced. The survival rate of several kinds of cancers like prostate, breast, colon, rectum and leukemia has increased because of early detection and diagnosis. There is no increase in survival rates for cancers like bronchus, pancreas and lung because early detection is very difficult.

Even though the growth of technology develops there is no proper solution to cure and diagnosis of treatment to cancer, the survival rate is low for lung cancer [4]. For diagnosing and treatment, the medical images like x-rays and CT scan images will give a clear picture to the physicians about the condition of the disease. CT scan images play an important role in cancer detection [3]. But at times the Lung CT scan images may suffer from noise and contrast. So, the images should be preprocessed. Lungs region can be segmented using the OTSU method. In classification technique is used to detect cancer could be benign or malignant.

The structure of the paper is formatted as Section I provide a general introduction related to lung cancer and image processing techniques. Section II discusses the proposed methodology of lung cancer detection. Section III discusses the Experimental Results of the detecting techniques for lung cancer Image classification on its earlier stage. The conclusion is defined as the next section IV.

2. PROPOSED METHOD

Lung cancer is one of the most common deadliest diseases; it affects both men and women throughout the world [5]. Early prediction of lung cancer can lead to better diagnosis. For accurate detection, it needs to identify efficient features and delete redundancy features among all features. Data mining techniques is the most important research tool for biomedical and bioinformatics researches to predict the disease outcome using the datasets. Feature selection is a promising application that is used to increase the speed of classification technique. Feature selection in medical prediction plays a major role as it identifies prominent features that influence prediction.

The main aim of this research work is to apply the machine learning-based data mining techniques to find the lung nodules which are in the CT scan cancer images. Without removing the lungs, the nodules are detected, and then only the position of nodules can be identified. The starting stage of cancer will be small nodules. To detect and classify the initial stage of nodules, data mining techniques are used to extract nodules from the CT scan images. The basis for developing a system is, it may be capable of detecting nodules from the starting stage in the CT scan image which is difficult for the radiologists to find because of a very small size. The complication of lung nodules detection is as follows:

- The size of the nodules may vary.
- There is a vast change in the nodule density.
- The nodules situated anywhere in the lung region, they may be covered by the ribs.

The lung nodules detection steps in this research work are shown in the following Figure 1.

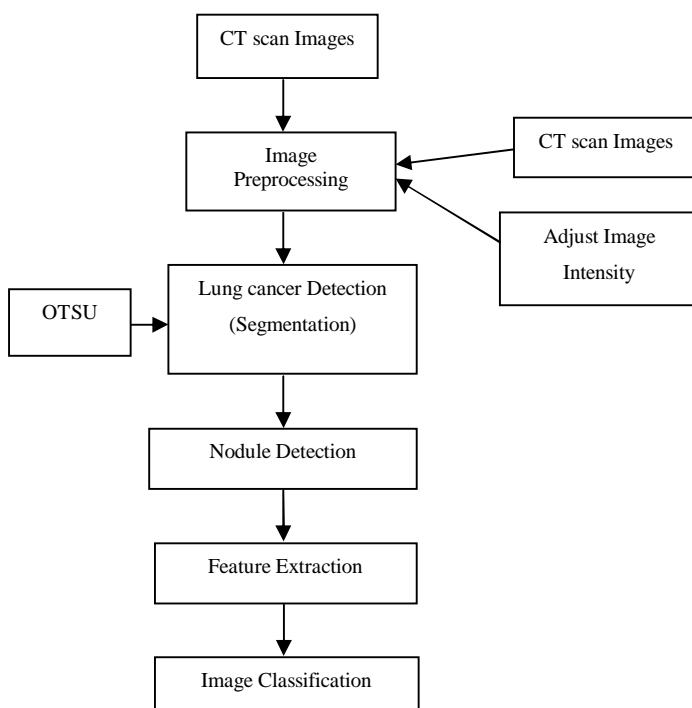


Figure 1: The step to detect the lung nodules

The first step is the preprocessing technique, which is used to remove the noises from the processed image and adjust the image intensity, etc. As a result of preprocessing, the image quality can be improved. Lung segmentation is used to detect the lung region from the whole chest x-ray image. Classification technique used to find the stage of cancer.

Data Preprocessing

In this research work the wiener filter which is used to remove noise from the image, improve the image quality and adjust the image contrast. A low contrast image shows the difference between the light and dark pixel values. Wiener filter is used to equalize the pixel values and brightening the image [6]. It is a low pass filter, which is used to reduce the noise from the image by a constant power. Wiener filter gives an estimation of the original clear image with minimum mean square error.

Image Segmentation

Image segmentation is an important step in most of the image applications [7]. The morphological operations in the image segmentation are fully based on the set theory. These operators were particularly used in the analysis of binary images and commonly used for edge detection, noise removal and image enhancement. This technique is used to segment a particular region from the given image. In this paper, the OTSU method is used for segmentation. The Otsu Segmentation is to obtain a binarization of the enhanced image. The OTSU's method automatically performs histogram shape-based image thresholding, or it will reduce the gray level image to a binary image [8]. The binarization is done on the histogram equalized image. The condition is 0 Or 1.1 represents the required region and 0 represents remaining. After the segmentation, the lung region is white and the rest is black.

Feature Extraction

Feature selection used for cancer classification technique that will give better accuracy. Feature extraction is used to extract a particular region using mathematical functions. It acts as a basis for the classification process. Four features are extracted in feature extraction such as area, shape, perimeter and centroid [9]. The definition of features as follows:

- 1) Area: Area is a scalar value which is obtained by summation of pixels [10]
- 2) Perimeter: Perimeter used to find a distance between the two interconnected pixels of the image.
- 3) Shape: Shape is a combination of area and perimeter, used to find the shape of the nodule [11].
- 4) Centroid: Points that may be considered as the center. Using centroid value from the center point nodules was detected [12].

3. EXPERIMENTAL RESULTS

The CT scan images are used for the proposed system. The lung nodules are detected using the processing techniques like preprocessing, lung segmentation and classification. The variation of classification accuracy is obtained for all the images. The accuracy is calculated using the support vector machine. In segmentation, CT scan lung cancer image has undergone template matching. The proposed system obtained with high accuracy and lower misjudgement and also the images are classified as normal, benign and malignant. The accuracy of CT scan lung cancer images is shown in the following table.

Table 1: Result obtained for lung cancer Image classification Accuracy

Image Category	No. of Samples	Accuracy
Normal	60	100
Benign	60	98
Malignant	60	97.5

4. CONCLUSION

The Perfect accuracy of cancer classification is important for human's life. Normally, using the diagnosis tool, most of the researchers are interested in AI classification techniques to classify cancer. If cancer has spread, treatment becomes more difficult and a person's chance of survival is less. The earlier cancer is found, the greater the chance of survival. According to the above concepts, this research paper proposed and developed a system for early detection of CT lung cancer in scan images. In the developed system MATLAB is used for the entire implementation. The OTSU segmentation is processed in this paper for segmenting the images. The experimental results show a better accuracy of the proposed method to achieve good accuracy results.

ACKNOWLEDGEMENT

The initial work of this paper is presented in the International Conference on Wireless Sensor Networks, Ubiquitous Computing and Applications (ICWUA) held in Malaysia on 03.09.2019.

This article has been published under RUSA Phase 2.0 grant sanctioned vide letter No. F. 24-51/2014-U, Policy (TN Multi-Gen), Dept. of Edn. Govt. of India, Date 09.10.2018.

REFERENCES

1. Anjali Sevani et al., "Study and Realization of Lung cancer identification and detection using Image Processing Techniques", *International Journal of Advance Engineering and Research Development*, Volume 5, Issue 01, e-ISSN: 2348 – 4470, 2018, pp. 792-797.
2. Abdelwadood M. Mesleh, "Lung Cancer Detection Using Multi-Layer Neural Networks with Independent Component Analysis: A Comparative Study of Training Algorithms", *Jordan Journal of Biological Sciences*, Volume 10, Number 4, ISSN 1995-6673, 2017, pp. 239 – 249.
3. Bariqi Abdilllah et al., "Image processing based detection of lung cancer on CT scan images", *The Asian Journal*, doi :10.1088/1742-6596/893/1/012063, 2016, pp. 1-8.
<https://doi.org/10.1088/1742-6596/893/1/012063>
4. Kale Vaishaw et al., "An Innovative Approach for Investigation and Diagnosis of Lung Cancer by Utilizing Average Information Parameters", *Elsevier Procedia of Computer Science*, 132, 2018, pp. 525-533.
<https://doi.org/10.1016/j.procs.2018.05.005>
5. Kamil Dimililer et al., "Tumor Detection On CT Lung Images Using Image Enhancement", *Journal of Science and Technology*, Volume 7, Issue 1, 2017, pp. 131-138.
6. Macedo Firmino et al., "Computer-aided detection system for lung cancer in computed tomography scans: Review and future prospects", *BioMedical Engineering*, Volume 13, Issue 41, 2014, pp 1- 16.
<https://doi.org/10.1186/1475-925X-13-41>
7. Mokhled S. Al-tarawneh, "Lung Cancer Detection Using Image Processing Techniques", *Leonardo Electronic Journal of Practices and Technologies*, ISSN 1583-1078, 2012, pp. 147-158.
8. A.Sakthivel and Dr.A.Nagarajan, "Cancer Detection Methodology Using Fuzzy Based Classification Techniques", *International Journal of Engineering Sciences & Research Technology*, ISSN: 2277-9655, Volume 7, Issue 3, 2018, pp. 727-733.
9. V.Vani and M. Kalaiselvi Geetha, "Implementation of Classification System for Brain Tumor using DWT", *International Journal of Advanced Trends in Computer Science and Engineering*, Volume 5, No.6, 2016, pp. 82-88.
10. G.D.Praveenkumar and G.Sathishkumar, "Application of Cancer Data Using Topsis Methods", *International Journal of Advanced Trends in Computer Science and Engineering*, Volume 5, No.4, 2016, pp. 17-20.
11. Sheetal V Prabhu and J.A.Shaikh, "Artificial Neural Network-Based Classification System for Lung Nodules on Computed Tomography Scans", *International Journal of Engineering Science and Computing*, March 2017, Volume 7, Issue 3, 2017, pp. 5097-5099.
12. Vishnu A V et al., "Early Prediction of Lung Cancer Using MRI Images", *International Journal of Innovations & Advancement in Computer Science IJIACS*, ISSN 2347 – 8616, Volume 6, Issue 11, November 2017, pp. 41-42.