e-ISSN: 2582-5208
International Research Journal of Modernization in Engineering Technology and Science
Volume:03/Issue:06/June-2021 Impact Factor- 5.354 www.irjmets.com

COMPARISON OF DIFFERENT CLASSIFIERS FOR LUNG CANCER DETECTION AND CLASSIFICATION

Aiswarya C V*[1]

*[1]M.Tech Student, Department of Electronics and Communication Engineering,
College Of Engineering Thalassery, Kerala, India

ABSTRACT

A cancer that occurs in the lungs and most often occurs in people who smoke is specified as lung cancer. Lung cancer is considered as the prime reason of cancer death. Detection of lung cancer in the earliest stage will be very useful to improve the survival rate of cancer patients. But diagnosis of cancer is one of the crucial challenging tasks for radiologists because most of the cancer cells are overlapped each other. An intelligent computer aided diagnosis system can be very much helpful for detecting and predicting lung cancer. This paper proposed a well-organized detection as well as classification of lung cancer by using both SVM and ANN classifiers. The proposed methodology for cancer detection consists of three phases; include pre-processing, segmentation and classification. CT images are converted into grayscale images in the pre-processing stage. Then segmentation is done to view the cancer affected region clearly. Finally classification is done using SVM and ANN classifiers and done their performance analysis. SVM performs certain data transformations and classifies data based on the labels which is already decided. ANN works similar to the information processing ability of human brain. This project is developed using MATLAB.

Keywords: Advanced lung cancer detection, ANN classifier, image processing, segmentation, SVM classifier

I. INTRODUCTION

Image processing gives a better platform for upgrading the human works. In this system for pre detection, prediction and classification stages of lung cancer image processing techniques are used. Pre-processing is used for the purpose of grayscale conversion of image. Classification is very important to detect whether it is malignant or benign. It is computational procedure to classify pro images in to categories based on their resemblances. It is done by using feature extraction. In proposed system CT image is converted to grayscale image in preprocessing stage and both SVM and ANN are used as the classifiers to classify whether the tumor is cancerous or not.

Doctors believe smoking is one of the main reasons behind lung cancer. Smoking damages the cells present in the lungs. When you inhale smoke produced from cigarette, which contains carcinogens (cancer-causing substances), leads to change in the lung tissue. Initially human body may have the ability to restore the changes. But in the case of too much damage, the cells which are normally present in the lungs are getting harmed. Which in turn converts the normal lung cells to behave abnormal and in due course it will turn to cancer cells. A person who smokes cigarette excessively has higher possibility of having lung cancer compared to a person who has never smoked. There are many other factors, such as excessive alcohol and environment pollution may cause lung cancer. Lung cancer is seen as uncontrolled growth of cells present in the lungs. Spreading of the cancer to other parts of the body, such as human brain leads to vision problem and body weakness. Some of the main manifestations of lung cancer comprise cough containing blood, Fluid in the chest (pleural effusion) causing chest pain and breathing problem.

Detecting the lung cancer in the earliest stage helps in improving the survival rate of patients. Lots of techniques are used for the diagnosis of lung cancer; they include Magnetic Resonance Imaging (MRI) scan, Computerized Tomography (CT) scan, Tissue Sample (biopsy) and Sputum Cytology. In MRI scan method, magnetic imaging methods and radiowaves are used to detect and also to get a digital image of lungs cells. In Computerized Tomography scan images of the lungs from several angles are captured and studied. A biopsy is a technique used for diagnosis of cancer where a small segment of the lung affected by the tumor taken and then examined the same using a microscope. One of the methods for performing Biopsy is bronchoscopy; it is performed by inspecting cancer affected portions of the lungs with a lighted tube inserted into the human lungs. Next method is mediastinoscopy, here tools required for surgery are inserted behind the breastbone through an opening. Needle biopsy is another technique in which CT images or X- ray images are used to insert needle for collecting tumor affected cells. Sputum Cytology is a method of observing the sputum under the

microscope. Majority of the above discussed methods are identifying the cancer in its final stages. Hence a better technique is required for predicting the lung cancer as early as possible to increase the survivability of patients. Image processing methods gives a better tool for upgrading the lung cancer detection system.

The purpose of proposed system is to classify the lung cancer as early as possible based on certain features.

Objectives of the proposed system are pointed below:

- To reduce the number of rules for tumor detection.
- To lower the cost and time needed for numerous Medical Checkups.
- To rise the accuracy of cancer diagnosis system
- Lung Cancer diagnosis can be done by using low amount of features.
- Early stage detection of lung cancer.
- To improve the survivability rate of the lung cancer patients.

II. METHODOLOGY

Method and analysis which is performed in the research work is included in this section. CT images are taken as the input image for tumor detection using image processing. The system depicted here consists of three phases. First grayscale conversion of the input CT image is performed and then segmentation is used to displays the tumor region vividly and classification algorithm is applied to classify whether the tumor present inside the lung is either benign or malignant. Classification can be done using both SVM and ANN classifiers and compare their performance based on accuracy. Here I am using Support Vector Machine algorithm that works on structural risk minimization to classify the images. ANN works like the way human brain process information.

The algorithm of the proposed system is discussed as follows:

Step 1: CT image of the lung is taken as the input.

Step 2: Perform pre-processing

Step 3: Done segmentation by otsu thresholding

Step 4: Certain features of the tumor region in the images are calculated

Step 5: Implemented Classification of brain tumor using both SVM and ANN classifiers

The above steps are described below.

A. Pre-processing Stage

Pre-processing means converting RGB images to Grayscale images: that means image comprise of grey shades. A 'gray' color has blue element, green element and red element have closer intensity values in RGB space. In this project CT images are taken as input image and some of them may comprise RGB (primary) content. Hence it must be converted to gray scale values which scale from 0 to 255 pixel value where 0 implies the black and 255 implies white color.

B. Segmentation

Segmentation of image is one of the important and the most crucial work in image processing technique. Its motive is to take out the information present in an image. Three major segmentation techniques broadly used include Thresholding (Boundary approach), edge-based approach and finally region-based approach. Here boundary approach is used for this project. Pixels are allocated to order based on the range of pixel values in which a value is present in case of thresholding. Which is one of the most popular and easiest method for image segmentation. Let g(x,y) be the position of the each single pixel of the grayscale image and the threshold value is T. T is already defined in this case. If the threshold value >= current pixel value f(x,y) then pixel g(x,y) is 0. Otherwise it is assigned 1. By displaying the values of "g" will get the segmented image. Three methods of threshold based segmentations are present they include global, multiple and variable thresholding. Which are explained below in detail.

1) Global Thresholding: In this case, all images are segmented by using a single threshold value. This type of thresholding is used in case of almost similar pixel values for both tumor and background region. It is mathematically expressed as (1) and (2) (T is the threshold value).

$$T: g(x, y) = 1, \text{whenever } T<f(x,y) \qquad (1)$$

e-ISSN: 2582-5208
International Research Journal of Modernization in Engineering Technology and Science
Volume:03/Issue:06/June-2021 Impact Factor- 5.354 www.irjmets.com

T: g(x, y) = 0, whenever T>=f(x,y) (2)

2) Variable Thresholding: In this case, T (threshold value) changes in every image. Two important methods of variable thresholding are explained here in detail.

a) Local or Regional Thresholding: In this case pre-processed image is taken as input and will get the black and white image as resulting image. T is taken based on the neighborhood of pixels located at (x, y).

b) Adaptive Thresholding: Here, T (threshold value) at every position of image will be based on the neighboring brightness values.

c) Multiple Thresholding: In this case more than one threshold values are calculated by using this method. Certain intensity regions of the image are extracted corresponding to one background.

g(x,y) = p , whenever T2<f(x,y) (3)

g(x,y) = q , whenever T2>=f(x,y)>T1 (4)

g(x,y) = r , whenever T1>=f(x,y) (5)

Here T1 and T2 are threshold values in (3), (4) and (5).

C. Classification

This is done to classify whether the tumor is benign or malignant using SVM algorithm. The complete flow of the implementation is as shown below

Step 1: Feature extraction

Step 2: Classification using classifiers

The above two steps are described below in detail.

1) Feature extraction: It is done to extract features for the purpose of classification. In this system consists of both intensity as well as GLCM features. Five intensity features calculated here include mean, standard deviation, variance, skewness and kurtosis. Along with that entropy and variance are extracted from the given input image as GLCM features.

2) Classification using classifiers: Lung cancer classification is performed by using both SVM and ANN classifier. SVM stands for Support Vector Machine, which works by examining the predefined label it returns. The basic SVM predicts which of the two possible classes forms the output from a set of input data. The classification process comprises of two stages, they include the training phase and the testing phase. In the first phase that is the training phase contains images with known data and image that must be classified should be given given to the testing phase.

An Artificial Neural Network is a group of connected nodes also known as artificial neurons, which is similar to the connections in a human brain. In the implementation of Artificial Neural Network, the "signal" should traverse from the input layer to the output layer through neumerous hidden layers. The link between the neurons is termed as edges. A weight is allotted for Neurons and edges and these weights get adjusted in the training phase. The strength(either increases or decreases) of the signal varies according to the weight. The accuracy of classifier depends on the efficiency of classification.

D. Measures of Performance Evaluation

The performance evaluation of the system is performed by calculating different measures. The measures used here include Classification Accuracy (AC), Sensitivity and Specificity. Confusion Matrix is used for calculating these measures. A confusion matrix comprise of information about actual as well as predicted classifications done by a classification system. The data in the matrix are used for the performance evaluation of such systems. The table given below shows the confusion matrix for a system with two expected results.

Table I. Confusion Matrix

		Predicted	
		Negative	***Positive***
Actual	**Negative**	TN	FN
	Positive	FP	TP

TN (True Negative) – Normal condition predicted correctly

FN (False Negative) – Normal condition predicted incorrectly

FP (False Positive) – Abnormal condition predicted incorrectly

TP (True Positive) – Abnormal condition predicted correctly

Accuracy (AC), sensitivity and specificity can be obtained:

$$\text{Accuracy} = (TP+TN) / (TP+FN+TN+FP) \quad (6)$$

$$\text{Sensitivity} = TP / (FN+TP) \quad (7)$$

$$\text{Specificity} = TN / (FP + TN) \quad (8)$$

III. IMPLEMENTATION AND RESULTS

Model and Material which are used for lung cancer detection is presented in this section. Table and model should be in prescribed format. Lung cancer identification and classification using both SVM and ANN classifier is done. The whole system has three stages. They are pre-processing, segmentation and classification. The feature extraction is done by training the dataset. Labeled images of benign tumor as 0 and labeled the image with malignant tumor as 1. The software used is MATLAB. The steps are as follows:

A. Pre-processing

CT images of lungs are taken as input image. Two classes are considered here, benign and malignant. Some CT images consist of primary content so it should be converted to gay scale. In the below figure we cannot see any difference because here the input image is a grayscale image.

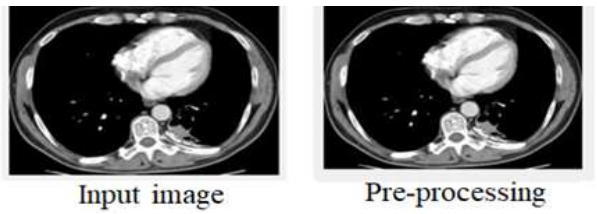

Figure 1: Pre-Processing

B. Segmentation

Segmentation is done to view the tumor region clearly. By using Otsu method of thresholding, image is converted to black and white. Next morphological operations are used to remove the noise present (removed area below 500 pixels). Now multiplied the noise removed image with the input image to get the segmented image.

Figure 2: Segmentation

C. Classification

Classification is done to classify whether the tumor is cancerous or non-cancerous using SVM algorithm. Which is done based on certain statistical features (mean, standard deviation, variance, skewness and kurtosis). The classification of tumor consists of two phases: the training phase and the testing phase. The training phase consists of known data and the testing phase consists of unknown data.

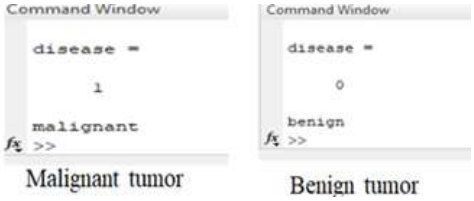

Figure 3: Classification

D. Performance Analysis

The principal objective of this project is mainly to compare the performance of two classifiers i.e. ANN (Artificial Neural Network) classifier and SVM (Support Vector Machine) classifier. For ANN classifier, the highest accuracy of 95%, with a sensitivity of 94%, and a specificity of 95% are obtained. In case of SVM classifier, the accuracy of 75%, with a sensitivity of 62.9%, and a specificity of 100% are obtained.

IV. CONCLUSION

In this work, CT images are preferred for identifying the classification as normal or abnormal. Here, analyzed different methods used in image processing and also explained its needs and importance for the purpose of lung cancer detection and classification on Computerized Tomography images. The tumor is segmented using Multi threshold based segmentation. Then both intensity features and GLCM features are calculated using MATLAB functions. Then SVM as well as ANN classifier are used to classify the tumor. Accuracy obtained for classification using SVM classifier is 62.5% and accuracy obtained using ANN classifier is 95%. Since the whole lung cancer diagnosis system is accessible online, it is advantageous for people from anywhere in the world.

ACKNOWLEDGEMENTS

I would like to express my deep gratitude to Professor Shabin P, for his patient guidance, enthusiastic encouragement and useful critiques of this project.

V. REFERENCES

[1] T. Sowmiya, M. Gopi, M. New Begin L. Thomas Robinson “Optimization of Lung Cancer using Modern data mining techniques” International Journal of Engineering Research ISSN:2319- 6890)(online), 2347-5013(print)Volume No.3, Issue No.5, pp : 309-3149(2014).

[2] 1Ayushi Shukla, 2Chinmay Parab, 3Pratik Patil, 4Prof. Savita Sangam “Lung Cancer Detection using Image Processing Techniques.” International Research Journal of Engineering and Technology (IRJET) e-ISSN: 2395-0056 Volume: 05 Issue: 04 | Apr-2018 www.irjet.net p-ISSN: 2395-0072.

[3] Jinsa Kuruvilla∗, K. Gunavathi “Lung cancer classification using neural networks for CT images” computer methods and programs in biomedicine 113 (2014) 202–209 www.intl.elsevierhealth.com/journals/cmpb

[4] RiriesRulaningtyas, KhusnulAin “Edge Detection and Brain Tumor Pattern Recognition”,IEEE International Conference on Instrumentation, Commmunication, Information Technology and Bioedical Engineering, Vol. 2, Issue 4, April 2013.

[5] Shraddha G. Kulkarni1, Sahebrao B. Bagal2 “Lung Cancer Tumor Detection Using Image Processing and Soft Computing Techniques” International Journal of Science Technology and Management Vol.No. 5, May 2016, www.ijstm.com.