

Computer Aided Lung Cancer Detection and Tumor Staging in CT image using Image Processing

Sruthi Ignatious

Dept. of Computer
Science and
Engineering,
Mar Baselios college of
Engineering and
Technology
Trivandrum, India

Robin Joseph

Dept. of Computer
Science and
Engineering,
Mar Baselios college of
Engineering and
Technology
Trivandrum, India

Jisha John

Dept. of Computer
Science and
Engineering,
Mar Baselios college of
Engineering and
Technology
Trivandrum, India

Dr. Anil Prahladan

Dept. of Imageology,
Regional Cancer Centre
Trivandrum, India

ABSTRACT

Lung cancer is one of the death threatening diseases among human beings. Early and accurate detection of lung cancer can increase the survival rate from lung cancer. Computed Tomography (CT) images are commonly used for detecting the lung cancer. Nowadays the lung cancer is staged according to the TNM staging method where T means Tumor, N means Nodule and M means Metastases. The existing lung cancer detection algorithms cannot stage cancer according to the TNM staging method. The proposed system can identify the T stage of the cancer accurately. The proposed system includes different stages such as pre-processing, segmentation, feature extraction, tumor detection and tumor stage identification. The proposed system promises better result than the existing systems, which would be beneficial for the radiologist for the accurate and early detection of cancer. The method has been tested on 200 slices of CT images of various stages of cancer obtained from Regional Cancer Centre Trivandrum and is found to give good results. The accuracy of the proposed method in this dataset is 94.4%

General Terms

Algorithms, Classification, Staging

Keywords

CT image, Pre-processing, Segmentation, Feature Extraction, TNM stage

1. INTRODUCTION

Medical imaging is one of the emerging fields of research. Medical imaging using computer aided techniques helps to detect the death threatening diseases such as cancer. The uncontrolled cell growth in the tissues is the main reason for the cancer. The cancer can affect different organs within the human body. Lung cancer is a common type of cancer. Cancer can be of two types, small cell lung cancer and non-small cell lung cancer. The main reason for cancer is smoking. The other reasons include air pollution, genital factors etc.. Computed Tomographic (CT) images are the important tool used for detecting the lung cancer. The lung cancer detection and staging is done by the radiologist using the CT image.

The lung cancer is staged according to the Tumor Node Metastases (TNM) staging method. This staging method describes the size of the Tumor (T), checks whether the cells of the cancer have spread into the lymph nodules (N) and

check whether the cancer have spread anywhere else in the body, metastases (M) [3]. The T stage of cancer is of different types. They include T1, T2, T3 and T4. The proposed computer aided detection technique stages the Tumor (T) portion within the lung.



Fig 1: Sample CT image of lung [1]

The figure 1 shows the CT image of lung which is affected by cancer. The marked portion within the image shows the tumor portion within the lung. In the above CT image the right lung is affected with cancer. 'The Hindu', one of the leading newspapers in India conducted a survey in 2010 about the cancer patients. According to the survey, cancer killed 5,56,400 people across the country. The 30-69 age group accounted for 71 per cent (3,95,400) of the deaths. In 2010, cancer alone accounted for 8 per cent of the 2.5 million total male deaths and 12 per cent of the 1.6 million total female deaths in this age group (30 to 69 years). Nearly 23 per cent, oral cancer caused the most number of deaths among men. It was followed by stomach cancer (12.6 per cent) and lung cancer (11.4 per cent). In the case of women, cervical cancer was the leading cause (about 17 per cent), followed by breast cancer (10.2 per cent) [2]. From the above survey it is clear that the lung cancer is a serious death threatening disease. The death rate for this cancer is increasing every year. If the disease is detected early and accurately proper medical support can be given to the patients by the doctor [4].

This work introduces a computer aided lung cancer detection system which will be highly beneficial for the medical field. Section 2 gives clear idea about the related works within this field, section 3 explains the proposed system in detail, section 4 gives the results and discussions of this work and section 5 gives the future enhancement within this work.

2. RELATED WORK

A computer aided lung cancer detection system includes three main processing steps; they are enhancement (pre-processing), segmentation and feature extraction [4]. These are the essential steps within the lung cancer detection system. Many lung cancer detection systems were developed for improving the accuracy of the detection. Some of the systems are explained below. S. Shakit Praveen and C. Kavitha [9] introduced a lung cancer detection system used for detecting the tumor. In that method, the CT slice of a patient is given as input and the best slice for the processing is chosen using the PSNR (Peak Signal to Noise Ratio) value. The chosen slice is enhanced using the median filter. From the enhanced image the outline of the lung is extracted. Using flood fill algorithm the extracted lung portion is filled. After that the lung filled portion is extracted. This extracted lung portion is segmented for identifying the nodules. The segmentation used is automatic region growing method. The main drawback of this system is, the features of the tumor are not identified and hence the tumor is not classified.

Anitha Chaudary and Sonit.S.Singh [5] introduced a lung cancer detection system using various image processing techniques. In that system, the CT image is enhanced using the gabor filter. The enhanced CT image is segmented using marker controlled watershed segmentation algorithm. The features within the segmented image are extracted. The features that are extracted are area, perimeter and eccentricity. Based on the threshold value and the extracted features the cancer is staged into stage 1 or stage 2. The main drawback of this system is, the cancer is staged according to the old staging method of cancer. Anam Tariq, M. Usman Akra and M. Younus JavedS [1] introduced a lung cancer detection system. In that method, the lung is segmented using the threshold segmentation. After the segmentation process the segmented image is post processed. The features within the post processed image are identified. The features that are extracted are area, energy, entropy, eccentricity, standard deviation and mean. These extracted features are given to the neuro fuzzy classifier for the classification of cancer. The main drawback of this system is no staging of cancer is done.

D.S. Elizabeth, H.K. Nehemiah, C.S. Retmin Raj, and A. Kanna [10] introduced a lung cancer detection system which uses the greedy snake algorithm for segmentation. From the segmented image the region of interest is identified. The feature of this region is identified and thus the cancer is detected from the extracted features. In this method the radial basis neural network is used for detecting the cancer. The main disadvantage of this method is, there is no staging of cancer.

3. PROPOSED SYSTEM

The proposed system includes five main processing steps; they are pre-processing, segmentation, feature extraction, tumor detection and tumor stage identification. The figure 2 shows the structure of the proposed lung cancer detection system.

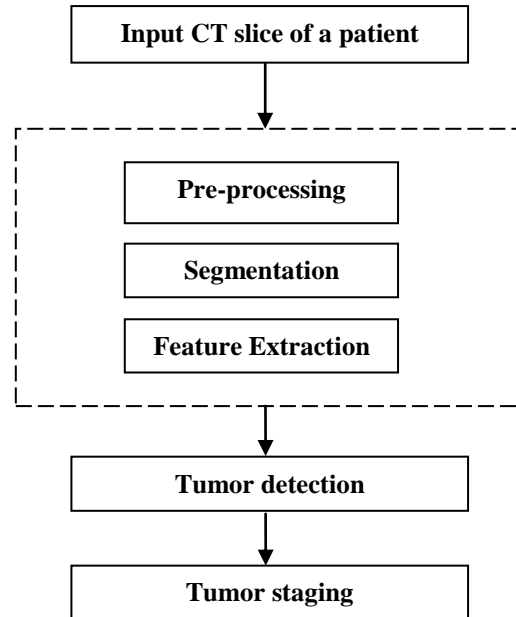


Fig 2: Block diagram of the proposed lung cancer detection system

The basic steps within the proposed system are pre-processing and segmentation. The segmented image can be given for feature extraction, thus tumor portion within the lung can be identified. Depending on the feature of the tumor, the stage of the lung cancer can be identified and the result can be given to the patients or doctors. The CT image slice is pre-processed for enhancing the image. This process sharpens the image for the processing easiness. The pre-processed image is segmented using the Marker Controlled Watershed segmentation. The output of the segmentation is the binary image showing the tumor portions. The segmented binary image is used for the extraction of features of the tumor portion. The features extracting for the identification of the stage of cancer are Area, Perimeter, Eccentricity, Convex area and Mean intensity. Based on the threshold value and the extracted features the tumor is staged. The threshold values have been obtained after experimental analysis on the dataset with the help of experts. Using this threshold values the tumor is staged into stage 1 tumor (T1) or stage 2 tumor (T2) or stage 3 tumor (T3).

The different steps involved in the proposed method are the following:

Step 1: Read in the CT image slice of a particular patient.

Step 2: Pre-process the CT image slice.

Step3: Segment the pre-processed image using Marker Controlled Watershed segmentation.

Step 4: Convert the segmented image into a binary image.

Step 5: Extract the features from the binary image.

Step 6: With the extracted features, identify the stage of the cancer.

3.1 Pre-processing

Pre-processing is an essential step for many of the image processing applications. The pre-processing technique used here is sharpening of the CT image. Thus the finer details within the image are enhanced [6]. The image obtained after

the pre-processing process is suitable for further processing. The sharpening method uses a convolution method for contrast enhancement. The mask for sharpening [11] is given below.

$$\frac{1}{(\alpha + 1)} \begin{bmatrix} -\alpha & \alpha - 1 & -\alpha \\ \alpha - 1 & \alpha + 5 & \alpha - 1 \\ -\alpha & \alpha - 1 & -\alpha \end{bmatrix}$$

Where value of α can vary from 0.0 to 1.0. The default value of α is 0.2. In this work the value of α is 0.2.

3.2 Segmentation

Image segmentation [7] is an important step in image analysis, object representation, visualization and many other image processing tasks [12]. The image segmentation process partitions the image into multiple segments. It is used to locate objects and boundaries within the image. Image segmentation helps medical field for detecting the death threatening diseases such as cancer [4].

There are different segmentation methods available. In this work Marker Controlled Watershed segmentation [5] is used. The main advantage of using Marker controlled watershed segmentation is it removes the over segmentation within the image. This method solves the over segmentation problem by using the markers. [8].

The different steps involved in Marker Controlled Segmentation [13] are the following:

Step 1: Read in the color image and convert it to gray scale image.

Step 2: Compute the Gradient Magnitude as the segmentation function.

Step 3: Mark the foreground objects within the image.

Step 4: Find out the background marker points within the image.

Step 5: Find out the watershed transform of the segmented function of the image.

Step 6: Resultant segmented binary image is obtained.

3.3 Feature Extraction

Feature extraction is an essential step for the analysis of image. This step determines the relevant information for the processing. Thus the normality or abnormality within the lung can be identified. The extracted features are used for detection and staging of the tumor. The different features that are extracting here are Area, Perimeter, Eccentricity, Convex Area and Mean Intensity. The features are explained in detail.

1. Area

Area is the summation of all the pixels within the tumor portion of lung. The area obtained is a scalar value.

2. Perimeter

Perimeter is the summation of the outline of interconnected tumor portion within the lung. The perimeter obtained is a scalar value.

3. Eccentricity

This is used for determining the roundness of the tumor. The value of eccentricity will be equal to 1 for a regular object and it is greater than or less than one for an irregular shape.

4. Convex Area

The convex hull specifies the smallest convex polygon which encloses the tumor portion within the lung. The number of pixels within this convex polygon will give the convex area. It is a scalar value.

5. Mean Intensity

It is a scalar value. It indicates the average intensity of pixels within a particular region of interest, which is the tumor portion.

3.4 Tumor detection and staging

The features extracted in the previous step used for the detection of the cancer. Based on the threshold value for the feature 'area' obtained from experimental analysis, the probable tumorous regions within the lung are detected. From the probable tumorous regions the individual tumor portions are selected by the experts. Depending on the features of the selected tumor portion the stage of the tumor is detected accurately by the classifier. The classifiers used for the study are Support Vector Machine (SVM), Naive Bayes Multinomial classifier (NBM), Naive Bayes Tree (NB tree) and Random tree.

SVM is a supervised learning algorithm for classification [14]. It works with a hyper plane. In SVM, during training, the data values are plotted in the multidimensional plane. Using a hyper plane, the data values are separated into binary classes. Set a margin in such a way that it should include maximum support vectors. When a new data is given for testing the data is plot into the training space and thus the class will be predicted. NBM and NB tree are the variations of Naive Bayes classifier [15]. This classifier uses a probabilistic model for classification. During training, the probability of occurrence of each data value in a particular class is calculated. When a new data set is given for testing the probability is calculated. The minimum probability difference class is selected and thus the class is predicted.

Random tree [16] is a classifier used for a stochastic process. Random tree includes a collection of decision tree. The root node for the random tree is selected randomly. During training, random instances are selected from the datasets. With the selected instances the decision trees are created. These decision trees belong to different classes. During testing, classifier takes the input feature vector. This vector is taken by every tree in the forest and thus the classification process is done. The output will be class label of majority votes.

4. RESULTS AND DISCUSSION

The proposed computer aided lung cancer detection system is trained with 200 slices of CT image of JPEG format. The dataset is obtained from the Regional Cancer centre (RCC). The trained system is tested by the experts of the institution. The proposed system can detect and stage the tumor. The figure 3 shows the resultant CT image obtained after tumor detection step.

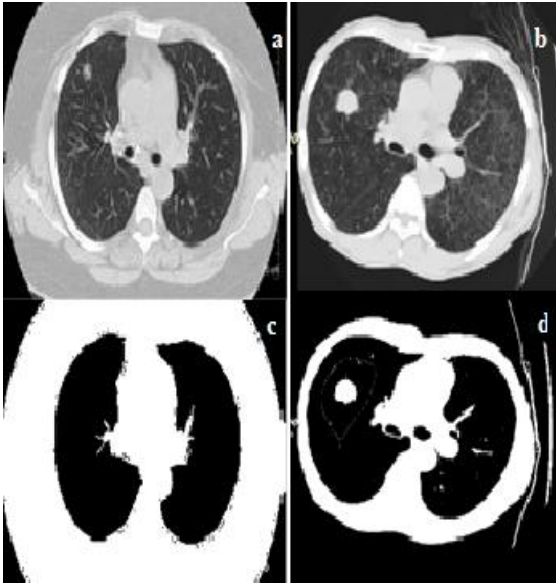


Fig 3: Resultant images obtained after tumor detection stage.

In figure 3, figure (a) is the CT image of healthy person and figure (c) is the segmented image binary image of healthy person. Since the input is the CT image of healthy person, there is no tumor is detected within the lung. The lung portion within the segmented binary image is black in color. The figure (b) is the CT image lung cancer affected person and figure (d) is the segmented binary image of lung cancer affected person. Since the input is the CT image of lung cancer affected person, there is white tumor portion within the lung. Thus tumor is detected within the lung. If the tumor is detected, proper staging of tumor is to be done.

The figure 4 shows the resultant image obtained after each processing steps of T1 stage, T2 stage and T3 stage. The T4 stage of the cancer has not been considered for our study since rest of the lung portion is infected by the tumor. T1 stage is the initial stage of cancer. It is the localised cancer. If it is diagnosed correctly and early, it can be treated by the doctor. Thus the life of the patient can be saved. T2 stage is the second stage of cancer. It results in the spreading of the cancer. It is a complicated stage than the T1 stage. If it is not diagnosed as correct stage, it cannot be treated by the doctor. This will cause the death of patients. It can be treated by the doctor with proper medical support. T3 stage is a complicated stage than the other two stages of cancer. In this stage, the tumor will spread to the nearby organs of lung. So it is very difficult to cure this stage of cancer than the other two stages

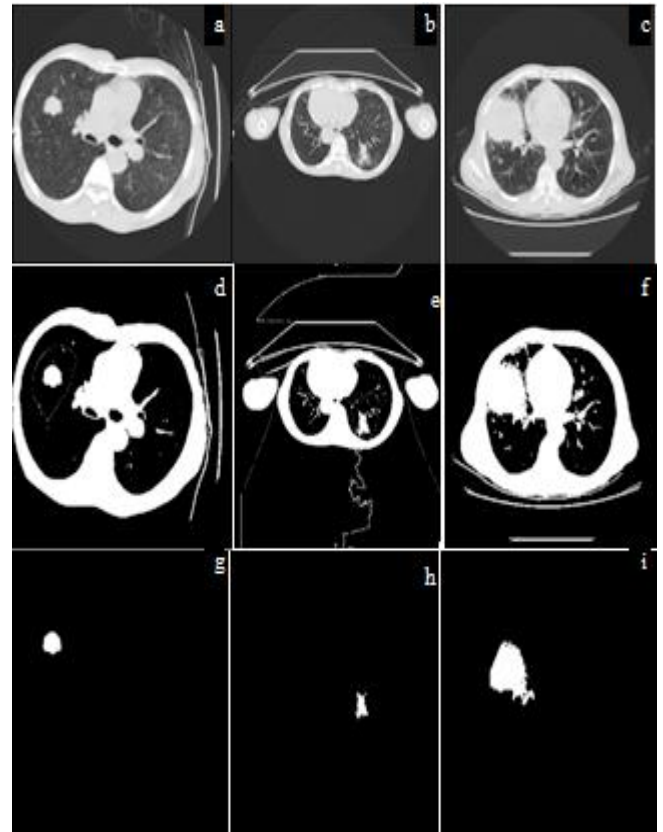


Fig 4: The resultant images of T1, T2 and T3 stages of cancer obtained after each processing steps.

In figure 4, figure (a) is the input CT image of lung of T1 stage, figure (b) is the input CT image of lung of T2 stage, figure (c) is the input CT image of lung of T3 stage, figure (d) is the segmented binary CT image of T1 stage, figure (e) is the segmented binary CT image of T2 stage, figure (f) is the segmented binary CT image of T3 stage, figure (g) is the image showing T1 stage of tumor portion, figure (h) is the image showing T2 stage of tumor portion, and figure (i) is the image showing T3 stage of tumor portion. Since the proposed system is used for the detection and staging of death threatening disease like lung cancer, accuracy is the performance measure used for deciding the classifier's performance.

Accuracy [17] is the ratio of the correctly classified cases within the dataset.

$$\text{Accuracy} = (\text{TP} + \text{TN}) / (\text{TP} + \text{TN} + \text{FP} + \text{FN}) \quad (1)$$

Where TP denotes True Positive, which gives correct instances which are classified correctly. TN denotes True Negative, which gives correct instances which are classified incorrectly. FP denotes False Positive, which gives wrong instances which are classified correctly. FN denotes False Negative, which gives wrong instances which are classified incorrectly.

The different classifiers used in the study are Support Vector Machine (SVM), Naive Bayes Multinomial (NBM), Naive Bayes tree (NBtree) and random tree. The accuracy of these classifiers are compared. The random tree classifier

out performs other classifiers in terms of accuracy. The table 1 shows the accuracy obtained from the different classifiers used for staging of cancer.

Table 1. Accuracy of different classifiers for the staging of tumor

Classifier	Accuracy
SVM	61.11%
NBM	75%
NB Tree	91.67%
Random Tree	94.4%

The accuracy rate of random tree classifier is 94.4%. So random tree classifier is used for the staging of cancer.

The figure 5 gives the bar graph, showing accuracy obtained from the different classifiers used for staging of cancer.

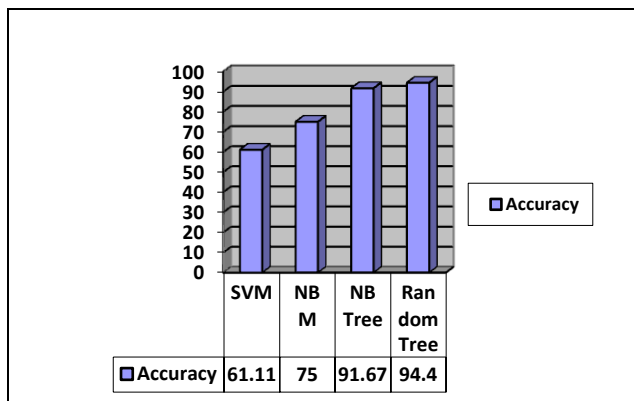


Fig 5: The graph showing the accuracy rate of different classifiers.

From the above graph it is clear that Random tree classifier is giving more accuracy than other classifiers. So for the staging of tumor the random classifier is used. Thus the proposed computer aided lung cancer detection and tumor staging promises accurate result.

5. FUTURE ENHANCEMENT

The proposed system cannot automatically detect the tumor along lung border. This can be improved by applying good segmentation technique. The accuracy can be increased by extracting more features of the tumor, increasing the size of the dataset.

6. CONCLUSION

The proposed lung cancer detection and tumor staging system identifies tumor and stages the tumor within the lung. The CT image is pre-processed and the pre-processed image is then subjected to segmentation by using Marker Controlled watershed segmentation. Segmented image is used for feature extraction. With the extracted features the tumor is detected within the lung. The random tree classifier is used for the staging of the cancer. The accuracy rate of the proposed system is 94.4%. Thus this system helps the radiologist to identify the stage of the tumor and increase the accuracy.

7. ACKNOWLEDGMENT

We express our sincere thanks to the Imageology department of Regional Cancer Centre (RCC) Trivandrum, India for giving the valuable data for our work. We also express our

sincere gratitude to Dr. Anil Prahladan who has helped us in the developing and testing phases of the work.

8. REFERENCES

- [1] Anam Tariq, M. Usman Akra and M. Younus JavedS, "Lung Nodule Detection in CT image using Neuro Fuzzy classifier", IEEE Transactions (2013).
- [2] R. Prasad, "Cancer killed 5.56 lakh in India in 2010", The Hindu News paper, March 30, 2012.
- [3] <http://www.cancer.gov/about-cancer/diagnosis-staging/staging/staging-fact-sheet>
- [4] Sruthi Ignatious, Robin Joseph, "Computer Aided Lung cancer Detection System", IEEE, Global Conference on communication and Technologies, (2015)
- [5] Anitha Chaudary, Sonit.S.Singh, "Lung cancer Detection on CT images by using image processing", IEEE, International Conference on Computer Sciences, (2012).
- [6] Henri Kivinen, "Automatic Image Enhancement Methods for reader reporters image", WP3 hyper local content d3.2.2.1.
- [7] Rafael C., Gonzalez & Woods R.E., "Digital Image Processing", Pearson Education.
- [8] Imzad Rizvi, B.K. Mohan, "Wavelet based Marker-Controlled Watershed Segmentation Technique for High Resolution Satellite Images", 2nd International Conference and workshop on Emerging Trends in Technology (ICWET) 2011
- [9] S.Shaik Parveen, Dr.C.Kavitha, "Review on Computer Aided Detection and Diagnosis of lung cancer nodules", International Journal of Computers & Technology, Volume 3 No. 3, Nov-Dec (2012).
- [10] D.S. Elizabeth, H.K. Nehemiah, C.S. Retmin Raj, A. Kanna, "Computer Aided diagnosis of lung cancer based on analysis of significant slice of chest computed tomography image", IET Image processing. Published on (2010) and revised on (2011)
- [11] <http://dali.feld.cvut.cz/ucebna/matlab/toolbox/images/fspacial.html>
- [12] <http://www.cs.uu.nl/docs/vakken/ibv/reader/chapter10.pdf>
- [13] <http://in.mathworks.com/help/images/examples/marker-controlled-watershed-segmentation.html>
- [14] <http://www.support-vector-machines.org/>
- [15] R. Kohavi, "Scaling up the accuracy of Naïve-Bayes Classifiers: a Decision-Tree Hybrid", Proceedings of the second International Conference on knowledge Discovery and Data Mining, 1996
- [16] <http://docs.opencv.org/modules/ml/doc/randommlrees.html>
- [17] Fatma Taher, Rachid Sammouda, "Identification of Lung cancer based on Shape and color", IEEE transactions on medical images (2008).