Computer Aided Detection Scheme To Improve The Prognosis Assessment Of Early Stage Lung Cancer Patients

Athira K V[1], Nithin S S[2]

[1]*Student, Dept. Of electronics & communication Engineering, NCERC, Thrissur, India*
[2]*Assistant Professor, Dept. of electronics & communication Engineering, NCERC, Thrissur, India*

Abstract - *To develop a computer aided detection scheme to predict the stage 1 non-small cell lung cancer recurrence risk in lung cancer patients after surgery. By using chest computed tomography images; that taken before surgery, this system automatically segment the tumor that seen on CT images and extract the tumor related morphological and texture-based image features. We trained a Naïve Bayesian network classifier using six image features and an ANN classifier using two genomic biomarkers, these biomarkers are protein expression of the excision repair cross-complementing 1 gene (ERCC1) & a regulatory subunit of ribonucleotide reductase (RRM1) to predict the cancer recurrence risk, respectively. We developed a new approach that has a high potential to assist doctors in more effectively managing first stage NSCLC patients to reduce the cancer recurrence risk.*

Key Words: **computer-aided diagnosis, lung tumor segmentation, image features analysis, fusion of image features and genomic biomarkers, prediction of lung cancer recurrence risk.**

1.INTRODUCTION

LUNG cancer is the second most commonly diagnosed cancer in both men and women. An estimated 222,500 new cases of lung cancer will be diagnosed in 2017, accounting for about 25%of all cancer diagnoses. It is the leading cause of cancer death in both men and women. With an estimate of 155,870 death in 2017, it will account for 1 in 4 cancer deaths. Over 85% of the lung cancers are non-small cell lung cancer (NSCLC). Based on the type, stage, and molecular characteristics of the cancer, treatment can include surgery, radiation therapy, chemotherapy, immunotherapy, and/or targeted therapy. For early stage non-small cell lung cancers, surgery is the usually the treatment of choice and advanced stages of non-small cell lung cancer patients are usually treated with chemotherapy, targeted drugs or Immunotherapy [1].

Lung cancer shows the highest mortality rate especially in men due to the over tobacco usage [2]. Even though early stage cancer detection and appropriate treatment can improve the survival rate of lung cancer patients, lung cancer recurrence rates after surgery of the malignant lung tumor can be from 30% to 60% is reported in the earlier studies [3]. As a result, mortality rate among the stage I NSCLC patients is also much higher than many other types of detected at an early stage. The most recent rates published for the current AJCC staging system are the 5-year survival rate for the people with stage 1A1 NSCLC is about 92%, for people with stage 1A2 NSCLC is about 83%, for people with stage 1A3 NSCLC is about 77% and also the 5-year survival rate for people with stage 1B NSCLC is about 68% [4]. Therefore, in order to more effectively treat and manage the stage I NSCLC patients, it is important to develop an proper prediction model to more accurately predict cancer prognosis after cancer surgery. Since there is no clinical standard for assessing the risk of cancer recurrence after surgery, researchers have tested with different genomic biomarkers to identify the different genomic defects in lung cancer development and prognosis [5]. Among them, two biomarkers, ERCC1 gene and RRM1, have been extensively investigated and reported as prognostic biomarkers of the NSCLC patients [6]–[7]. Many advanced imaging methods such as high-resolution CT, positron emission tomography (PET), PET-CT, and magnetic resonance imaging (MRI) have been used in lung cancer imaging. Among them, CT remains the most popular imaging method due to its higher accuracy, wide accessibility, and cost effectiveness.

In this study, we developed a new quantitative image feature analysis method to predict the risk of lung cancer recurrence of the stage I NSCLC patients after lung cancer surgery. For this purpose, we developed a new computer-aided detection (CAD) scheme to automatically segment lung tumors depicting on CT images acquired before surgery and compute tumor-related morphological and texture-based image features. Using a set of the selected image features, we trained a Naive Bayes classifier to predict the risk of cancer recurrence of the stage I NSCLC patients after surgical treatment. We also trained a classifier to combine two genomic biomarkers (ERCC1 and RRM1).finally applying a fusion method to combine both result from 2 classifiers and predict the cancer recurrence risk.

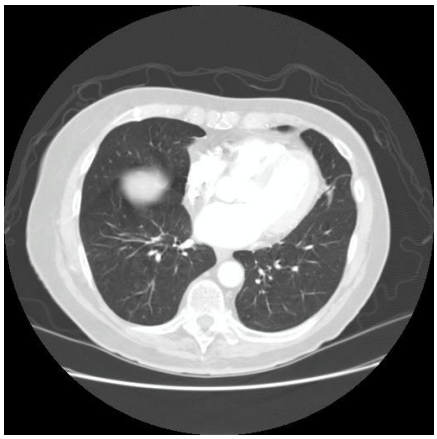

Fig -1: CT image of one patient.

1.1 BLOCK DIAGRAM

Given below shows the block diagram of our new CAD scheme to predict the cancer recurrence risk.

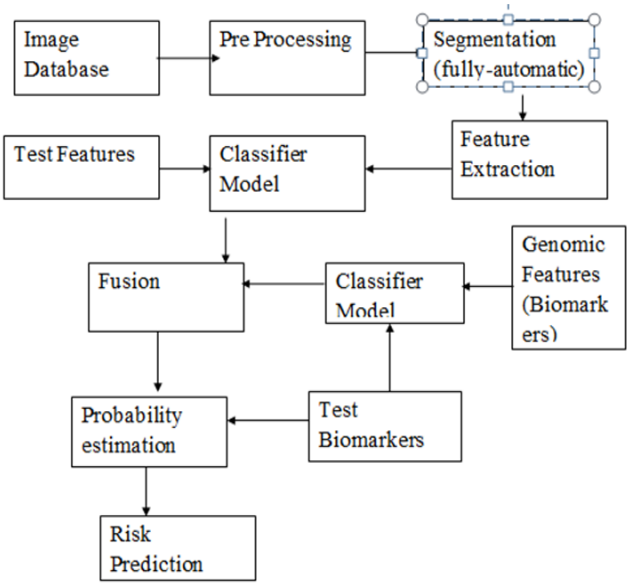

Fig -2: CAD system

1.2 MATERIALS AND METHOD

A) Test Dataset

The Test dataset includes the thoracic CT image examinations of 32 patients; who underwent lung cancer diagnosis and treatment in the hospital. All of these patients were diagnosed with stage I NSCLC. These Computed tomography images were collected from The Cancer Imaging Archive (TCIA) site. Based on the current clinical guideline, a lung surgery was performed on each patient to remove the malignant tumors. After surgery, the tumor specimens were extracted. Two genomic biomarkers, ERCC1 and RRM1, were evaluated from the selected tumor specimen and a standard IHC-based analytic method in the pathology laboratory of the hospital.

B) Lung Tumor Segmentation

In the previous studies used a semi-automatic computer-aided detection scheme to segment lung tumors. In that scheme, tumor center and diameter have been marked previously by a radiologist in the original clinical CT image reading were used as the initial segmentation Our new method uses fully automatic segmentation of malignant tumor. For this purpose it uses K-Means clustering algorithm and segment the tumor in the CT images. The main advantages of K-means clustering algorithm are ease of implementation and high speed performance and also measurable and efficient in large data collection. The next paragraph gives the detailed procedure of doing the lung tumor segmentation.

Here first we gave the image input and then converted the image in to labspace for better human vision, after that does the k-means clustering to make 2 clusters. Where, one cluster includes the lung portions and the second cluster includes chest wall, vascular tissues and nodule. here we have to select the one cluster based on low pixel sum value, that is the cluster having lung portion(fig 3). Then does the morphological operation on selected cluster. And next is the mask creation, which produces mask and border corrected segmented lung (fig 4) and then apply this mask in to images. Again apply the k-means clustering on border corrected segmented lung with 3 clusters and first cluster contain background, second one contain lung and third one contain nodule. Here we select the cluster having nodule and produced another mask and finally apply this mask in to the image.

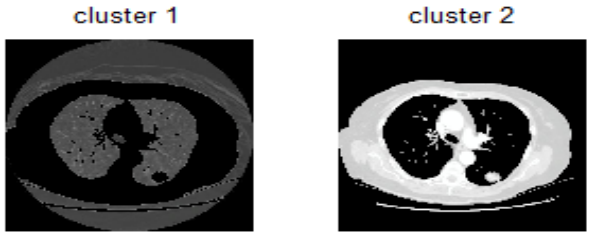

Fig -3: First Produced 2 Clusters

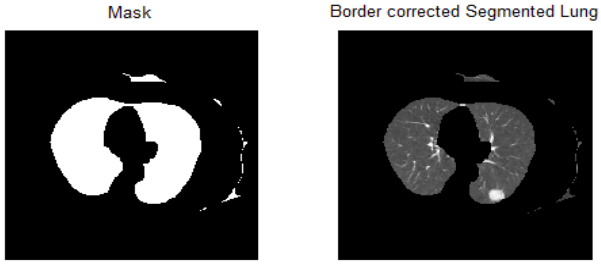

Fig -4: Mask and border corrected segmented lung

C) Image Feature Extraction

Here we extract and compute the six image features from the segmented tumor image. The features are given below,

1) standard deviation of tumor pixel value: It describe the degree of tumor density distribution. Larger this value indicates greater tumor heterogeneity.
2) The maximum tumor diameter: It represents maximum tumor size in the tumor central slice.
3) A tumor shape factor :here first counts the no. Of pixels located in the tumor boundary of all segmented tumor regions on the CT image. The tumor shape factor is computed by multiplying this number to the unit of pixel size of the CT images and dividing it by the tumor volume. Larger value indicates the higher level of irregularity on the tumor boundary surface.

Three texture features were computed from a gray-level run-length (GLRL) matrix of a tumor region based on the run-length statistics.

4) Gray-level nonuniformity(GLN): Measures the similarity of gray level values throughout the image. The GLN is expected small if the gray level values are alike throughout the image.
5) Short-run low gray-level emphasis (SRLGE): Measures the joint distribution of short runs and low gray level values. The SRLGE is expected large for the image with many short runs and lower gray level values.
6) Long-run low gray-level emphasis (LRLGE): Measures the joint distribution of long runs and low gray level values. The LRLGE is expected large for the image with many long runs and low gray level values

D) Two Classifiers And Data Analysis

This system used two classifiers. The naive Bayesian network based classifier is trained and tested with the dataset in order to combine the six image features on the dataset to predict the first stage NSCLC patients cancer recurrence risk. We can use different types of classifiers for this purpose, but due to the limited dataset, we used this simple classifier. Also used another machine learning classifier called multi-layer perceptron based classifier to combine the two genomic biomarkers such as ERCC1,RRMI for predicting the cancer recurrence risk. Finally we combine the both results from the two classifiers by using simple fusion method, weighted-average to make a final result that is, higher risk or lower risk of cancer recurrence.

2. PERFORMANCE

A receiver operating characteristic (ROC) curve is a way to compare diagnostic tests. It is a plot of the true positive rate against the false positive rate. Test accuracy is also shown as the area under the curve (which you can calculate using integral calculus). The greater the area under the curve, the more accurate the test. A perfect test has an area under the ROC curve (AUROCC) of 1. The diagonal line in a ROC curve represents perfect chance. In other words, a test that follows the diagonal has no better odds of detecting something than a random flip of coin. The area under the diagonal is .5 (half of the area of graph). Therefore, a useless test (one that has no better odds than chance alone) has a AUROCC of .5.

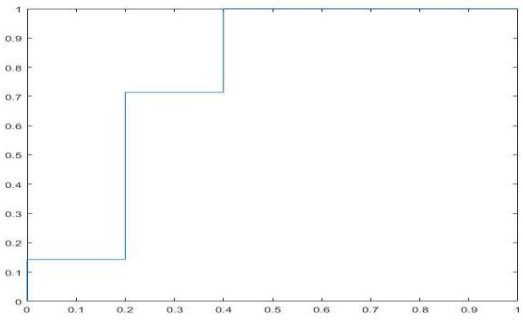

Chart -1: ROC curve

3. CONCLUSIONS

In order to more effectively treat and manage first stage non small cell lung cancer patients, it is important to develop a prediction model to more accurately predict the cancer recurrence risk after surgery. Our system will counter these limitation. This CAD scheme can be a really helpful system for doctors to predict the cancer recurrence after surgical treatment. This scheme provides better performance about 80% accuracy, so it will surely assist the doctors in determining the cancer recurrence in first stage NSCLC after surgery.

REFERENCES

[1] "Cancer facts and figures 2017", American cancer society annual publication, Atlanta, Georgia.

[2] Rebecca L. Siegel, MPH1 ; Kimberly D. Miller, MPH2 ; Ahmedin Jemal, DVM, PhD, "cancer statistics 2017," CA, Cancer J. Clin., vol. 67,Jan./Feb. 2017

[3] S. Singhal et al., "Prognostic implications of cell cycle, apoptosis, and angiogenesis biomarkers in non-small cell lung cancer: A review," Clin.Cancer Res., vol. 11, pp. 3974–3986, 2005.

[4] American Cancer Society. (2017, dec.). Non-Small Cell Lung Cancer Survival Rates by Stage. [Online]. Available:http://www.cancer.org/cancer/lungcancer-nonsmallcell/detailedguide/non- small -cancer-survival-rates

[5] R. Rosell et al., "Screening for epidermal growth factor receptor mutations in lung cancer," N. Engl. J. Med., vol. 361, pp. 958–967, Sep. 2009.

[6] G. Bepler et al., "ERCC1 and RRM1 in the international adjuvant lung trial by automated quantitiative in situ analysis," Amer. J. Pathol., vol. 178, no. 1, pp. 69–78, Jan. 2011.

[7] G. Bepler et al., "RRM1 and PTEN as prognostic parameters for overall and disease-free survival in patients with non-small-cell lung cancer," J. Clin. Oncol., vol. 22, no. 10, pp. 1878–1885, May 2004.