

The Identification of EGFR Mutation Based on the CT Image of Lung Adenocarcinoma

Ting-Chen Tsan, Hao-Hung Tsai, and Chiun-Li Chin

Abstract—Lung adenocarcinoma is the most common cause of lung cancer in Taiwan. However, it is difficult to detect lung adenocarcinoma at early stage. Most radiologists usually decide the treatment strategies by screening whether the patients have the epidermal growth factor receptor (EGFR) gene mutations or not. However, the screen is expensive and time-consuming. Therefore, the aim of the present paper is to identify whether there are EGFR mutations in tumors with CT image. We segment the whole region of lung adenocarcinoma and extract the texture features in spatial domain and frequency domain using GLCM and Gabor Wavelet. Finally we use Neuro-fuzzy based AdaBoost to classify and identify the relationship between the texture features of lung adenocarcinoma CT images and EGFR mutations. The experimental result in this paper shows that our success rate was 92%, and we will improve our methods to increase the recognition rate in the future.

Index Terms—Lung Adenocarcinoma, Epidermal growth factor receptor, Gabor Wavelet, Neuro-fuzzy based AdaBoost

I. INTRODUCTION

Lung cancer is a malignant lung tumor characterized by uncontrolled cell growth in tissues of the lung. If left untreated, this growth can spread beyond the lung by process of metastasis into nearby tissues or other parts of the body. Most cancers that start in the lung, known as primary lung cancers, are derived from epithelial cells. The two main types of lung cancer are small cell lung carcinoma (SCLC) [1] and non-small cell lung carcinoma (NSCLC) [2] which accounts for 80%. NSCLC includes 35–40% of lung adenocarcinoma, 25–30% of squamous cell carcinoma, and 10–15% of large cell carcinoma. According to the statistics in recent years, lung adenocarcinoma patients are increased and most of them are female and people who smoke. Lung adenocarcinoma at early-stage is asymptomatic. More than half of patients find it so late that it is very difficult to treat. If tumors can be detected and removed early, the survival rate of patients has the opportunity to reach more than 40%; but if tumors are discovered too late, the proportion of patients who can survive over one year is less than 50%. There is a mutation related to lung adenocarcinoma called epidermal growth factor receptor (EGFR) mutation. A tyrosine kinase made from EGFR mutation will influence the intracellular signal transduction and let cancer cells continue to grow. EGFR

gene is present in cancer cells and scarcely present in normal cells so that it is regarded as an important target of targeted therapy in recent years. The targeted drug Iressa which is used to treat lung cancer affects directly on intracellular EGFR and makes EGFR lose the ability of proliferation and metastasis. The further research found that Iressa have the best effect on lung adenocarcinoma patients who have EGFR mutation and most of the patients are Asians. 85% of patients are under stable conditions and the average life is prolonged for 9 months. For lung adenocarcinoma patients at late-stage, it is important to screen the EGFR gene. Targeted therapy can improved the survival rate of patients with the positive result of screen substantially but have no therapeutic effect on patients with the negative result. Many medical centers have offered the screen of EGFR gene at present, but it is expensive for lung adenocarcinoma patients to screen the EGFR gene and the patients should wait for the report of the result for a long time. For the current treatment strategies, the patients without EGFR mutation are limited to choose the drugs indeed. Therefore, whether EGFR mutates or not has been an important basis for radiologists to decide the treatment strategy now.

However, whether lung adenocarcinoma patients have to screen the EGFR gene is assessed by radiologists. After observing the CT images, if radiologists determine the probability of EGFR mutation is high, they will suggest the patients screen the EGFR gene; otherwise, radiologists will decide other treatment strategy. After a period of time, if the situation of poor recovery is happen to the patients, radiologists will reappraise whether the patients have EGFR mutation, suggest them screen the EGFR gene and decide the treatment strategy according to the screening results. This diagnostic strategy relies on the diagnostic experience of radiologists. If lung adenocarcinoma patients can receive treatment in the shortest time are closely related to the assessment from radiologists.

Therefore, we hope to identify whether EGFR mutates from the texture feature of lung adenocarcinoma CT image. It can not only enhance the radiologists' diagnostic speed, but reduce the screening cost and time for the screen of EGFR gene.

The rest of this paper has five sections which are described as follow: we will discuss the related literature in section II, and describe the proposed flow and method in section III. Section IV is the experimental results and Section V is discussions and conclusions.

II. RELATED WORKS

In recent years, about 9,000 lung cancer patients are

Manuscript received May 27, 2016; revised September 13, 2016.

Ting-Chen Tsan and Chiun-Li Chin are with the Department of Medical Informatics, Chung Shan Medical University, Taichung, Taiwan (e-mail: tusbasa1@gmail.com, ernstli@csmu.edu.tw).

Hao-Hung Tsai is with the Department of Medical Imaging, Chung Shan Medical University Hospital, Taichung, Taiwan (e-mail: hao620620@gmail.com).

increased in Taiwan per year with more than 5,000 lung adenocarcinoma patients. More than 3,000 cases of lung adenocarcinoma are discovered at late-stage. In order to treat early and effectively for patients, it is important to screen the EGFR gene. If we can establish a system to help radiologists judge whether the patients have EGFR mutation, it can not only reduce the screening time but enhance the accuracy of the diagnostic results. In order to discuss the relation between lung adenocarcinoma CT images and EGFR mutation, we will focus on feature extraction and classification in this paper, and discuss the related literature.

In 2015, the scholars, Nooshin Nabizadeh and Miroslav Kubat, published a paper about brain tumor detection and segmentation based on MR images [3]. They compared the capability and efficacy of two different feature sets – Gabor wavelets and statistical features – in automated segmentation of brain tumor lesions in MRI images. The set of statistical features were extracted using different texture-based feature extraction techniques: first-order statistical descriptors, GLCM, GLRLM, HOG, and LBP methods. After feature aggregation and feature dimensionality reduction relied on principal component analysis (PCA), two sets of features were used to train the classifier, such as SVM, KNN, SRC, NSC, and k-means clustering. The comparison results indicated that statistical features usually offered higher accuracy than Gabor wavelet features. Moreover, statistical features which had much smaller dimensionality than Gabor wavelet-based feature could occupy less amount of memory and avoided highly redundant and high computational costs.

In 2015, the scholars published a paper about mammogram image visual enhancement, mass segmentation and classification [4]. First, the noise was removed and the breast region was segmented using contour segmentation. Second, the enhancement methods – CLAHE, median filtering and Gaussian smoothing – were applied to enhance the mammogram images. Third, the boundary of the region of interest (ROI) was traced using the boundary extraction morphological algorithm. Finally, each segmented mass was classified into one of the following four classes: benign, probable benign and possible malignant, probable malignant and possible benign, or malignant by calculating the object roundness and the specific threshold. The results indicated that combining CLAHE and Median filtering methods gave the best results amongst the rest of the proposed hybrid methods. However, the classification method which only used the roundness of mass to analyze benign and malignant is too simple and the accuracy is questionable.

In 2014, the scholars, Yoganand Balagurunathan and other scholars, published a paper about reproducibility and prognosis of quantitative features extracted from CT images [5]. The tumors were segmented using both manual and ensemble methods. A set of features (219 three-dimensional and 110 two-dimensional) was computed, and quantitative image features were statistically filtered to identify a subset of reproducible and non-redundant features. Then extract the features from the image and divided into the Non-Texture and Texture. For tumor size, shape or position belong to the Non-Texture features. On the other hand, if it is the Texture features, using the histogram, run-length, co-occurrence, Laws kernel and the method of wavelets to classify. After

feature extraction, the project use the representative features to predict the prognostic score. The advantages of this study are to extract the sundry features. What's more, it can improve the efficiency and the accuracy of identification through finding the representative features.

In 2013, the scholars published a paper about region-based NPC lesion segmentation from MRI using clustering and classification-based methods with learning [6]. Computer-aided tumor segmentation methods (either contour-based or region-based) are necessary to alleviate clinicians' workload. For contour-based methods, a minimal user interaction to draw an initial contour inside or outside the tumor lesion for further curve evolution to match the tumor boundary is preferred, but parameters within most of these methods require manual adjustment, which is technically burdensome for clinicians without specific knowledge. Therefore, segmentation methods with a minimal user interaction as well as automatic parameters adjustment are often favored in clinical practice. In this paper, two region-based methods with parameters learning are introduced for NPC segmentation. It can improve the inconvenience of the traditional segmentation methods. But this study only discussed the image segmentation, it did not research the subsequent feature extraction and classification.

According to the aforementioned, we found that most literatures proposed the single feature for feature extraction to describe the image and it could cause the inaccuracy identified result of system. In addition, the classify algorithms they proposed have the problem of innovation. Therefore, we present an improved system to analyze the relation between lung adenocarcinoma CT images and patients with EGFR mutation.

III. MATERIALS AND METHODS

In this section, we will introduce the algorithms and methods of the proposed system. The functions and operations of each part are described below in detailed. The processing flowchart is shown in Fig. 1.

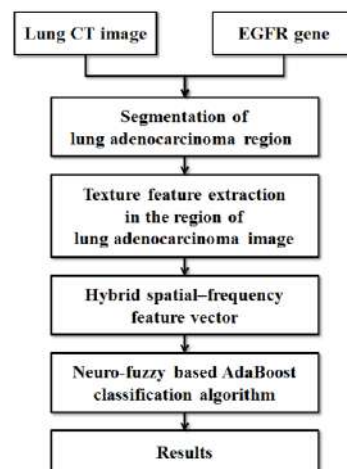


Fig. 1. The processing flowchart of the proposed system.

A. Segmentation of Lung Adenocarcinoma Region

Because lung adenocarcinoma is possible to spread to other tissues and organs like lymph and liver, we develop a

semiautomatic segmentation algorithm to segment the region of lung adenocarcinoma. The processing flowchart is shown in Fig. 2.

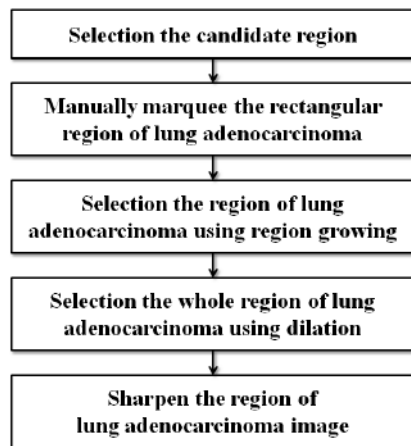


Fig. 2. The processing flowchart of the proposed system.

1) Selection the candidate region

To speed up the interpretation from radiologists, we will remove the region without lung adenocarcinoma in this step. Fig. 3(a) shows the original CT image. Through observing the histogram of the image shown in Fig. 3(b), we found that the distribution of gray level is divided into two groups obviously. Therefore, we get the threshold using Otsu's method and adjust the pixels of which the gray level is less than the threshold to remove the regions which are not the parts of lung adenocarcinoma. It can avoid disturbing the judgments of system and enhance the accuracy of segmentation. The result is shown in Fig. 3(c).

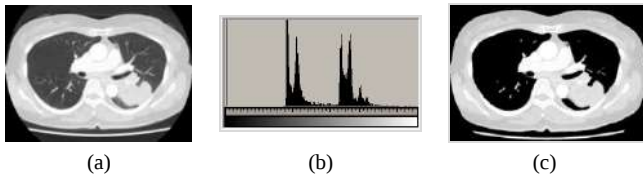


Fig. 3. (a) The original CT image, (b) the histogram of the CT image, (c) the candidate region of the CT image.

2) Manually marquee the rectangular region of lung adenocarcinoma

After the processing of last step, we will let radiologists marquee the largest external rectangle of lung adenocarcinoma manually and segment this region. The result is shown in Fig. 4(a) and Fig. 4(b).

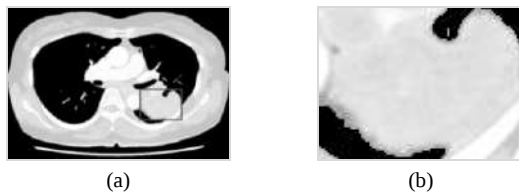


Fig. 4. (a) Manually marquee the largest external rectangle of lung adenocarcinoma, (b) the rectangular region of lung adenocarcinoma.

3) Selection the region of lung adenocarcinoma using region growing

In order to segment the region of lung adenocarcinoma

accurately, after observing the histogram of the rectangular region shown in Fig. 5(a), we found that the distribution of gray level is crowded in a range. Therefore, we apply the region-based image segmentation method – Region Growing. We set the largest number of gray level, the peak of the histogram, as the initial seed point p and allowable range α to select the gray level of pixels which conform $p \pm \alpha$. The result is shown in Fig. 5(b).

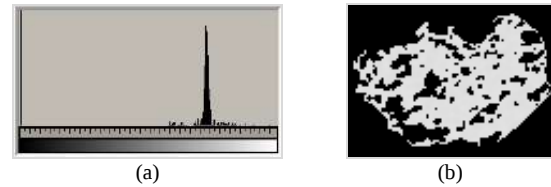


Fig. 5. (a) The histogram of the rectangular region of lung adenocarcinoma, (b) The selected region of lung adenocarcinoma based on the peak as the initial seed point.

4) Selection the whole region of lung adenocarcinoma using dilation

The region of lung adenocarcinoma gotten from last step has fragmentary and incomplete block, so we fill the holes using dilation and get the whole region of lung adenocarcinoma. The result is shown in Fig. 6.



Fig. 6. The region of lung adenocarcinoma image.

5) Sharpen the region of lung adenocarcinoma image

Finally, we sharpen the region of lung adenocarcinoma image to make the texture feature more obvious and be conducive to extracting and recognizing the texture feature in next step. The result is shown in Fig. 7.



Fig. 7. The region of lung adenocarcinoma image after sharpening.

B. Texture Feature Extraction in the Region of Lung Adenocarcinoma Image

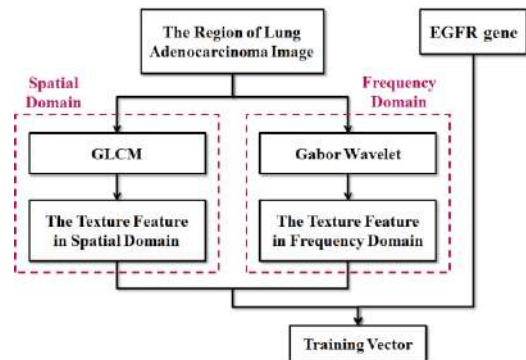


Fig. 8. The processing flowchart of texture feature extraction in the region of lung adenocarcinoma image.

After sharpening the region of lung adenocarcinoma image, we will extract the texture features in spatial domain by GLCM [7]–[9] and in frequency domain by Gabor Wavelet [10]–[13], and then fuse the feature to a feature vector. The processing flowchart is shown in Fig. 8. We will record the feature vectors we get as the basis of feature fusion to classify and identify the EGFR mutation.

1) The texture feature in spatial domain

In spatial domain, we use Gray Level Co-occurrence Matrix (GLCM) to extract the texture feature. There is a relationship of gray level between two pixels separated by a certain distance in image space. GLCM is used to describe the common characteristics texture by spatial correlation characteristics of gray level. The quantitative methods of GLCM are the following five features: Contrast, Corrglation, Energy, Homogeneity, and Entropy. They are calculated as (1) to (5):

$$\text{Contrast: } \sum_{i,j} |i - j|^2 p(i, j) \quad (1)$$

$$\text{Corrglation: } \sum_{i,j} \frac{(i - \mu_i)(j - \mu_j)p(i, j)}{\sigma_i \sigma_j} \quad (2)$$

$$\text{Energy: } \sum_{i,j} p(i, j)^2 \quad (3)$$

$$\text{Homogeneity: } \sum_{i,j} \frac{p(i, j)}{1 + |i - j|} \quad (4)$$

$$\text{Entropy: } \sum_{i,j} p(i, j) \log p(i, j) \quad (5)$$

where $p(i, j)$ is the Gray Level Co-occurrence Matrix, μ_i is the average of i , and μ_j is the average of j .

2) The texture feature in frequency domain

In frequency domain, we use Gabor Wavelet to extract the texture feature. Gabor Wavelet can capture the local structure of the image corresponding to spatial frequency, spatial localization and orientation selectivity, so it is usually applied to texture analysis and image segmentation [3]. In the spatial domain, a two-dimensional Gabor filter is a Gaussian kernel function modulated by a complex sinusoidal plane wave, defined as (6):

$$\varphi_{\omega, \theta}(x, y) = \frac{\omega^2}{\pi \eta} \exp\left(-\frac{x'^2 + \gamma^2 y'^2}{2\sigma^2}\right) \exp(j2\pi \omega x') \quad (6)$$

where x' and y' are defined as (7) and (8):

$$x' = x \cos \theta + y \sin \theta \quad (7)$$

$$y' = -x \sin \theta + y \cos \theta \quad (8)$$

where $\varphi_{\omega, \theta}(x, y)$ is a Gabor filter, x and y are the coordinate position in the spatial domain, ω is the frequency of the sinusoid, θ is the orientation, σ is the standard deviation, and γ is the spatial aspect ratio.

The new images are obtained after Gabor Wavelet transform by convolving an image $f(x, y)$ with $\varphi_{\omega, \theta}(x, y)$ as (21), and we can calculate the following four features:

Mean, SD, Skewness, and Kurtosis. Formulae for these features are listed as (9) to (13):

$$G_{\omega, \theta}(x, y) = f(x, y) * \varphi_{\omega, \theta}(x, y) \quad (9)$$

$$\text{Mean: } \mu = \frac{1}{XY} \sum_{x=1}^X \sum_{y=1}^Y G_{\omega, \theta}(x, y) \quad (10)$$

$$\text{SD: } \sigma = \sqrt{\frac{1}{XY} \sum_{x=1}^X \sum_{y=1}^Y \|G_{\omega, \theta}(x, y) - \mu\|^2} \quad (11)$$

$$\text{Skewness: } \frac{1}{XY} \sum_{x=1}^X \sum_{y=1}^Y \frac{[G_{\omega, \theta}(x, y) - \mu]^3}{\sigma^3} \quad (12)$$

$$\text{Kurtosis: } \frac{1}{XY} \sum_{x=1}^X \sum_{y=1}^Y \frac{[G_{\omega, \theta}(x, y) - \mu(x, y)]^4}{\sigma^4} - 3 \quad (13)$$

where X is the image width, Y is the image height, μ is the mean, and σ is the standard deviation.

C. Hybrid Spatial–Frequency Feature Vector

After extracting the texture feature in the region of lung adenocarcinoma image, we will fuse the features, compose to a feature vector representing the image, and let the Neuro-fuzzy based AdaBoost train and classify the data. The features and the feature vector we get are listed as follows:

- f_1 = Contrast of GLCM
- f_2 = Corrglation of GLCM
- f_3 = Energy of GLCM
- f_4 = Homogeneity of GLCM
- f_5 = Entropy of GLCM
- f_6 = Mean of GW
- f_7 = SD of GW
- f_8 = Skewness of GW
- f_9 = Kurtosis of GW
- $F = [f_1, f_2, f_3, f_4, f_5, f_6, f_7, f_8, f_9]$

After fusing the texture features, we can obtain a nine-dimensional feature vector. Next, we will train and test the feature vector using Neuro-fuzzy based AdaBoost classification algorithm to ensure the feature vector can identify the EGFR mutation effectively.

D. Neuro-fuzzy Based AdaBoost Classification Algorithm

After we get the feature vector, we use Neuro-fuzzy based AdaBoost [14], [15] classification algorithm as our classifier to analyze whether EGFR mutates or not. Fuzzy theory is based on fuzzy set. The basic spirit is to accept the fact that vague phenomenon exists, and the concept of fuzzy uncertainty in transaction processing is regarded as the research goals. Neuro-fuzzy based AdaBoost is a supervised learning algorithm. The input is the feature vector and output is the EGFR gene. The algorithm is shown in Fig. 9.

The algorithm uses SONFIN as the weak learner in AdaBoost algorithm. SONFIN for neural fuzzy network is to architect the fuzzy inference system on the neural network. It makes itself have the thinking of human, and has the self-learning and the ability of organization to make up for lack of the processing capability of the weak classifier which Adaboost chooses at the same time.

Input:

For A Dataset

- Training Data " TD_j " of size "M" with correct labels
- Adaptive Neural-Fuzzy based AdaBoost as Supervised algorithm base classifier
- Number of iterations or classifiers(T)

Feature combination:

- $F_{es}=M$ //Number of feature
- $F_{sc}=2.3.4...M$ //Number of feature combination
- $P_i = C_2^M + C_3^M + ...C_M^M$ //Total

Initialize:

- $L=1...C$ //Number of Classes
- $i=1...N$ //Number of Samples
- $d(i) = \frac{1}{N}, i=1, ..., N$ //Weight of Sample
- $\mu=0.5$ //False Alarm Threshold
- $y_i \in Y = \{1, ..., C\}$ // y_i is known classification results

Training:

For $L = 1:C$

For $i = 1:N$

if $y_i = L$

$\begin{cases} \text{then } y_i^L = 1 \\ \text{else } y_i^L = -1 \end{cases}$

end

For $t=1, ..., T$

For $f=1, ..., P_i$

a. Train Data f by SONFIN obtain result

b. Compute error of h_i^f :

$$\varepsilon_i^f = \sum_{i=1}^n [d_i^L(i) \times (h_i^f(x_i) \neq y_i^L)]$$

End

c. The best feature selection's model:

$$h_i^L = \arg \min(\varepsilon_i^f)$$

Assign weight to the Classifier:

$$\alpha_i^L = 0.5 \times \log\left(\frac{1 - \varepsilon_i^L}{\varepsilon_i^L}\right)$$

d. Update Distribution of Instances:

for $i = 1:N$

if $h_i^L(x_i) \neq y_i^L$

$\begin{cases} \text{then } d_{i-1}^L(i) = d_i^L(i) \times \alpha_i^L \\ \text{else } d_{i-1}^L(i) = d_i^L(i) \end{cases}$

end

$$d_{i-1}^L(i) = \frac{d_i^L(i)}{\sum_{i=1}^n d_{i-1}^L(i)}$$

End

$$H_L = \arg \max\left(\sum_{i=1}^T \alpha_i^L \times h_i^L(x_i)\right)$$

End

e. Obtain Composite Hypothesis:

$$H = \arg \max(H_L)$$

Fig. 9. Neuro-fuzzy based AdaBoost algorithm.

First, it builds the sample weights initialization, and then comes to the stage for the run AdaBoost. In AdaBoost algorithm iterations which made weak classifier is produced by fuzzy set out fuzzy inference. The methods of reference are as (14) to (15):

$$H_L = \arg \max\left\{\sum_{t=1}^T \alpha_t^L h_t^L(x_i)\right\} \quad (14)$$

$$H(x) = \arg \max(H_L) \quad (15)$$

where $H(x)$ is the final draw of the strong classifier, T is the total number of iterations, $\alpha(t)$ iteration to t followed by $h_t(x)$ of weak classifiers weights, and $h_t(x)$ is SONFIN to the t iteration and calculated as (16):

$$h_t(x) = \arg \min(\varepsilon_t^f) \quad (16)$$

where ε_t^f is the error. It is regarded as weak learner. We will select the representative weak learner which has minima error value.

The SONFIN is a general connectionist model of a fuzzy logic system. In addition to the uncertainty and imprecision of processing power, it is also able to find its optimal structure and parameters automatically.

IV. EXPERIMENTAL RESULTS

A. Image Dataset

To validate the proposed method, we relied on the Chung Shan Medical University Hospital dataset during January 2014 to December 2014. This dataset has 34 lung adenocarcinoma patients, including 29 women and 5 men, and there are 19 patients who have EGFR mutations and 15 patients without EGFR mutations.

The images used in this paper are CT images in DICOM format with the size of 512×512 and the slice thickness of 5.0mm. We observed the CT images using lung window (window width 1500 and window level -500) and adopted the images with radiocontrast agents, including various types of lung adenocarcinoma CT images, as shown in Fig. 10.

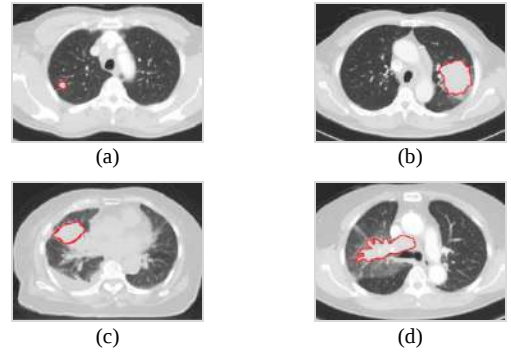


Fig. 10. (a) Under 3cm lung adenocarcinoma CT images (without EGFR mutations), (b) more than 3cm lung adenocarcinoma CT images (without EGFR mutations), (c) spread to the front of the liver CT imaging of lung adenocarcinoma (with EGFR mutations), (d) spread to the lymph tissue CT images of lung adenocarcinoma (with EGFR mutations).

B. The Segmentation Results of Lung Adenocarcinoma Images

Fig. 11 shows the region of lung adenocarcinoma images after segmentation and sharpening. Observing the images we can find that the sharp and size of lung adenocarcinoma images are irregular so that it is difficult to identify whether EGFR mutates or not by extracting the geometric features. Therefore, we proposed the method of extracting the texture features in spatial domain and frequency domain to achieve our goal.

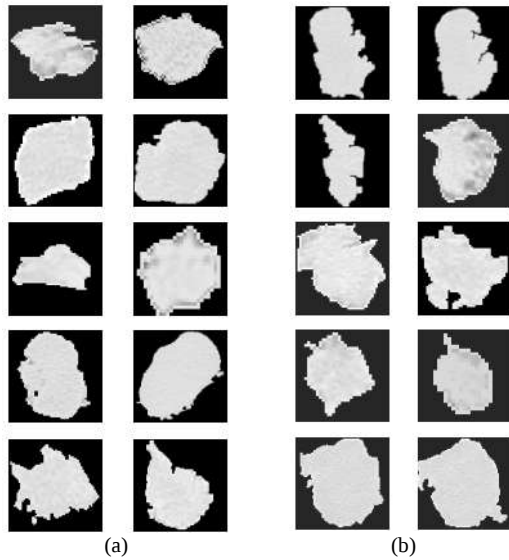


Fig. 11. Examples of the segmentation images of lung adenocarcinoma. (a) The images with EGFR mutations, (b) the images without EGFR mutations.

C. The Results of Classification

The window size of extracting GLCM features and Gabor wavelet features is 12×12 . Besides, referring to the system offered by [16] and created on the MATLAB application, we used different Gabor wavelet filters in five scales and four orientations (0° , 45° , 90° , 135°) to extract the frequency texture features. We compared the success rate of classification if the Gabor wavelet features are fused with GLCM features or not. The results are shown in Table I and Table II.

TABLE I: THE SUCCESS RATE OF CLASSIFICATION BY ONLY GW FEATURES IN DIFFERENT GABOR WAVELET FILTERS

scale orientation	2	3	4	5	6
0°	80%	86%	82%	76%	68%
45°	76%	74%	78%	78%	70%
90°	80%	82%	74%	72%	66%
135°	86%	82%	72%	66%	62%

TABLE II: THE SUCCESS RATE OF CLASSIFICATION BY FUSING GLCM AND GW FEATURES IN DIFFERENT GABOR WAVELET FILTERS

scale orientation	2	3	4	5	6
0°	88%	90%	90%	90%	90%
45°	90%	88%	90%	88%	90%
90°	90%	88%	88%	90%	92%
135°	88%	90%	90%	90%	90%

V. DISCUSSIONS AND CONCLUSIONS

Lung adenocarcinoma in non-smokers is more common than many people realize. Two-thirds of the non-smokers who get lung adenocarcinoma are women. This percentage is significantly higher in Asian women. For cancer patients, chemotherapy is necessary, but it will also kill the good and bad cells, often resulting in patient physical weakness, and even infection. Therefore medical experts developed the targeted therapy which is growing in importance to reduce side effects and prolong the life of terminal patients at late-stage. For the purpose of identifying the relationship

between CT image and EGFR mutations, we hope to reduce the cost and time spent on waiting, help doctors to make a quick and accurate diagnosis, and prescribe immediate treatment to patients with drug targets.

In the part of segmentation, we want to select the complete region of lung adenocarcinoma, but some carcinoma cells have spread to other tissues and organs, like the front of the liver or the lymph tissue, so that it is difficult to define the boundary of the lung adenocarcinoma clearly. We will be as far as possible to complete the segmentation by asking the experienced radiologists or improving the method.

The experimental result paper showed that our highest success rate was 92%. According to the Table I, we could find that if the scale is lower, the success rate is higher and the difference of the success rates in five scales is more obvious. Otherwise, we could find that the success rates of classification are increase obviously by fusing GLCM and Gabor wavelet features. It means that it is effective to identify EGFR mutations by combining the features in spatial and frequency domain. However, because of the nine-dimensional feature vectors, it spent much time running the classification. Besides, the difference of the success rates in five scales and four orientations is lower so that it is unclear to know which factor is better to identify EGFR mutations or make no difference. Therefore, we will work on selecting the representative features from these nine features to reduce the time complexity and getting more lung adenocarcinoma images to let the recognition more accurate.

REFERENCES

- [1] Y. Tsutani *et al.*, "Sublobar resection for lung adenocarcinoma meeting node-negative criteria on preoperative imaging," *The Annals of Thoracic Surgery*, vol. 97, pp. 1701-1707, Feb. 2014.
- [2] A. Haro, T. Yano, M. Kohno, T. Yoshida, T. Okamoto, and Y. Maehara, "Ground-glass opacity lesions on computed tomography during postoperative surveillance for primary non-small cell lung cancer," *Lung Cancer*, vol. 76, pp. 56-60, Apr. 2012.
- [3] N. Nabizadeh and M. Kubat, "Brain tumors detection and segmentation in MR images: Gabor wavelet vs. statistical features," *Computers and Electrical Engineering*, vol. 45, pp. 286-301, July 2015.
- [4] N. A. Najdawi, M. Biltawi, and S. Tedmori, "Mammogram image visual enhancement, mass segmentation and classification," *Applied Soft Computing*, vol. 35, pp. 175-185, Oct. 2015.
- [5] Y. Balagurunathan *et al.*, "Reproducibility and prognosis of quantitative features extracted from CT images," *Translational Oncology*, vol. 7, pp. 72-87, Feb. 2014.
- [6] W. Huang, K. L. Chan, and J. Zhou, "Region-based nasopharyngeal carcinoma lesion segmentation from MRI using clustering and classification-based methods with learning," *Journal of Digital Imaging*, vol. 26, pp. 472-482, Aug. 2013.
- [7] Q. Wu, Y. Gan, B. Lin, Q. Zhang, and H. Chang, "An active contour model based on fused texture features for image segmentation," *Neurocomputing*, vol. 151, pp. 1133-1141, Mar. 2015.
- [8] X. Ou, W. Pan, and P. Xiao, "In vivo skin capacitive imaging analysis by using grey level co-occurrence matrix (GLCM)," *International Journal of Pharmaceutics*, vol. 460, pp. 28-32, Jan. 2014.
- [9] G.-M. Xian, "An identification method of malignant and benign liver tumors from ultrasonography based on GLCM texture features and fuzzy SVM," *Expert Systems with Applications*, vol. 37, pp. 6737-6741, Oct. 2010.
- [10] N. Kumaran and R. Bhavani, "MRI Image Retrieval Using Gabor Wavelet Based Texture Features," *International Journal of Advanced Research in Science and Technology*, vol. 2, pp. 40-45, Aug. 2013.
- [11] N. Nabizadeh *et al.*, "Efficacy of Gabor-Wavelet versus statistical features for brain tumor classification in MRI: A comparative study," in *Proc. IP, Computer Vision, and Pattern Recognition: The 2013 WorldComp International Conference*, 2013.

- [12] X. Wang, L. Lei, and M. Wang, "Palmpoint verification based on 2D – Gabor wavelet and pulse-coupled neural network," *Knowledge-Based Systems*, vol. 27, pp. 451-455, Mar. 2012.
- [13] L. Shena, L. Baia, and M. Fairhurst, "Gabor wavelets and general discriminant analysis for face identification and verification," *Image and Vision Computing*, vol. 25, pp. 553-563, May 2007.
- [14] S. Z. Li, B. Yu, W. Wu, S. Z. Su, and R. R. Ji, "Feature learning based on SAE–PCA network for human gesture recognition in RGBD images," *Neurocomputing*, vol. 151, pp. 565-573, Mar. 2015.
- [15] D. G. Stavrakoudis, J. B. Theocharis, and G. C. Zalidis, "A boosted genetic fuzzy classifier for land cover classification of remote sensing imagery," *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 66, pp. 529-544, July 2011.
- [16] D. Kato and I. Ohzawa, "2D Gabor wavelet transform and inverse transform (reconstruction) Demo using Matlab," Visiome Platform.



Ting-Chen Tsan is currently studying in Chung Shan Medical University and doing her B.S. degree. Her research interests include medical image processing and pattern recognition. She took part in many competitions in informatics and got good grades.



radiology.

Hao-Hung Tsai received the B.S. degree from the Faculty of Post-Baccalaureate Medicine Department, Kaohsiung Medical University, Kaohsiung, Taiwan. He worked at the Department of Medical Imaging in Taichung Veterans General Hospital from 2003 to 2007. He is currently as visiting staff at the Department of Medical Imaging in Chung Shan Medical University Hospital. His research interests include medical image interpretation and diagnostic



Chiun-Li Chin received the B.S. degree in computer science and information engineering and the M.S. degree in electrical and control engineering from Chung Hua University, Hsin Chu, Taiwan, and the Ph.D. degree in electrical and control engineering from National Chiao Tung University, Hsin Chu, Taiwan. From 2014 to 2015, he was a chair of Department of Medical Informatics in Chung Shan Medical University. He is currently an associate professor of the Department of Medical Informatics in Chung Shan Medical University. His research interests include medical image processing, computer vision, pattern recognition, and mHealthcare software design.