This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TBME.2015.2477688, IEEE Transactions on Biomedical Engineering



# Fusion of Quantitative Image and Genomic Biomarkers to Improve Prognosis Assessment of Early Stage Lung Cancer Patients

Nastaran Emaminejad, Wei Qian, Yubao Guan, Maxine Tan, Yuchen Qiu, Hong Liu, and Bin Zheng*

***Abstract* — *Objective:* This study aims to develop a new quantitative image feature analysis scheme and investigate its role along with 2 genomic biomarkers namely, protein expression of the excision repair cross-complementing 1 (ERCC1) genes and a regulatory subunit of ribonucleotide reductase (RRM1), in predicting cancer recurrence risk of Stage I non-small-cell lung cancer (NSCLC) patients after surgery. *Methods:* By using chest computed tomography images, we developed a computer-aided detection scheme to segment lung tumors and computed tumor-related image features. After feature selection, we trained a Naïve Bayesian network based classifier using 8 image features and a Multilayer Perceptron classifier using 2 genomic biomarkers to predict cancer recurrence risk, respectively. Two classifiers were trained and tested using a dataset with 79 Stage I NSCLC cases, a synthetic minority oversampling technique and a leave-one-case-out validation method. A fusion method was also applied to combine prediction scores of two classifiers. *Results:* AUC (areas under ROC curves) values are 0.78±0.06 and 0.68±0.07 when using the image feature and genomic biomarker based classifiers, respectively. AUC value significantly increased to 0.84±0.05 ($p$<0.05) when fusion of two classifier-generated prediction scores using an equal weighting factor. *Conclusion:* A quantitative image feature based classifier yielded significantly higher discriminatory power than a genomic biomarker based classifier in predicting cancer recurrence risk. Fusion of prediction scores generated by the two classifiers further improved prediction performance. *Significance:* We demonstrated a new approach that has potential to assist clinicians in more effectively managing Stage I NSCLC patients to reduce cancer recurrence risk.**

***Index Terms* — Computer-aided diagnosis, Fusion of image features and genomic biomarkers, Quantitative image feature analysis, Prediction of lung cancer recurrence risk, Radiomics.**

This work was supported in part by TSET Cancer Center Program, Oklahoma Tobacco Settlement Endowment Trust, Peggy and Charles Stephenson Cancer Center, University of Oklahoma. *Asterisk indicates corresponding author.*

N. Emaminejad is with the School of Electrical and Computer Engineering, University of Oklahoma, Norman, OK 73019, USA (email: emami.nastaran@ou.edu).

W. Qian is with the Department of Electrical and Computer Engineering, University of Texas, El Paso, TX 79968, USA and Sino-Dutch Biomedical and Information Engineering School, Northeastern University, Shenyang 110819, China (email: wqian@utep.edu).

Y. Guan is with Department of Radiology, Guangzhou Medical University, Guangzhou, China (email: yubaoguan@163.com ).

M. Tan, Y. Qiu, H. Liu and *B. Zheng are with the School of Electrical and Computer Engineering, University of Oklahoma, Norman, OK 73019, USA (email: Maxine.Y.Tan-1@ou.edu, qiuyuchen@ou.edu, liu@ou.edu and Bin.Zheng-1@ou.edu)

Digital Object Identifier

## I. Introduction

Lung cancer is the second most prevalent cancer in both men and women. It also has the highest mortality rate, which accounts for more than a quarter of all cancer related deaths in the United States (i.e., 28% for male and 26% for female cancer patients [1]). In lung cancer, over 75% cases are non-small-cell lung cancer (NSCLC) cases. Since lung is a relatively large human organ that may involve many other chronic lung diseases, lung tumors can often grow for a long time before they are found. Hence, the majority of lung cancers are currently detected and diagnosed at the advanced stages with a lower survival rate. Therefore, great research and promotion effort has been made to implement lung cancer screening programs using the low-dose computed tomography (CT) imaging modality for the last decade [2-3]. In July 29, 2013, the U.S. Preventive Services Task Force issued a draft recommendation in favor of lung cancer screening for long-term smokers using low-dose CT tests [4]. Meanwhile, the trend of using CT for screening and/or detecting other lung diseases (e.g., chronic obstructive pulmonary diseases) is also on the rise. As a result, more early stage lung cancers (i.e., stage I NSCLC) are detected during the regular lung cancer screening or other incident findings.

Although early cancer detection and treatment can improve the survival rate of lung cancer patients, lung cancer recurrent rates after surgical resection of the malignant tumors can range from 30% to 60% as reported in the previous studies [5]. As a result, mortality rate among the stage I NSCLC patients is also much higher than many other types of cancers (e.g., breast and colon cancer) detected at an early stage. According to the statistical data reported from the National Cancer Institute's Surveillance, Epidemiology, and End Results database, current 5-year survival rates are 49% and 45% for Stage IA and Stage IB NSCLC patients, respectively [6]. Therefore, in order to more effectively treat and manage the stage I NSCLC patients, it is important to develop an effective clinical marker or prediction model to more accurately predict cancer prognosis after cancer surgery. For the patients identified with the higher risk of cancer recurrence, the specific chemotherapy should be applied after surgery to minimize cancer recurrence risk.

Since there is no clinical standard for assessing the risk of post-surgery cancer recurrence to date, researchers have explored different genomic biomarkers to identify the

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TBME.2015.2477688, IEEE Transactions on Biomedical Engineering


molecular fingerprints of different genomic defects in lung cancer development and prognosis [7]. Among them, two biomarkers namely, *excision repair cross-complementing 1* (ERCC1) gene product and the *regulatory subunit of ribonucleotide reductase* (RRM1), have been extensively investigated and reported as prognostic biomarkers of the NSCLC patients [8-10]. ERCC1 is a rate limiting enzyme in the nucleotide excision repair and RRM1 functions as a regulatory subunit of ribonucleotide reductase that controls the substrate specificity and the on/off function of ribonucleotide reductase. In the clinical practice, these two biomarkers are examined and measured by applying a standard Immune-histo-chemical (IHC) method to analyze the surgery-resected tumor tissue specimen. Studies have shown that high expression of ERCC1 together with RRM1 often indicated an improved outcome in NSCLC patients who are treated only with surgical resection [11]. Thus, the researchers believe that comparative analyses between gene and IHC protein expression of ERCC1 and RRM1 might define a useful genomic biomarker to predict prognosis of the NSCLC patients [12].

Despite the promising results, using genomic biomarkers still faces multiple challenges and has a number of limitations, such as the higher cost, processing error in signal probing and/or subjective rating, as well as the lower specificity [13-14]. Hence, radiographic imaging examinations still play a critical role in current clinical practice for lung cancer detection, diagnosis, and prognosis assessment. Many advanced imaging modalities including high-resolution CT, Positron Emission Tomography (PET), PET-CT and Magnetic Resonance Imaging (MRI) [15-17] have been investigated and used in lung cancer imaging. Among them, CT remains the most popular imaging modality due to its higher accuracy, wide accessibility and cost-effectiveness. However, reading and interpreting a large number of CT images of a lung cancer case is difficult for radiologists. Using Response Evaluation Criteria in Solid Tumors (RECIST) guideline [18] to subjectively measure tumor size in one-dimension and evaluate tumor size change during sequential CT examinations is often not reliable (due to the inter-reader variability) and has lower correlation to the clinical outcomes of the patients. Therefore, there is an increasing interest in identifying novel quantitative image markers computed from the CT images to predict cancer prognosis [19-20].

In this study, we investigated a new quantitative image 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 malignant tumors depicting on CT images acquired before surgery and compute tumor-related morphological, density and texture based image features. Using a set of the selected non-redundant image features, we trained a Naïve Bayesian network based 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). Using a receiver operating characteristic (ROC) evaluation method, we assessed prediction performance of using each of the image features and genomic biomarkers as well as two multivariable based machine learning classifiers. Finally, we investigated the potential to further improve the prediction accuracy by applying a simple fusion method to combine the prediction scores generated by the image feature and genomic biomarker based classifiers.

## II. Materials And Methods

### A. A Test Dataset

The test dataset for this study was retrospectively acquired under an institutional review board approved data collection protocol from the First Affiliated Hospital of Guangzhou Medical University, Guangzhou, China. The dataset includes images of thoracic CT examinations of 79 patients who underwent lung cancer diagnosis and treatment in the hospital. All of these patients were diagnosed with Stage I NSCLC. Based on the current clinical guideline, a lung surgery was performed on each patient to resect malignant tumors. After surgery, the tumor specimens were extracted. Two genomic biomarkers, ERCC1 and RRM1, were evaluated using the selected tumor specimen and a standard IHC based analytic method in the pathology laboratory of the hospital.

In the ERCC1 protein expression test, the cells showed stained yellow to brown color is defined as positive cell. Using a microscope, a technologist in the pathology laboratory examines the IHC processed specimen to detect the colored cells and estimates the percentage of positive cells within the examined specimen using a 4-scale grading system in which white cell is scored as 0, light yellow cell is 1, brown cell is 2, and dark yellow cell is 3. Finally, based on the visually estimated percentage of stained colored cells, the technologist assigns an ERCC1 protein expression score for the case from 0 to 3, which indicate that the percentage of estimated colored cells is (1) < 10%, (2) 10% to 25%, (3) 26% to 50%, and (4) > 50% for the four scoring categories, respectively. The technologists also visually examine the specimen IHC probed for the RRM1 test and estimate the percentage of cells that show the color sign of RRM1 expression. Each case is graded into two categories represented by the technologists as either high (≥50%) or low (<50%) in RRM1 cell expression. Table I shows the demographic information of these 79 patients along with the corresponding subjectively assigned ERCC1 and RRM1 scores in the cases of this testing dataset.

Among these 79 stage I NSCLC patients, 46 are male and 33 are female. The age of these patients ranged from 39 to 85 years old. The average age is 61 years old with a standard deviation of 9.05. Among these patients, 32 are younger than 60 years old and 47 are older than 60 years old. Based on the available patient follow-up data, these 79 patients were divided into two groups of with and without cancer recurrence within the 3 years after the lung cancer surgery. Specifically, 21 patients had cancer recurrence and 58 remained progression or disease-free survival (3-year DFS). Table II summarizes the distribution of these two groups of cases based on two genomic biomarker values, identified tumor and/or cancer cell characteristics.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TBME.2015.2477688, IEEE Transactions on Biomedical Engineering



TABLE I
DEMOGRAPHIC INFORMATION OF 79 NSCLC PATIENTS IN THE DATASET

| | Gender | | Age (Years old) | |
|---|---|---|---|---|
| | Male | Female | ≤60 | >60 |
| ERCC1 | | | | |
| 0 | 14 | 7 | 12 | 9 |
| 1 | 16 | 14 | 13 | 17 |
| 2 | 9 | 7 | 6 | 10 |
| 3 | 7 | 5 | 1 | 11 |
| RRM1 | | | | |
| 1 | 12 | 16 | 15 | 13 |
| 2 | 32 | 19 | 21 | 13 |

TABLE II
THE DISTRIBUTION OF TWO BIOMARKERS AND TUMOR OR CANCER CELL CHARACTERISTICS BETWEEN THE TWO GROUPS OF CASES WITH AND WITHOUT CANCER RECURRENCE

| Cancer Recurrence | No | Yes | Total cases |
|---|---|---|---|
| ERCC1 | | | |
| 0 | 14 | 7 | 21 |
| 1 | 20 | 11 | 31 |
| 2 | 13 | 2 | 15 |
| 3 | 11 | 1 | 12 |
| RRM1 | | | |
| ≤ 50% | 23 | 5 | 28 |
| > 50% | 35 | 16 | 51 |
| Cell type | | | |
| Squamous Cell carcinoma | 13 | 3 | 16 |
| Adenocarcinoma | 40 | 17 | 57 |
| Other | 5 | 1 | 6 |
| Tumor size (Diameter) | | | |
| ≤ 3cm | 46 | 12 | 58 |
| > 3cm & ≤ 5cm | 12 | 9 | 21 |

All CT examinations were acquired using a 16-detector based Toshiba Aquilion CT machine. In CT image scanning protocol, X-ray tube voltage ranged from 120 to 140KV and the current ranged from 140 to 340mAs depending on patient body size. The CT image slices were reconstructed with an image size of 512×512 pixels with pixel size ranged from 0.51 to 0.74mm also depending on the patient body size. The image slice thickness of all CT images in this dataset was 2mm.

### *B. Lung Tumor Segmentation*

We used a semi-automated computer-aided detection (CAD) scheme to segment lung tumors. In brief, the tumor center and diameter that have been marked previously by a radiologist in the original clinical CT image reading were used as the initial segmentation seed in our CAD scheme. In each case, the image processing of the scheme started from segmenting the tumor area from the CT slice with the marks of the radiologist. The marked tumor seed was mapped into next adjunct CT image slices to segment the tumor area depicting on the next slices. This segmentation process was iteratively performed until the scheme reached a slice without remaining tumor area being detected. Specifically, in each involved image slice, CAD scheme first applied a conventional region growing algorithm with an empirically selected threshold of CT number (-450 HU). This step generally works well in segmentation of well-circumscribed lung nodules.

However, this fixed region growing often did not provide an accurate result in segmenting juxtapleural and vascularized tumors, which connect with lung boundary and/or vessels. Thus, the following two image processing steps were applied. First, as shown in Fig. 1(a), by applying an initial region growing algorithm, the segmented tumor has a leakage to outside lung area (Fig. 1(b)). To segment this tumor, our scheme applied a modified convex hull function based algorithm introduced by Kuhgnik *et al*. [21] to stop the leakage of the segmented tumor area to the normal lung tissues and also smooth tumor boundary. Based on an anatomical fact that a lung is mostly convex, this algorithm is efficient and has good capability of removing the thoracic lesions from the chest wall (Fig. 1(c)). However, applying the convex hull function algorithm is also likely to generate a few minor (isolated) regions due to image noise. The scheme then applied a region labeling algorithm to remove small regions while maintaining the segmented tumor region (Fig. 1(d)).

Second, in order to cut and remove the connected vascular structures from the tumor boundary, our scheme applied a distance map based morphological operation as proposed by Kuhgnik *et al*. [21]. As shown in Fig. 2, after a juxtapleural tumor (Fig. 2(a)) was segmented using the initial region growing algorithm and convex hull function, the scheme fitted a rectangular window to the initial tumor boundary (Fig. 2(b)). The window was also centered on tumor center. The scheme applied an *Euclidean distance transform* to convert the image inside the window into a distance map $E$ that contains the minimum distance of each pixel of the tumor region to the tumor boundary pixels. Then, a seed optimization was done by searching for the pixel $C$ with the longest distance in the neighborhood of the initial given seed, and a new radius for the tumor was calculated based on this new seed.

$$\boldsymbol{r} = \boldsymbol{E}(\boldsymbol{C}) \tag{1}$$

Using this process, a normalized distance map $\acute{E}$ was obtained

$$\acute{\boldsymbol{E}} = \boldsymbol{E}/\boldsymbol{r} \tag{2}$$

As shown in Fig. 2(b), a small window was placed to cover the targeted tumor on the CT image. If the region growing starts from the center $C$, it reaches the boundary of the image through the vasculature and the result would not be a smoothly detected tumor. In order to remove vascular connections, a distance based morphological operation was performed on the distance map. The dilation and erosion were based on the shortest distance of each pixel to the tumor boundary. To perform dilation and erosion, an adaptive threshold was calculated from the normalized distance map $\acute{E}$ with the initial value of 1. The threshold was lowered from 1 to a value with which the boundaries of the window were reached. The last threshold $\acute{\theta}$ was used to calculate the threshold $\theta$ for dilation and erosion.

$$\boldsymbol{\theta} = \boldsymbol{r} * \acute{\boldsymbol{\theta}} \tag{3}$$

Fig. 2(c) shows the eroded tumor $N_{-}$ and Fig. 2(d) shows the dilated tumor after erosion. The erosion was done by

$$\boldsymbol{N}_{-} = \boldsymbol{E} > \boldsymbol{\theta} \tag{4}$$

Then a second distance transform D was performed on $N_{-}$ and the dilation was done by

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TBME.2015.2477688, IEEE Transactions on Biomedical Engineering




**Fig. 1.** (a) A case with Juxtapleural tumor, (b) Leakage of region growing, (c) Chest wall connection is removed by convex hull function, (d) Final segmented tumor region.

$$\boldsymbol{N_+ = D < \theta} \quad (5)$$

Last, the scheme refined the tumor segmentation boundary. The image after the first segmentation was considered as $N_0$ and the image after convex hull function was considered as $N_2$. Then, $I = N_0 \cap \partial N_+$ was calculated and dilated with a value of $\delta = 3$, ($\partial N_+$is the boundary of $N_+$). As a result, the tumor was segmented from the connected vascular trees while the unsmoothed tumor boundary contour or surface is largely preserved (Fig. 2(e)):

$$\boldsymbol{N_* = (N_+ \cap N_2) \backslash I_\delta} \quad (6)$$

After tumor segmentation, results were also visually reviewed slice by slice. If an obvious leakage or any unremoved structures attached to the tumor region were observed, a manual correction was applied in this tumor to set up a new initial boundary condition and the region growing algorithm was reapplied on this slice to segment the tumor region. The similar semi-automated tumor segmentation and performance evaluation method has been reported in the previous studies to segment lung nodules and breast masses to reduce or minimize the erroneous measurement results of the feature values computed from the images [22-23].

### *C. Quantitative tumor image features*

Our CAD scheme initially computed a set of 35 three-dimensional tumor volume related morphological, CT number distribution and texture features from all involved CT image slices. In order to reduce the dimensionality of the feature space by removing redundant features and increase robustness of the multi-feature based risk prediction model or classifier developed later in this study, we used a CFS Subset Evaluator

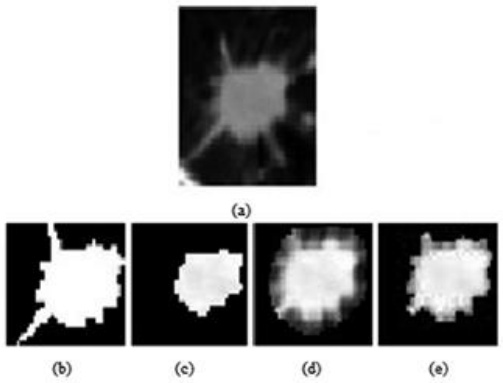


**Fig. 2.** (a) A Vascularized tumor. (b) Binary image after region growing and convex hull fitted in a window that touches the boundary ($N_0$). (c) Binary image of the region after erosion ( $N_-$). (d) The image after dilation ($N_+$). (e) Final image with removed vascular connections ($N_*$).

attribute selection in WEKA data mining software package [24] with a Best-First heuristic feature search and selection method [25] to examine and filter all features. We then selected the effective and non-redundant features from the initial pool of 35 image features based on the importance sorting of features in predicting the risk of cancer recurrence. From the initial feature pool, 8 image features were selected by this feature selection method. The selected features can be divided into 2 subgroups of 5 morphological and CT number distribution based image features and 3 tumor texture based image features. The first subgroup includes the following 5 image features:

1) *Standard deviation of tumor pixel value (CT number):* It describes the degree of tumor density distribution and heterogeneity. The larger standard deviation of tumor pixel values indicates greater tumor heterogeneity.
2) *The maximum tumor diameter computed from the targeted CT slice*: This feature represents the maximum tumor size in the tumor central slice, which is similar to that assessed and measured using the RECIST guideline.
3) *A tumor shape factor*: The scheme first counts the number of pixels located in the tumor boundary contour of all segmented tumor regions on the related CT image slices. Then, 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. The larger value of this feature indicates the higher level of fluctuation or irregularity on the tumor boundary surface.
4) *The maximum CT number within the tumor volume*: This feature represents the highest density level of the tumor and correlates to the highest tumor stiffness level, which is another potentially important image feature associated to the tumor prognosis as reported in previous study [26].
5) *Standard deviation of pixel values in central volume of tumor*: To compute this feature, we defined a tumor central volume as a cubic volume that is centered in the tumor center and has the length equal to the shortest radius of the tumor. This feature quantifies the heterogeneity of tumor density in the tumor center region.

Three texture features were computed from a Gray Level Run Length (GLRL) matrix of a tumor region based on the

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TBME.2015.2477688, IEEE Transactions on Biomedical Engineering


run-length statistics. To compute GLRL features, the image grayscales were reduced from 12 bits (0 to 4095) resolution to 8 bits (0 to 255). A run is defined as a sequence of pixels with intensity in certain range of each bin, and run length is the length of that run [27]. The GLRL was computed through a package developed by Wei [28] using a zigzag method in four steps to: 1) determine direction, 2) perform zigzag scan, 3) obtain new sequences, and 4) calculate run-length matrix. Specifically, the 3 texture features selected by the Best-first heuristic feature search and selection method are:

6) *Gray-Level Non uniformity (GLN),* $F_6 = GLN = \sum_{l=1}^{L}[\sum_{g=0}^{G-1} P(g,l)^2]$, where $l$ is the length of the run, $L$ is the maximum run length, $g$ is gray level bin, $G$ is the maximum number of gray level bins, and $P(g,l)$ is the probability of the specific run length, respectively. This feature value increases as the gray-level outliers dominate the histogram.
7) *Short Run Low Gray-Level Emphasis (SRLGE),* $F_7 = SRLGE = \sum_{g=0}^{G-1}\sum_{l=1}^{L}\frac{P(g,l)}{l^2(g+l)^2}$. It is a diagonal metric that increases when the texture is dominated by many short runs of low gray value.
8) *Long Run Low Gray-Level Emphasis (LRLGE),* $F_8 = \sum_{g=0}^{G-1}\sum_{l=1}^{L}\frac{P(g,l)l^2}{(g+l)^2}$. It increases when the texture is dominated by long runs that have low gray levels.

Since each texture feature had different values in 4 scanning directions ($\alpha = 0^\circ, 90^\circ$ and $\pm 45^\circ$), each texture feature computed in one CT image slice was represented by a mean of the 4 feature values calculated along 4 directions. Finally, a texture feature in a 3D tumor volume was computed by a tumor region weighted summation method, $F^{3D} = \sum_{i=1}^{N} w_i F_i$, where $N$ is the total number of CT image slices depicting the segmented tumor regions and $w_i$ is a weighting factor represented by the size of tumor region in one CT slice.

### D. Two Classifiers and Data Analysis

First, besides RRM1 that only has two binary (low or high) groups, each of 8 selected image features and ERCC1 were first individually processed by using a maximum likelihood based receiver operating characteristic (ROC) fitting program (ROCKIT, http://www-radiology.uchicago.edu/krl/, University of Chicago) to compute the area under a ROC curve (AUC). We used AUC as an assessment index to analyze and compare the performance of each feature to predict or classify the test cases of our dataset into two classes namely, positive (+) for cases with cancer recurrence after surgery within 3 years and negative (-) for cases without cancer recurrence within 3 years. The correlation coefficients of AUC values between applying these image features and ERCC1 were also computed and compared.

Second, we built a statistical machine learning classifier to combine the 8 selected image features and predict the risk of a Stage I NSCLC patient having cancer recurrence after cancer surgical treatment. Although many different machine learning classifiers [29] can be used for this purpose, in this study, due to the limited test dataset, we built a simple Naïve Bayesian network based classifier to reduce the potential risk of overtraining problem. The primary advantage of a Bayesian network is that it can provide the flexibility for specifying the joint probability distribution of a problem domain by exploiting the dependencies between variables and capturing the uncertain knowledge of a given problem in a natural and efficient way. The Bayesian learning method has a long history to build a classifier by combining quantitative image and non-image features to assist cancer diagnosis [30-31]. In this study, we used WEKA data mining software package [24] to build the classifier. The classifier was trained and tested using a standard leave-one-case-out validation method [32]. Due to the imbalanced number of cases in the two classes (21 "positive" cancer recurrence cases vs. 58 "negative" without cancer recurrence cases), in order to achieve a more balanced optimization result in predicting the cases in two classes, we applied a synthetic minority oversampling technique named SMOTE [33] to add synthetic data to double the test cases in the "positive" class. Specifically, when using the leave-one-case-out training and testing method, the classifier was trained and tested 79 times. In each training/testing process cycle, one case was removed from the training dataset. The SMOTE algorithm was applied to double the number of "positive" cases based on the distribution of 78 cases in the training dataset. The trained/optimized classifier was then applied to a test case that was not involved in the SMOTE based synthetic data generation and classifier training process. This training and testing process was iteratively performed 79 times, which thus produces risk prediction scores (which are normalized in a range from 0 to 1) for all 79 cases in our dataset. The higher score indicates the higher risk of the patient having cancer recurrence within the 3 years after lung cancer surgery. Finally, we computed AUC value using these 79 prediction scores and the ROCKIT program.

Third, we also applied a machine learning classifier to combine two genomic biomarkers of ERCC1 and RRM1 for predicting lung cancer recurrence risk. We trained and tested the same Naïve Bayesian network based classifier as used for CT image features. Probably due to only two variables and limited discrete status (2 for RRM1), the Naïve Bayesian network classifier yielded a significantly lower prediction performance as compared to using ERCC1 only (which will be reported in section III). Thus, we also built and tested a simple Multilayer perceptron based classifier to combine the two genomic biomarkers, which yielded significantly higher performance. The training and testing process was also used by the same WEKA data mining software package, SMOTE synthetic data generation algorithm, leave-one-case-out validation method, and the ROCKIT program as discussed in training and testing the image feature based classifiers.

Fourth, after comparing the prediction performance between the quantitative image feature-based classifier and the two genomic biomarker based classifiers, we applied several data fusion methods [34] (including the maximum, minimum, and

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TBME.2015.2477688, IEEE Transactions on Biomedical Engineering


weighting average) to combine the prediction risk scores generated by these two classifiers and test the feasibility of further improving performance in predicting risk of cancer recurrence. The fusion result was also compared with those yielded using two individual classifiers.

Fifth, since in clinical practice, the radiologists and/or oncologists make a binary decision on image reading, cancer diagnosis, prognosis assessment, and selection of treatment option, besides AUC values, we also assessed and compared the prediction accuracy, positive and negative predictive values based on the confusion matrix built by selecting an optimal operation threshold to divide the cases into two binary categories (with and without cancer recurrence). For this purpose, we grouped the cases with lower ERCC1 gene expression (0 and 1) into one group and cases with higher ERCC1 (2 and 3) into another group. We also applied a default operation threshold of 0.5 on the middle point of the prediction scores (0 to 1) generated by the image feature based classifier to divide the cases into the high and low risk case groups. As a result, we built three confusion matrices based on ERCC1, RRM1, and quantitative image feature analysis classifier, respectively. We then compared the prediction accuracy between using the quantitative image features and the two genomic biomarkers.

Last, we also applied and tested several other statistical data analysis models to measure and assess the association between the model predicted high or low risk scores in cancer recurrence and patients' DFS, which include (1) a Chi-squared test, (2) Kaplan-Meier plots and (3) Cox proportional hazards regression models. In the Chi-squared test, we divided cases into two groups by applying a threshold of 18 months, which is the middle point of 36 months (3 year) tracking period, while in using Kaplan-Meier and Cox regression models, the cases were divided into two groups of low and high risk in cancer recurrence by applying an operation threshold of 0.5 to the prediction scores generated by (1) the image feature based classifier, (2) the genomic biomarker based classifier or (3) the fusion model of two classifiers.

## III. Results

Table III summarizes and compares the computed AUC values using each of the 8 image features and one biomarker of ERCC1. The results show that image feature 2 (The maximum tumor diameter) and ERCC1 yielded the maximum and quite comparable AUC values. Table IV summarizes the correlation coefficients between each pair of image features, as well as the correlation coefficients between each image feature and each genomic biomarker (ERCC1 and RRM1). Except image features 1 and 2 (tumor CT value standard deviation and tumor size), the absolute values of all other correlation coefficients are much smaller than $r < 0.5$, which indicates that these features are not highly correlated or redundant. Thus, the result supports that the combination of this set of image features has potential to add supplementary information or discriminatory power to significantly increase performance of predicting the risk of lung cancer recurrence than using a single image feature or biomarker.

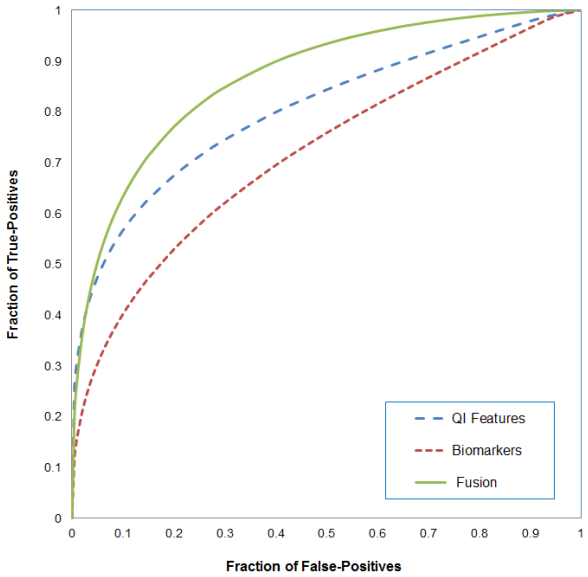


**Fig. 3.** Comparison of 3 ROC curves generated using quantitative tumor image (QI) feature based classifier, genomic biomarker based classifier, and a fusion method to predict the risk of lung cancer recurrence.

The image feature based Naïve Bayesian network classifier yielded an AUC = 0.78 with a 95% confidence interval (CI) of [0.66, 0.88], while the two genomic biomarker based classifier yielded an AUC = 0.68 with a 95% CI of [0.56, 0.79]. Table V shows and compares the AUC values obtained by choosing the maximum or minimum of the prediction scores generated by the two image feature and genomic biomarker based classifiers, as well as the fusion scores produced by systematically increasing the weighting factor from 0.1 to 0.9 applied to the prediction scores generated by the image feature based classifier (or 0.9 to 0.1 applied to the prediction scores generated by the two genomic biomarker based classifier). The test results showed that the maximum AUC = 0.84±0.05, which is significantly higher than the AUC values generated using either image feature or the genomic biomarker based classifier ($p < 0.05$) as computed by ROCKIT program. Fig. 3 shows and compares three ROC curves generated using the two classifiers and the optimal fusion method tested in this study.

TABLE III
SUMMARY OF THE COMPUTED AREAS UNDER ROC CURVES USING INDIVIDUAL IMAGE FEATURES AND ERCC1

| Image Feature | AUC | STD | 95% CI |
|---|---|---|---|
| 1[a] | 0.55 | 0.08 | [0.40, 0.70] |
| 2[b] | 0.64 | 0.07 | [0.49, 0.76] |
| 3[c] | 0.56 | 0.08 | [0.40, 0.70] |
| 4[d] | 0.58 | 0.08 | [0.44, 0.71] |
| 5[e] | 0.51 | 0.07 | [0.37, 0.65] |
| 6[f] | 0.59 | 0.07 | [0.45, 0.72] |
| 7[g] | 0.57 | 0.08 | [0.41, 0.72] |
| 8[h] | 0.59 | 0.07 | [0.45, 0.71] |
| ERCC1 | 0.63 | 0.07 | [0.48, 0.75] |

[a] Standard deviation of tumor pixel value, [b] The maximum tumor diameter, [c] Tumor shape factory, [d] Maximum CT number, [e] Standard deviation of pixel values in central volume, [f]GLN, [g]SRLGE, [h]LRLGE.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TBME.2015.2477688, IEEE Transactions on Biomedical Engineering



TABLE IV
SUMMARY OF THE CORRELATION COEFFICIENTS AMONG 8 SELECTED TUMOR IMAGE FEATURES AND TWO GENOMIC BIOMARKERS

| Image Feature | 1[a] | 2[b] | 3[c] | 4[d] | 5[e] | 6[f] | 7[g] | 8[h] | ERCC1 |
|---|---|---|---|---|---|---|---|---|---|
| 2 | -0.52 | | | | | | | | |
| 3 | -0.20 | 0.32 | | | | | | | |
| 4 | -0.22 | 0.12 | 0.24 | | | | | | |
| 5 | 0.20 | -0.32 | -0.14 | -0.07 | | | | | |
| 6 | -0.15 | 0.04 | 0.33 | -0.38 | 0.12 | | | | |
| 7 | 0.08 | 0.11 | 0.07 | -0.22 | -0.02 | -0.07 | | | |
| 8 | 0.27 | -0.10 | 0.01 | -0.30 | -0.01 | -0.11 | 0.17 | | |
| ERCC1 | -0.08 | -0.15 | -0.09 | 0.26 | -0.04 | -0.20 | -0.21 | -0.18 | |
| RRM1 | -0.12 | -0.14 | 0.04 | 0.01 | 0.09 | 0.08 | 0.17 | 0.25 | 0.07 |

[a] Standard deviation of tumor pixel value, [b] Maximum tumor diameter, [c] Tumor shape factor, [d] Maximum CT number, [e] Standard deviation of pixel values in central volume, [f] GLN, [g] SRLGE, [h] LRLGE.

TABLE V
SUMMARY OF THE PERFORMANCE (AUC VALUES) USING DIFFERENT FUSION METHODS TO COMBINE PREDICTION SCORES GENERATED BY TWO CLASSIFIERS

| Fusion Method /Weighting factor | AUC | STD | 95%CI |
|---|---|---|---|
| Maximum | 0.78 | 0.06 | [0.65, 0.88] |
| Minimum | 0.79 | 0.06 | [0.66, 0.89] |
| 0.1*ICS[a] + 0.9*GCS[b] | 0.73 | 0.07 | [0.62, 0.84] |
| 0.2*ICS + 0.8*GCS | 0.76 | 0.07 | [0.63, 0.86] |
| 0.3*ICS + 0.7*GCS | 0.78 | 0.06 | [0.66, 0.89] |
| 0.4*ICS + 0.6*GCS | 0.80 | 0.06 | [0.68, 0.90] |
| 0.5*ICS + 0.5*GCS | 0.84 | 0.05 | [0.74, 0.93] |
| 0.6*ICS + 0.4*GCS | 0.82 | 0.06 | [0.70, 0.91] |
| 0.7*ICS + 0.3*GCS | 0.80 | 0.06 | [0.65, 0.89] |
| 0.8*ICS + 0.2*GCS | 0.79 | 0.06 | [0.64, 0.88] |
| 0.9*ICS + 0.1*GCS | 0.78 | 0.06 | [0.63, 0.89] |

[a] Risk prediction scores generated by image feature based classifier and [b] Risk prediction score generated by genomic biomarker based classifier.

Table VI summarizes and compares the overall prediction accuracy, the positive and negative predictive values of using ERCC1, RRM1, and the tumor image feature based classifier. The results show that by using the quantitative image feature based classifier and a 0.5 operation threshold on the prediction scores, the overall cancer recurrence prediction accuracy increased by more than 30% as compared to the use of two genomic biomarkers.

TABLE VI
COMPARISON OF THE PREDICTION ACCURACY, POSITIVE PREDICTIVE VALUE (PPV), AND NEGATIVE PREDICTIVE VALUE (NPV) USING TWO GENOMIC BIOMARKERS AND QUANTITATIVE IMAGE FEATURE BASED CLASSIFIER

| | ERCC1 | RRM1 | Image feature |
|---|---|---|---|
| Accuracy | 53% | 51% | 80% |
| PPV | 0.35 | 0.18 | 0.65 |
| NPV | 0.89 | 0.69 | 0.84 |

Chi-squared test showed a significant association between the two groups of DFS and the prediction scores generated using our fusion model of two classifiers with the coefficients of ($df$ = 1, $\rho$ < 0.05, $\chi^2$=18.65). Table VII summarizes the Kaplan-Meier data analysis results, which compares 3-year DFS rates for each prediction method as well as the computed log-rank test *P*-value for the significant difference between the DFS of the two low and high risk case groups. Meanwhile, Fig. 4 plots the predicted DFS trends in the two groups using (1) image feature based classifier, (2) genomic biomarker based classifier and (3) the fusion model of two classifiers. Fig. 5 plots and compares the actual DFS or the recorded cancer recurrence trend during the 3 year period after lung cancer surgery with the 3 prediction classifiers or methods.

TABLE VII
KAPLAN-MEIER ANALYSIS OF DISEASE-FREE SURVIVAL AND OVERALL SURVIVAL IN TWO GROUPS OF LOW RISK AND HIGH RISK OF CANCER RECURRENCE PROPOSED BY EACH PREDICTOR

| | DFS rate at 3 years | [a] P-value |
|---|---|---|
| Image feature based classifier<br>High risk group(n=24)<br>Low risk group(n=55) | <br>25%±0.12<br>76%±0.14 | <0.001 |
| Genomic biomarker based classifier<br>High risk group(n=22)<br>Low risk group(n=57) | <br>43%±0.1<br>77%±0.08 | <0.001 |
| Fusion method<br>High risk group(n=22)<br>Low risk group(n=57) | <br>20%±0.1<br>76%±0.07 | <0.001 |

[a] Log-rank test for the difference between Disease Free Survival curves of proposed low and high risk groups by each predictor.

Table VIII compares three sets of the Cox regression data analysis results. The table compares the hazard ratios (HR) computed using 3 Cox regression models that compare each pair of three different sets of prediction scores generated using (1) image feature based classifier, (2) genomic biomarker based classifier and (3) the fusion method of the two classifiers. The data analyses 1 and 2 show that the hazard ratios computed using the fusion method of the two classifiers is significantly higher than using prediction scores generated using each of the two classifiers alone. In addition, analysis 3 shows that the prediction result generated using the image feature-based classifier has a significantly higher prediction power than using the genomic biomarker based classifier.

Table VIII
SUMMARY OF DATA ANALYSES ON THREE COX REGRESSION MODELS

| | HR | 95% CI | P-value |
|---|---|---|---|
| Analysis 1 | | | |
| Fusion of two classifiers | 3.27 | [1.95, 5.48] | <0.001 |
| Genomic biomarker based classifier | 0.89 | [0.5, 1.56] | 0.68 |
| Analysis 2 | | | |
| Fusion of two classifiers | 3.15 | [1.34, 7.39] | <0.001 |
| Image Feature based classifier | 0.98 | [0.49, 1.97] | 0.9 |
| Analysis 3 | | | |
| Genomic biomarker based classifier | 1.78 | [1.10, 2.97] | 0.019 |
| Image Feature based classifier | 2.25 | [1.55, 4.26] | <0.001 |

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TBME.2015.2477688, IEEE Transactions on Biomedical Engineering


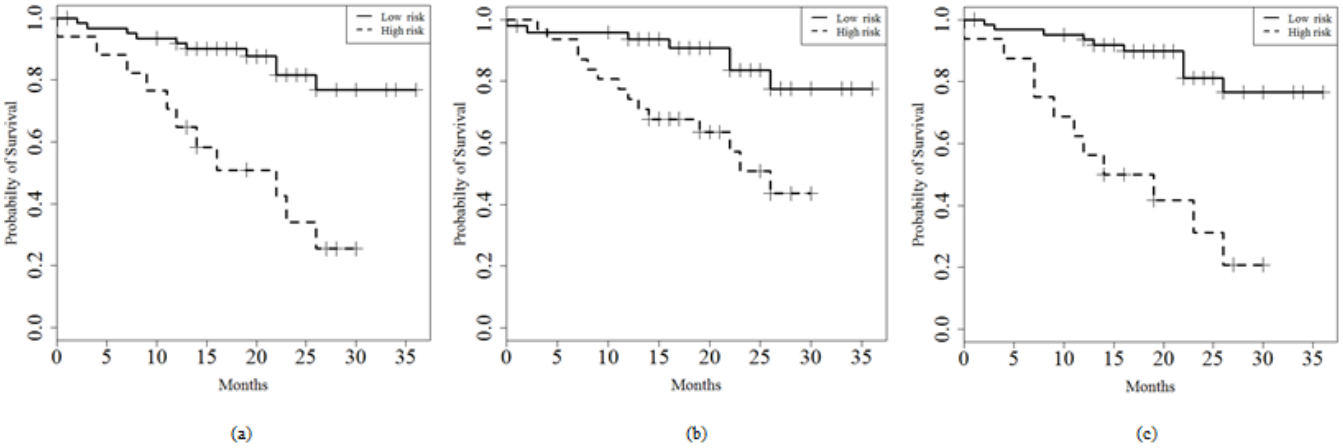


**Fig. 4.** Comparison of Kaplan-Meier disease-free survival estimates from 79 patients using (a) image feature based classifier, (b) two genomic biomarker based classifier and (c) fusion of two classifiers. The predicted low risk group is presented using a solid line with prediction scores < 0.5, while the high risk group is presented in the dash line with prediction scores ≥ 0.5.

## IV. DISCUSSION

The study has a number of unique characteristics. First, although many different types of CAD schemes of chest CT images has been previously developed and tested (e.g., [35-37]), in this study we explored a new approach of applying a CAD-based quantitative image analysis concept to a new application field of predicting cancer recurrence risk. Since the ultimate goal of current promotion of lung cancer screening using low-dose CT imaging examinations is to reduce cancer mortality rate, this study investigated a clinically significant issue of how to more accurately predict cancer recurrence risk after initial surgery among the Stage I NSCLC patients. Our study results demonstrated the feasibility of developing a new CAD scheme to predict risk of cancer recurrence, which may eventually help improve the efficacy of lung cancer screening programs, cancer treatment and patient management.

Second, we investigated a number of new image features computed from CT images and built a machine learning classifier to predict lung cancer recurrence risk. Then, using a best-first heuristic feature searching algorithm, 8 image features were selected from an initial pool of 35 features, which include 5 morphological and pixel value related features and 3 texture-based features. The physical meaning of these 5 morphological and pixel value based image features can be well described to mimic the image features used by radiologists to assess and interpret tumors in the clinical practice. In our study, the tumor diameter (feature 2) yielded the highest AUC value (Table III), which correlates well with the RECIST guideline. However, the study also showed that optimally combining multiple image features using a classifier can further significantly increase prediction performance (AUC value). This is a major advantage of applying a CAD scheme over the human observers. Besides the tumor size, the classifier also includes a number of other sensitive image features to the human observers, such as irregularity of tumor boundary surface, and the heterogeneity of tumor density (or stiffness). Both features are important, in particular, tumor stiffness has recently been attracting research interests because cancer biology and oncology studies have identified that tumor stiffness is a very important biomechanical property of tumors, which relates to the elasticity level of the tumor tissue and affects tumor progression or response to chemotherapy [26, 38]. Although human observers can also subjectively assess many of tumor-related image features individually, it is much more difficult for them to consistently integrate these features into a decision making process. Quantitative image feature analysis can be more efficient and reliable to integrate these features by eliminating inter- and intra-observer variations. Also, due to the difference between human vision and computer vision, the meaning of many texture based features (e.g., the GRLR features) may not easily correlate to human perception. However, texture based image features have been widely used in many CAD schemes to detect and/or classify between malignant and benign or false-positive lesions. Thus, adding the texture based features also expands the capability of assessing cancer prognosis in CAD schemes.

Third, in the previously reported studies that aim to predict cancer prognosis, researchers have investigated a large number of genomic biomarkers, demographic and other clinical factors to predict cancer prognosis or patient survival rates. Among them, ERCC1 and RRM1 are two popular biomarkers that have been investigated to assess lung cancer prognosis including Stage I NSCLC patients. In this study, we compared the risk prediction performance of our new quantitative image analysis results with ERCC1 and RRM1 biomarkers. When applying to the same testing dataset, we found that ERCC1 yielded a quite comparable AUC value in predicting cancer recurrence as the "highest performed" image feature selected in this study (Table III). Although our study results support that these 2 genomic biomarkers have statistically significant association with the prognosis of the lung cancer patients (DFS), their absolute prediction accuracy or discriminatory power (i.e., positive and/or negative predictive value) on the individual patient level remains lower (Table VI). Our study results demonstrated that using a quantitative image feature based prediction model yielded significantly high prediction accuracy than using each of these 2 genomic biomarkers.

Fourth, currently, the biomedical imaging research and the biological or genetic research are typically conducted separately by the researchers in different fields. Identifying the association between the quantitative image features and the

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TBME.2015.2477688, IEEE Transactions on Biomedical Engineering


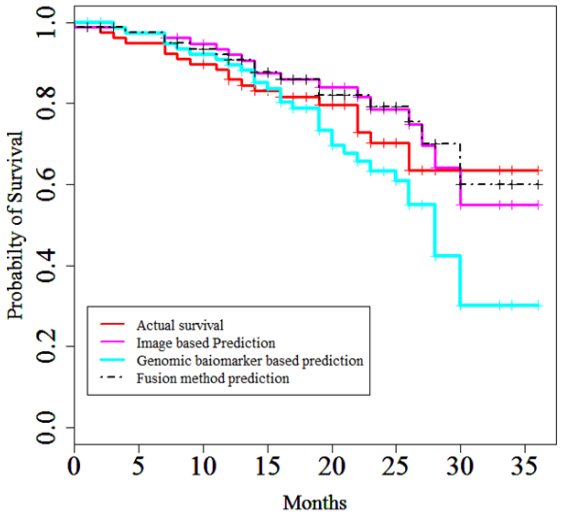


**Fig. 5.** Comparison of Kaplan-Meier analysis between actual disease-free survival and predicted disease-free survival by 3 different methods.

genomic biomarkers has attracted wide research interest in an emerging developing field of Radiomics recently [20]. In this study, we investigated the correlations between the image features and two genomic biomarkers. We applied a simple fusion method to combine the image phenotype and genotype features or biomarkers to predict lung cancer recurrence risk. The study results demonstrated that the quantitative CT image features and 2 popular genomic biomarkers provided supplementary discriminatory information in cancer prognosis assessment. As a result, combining the prediction results generated using the image feature and genomic biomarker based classifiers further improved the risk prediction performance than using either of the single classifiers.

From this study, we also performed several additional experiments and observed a number of interesting results. First, optimal feature selection is important in developing a CAD-based quantitative image feature analysis scheme. Training a Naïve Bayesian network with all 35 features yielded a substantially lower performance (AUC = 0.54 with 95% CI of [0.36, 0.71]) than the classifier trained using 8 selected image features (as shown in Fig. 3).

Second, when using the original dataset of 79 cases to train two classifiers, the image feature based classifier yielded AUC = 0.71 with a 95% CI of [0.57, 0.82] and the genomic biomarker based classifiers yielded AUC = 0.54 with 95% CI of [0.40, 0.67]. The results showed a significant advantage of using a SMOTE method to generate synthetic data and balance the number of training cases in the two risk classes in training and testing a machine learning classifier.

Third, we tested the performance by fusion of the normalized ERCC1 scores (from 0 to 1) and the prediction scores generated by the image feature based classifier. The highest obtained AUC value was 0.78 with 95% CI of [0.66, 0.88], which is lower than the highest AUC value yielded by fusion of the prediction scores of two classifiers (Table V).

Fourth, we also trained and tested two classifiers that used 8 image features and 2 genomic biomarkers together as the input features. The results showed that a Naïve Bayesian classifier yielded an AUC of 0.77 with 95% CI of [0.64, 0.88] and a Multilayer perceptron classifier yielded an AUC of 0.71 with 95% CI of [0.58, 0.82]. These AUC values were also lower than the highest AUC value yielded by fusion of the prediction scores of two classifiers as reported in Table V.

Despite the promising results, we also recognized that this was our first preliminary study with a number of limitations. First, this was a retrospective study with a small dataset. The retrieved study materials and information only included the existing CT images, two genomic biomarker testing results, and 3-year DFS. Hence, the robustness of the reported results in this study needs to be tested and validated in future with both the retrospective studies with large image datasets collected from different medical institutions and the prospective studies that include more detailed clinical information (i.e., clear margins and/or extent of the surgery of the tumors) and the longer DFS data (e.g., 5 years).

Second, we only explored and computed 35 tumor-related image features in this study. Other lung tissue background image features (i.e., features related to the chronic obstructive pulmonary disease or other lung diseases), which may also provide supplementary prediction information, have not been incorporated into our image feature analysis scheme.

Third, from the existing genomic biomarkers of lung cancer, our dataset only has the test records of two popular biomarkers (ERCC1 and RRM1). Many of the other existing genomic biomarkers, the potentially useful clinical markers (i.e., the clearance of tumor margin by the surgeons) and epidemiology based risk factors have not been compared with our image features or integrated into our fusion models. Hence, more studies are needed to more effectively integrate image features and other useful biomarkers or risk factors.

Fourth, in this study we performed a semi-automatic scheme to segment lung tumors. The segmentation results were visually examined. As a result, a small fraction of tumor segmentation boundary contours (<5%) were manually corrected when the substantial errors were visually detected. Due to the lack of a "ground-truth" and the potential inter-observer variability, we recognize that this is not an optimal tumor segmentation method and may create errors in computing tumor-related image features. However, we believe that this semi-automated tumor segmentation is an efficient approach to perform this proof-of-concept study. Developing more accurate and reliable CAD schemes for tumor segmentation is still one of research focuses in future studies.

Last, this is only a technology development study. Its clinical relevance or whether and how to optimally apply this new CAD-based quantitative image feature analysis scheme to assist decision-making of the clinicians (i.e., radiologists and/or oncologists) in the clinical practice has not been investigated and tested to date.

## V. Conclusion

Developing precision medicine [39] or a more effective personalized strategy for treating and managing Stage I NSCLC patients requires a more accurate clinical marker and/or assessment tool to predict cancer prognosis (including

cancer recurrence risk). Current studies mainly focus on identifying more effective genomic biomarkers, demographic factors, and other clinical variables. In this study, we investigated a new quantitative image feature analysis approach using chest CT images and demonstrated two new study results. First, an image feature based classifier yielded significantly higher performance than two popular genomic biomarkers in predicting cancer recurrence risk. Second, the image features and genomic biomarkers are not highly correlated and provide supplementary information. As a result, fusion of these two types of features and biomarkers further improved prediction performance.

Despite the limitations, this preliminary study provides us a valid foundation to continue working in this new and promising CAD field to develop and optimize highly performed and robust risk prediction schemes that may have potential to eventually assist clinicians in more accurately identifying the patients with a higher risk of lung cancer recurrence after surgery. Therefore, for these high risk patients, the post-surgery chemotherapy is required to prevent or minimize the risk of cancer recurrence and thus increase their disease-free survival and overall survival time.

## REFERENCES


[1] R. Siegel, *et al.*, "Cancer statistics, 2015," *CA Cancer J Clin.*, vol. 65, pp.5-29, Jan/Feb., 2015.

[2] S. Swensen, *et al.*, "CT screening for lung cancer: five-year prospective experience", *Radiology*, vol. 235, no. 235, pp. 259-265, APR. 2005.

[3] P. Bach, *et al.*, "Computed tomography screening and lung cancer outcomes", *J. Amer. Med. Assoc.*, vol. 297, no. 9, pp. 953-961, Mar. 2007.

[4] L. Humphrey, *et al.*, "Screening for lung cancer with low-dose computed tomography: a systematic review to update the U.S. Preventive Services Task Force Recommendation", *Ann. Intern. Med.*, vol. 159, no. 6, pp. 411-420, Sep. 2013.

[5] 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.

[6] American Cancer Society, (2014, Nov.). *Non-small cell lung cancer survival rates by stage*. http://www.cancer.org/cancer/lungcancer-nonsmallcell/detailedguide/non-small-cell-lung-cancer-survival-rates

[7] 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.

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

[9] J.C. Soria, "ERCC1-tailored chemotherapy in lung cancer: the first prospective randomized trial", *J. Clin. Onc.*, vol. 25, pp. 2648-2649, 2007.

[10] 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.

[11] C. Poleri, *et al.*, "Risk of recurrence in patients with surgically resected stage I non-small cell lung carcinoma: histopathologic and immunohistochemical analysis," *Chest* vol. 123, no. 6, pp. 1858-1867, Jun. 2003.

[12] G. Loannidis, *et al.*, "How close are we to customizing chemotheraphy in early non-small-cell lung cancer," *Ther. Adv. Med. Oncol.*, vol. 3, no. 4, pp. 185-205, Jul. 2011.

[13] E. P. Diamandis, "Mass spectrometry as a diagnostic and a cancer biomarker discovery tool: opportunities and potential limitations," *Mol Cell Proteomics*, vol. 3, no. 4, pp. 367-378, Apr. 2004.

[14] H. Li, *et al.*, "Cost-effectiveness of a novel molecular test for cytologically indeterminate thyroid nodules," *J .Clin. Endocrinol Metab.*, vol. 96, no.11, pp. E1719-1726, Nov. 2011.

[15] R. Wahl, *et al.*, "Staging of mediastinal non-small cell lung cancer with FDG PET, CT, and fusion images: preliminary prospective evaluation", *Radiology*, vol. 191, no. 2 pp. 371-377, May. 1994.

[16] D. Lardinois, *et al.*, "Staging of non-small cell lung cancer with integrated positron-emission tomography and computed tomography", *N. Engl. J. Med.*, vol. 348, no. 25, pp. 2500-2507, Jun. 2003.

[17] K. Hisanobu, *et al.*, "Magnetic resonance imaging for lung cancer", *J. Thoracic Imaging*, vol. 28, no. 3, pp. 138-150, May. 2013.

[18] E. Eisenhauer, *et al.*, "New response evaluation criteria in solid tumors: Revised RECIST guideline (version 1.1)", *Euro. J. Cancer*, vol. 45, no. 2, pp. 228-247, Jan. 2009.

[19] J.O'conner, *et al.*, "Quantitative imaging biomarkers in the clinical development of targeted therapeutics: current and future perspectives," *The Lancet Oncology*, vol. 9, no. 8, pp. 766-776, Aug. 2008.

[20] H. Aerts, *et al.*, "Decoding tumor phenotype by noninvasive imaging using a quantitative Radiomics approach," *Nat. Commun.*, vol. 5, pp. 4006-1-8, Aug. 2014.

[21] J. Kuhnigk, *et al.*, "Morphological segmentation and partial volume analysis for volumetry of solid pulmonary lesions in thoracic CT scans", *IEEE Tran.s Med. Imag.*, vol. 25, no. 4, pp. 417-434, Apr. 2006.

[22] S. Iwano, *et al.*, "Semi-automatic volumetric measurement of lung cancer using multi-detector CT", *Acad. Radiol.*, vol. 16, no. 10, pp. 1179–1186, Oct. 2009.

[23] B. Zheng, *et al.*, "A method to improve visual similarity of breast masses for an interactive computer-aided diagnosis environment," *Med. Phys.* vol. 33, no. 1, pp.111-117, Jan. 2006.

[24] I. Witten. (2011). *Data mining: practical machine learning tools and techniques*. Available: http://www.cs.waikato.ac.nz/ml/weka/.

[25] E. Burns, *et al.*, "Best-first heuristic search for multicore machines," *J. Artificial Intelligence Research* vol. 39, pp. 689-743, Dec. 2010.

[26] J. Schrader, et al., "Matrix stiffness modulates proliferation, chemotherapeutic response, and dormancy in hepatocellular carcinoma cells," *Hepatology* vol 53, no. 4, pp.1192-1205, Apr. 2011

[27] M. Haidekker, *Image Analysis: A Perspective, in Advanced Biomedical Image Analysis*. Hoboken: Wiley, 2010.

[28] X. Wei. *Gray Level Run Length Matrix Toolbox v1.0*. Available: http://www.mathworks.com/matlabcentral/fileexchange/17482-graylevel-run-length-matrix-toolbox. Nov. 2007.

[29] T. M. Mitchell, *Machine Learning*. Boston, MA: WCB/McGraw-Hill, 1997.

[30] C.E. Kahn, *et al.*, "Construction of a Bayesian network for mammographic diagnosis of breast cancer", *Comput. Biol. Med.*, vol. 27, no. 1, pp. 1929-1940, Jan. 1997.

[31] X. Wang, *et al.*, "Computer-assisted diagnosis of breast cancer using a data-driven Bayesian belief network", *International J. Med. Informatics*, vol. 54, no. 2, pp. 115-126, May. 1999.

[32] Q. Li, K. Doi, "Reduction of bias and variance for evaluation of computer-aided diagnosis schemes," *Med. Phys.*, vol. 33, no. 4, pp. 868-875, Apr. 2006.

[33] N. V. Chawla, *et al.*, "SMOTE: synthetic minority over-sampling technique," *J. Artificial Intelligent Research* vol. 16, pp. 321-357, Jun. 2002.

[34] D. Lederman, *et al.*, "Improving breast cancer risk stratification using resonance-frequency electrical impedance spectroscopy through fusion of multiple classifiers," *Annals of Biomed. Engr.*, vol.39, no. 3, pp. 931-945, Mar. 2011.

[35] S. Park, *et al.*, "Computer-aided detection of early interstitial lung diseases using low-dose CT images," *Phys. Med. Biol.*, vol. 56, no. 2, pp. 1139-1153, Jan. 2011.

[36] S. Park, *et al.*, "A multistage approach to improve performance of computer-aided detection of pulmonary embolisms depicted on CT images: Preliminary investigation," *IEEE Trans. Biomed. Eng.*, vol. 58, no. 6, pp. 1519-1527, June 2011.

[37] M. Tan, *et al.*, "A novel computer-aided lung nodule detection system for CT images," *Med. Phys.*, vol. 38, no. 10, pp. 5630-5645, Sep. 2011.

[38] D. Butcher, *et al.*, "A tense situation: forcing tumor progression," *Nat. Rev. Cancer*, vol. 9, no. 2, pp. 108–122, Feb. 2009.

[39] F. Collins, H. Varmus, "A new iniative on precision medicine," *N. Engl. J. Med*, vol. 372, no. 9, pp. 793-795, Feb. 2015.