

## Monitoring Pulmonary Fibrosis by Fusing Clinical, Physiological, and Computed Tomography Features

Jesus J Caban<sup>1</sup>, Jianhua Yao<sup>2</sup>, Ulas Bagci<sup>2,3</sup>, and Daniel J Mollura<sup>2,3</sup>

**Abstract**—Advances in medical imaging and screening tests have made possible the detection and diagnosis of many diseases in their early stages. Those advances have enabled more effective planning, execution, and monitoring of a treatment plan. However, early detection has also resulted in an increase of the number of longitudinal radiographs requested for most patients, thus increasing the risk for potential long-term effects of ionizing radiation exposure and increasing the cost associated with a specific treatment plan. The aim of this paper is to study the associations between clinical measurements and quantitative image features in patients with pulmonary fibrosis. The association between these multi-modal features could be used to more accurately determine the state of the disease and could potentially be used to predict many of the longitudinal image features when CT images are not available. Our results show how textural image features are highly correlated with the severity of fibrosis, how clinical variables can be combined to monitor progression, and how simple blood features can be used to predict statistical image attributes of the lungs.

### I. INTRODUCTION

DURING the last ten years, advances in imaging acquisition devices and systems biology have helped physicians and radiologists to more effectively diagnose diseases in their early stages. With the use of high-resolution imaging devices and computer-aided diagnosis (CAD) systems, radiologists are now able to quantify, measure and diagnose diseases from very subtle abnormal patterns. Along similar lines, medical research has sought to discover biomarkers for measuring disease progression and recurrence. In this quantification of biomarkers and imaging features

raises the possibility of integrating these variables for optimizing the yield of diagnostic information. In addition to heightening the utility of imaging via computer-based fusion of imaging features with biomarker measurements, adverse attributes of radiographic examinations (such as cost and exposure to ionizing radiation [1]) could potentially be reduced.

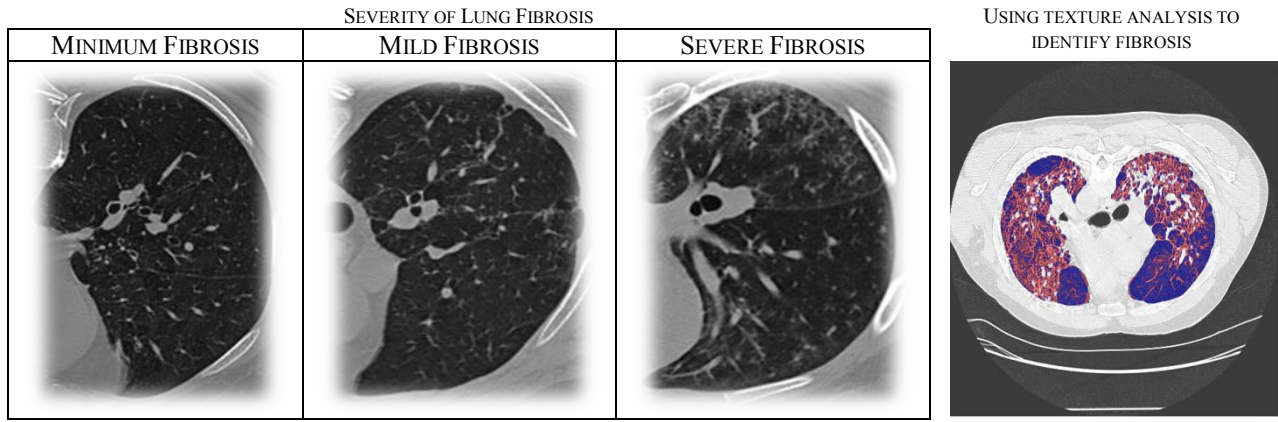
Recently, advances in systems biology have shown how biomarkers can be used to better support diagnostic and therapeutic decisions. Specifically, we have seen how simple blood tests can aid in monitoring, assessing, and testing for early detection of lung cancer in high-risk patients. Most of the existing protocols use blood samples to detect autoantibodies which are often involved in the development of lung cancer [2].

Although it has been found that many biomarkers are highly correlated with cancer, it is still an open problem to determine physiological biomarkers that are correlated with other lung diseases such as pulmonary fibrosis. Pulmonary fibrosis, characterized by chronic scarring of lung tissue and development of excess fibrotic tissue, is a potentially life threatening and, as of now, an incurable disease. Annually, an estimated 40,000 people die of pulmonary fibrosis and over 128,000 are affected by this disease in the United States [3]. Despite many advances in the detection of pulmonary disease, radiologists mainly rely on images to diagnose and monitor progression of fibrosis. Figure 1 illustrates some of the patterns often seen among patients with pulmonary fibrosis.

Recently, researchers have been exploring new ways to detect pulmonary fibrosis. Kumánovics et al. demonstrated how physiological measurements such as the levels of serum are correlated with lung fibrosis [4]. Depeursinge et al. showed how clinical and visual features could be fused to

Manuscript received April 15, 2011. This research is supported in part by the Imaging Sciences Training Program (ISTP), the Center for Infectious Disease Imaging Intramural program in the Radiology and Imaging Sciences Department of the NIH Clinical Center, the Intramural Program of the National Institutes of Allergy and Infectious Diseases, and the Intramural Research Program of the National Institutes of Bio-imaging and Bioengineering.

<sup>1</sup>National Library of Medicine (NIH), <sup>2</sup>Department of Radiology and Imaging Sciences (CC/NIH), <sup>3</sup>Center for Infectious Disease Imaging (CC/NIH), (e-mail: jesus dot caban at nih.gov)



**Figure 1:** (Left) A set of images highlighting the differences between each of the different stages. Note that as pulmonary fibrosis progresses, more fibrotic tissue appears within the lungs. Computer-aided diagnosis systems often use those patterns to quantify and score the severity of the disease. (Right) To test the effectiveness of our textural features, a model was trained to automatically identify fibrosis.

enhance image classification models [5]. Although radiological features are important for assessing pulmonary fibrosis, integration of quantified imaging variables with clinical variables could further optimize the diagnostic process.

The aim of this paper is to study the associations between clinical and image features and to explore how physiological attributes could be used to predict some of the statistical image patterns often seen in patients with lung fibrosis. If by using a set of physiological markers we could predict some of the statistical image patterns (not the actual image), we could help radiologists better determine the progression of a particular disease when a CT study is not available and better justify when a radiology scan is needed.

## II. APPROACH

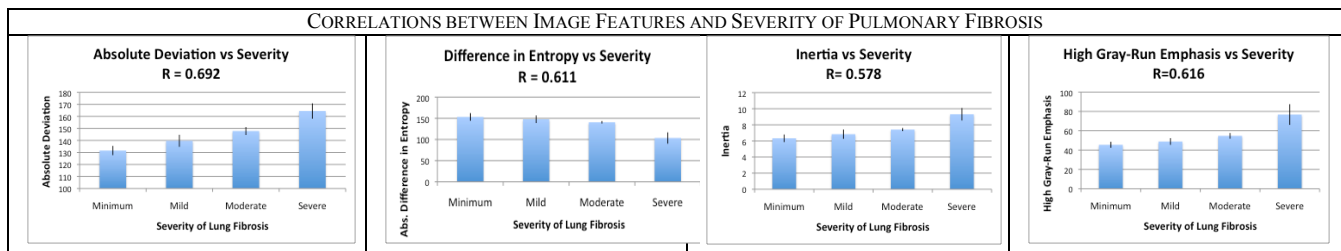
To study the associations between different multi-modal features, a study has been carried out in which blood work, pulmonary function tests (PFTs), and EKG studies were performed on 66 subjects within five days of capturing a high resolution CT (HRCT) of their lungs. Each of the 66 patients had a history of pulmonary fibrosis. A group of expert radiologists scored the severity of each patient by using a previously reported method and labeled the severity of the disease as being: *none*, *minimum*, *mild*, *moderate*, and *severe*.

The first step of our project was to determine if textural features are effective image descriptors for the task of identifying pulmonary fibrosis. To

accomplish this, a clinically validated fuzzy-connectedness segmentation approach was used to segment the lungs [6]. A multi-scale texture analysis approach was then used to obtain 26 quantitative image features describing the statistical properties of the lung regions. The extracted image descriptors were features from histogram statistics, co-occurrence matrices, and run-length matrices. Histogram statistics, also known as first-order statistics, are those features that can be extracted from an image histogram including mean, standard deviation, skewness, variance, and kurtosis [7]. Co-occurrence matrices are textural representations that can be used to capture second-order statistics such as entropy, energy, contrast, inertia, and many others [8]. Finally, run-length matrices are one of the techniques that can be used to capture high-order statistics such as the frequency of short or long runs of pixels with similar intensity values [9].

The second step was to estimate a set of clinical values that are associated with the severity of pulmonary fibrosis. Note that, as of today, there is not a specific biomarker that can be used to diagnose fibrosis, instead physicians use a set of clinical attributes known to be somehow related with lung disease and the radiologist report to determine progression. In our case, clinical and physiological variables were treated as soft or weak markers.

Finally, in the case that an imaging study is not available or cannot be performed, we studied how



**Figure 2:** A subset of the image features that were found to be highly correlated with the severity of pulmonary fibrosis.

uni- or multi-variate models of clinical attributes can be used to estimate some of the quantitative image features characteristic of pulmonary fibrosis.

### III. RESULTS

To capture the differences, significance, and correlations between individual features, multiple statistical analysis methods were used including t-tests, ANOVA, Pearson correlation, and linear regressions.

#### A. Correlation: Image Features and Severity

When analyzing the correlation between image features and the severity of pulmonary fibrosis, multiple textural features were found highly correlated. Figure 2 shows some of those associations. *Absolute deviation* from first order statistics was found highly correlated. This association is the results of an increase in the dispersion from the mean intensity value as the disease progresses. Within second order statistics, *inertia* and *entropy* (among others) were also found to be highly correlated. This is, as fibrosis progresses, the areas of the lungs are less uniform/similar, thus causing a larger difference in entropy. From run-length matrices, *high gray-run emphasis* – a measure of the frequency of observing gray runs (i.e. fibrotic tissue) – was shown to be highly correlated.

These findings and correlations confirm some of the previous work in detection of fibrosis using texture analysis [10].

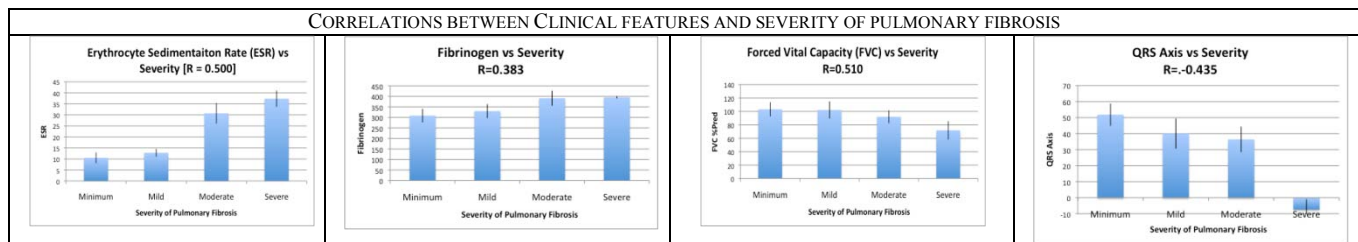
To validate our image features, a support vector machine (SVM) model to automatically identify fibrotic regions was created. A 90% accuracy was observed when using statistical textural features estimated from 8x8 patches. Figure 1(right) shows the results of an automatic identification process

of a severe case. The fibrotic tissue is automatically identified and highlighted in red.

#### B. Correlation: Clinical values and Severity

Now that we have determined that texture analysis is useful for quantifying and measuring pulmonary fibrosis, we wanted to explore which clinical and physiological value are associated with the severity of pulmonary fibrosis. At this step, multiple significant ( $p < 0.01$ ) correlations were found, but -- as expected -- the correlations were not as strong as when using image features given that (as of today) pulmonary fibrosis does not have unique biomarkers. Figure 3 shows some of our results. *Erythrocyte Sedimentation Rate* (ESR) -- a blood value that measurements inflammation - was found correlated. Another blood feature found correlated with fibrosis disease was *Fibrinogen*, a measure of the blood clotting ability. Multiple values obtained from pulmonary function tests were found correlated including *Forced Vital Capacity* (FVC) -- the volume of air that can be blown out after full inspiration and *Total Lung Capacity* (TLC) which measure the maximum volume of air present in the lungs. When analyzing the correlation with EKG, multiple values were correlated including the QRS axis of the heart.

Since the clinical features are relative weak markers of disease, they are not generally used independently to predict the stage of pulmonary fibrosis. They can instead be combined in a multi-variate fashion and used to better estimate and/or monitor the severity of pulmonary fibrosis. By combining the four weak markers shown in Figure 3, we were able to create a multi-variate model with correlation coefficient of  $R=0.791$  and  $R^2 = 0.626$ . This shows that multi-variate models of weak clinical markers could potentially be used to accurately predict the disease when image studies



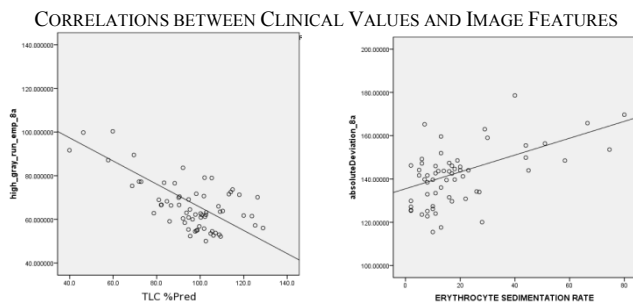
**Figure 3:** A subset of the clinical and physiological features that were found to be correlated with the severity of pulmonary fibrosis. In our framework, such features are used as weak markers to create multi-variate models capable of predicting the stage of the disease.

are not available.

### C. Correlations: Clinical Values and Statistical Images Features of Lungs with Fibrosis

In situations where CT images are not available or can't be captured, a set of multi-modal features could be used to provide general statistical information about the scarring within the lungs.

By analyzing some of the correlations between clinical variables and quantitative image features, we found many associations. Figure 4 shows the linear regression for two correlated variables. First, in Section 3(A) we saw that the frequency of high gray-runs (HGR) is highly correlated with the amount of scarring in the lungs. Now, when correlating clinical and CT features we found that the Total Lung Capacity (TLC) has a good correlation ( $R = -0.567$ ) with the frequency of observing high gray-runs within the lungs.



**Figure 4:** (left) Linear regression between TLC and the frequency of high gray-run within the lungs. (right) Linear regression between ESR and absolute deviation of the image densities.

Figure 4(right) also shows that there's a good correlation ( $R = 0.402$ ) between the blood value ESR and the deviation of the intensity within the lungs.

These results underscore the importance of using multi-modal features and justify the possibility of using clinical features to predict

some general statistical properties of the image. When combining multiple features such as QRS, ESR, and TLC, we were able to create a model to predict HGR with  $R^2=0.690$ .

## IV. CONCLUSION

Our study presents multiple associations between clinical laboratory data and quantified CT data, which can be statistically integrated for assessing patients with pulmonary fibrosis. This method of connecting laboratory biomarkers and radiologic data could potentially increase the yield and usefulness of diagnostic information in measuring disease progression. Further research will seek to expand/improve the statistical image features and biomarkers extracted for these correlations to improve the predictive utility of this method.

## REFERENCES

- [1] F. Korley; J. Pham; T. Kirsch, "Use of Advanced Radiology During Visits to US Emergency Departments for Injury-Related Conditions, 1998-2007", *JAMA*. 2010;304(13):1465-1471.
- [2] L. Peek, S. Lam, G. Healey, H. A. Fritsche, C. Chapman, A. Murray, P. Maddison, J. F. Robertson and W. Wood, "Use of serum autoantibodies to identify early-stage lung cancer."
- [3] The Coalition for Pulmonary Fibrosis (CPF), 2010 [www.coalitionforpf.org](http://www.coalitionforpf.org)
- [4] G. Kumánovics, T. Minier, J. Radics, L. Pálincás, T. Berki, L. Czirják, Comprehensive investigation of novel serum markers of pulmonary fibrosis associated with systemic sclerosis and dermato/polymyositis, *Clin Exp Rheumato*, Volume 26, No. 3, pp. 414-420, 2008.
- [5] A. Depeursinge et al, "Fusing visual and clinical information for lung tissue classification in high-resolution computed tomography", *Artificial Intelligence in Medicine*, vol. 50, pp 13-21, 2010
- [6] Saha, P.K., et al., "Scale-based fuzzy connected image segmentation: Theory, algorithms, and validation". *Computer Vision Image Understanding*, 77, 145-174 (2000).
- [7] Crosier, M. and L.D. Griffin, "Using Basic Image Features for Texture Classification". *International Journal of Computer Vision*, 88(3): p. 447-460 (2010)
- [8] R Haralick and K Shanmugam and I Dinstein, "Textural Features for Image Classification", *IEEE Transactions on Systems, Man, and Cybernetics*, 3(6), pp. 610-612 (1973)
- [9] Tang, X.O., Texture information in run-length matrices. *IEEE Transactions on Image Processing*, 7(11), p: 1602-1609 (1998)
- [10] M. K. Pietikainen, "Texture Analysis in Machine Vision", World Scientific Publishing Company, 2000