The Lung Tissue Research Consortium: An extensive open database containing histological, clinical, and radiological data to study chronic lung disease

David R Holmes III[1], Brian J Bartholmai[2], Ronald A Karwoski[1], Vanessa Zavaletta[1], and Richard A Robb[1]

August 17, 2006

[1]Biomedical Imaging Resource, Mayo Clinic
[2]Department of Radiology, Mayo Clinic

Abstract

The Lung Tissue Research Consortium (LTRC) is an NHLBI sponsored project which is studying patients with Chronic Obstructive Pulmonary Disease (COPD) and idiopathic pulmonary fibrosis (IPF) by creation of a large public repository of histological, radiological, and clinical data. At completion, over 1200 subjects will be enrolled. The goal is to have comprehensive data for the vast majority of enrollees, including a volumetric high-resolution CT of the chest, extensive clinical history and questionnaire results, pulmonary function testing, genetic and laboratory testing, stored serum, blood and lung tissue. The pathological specimens and CT scans will each have a corresponding structured semi-quantitative report and coded diagnostic assessment.. The CT scan reports will include subjective assessment of the regional distribution of specific named radiologic signs and visual characteristics. As required by the NHLBI, the data collected through the LTRC be will be anonymized and available to qualified independent researchers, upon request. The LTRC imaging database will be one of the largest single databases of lung CT data available to date, will have many CT imaging studies with very high volumetric resolution (isotropic 1.25mm pixel size or less) and the methods of image acquisition, processing and multicenter quality control will serve as a model for imaging studies of chronic lung disease for several years to come. It is important that the medical image analysis community investigate this database with the intent to develop new methods of registration, segmentation, and classification that will be used to diagnosis and treat patients with diffuse lung disease.

1 The Lung Tissue Research Consortium

In March 2003, the National Heart, Lung, and Blood Institute (NHLBI) released an announcement for the development of a Lung Tissue Research Consortium (LTRC) [1]. As stated in the RFA, the objective of the Lung Tissue Resource Consortium (LTRC) is to enable better management of lung diseases by increasing understanding of the pathogenetic mechanisms of these diseases. In particular, the RFA targets the collection of data for Chronic Obstructive Pulmonary Disease (COPD) and idiopathic pulmonary fibrosis (IPF). COPD alone afflict over 15 million Americans and are a major cause of morbidity and mortality [2].

In the request, NHBLI solicited the participation of several sites to manage different aspects of the LTRC. Clinical centers would recruit participants for the LTRC, collect clinical data, radiologic imaging, physiologic data and tissue/blood specimens. In addition, each clinical center would perform research on the data collected through the LTRC. A pathology core would define the tissue sampling protocols and train the clinical centers to acquire the samples. Tissue samples would be processed by the pathology core and an independent diagnostic assessment would be made by the core lab investigators as well. The radiology core laboratory (RCL) would define the image scanning protocols used to collect LTRC CT data, implement procedures to assure the quality and consistency of data acquired as well as provide expert subspecialty thoracic radiologist review of the image. Specialized software developed by the RCL would also provide quantitative assessment of the lung image data. A Data coordination center would provide oversight of the patient recruitment, data collection and coordination for the re-distribution of collected data to other researchers. The LTRC was awarded as follows. The NIH/NHLBI study chairman of the project is Dr. Robert Wise from Johns Hopkins. The Pathology Core Laboratory is headed by Dr. Carlyne Cool of University of Colorado. The Radiology Core Laboratory is directed by Dr. Brian Bartholmai from Mayo Clinic Rochester. The four clinical centers of the project are University of Colorado (Dr. Marvin Schwarz), Mayo Clinic (Dr. Andrew Limper), University of Michigan (Dr. Fernando Martinez), and University of Pittsburgh (Dr. Frank Sciurba). The Clinical Trials and Surveys Corporation is the Data Coordinating Center (Bruce Thompson, Ph.D).

Members of the LTRC are actively studying the data to develop a better understanding the chronic lung diseases; however, because the database has broad application ranging from disease progress through diagnosis, distribution of the data to other investigators is important. Accordingly, NHLBI has required that this database be open to other researchers. Opening the database will allow independent investigators to study all aspects of the data in hopes that new approaches to the diagnosis and treatment of chronic lung disease can be developed. One important component of the LTRC project is the vast image database that will be available to the research community to further understand how lung disease is manifested in image data. In addition, the large collection of image data will provide the foundation for the development of new image analysis methodologies. Differing analysis techniques by different laboratories may be validated through direct comparison of results and correlation with physiologic, historical, pathological, and biochemical data as well as the visual radiological assessment of disease provided by the RCL.

2 Methods

The LTRC radiology core has been tasked with the creating a repository of consistent, high-quailty, high-resolution correlative imaging data from subjects in the cohort. Computed tomography (CT) is commonly used in the evaluation of lung disease, and can demonstrate some of the structural changes in the lung associated with pulmonary disease. Emphysema, for example, is parenchymal distruction of the lung tissue, which is replaced by non-functional air spaces. Radiographically, these abnormal regions are identified in CT data as voxels with a density value less than that of normal lung, with typically a threshold of = -950 Hounsfield Units (HU) for high-resolution CT images less than 2mm thick[3]. To obtain CT datasets on LTRC subjects that would potentially visualize small anatomic abnormalities as well as abnormalities due to regional changes in pulmonary elasticity, air trapping or be potentially obscured by gravitational effects, the recommended CT scan protocol is a Three Phase Volumetric high-resolution CT (HRCT) scan of the chest. This examination involves acquisition of CT images in the supine position during a single suspended/held full inspiration, at suspended full expiration, and in the prone position during full inspiration. Images are acquired on scanners with 8 detectors or more in less than 18 seconds for each phase, and reconstructed at 1.0 to 1.25 mm thickness with 50% overlap with a high-frequency sparing algorithm. There are considerable variations in in-plane spatial resolution, edge enhancement and frequency response characteristics between

different scanner models and brands, but imaging characteristics for each scanner were taken into account and scanner-specific protocol changes were utilized to make data acquisition as consistent as possible across the LTRC sites.

The the LTRC Steering Committee and Clinical Centers intend that, whenever possible, subjects will be scanned using a full LTRC Three Phase Volumetric HRCT. These data will provide high-quality contiguous overlapping data of the chest and use of a non-edge-enhancing kernel should preserve HU accuracy over both large volumes and at the borders of anatomic structures in the lung where there may be high density gradients . If a full three-phase protocol cannot be obtained due to limitations of scanner technology available, a second Limited Three-Phase LTRC protocol might be obtained. This imaging protocol utilizes similar three-phase acquisition and reconstruction algorithms, but 1.0 to 1.25mm images are be obtained non-contiguously, spaced 10mm apart. If neither LTRC-specific protocol could be obtained for whatever reason, another historical chest CT obtained within 6 months prior to tissue sampling, if available, would be included in the repository.

Due to normal scanner-to-scanner variations and changes in even the same CT scanner characteristics over time, a multi-center CT QA protocol was established. This QA program requires scanning an ACR accredited phantom (Gammex 464 Phantom) with a standardized protocol to evaluate image quality with respect to CT number accuracy, slice width/thickness, image uniformity, image noise, and in-plane spatial resolution of the reconstruction algorithm. Each acquisition center is required to complete the QA protocol on each scanner utilized for LTRC scan acquisition on a regular basis.

Through the core lab, the collected data is evaluated by expert radiologists and a radiology report is included in the repository. This structured report contains semi-quantitative reporting of 15 visually characteristic abnormalities of the lung that may correspond to specific pulmonary pathology or regional involvement of they lungs by disease. These visual features are determined by a thoracic radiologist, and include air trapping, bronchial thickening, bronchiectasis, bullae, consolidation, crazy paving pattern, emphysema, ground glass infiltrate, honeycombing, micronodules ($<$5mm), mosaic attenuation, pulmonary cysts, reticular infiltrates, septal thickening and tree-in-bud pattern. These features are graded from none/normal to severe in a 5-step semi-quantitative method and the distribution of these findings are noted for each anatomic lobe of the lung and the central vs. peripheral region of each lobe separately. Also, the general distribution of emphysema and/or interstitial lung disease throughout the lung is noted. In addition, the reviewer provides overall presumed diagnoses and additional anomalies and other ancillary findings including the size and location of pulmonary nodules or masses noted in the data. A structured form is used to record the radiological findings in a consistent manner, and these findings are stored in a database for later recall or analysis. In addition, specific quantitative measures of pulmonary pathology performed by the core lab include emphysema quantification, lung volume analysis (with lobar, central and peripheral segmentation), tracheal measurements (cross-sectional area and wall thickness) and pulmonary parenchymal analysis through experimental texture- and histogram-based techniques designed to detect and quantify specific types of pulmonary pathology (such as honeycombing, ground glass infiltrates, reticular opacities etc.).

3 Results

As of May 2006, over 425 subjects were enrolled in the LTRC. Of all of the subjects, approximately 42% have been scanned according to the Full Three-Phase LTRC protocol (Figure 1B). A quantitative analysis of the high-resolution CT data is underway at Mayo. This analysis collects specific measurements from each of the datasets. The measurements include a volume of emphysema, lung/lobe volume measurements, tracheobronchial cross-sectional and wall thickness measurements, and amount of detectable pulmonary

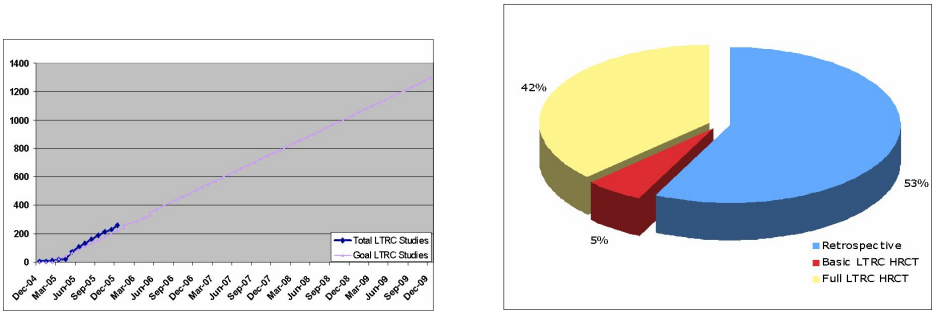

Figure 1: Enrollment for the LTRC is on schedule to include more than 1200 subjects by the end of 2009. Currently, there are over 230 subjects enrolled. Over 40% of patients enrolled thus far have been scanned with the full LTRC three phase protocol.

pathology including emphysema. The analysis workflow includes semi-automatic anatomic segmentation of the trachea, lungs, and lobes. In addition, the central and peripheral regions are defined. Emphysema regions are automatically segmented using a threshold = -950 HU. Measurements of texture are an active area of research.

4 Discussion

The data coordination center was specifically setup to oversee the distribution of data to researchers. The LTRC database will provided the most comprehensive collection of data on COPD and IPF patients to date. It includes histological, radiological, and clinical data. Currently, there are a limited number of CT scans and pathology slides available at the website (www.ltrcpublic.com); however, full datasets should be available by the end of 2006. A subset of the entire database will have public real-time query capabilities for investigators to identify particular subsets of subjects for study by diagnoses or other clinical variables. The process to request particular sets of data or tissue specimens requires a brief application process and review of the proposed experimental protocol by the LTRC Protocol Review Committee, and the application material can be found at www.ltrcpublic.com/forms.

The data available in the LTRC can be used to study many aspects of chronic pulmonary disease. From an image analysis perspective, the database provides an extensive collection of image datasets which can be utilized for image registration, automatic segmentation of anatomy, and texture analysis of lung tissue for disease classification. The availability of multiple patient postitions and phases of respiration may allow for interrogation of regional anatomic changes during respiration, due to change of position or perhaps detection of changes from normal in areas of known pulmonary pathology. Mayo researchers [4], for example, have studied the functional variability of these patients in a limited number of datasets (Figure 2A). Zavaletta and colleagues [5] are currently investigating texture analysis methods for differentiating different forms of diffuse disease (Figure 2B). One unique advantage of this database is that several forms of gold standards are available.. Specifically tissue samples, radologist interpretation of findings, pathological diagnosis, accepted clinical diagnosis and numerous physiological parameters are also available and can be used to validate classification schemes.

Open image databases provide a common collection of data for investigators to study. Each new investigation of the data can provide new insight into both imaging methodologies and specific diseases (as manifested through the image data). Broad databases, such as the NLM-Mayo Image Collection [6], are

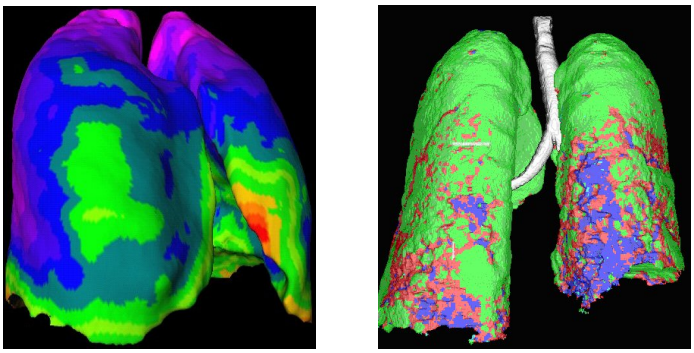

Figure 2: On the left, a color coded rendering of the lungs shows regions of high (green) and low (magenta) deformation through the respiratory cycle. On the right, a dataset from the LTRC has been classified according to common texture features consistent with diffuse lung disease.

a collection of diverse data which is useful for developing robust tools. In contrast, the LTRC database is a very extensive database of patients with specific lung disease which will be useful for validating the sensitivity and specificity of an imaging algorithm. As the collection grows over the next five years, it will be able to provide consistent, high-quality images of a broad range of diffuse lung diseases which might provide the basis for the development of new image analysis methodologies.

5 Acknowledgments

The authors would like to acknowledge all of the leadership and investigators associated with the LTRC project. Their combined effort will generate the foundation for future research. The authors would also like to recognize the Mayo participants in the development and support of the LTRC Radiology Core. Both members of Radiology and the Biomedical Imaging Resource bring unique expertise to this project. This work has been funded by the NIH/NHLBI Grant # N01-HR-46158.

References

[1] National Heart, Lung, and Blood Institute (NHLBI), RFP ANNOUNCEMENT: LUNG TISSUE RESEARCH CONSORTIUM, http://grants.nih.gov/grants/guide/notice-files/NOT-HL-03-009.html, 2003

[2] http://www.nlm.nih.gov/medlineplus/copdchronicobstructivepulmonarydisease.html

[3] Bergin C. Muller N. Nichols DM. Lillington G. Hogg JC. Mullen B. Grymaloski MR. Osborne S. Pare PD. The diagnosis of emphysema. A CT-pathologic correlation. Am Rev Resp Dis. 133(4):541-6, 1986.

[4] Haider CR. Bartholmai BJ. Holmes DR. Camp JJ. Robb RA. Quantitative characterization of lung disease. Medical Imaging Graphics. 29(7):555-63, 2005 Oct.

[5] Zavaletta, V., Karwoski, R.A., Bartholmai, B., Robb, R.A.: High Resolution Multi-Detector CT Aided Tissue Analysis and Quantification of Lung Fibrosis. Proc. SPIE Medical Imaging 2006, 2006.

[6] Holmes, D.R. III, Workman, E.L., Robb, R.A.: The NLM-Mayo Image Collection: Common Access to Uncommon Data. Medical Image Computing & Computer Assisted Intervention (MICCAI 2005), ISC/NAMIC/MICCAI Workshop on Open Source Software, 2005.

The Lung Tissue Research Consortium: An extensive open database containing histological, clinical, and radiological data to study chronic lung disease

David R Holmes III[1], Brian J Bartholmai[2], Ronald A Karwoski[1], Vanessa Zavaletta[1], and Richard A Robb[1]

July 10, 2006

[1]Biomedical Imaging Resource, Mayo Clinic
[2]Department of Radiology, Mayo Clinic

Abstract

The Lung Tissue Research Consortium (LTRC) is an NHLBI sponsored project which is studying histological, radiological, and clinical data on patients with Chronic Obstructive Pulmonary Disease (COPD) and idiopathic pulmonary fibrosis (IPF). At completion, over 1200 subjects will be enrolled. Each subject will have a minimum of one CT scan; however, many will likely have 3 CT scans. Each scan will have a corresponding radiological report, clinical report, and histological report. The NHLBI requires that the data collected through the LTRC be anonymized and available to independent researchers. The LTRC imaging database will be the most comprehensive database of lung CT data available to date and will serve as the basis for imaging studies of chronic lung disease for years to come. It is a significant opportunity for the medical image analysis community to investigate this database with the intent to develop new methods of registration, segmentation, and classification that may be used to diagnosis and treat patients with diffuse lung disease.

1 The Lung Tissue Research Consortium

In March 2003, the National Heart, Lung, and Blood Institute (NHLBI) released an announcement for the development of a Lung Tissue Research Consortium (LTRC) [1]. As stated in the RFA, the objective of the Lung Tissue Resource Consortium (LTRC) is to enable better management of lung diseases by increasing understanding of the pathogenetic mechanisms of these diseases. In particular, the RFA targets the collection of data for Chronic Obstructive Pulmonary Disease (COPD) and idiopathic pulmonary fibrosis (IPF). COPD alone afflict over 15 million Americans and are a major cause of morbidity and mortality [2].

In the request, NHBLI solicited the participation of several sites to manage different aspects of the LTRC. Clinical centers would recruit participants for the LTRC. In addition, each clinical center would perform research on the data collected through the LTRC. A pathology core would define the tissue sampling protocols and train the clinical centers to acquire the samples. Tissue samples would be processed by the pathology core as well. The radiology core would define the image scanning protocols used to collect LTRC CT data

as well as provide radiological review of the data and quantitative assessment. The Data coordination center would provide oversight of the data collection and recruitment. The data would also be coordinated for distribution to other researchers.

The LTRC was awarded as follows. The study chairman of the project is Dr. Robert Wise from Johns Hopkins. The Pathology Core Laboratory is headed by Dr. Carlyne Cool of University of Colorado. The Radiology Core Laboratory is directed by Dr. Brian Bartholmai from Mayo Clinic Rochester. The four clinical centers of the project are University of Colorado, Mayo Clinic, University of Michigan, and University of Pittsburgh. A company, Clinical Trials and Surveys Corporation, is the Data Coordinating Center.

Members of the LTRC are actively studying the data to develop a better understanding of chronic lung diseases; however, because the database has broad application ranging from disease progression through diagnosis, distribution of the data to other investigators is important. Accordingly, NHLBI has required that this database be open to other researchers. Opening the database will allow independent investigators to study all aspects of the data in hopes that new approaches to effective diagnosis and treatment of chronic lung disease can be developed. One significant feature of the LTRC project is the vast image database that will be available to the research community to further understand how lung disease is manifested in image data. In addition, the large collection of image data will provide the foundation for the development of new image analysis methodologies.

2 Methods

The LTRC radiology core has been tasked with creating a high-quality repository of CT data from subjects in the cohort. CT is commonly used in the evaluation of lung disease. Emphysema, for example, is radiographically identified in CT data as voxels with a value = -950 Hounsfield Units (HU) [3]. The core lab has determined that, whenever possible, subjects will be scanned using a full LTRC Three Phase Volumetric CT. The protocol includes a supine full inspiration, a supine full expiration, and a prone inspiration CT scan. The data will be contiguous with a slice thickness less than 1.25mm and 50% overlap between slices. The data will be reconstructed using a non-edge-enhancing kernel which preserves HU accurately. If a full three-phase protocol cannot be obtained, a second protocol was defined for a limited high-resolution CT scan of the chest with 1mm thick slices spaced 10mm apart. Other historical chest CT obtained within 6 months prior to tissue sampling would be included in the repository.

Due to scanner variation, a multi-center CT QA protocol was established. The QA program required scanning an ACR accredited phantom (Gammex 464 Phantom) to evaluate image quality with respect to CT number calibration, slice width/thickness, image uniformity, image noise, and spatial resolution. Each acquisition center is required to complete the QA protocol on a regular basis.

Through the core lab, the collected data is evaluated by expert radiologists and a radiology report is included in the repository. The semi-quantitative reporting contains specific regional findings of disease such as emphysema, bronchiectasis, honeycombing, and micro-nodules, as well as the general distribution of disease throughout the lung (e.g. by lobe and region). In addition, the reviewers provide overall patient diagnoses, any additional anomalies, and other ancillary findings in the data. A specific form is used to record the radiological findings in a consistent manner.

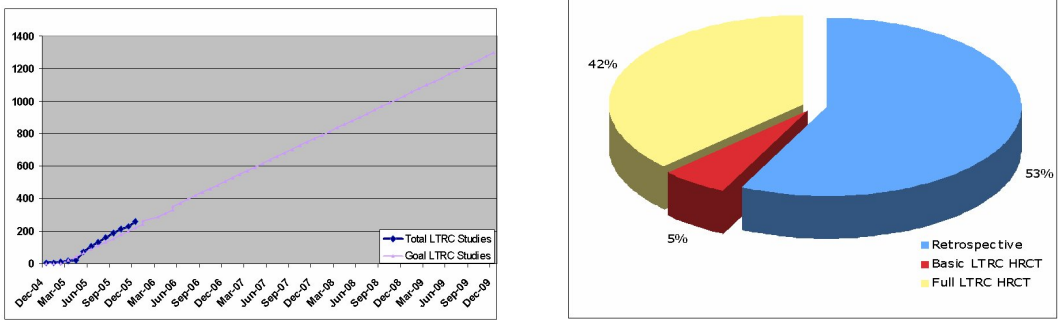

Figure 1: Enrollment for the LTRC is on schedule to include more than 1200 subjects by the end of 2009. Currently, there are over 230 subjects enrolled. Over 40% of patients enrolled thus far have been scanned with the full LTRC three phase protocol. Other subjects have either retrospective or basic LTRC CT scans.

3 Results

The LTRC has a goal of recruiting over 1200 subjects for the databases. As of December 2005, nearly 250 subjects were enrolled (Figure 1A). Of all of the subjects, over 40% have been scanned according to the Full LTRC protocol (Figure 1B). A quantitative analysis of the high-resolution CT data is underway at Mayo. This analysis performs specific measurements on each of the datasets. The measurements include total volume of emphysema, lung/lobe volume measurements, and amount of detectable pulmonary pathology, including common texture measures such as honeycombing and ground glass. The analysis workflow includes automatic segmentation of the trachea, lungs, and lobes. In addition, the central and peripheral regions are defined. Emphysema is automatically segmented using a threshold = -950 HU. Robust measurements of texture are an active area of core lab research.

4 Discussion

The data coordination center was specifically setup to oversee the distribution of data to researchers. The LTRC database will provided the most comprehensive collection of data on COPD and IPF patients to date. It includes histological, radiological, and clinical data. Currently, there are a limited number of CT scans and pathology slides available at the website (www.ctascstudies.com); however, full datasets should be available by the end of 2006. When complete, the database will have real-time query capabilities to identify particular subsets of subjects by diagnoses or other clinical variables.

The data available in the LTRC can be used to study many aspects of chronic pulmonary disease. From an image analysis perspective, the database provides an extensive collection of image datasets which require image registration, automatic segmentation of anatomy, and texture analysis of lung tissue for disease classification. Mayo researchers [4], for example, have studied the functional variability of these patients in a limited number of datasets (Figure 2A). Zavaletta and colleagues [5] are currently investigating texture analysis methods for differentiating different forms of diffuse disease (Figure 2B). One unique advantage of this database is that several forms of gold standards are available. Specifically, the radiological report includes multiple radiology reviews for each case. In addition, tissue samples are also available and can be used to validate classification schemes.

Open image databases provide a common collection of data for investigators to study. Each new inves-

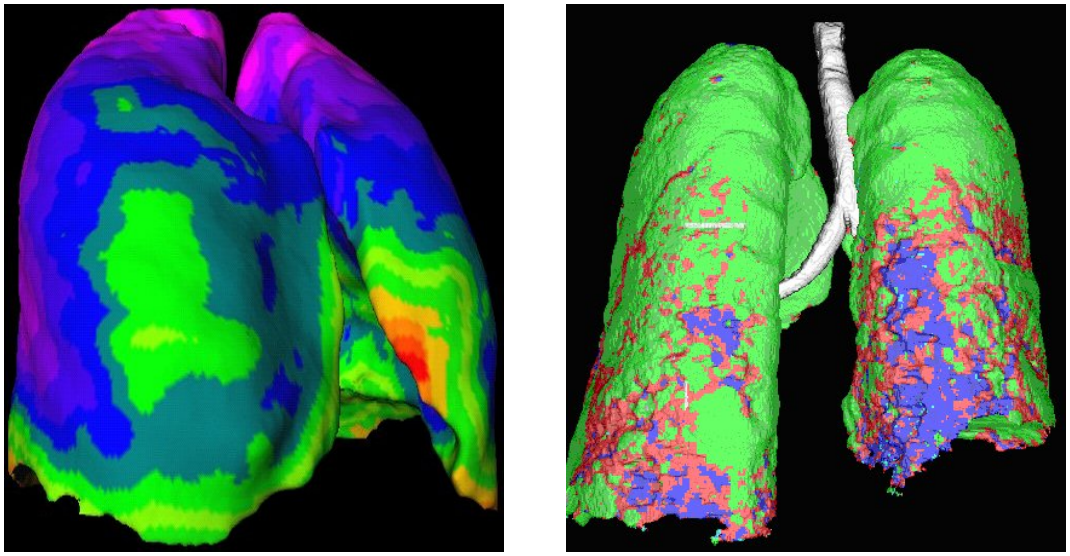

Figure 2: On the left, a color coded rendering of the lungs shows regions of high (green) and low (magenta) deformation through the respiratory cycle. On the right, a dataset from the LTRC has been classified according to common texture features consistent with diffuse lung disease.

tigation of the data can provide new insight into both imaging methodologies and specific diseases (as manifested through the image data). Broad databases, such as the NLM-Mayo Image Collection [6], are a collection of diverse data which is useful for developing robust tools. In contrast, the LTRC database is a very extensive database of patients with specific lung disease which will be useful for validating the sensitivity and specificity of an imaging algorithm. As the collection grows over the next five years, its use will be able to provide new insights into the progression of diffuse lung diseases while providing the basis for the development of new image analysis methodologies.

5 Acknowledgments

The authors would like to acknowledge all of the leadership and investigators associated with the LTRC project. Their combined effort will generate the foundation for future research. The authors would also like to recognize the Mayo participants in the development and support of the LTRC Radiology Core. Both members of Radiology and the Biomedical Imaging Resource bring unique expertise to this project. This work has been funded under NHLBI Grant #AR46158-01 .

References

[1] National Heart, Lung, and Blood Institute (NHLBI), RFP ANNOUNCEMENT: LUNG TISSUE RESEARCH CONSORTIUM, http://grants.nih.gov/grants/guide/notice-files/NOT-HL-03-009.html, 2003

[2] Lung Disease Reference

[3] Bergin C. Muller N. Nichols DM. Lillington G. Hogg JC. Mullen B. Grymaloski MR. Osborne S. Pare PD. The diagnosis of emphysema. A computed tomographic-pathologic correlation. American Review of

Respiratory Disease. 133(4):541-6, 1986 Apr.

[4] Haider CR. Bartholmai BJ. Holmes DR. Camp JJ. Robb RA. Quantitative characterization of lung disease. Computerized Medical Imaging Graphics. 29(7):555-63, 2005 Oct.

[5] Zavaletta, V., Karwoski, R.A., Bartholmai, B., Robb, R.A.: High Resolution Multi-Detector CT Aided Tissue Analysis and Quantification of Lung Fibrosis. Proc. SPIE Medical Imaging 2006, 2006.

[6] Holmes, D.R. III, Workman, E.L., Robb, R.A.: The NLM-Mayo Image Collection: Common Access to Uncommon Data. Medical Image Computing & Computer Assisted Intervention (MICCAI 2005), ISC/NAMIC/MICCAI Workshop on Open Source Software, 2005.