An Adaptive, Knowledge-Driven Medical Image Search Engine for Interactive Diffuse Parenchymal Lung Disease Quantification

Yimo Tao[a,b], Xiang Sean Zhou[a], Jinbo Bi[a], Anna Jerebko[a], Matthias Wolf[a], Marcos Salganicoff [a], Arun Krishnan[a]

[a] Siemens Medical Solutions Inc. USA, 51 Valley Stream Parkway, Malvern, PA
[b] Department of Electrical and Computer Engineering, Virginia Tech, Arlington, VA 22203, USA

Abstract

Characterization and quantification of the severity of diffuse parenchymal lung diseases (DPLD) using CT is an important issue in clinical research. Recently, several classification-based computer-aided diagnosis (CAD) systems for DPLD [1, 2] have been proposed. For some of those systems, a degradation of performance [2] was reported on unseen data because of considerable inter-patient variances of parenchymal tissue patterns.

We believe that a CAD system of real clinical value should be robust to inter-patient variances and be able to classify unseen cases online more effectively. In this work, we developed a novel adaptive knowledge-driven CT image search engine that combines offline learning aspects of classification-based CAD systems with online learning aspects of content-based image retrieval (CBIR) systems. Our system could seamlessly and adaptively fuse offline accumulated knowledge with online feedback, leading to an improved online performance in detecting DPLD in both accuracy and speed aspects. Our contribution lies in: (1) newly developed 3D texture-based and morphology-based features; (2) a multi-class offline feature selection method; and, (3) a novel image search engine framework for detecting DPLD. Very promising results have been obtained on a small test set.

Keywords: CT; texture analysis; interstitial lung disease; relevance feedback; content-based image retrieval; computer aided detection; CAD.

1. Methods
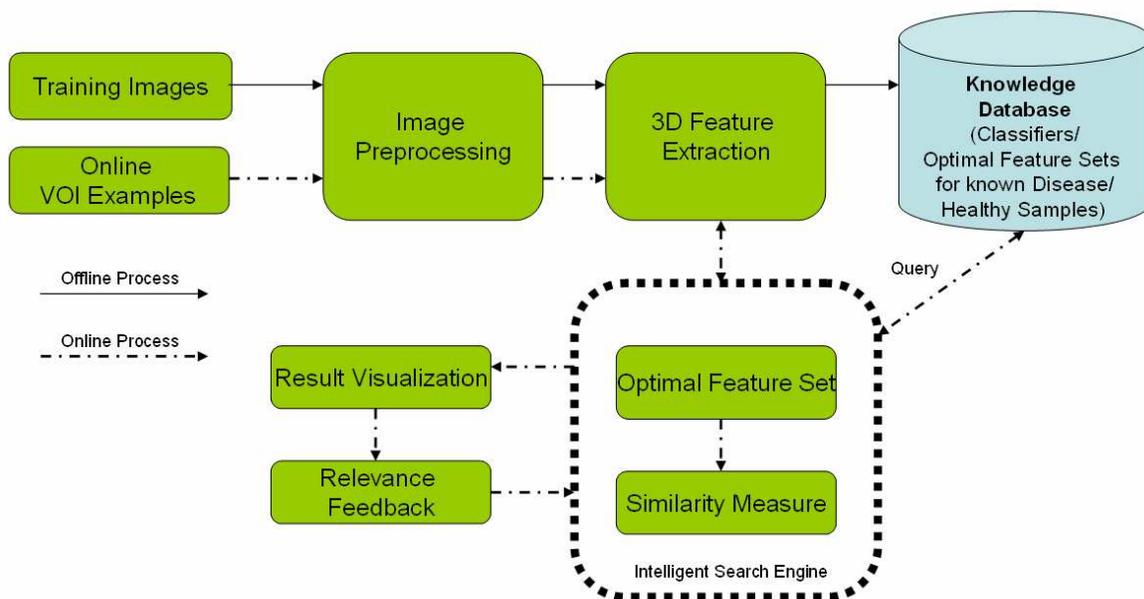
1.1 System Overview

Fig. 1. System Overview

Fig.1 shows the basic overview of our system. We first select training images and ask experienced experts to mark volumes of interest (VOIs) on CT images and label them as containing healthy or DPLD parenchymal tissues. The VOIs

are then processed to eliminate voxels characterized as being airways, after which volumetric features are extracted. Parametric statistical model for each disease class along with its optimal feature sets is learned and stored in our knowledge database.

For online DPLD detection, the user first marks/scribbles a partial VOI. Our system will extract volumetric features from voxels within VOI. By comparing the extracted volumetric features with disease classes in the knowledge database, our intelligent search engine is able to decide the possible disease classes and compute the features optimized for the specific disease class on the whole image volume. These features are then computed at all the other voxels in the image, with each voxel in the image then represented as a feature vector. The similarity between each voxel and the VOI voxels is measured by Mahalanobis distance. Voxels with high similarity will then be highlighted so that the user can provide feedback for further refinement, which requires the online learning of standard CBIR system.

1.2 Volumetric Feature Extraction

Adaptive histogram binning based 3D local binary patterns: Local binary patterns (LBP) have been successfully extended by Zhao [3] to 3D texture analysis. In this work, we employ it to analyze parenchymal texture in CT image. First, we use multi-level thresholding Otsu method to adaptively merge image regions of approximately the same gray level properties. Then the LBP-Top algorithm [3] was used to extract volumetric texture-based features.

3D bullae size histogram: We propose a 3D bullae size histogram feature to quantify morphology patterns of low-attenuation areas (referred as "bullae") in the lung. First, all air-filled areas are marked with a simple thresholding algorithm. Then we use a 3D morphological filter and component labeling algorithm to locate disconnected bullaes. Size histograms are constructed by distributing bullaes into four size ranges of normal, small sized, medium-sized and large-sized bullaes.

Other features: First order and second order intensity-based and texture-based statistical features, including co-occurrence matrix based texture (GLCM), intensity statistics, intensity ratio and wavelet texture, are computed.

1.3 Intelligent Search Engine

Our intelligent search engine is composed of four major components: (1) models offline learned for known diseases and healthy parenchymal tissues; (2) classification of user scribbled VOIs according to known disease models; (3) online feature selection based on dissimilarity measure between user scribbled VOIs and the healthy model; (4) similarity measure between each voxel in the image and user scribbled VOIs using the set of features selected both offline and online.

1.3.1 Offline Learning for Diseases and Healthy Parenchymal Tissues

We collect image data for certain DPLDs, and employ supervised learning techniques to construct classifiers for each disease, which can later be deployed in the on-line detection system. It is essential to select the best set of features to characterize each DPLD from a variety of low-cost image features. We have adopted both filtering methods that rank features according to a statistic score and wrapper methods that select a set of features on which the detection accuracy is optimal.

Various scoring functions exist to rank features in terms of discriminant capacity. The so-called Fisher score is used in our setting. Denote a feature *f*, and its values on class 1 *f1*, and values on class 2 *f2*, then

$$\text{Fisher score} = \left|mean(f1) - mean(f2)\right| / (stddev(f1) + stddev(f2))$$

We also investigate a novel logistic-regression multi-class feature selection (LRMCFS) approach that takes into account the relatedness of all the classes. For many standard classification approaches such as SVM, logistic regression, a multi-class problem is decomposed into multiple binary classification problems by taking a scheme like one-versus-all. Unlike the common feature selection process where each binary classifier selects its own features, our approach eliminates irrelevant features for all classes and identifies discriminant features for each of the classes. Hence the features are selected across all binary classifiers. For example, an individual binary classifier aims to separate emphysema from other diseases. It treats all other diseases as in one class and neglects there are actually multiple types of diseases. By learning all the binary classifiers together, our approach interacts between different classes for the best possible set of features. We formulate our approach based on logistic regression. In other words, we seek to optimize the following cost function:

$$\sum_{m=1}^{M}\left(\sum_{i=1}^{n}\ln(1+\exp(-y_i^m w_m^T S x_i)) + \left\|w_m\right\|^2\right) + \lambda\left\|s\right\|^2$$

where M binary decision functions $f(x) = w_m^T S x$ are to be constructed, S is a diagonal matrix with its diagonal vector equal to s, and n denotes the number of total training examples, and y_i^m is the binary labels of x_i, which equals to 1 if x_i

is in class *m*, -1 otherwise. The resulting one-versus-rest logistic regression classifiers based on selected features are stored in the knowledge database.

1.3.2 Online VOI Classification

After user marked VOI examples are available, we compute volumetric features for sampled voxels within VOI. Voxels are then classified by the multi-class classifier in knowledge database. Majority vote across all voxels in the VOI is used to determine the final VOI label.

1.3.3 Online Feature Selection based on User Scribble or Relevance Feedback

In order to enhance the flexibility in dealing with inter-patient variance, an online feature selection component was included in our system through on-line user interaction. The initial user scribble is used to construct the positive example set. We randomly sample healthy voxels from our knowledge database to form the negative example set. Rank features according to Fisher scores. Note that this feature set is specific to this particular patient. This is important and meaningful, because, for example, emphysema has high inter-patient variability and we may not find a good feature set for all possible variations offline. But for the current patient, emphysema patterns vary much less and a good, discriminative feature set may exist. This step is to supplement the generic, cross-patient disease-specific image features extracted off-line with *patient-specific and disease-specific* image features to achieve better retrieval results. We merge the top 10 online selected features with the offline feature set as the final feature set. Finally, we calculate these features for every voxel in the whole image.

1.3.4 Similarity Measurement

Mahalanobis distance is used as the similarity measurement, which is defined as: $D_M(x) = \sqrt{(x-\mu)^T \Sigma^{-1}(x-\mu)}$ where x is the feature vector of a voxel based on the selected feature set, μ is the mean feature vector computed using the VOI examples, Σ are covariance matrix. Our system allows users to define the similarity threshold through friendly user interface.

2. Results

2.1 Results of Offline Classification

Thirty CT scans of Patients with DPLDs and healthy patients were used in this study. Fifteen scans were from normal smokers/nonsmokers. Other fifteen scans contained DPLDs including emphysema and fibrosis. Table 1 shows the results of the error rate for a 3-fold stratified cross-validation experiment using different types of groups of features and multi-class LDA classifier. Note that the test set is completely independent from the training patients. The precision using the logistic regression classifier by combining all classifiers is 75.27%

Intensity-based Feature	Precision (%)	Texture-based Feature	Precision (%)
Intensity Ratio	60.17	LBP	61.38
Intensity Statistics	50.72	GLCM	67.72
Intensity Histogram	47.5	Wavelet	65.89
Bullae Index	53.34		

Table 1. Offline classification precision of different group of features using multi-class LDA classifier

2.2 Final System Illustration

Fig. 2(a) and (b) shows two axial slices of the testing images and VOI examples scribbled by radiologists along with the color-coded outputs from our system. It can be seen that our system was able to classify various diffuse lung diseases on the fly from initial VOI examples. The output of our system aligned well with the radiologist's markings.
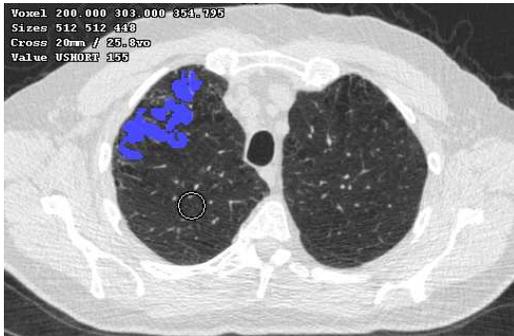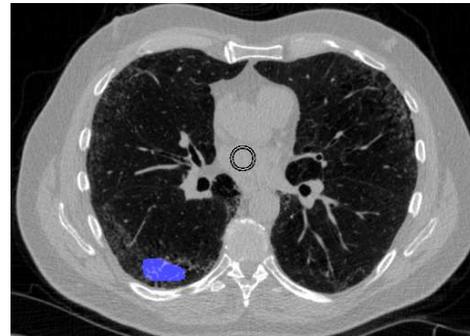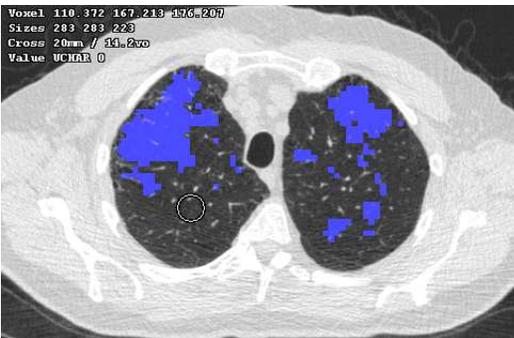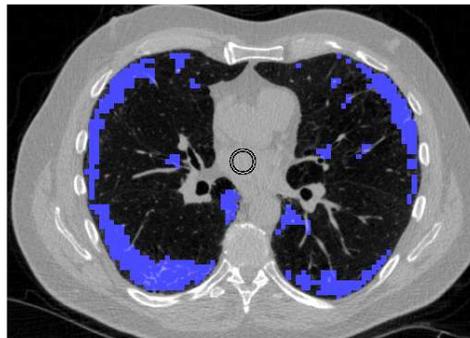
Fig.2 (a.1) A CT slice with emphysema in the axial view, (a.2) Partial VOI example marked by the radiologist, (a.3) The color-coded output by our system. (b.1) A CT slice with fibrosis in the axial view, (b.2) Partial VOI example marked by the radiologist, (b.3) The color-coded output by our system.

Conclusion

We have developed a novel adaptive knowledge-driven image search engine that combines offline learning aspects of classification-based CAD systems along with online learning aspects of CBIR system. Our system could seamlessly and adaptively fuse offline accumulated knowledge with online feedback knowledge, leading to an improved online performance in detecting DPLDs in terms of both accuracy and speed. Future steps are to optimize the processing time and to clinically evaluate the VOI marked by the system.

Reference

1. Xu, Y., et al., *MDCT-based 3-D texture classification of emphysema and early smoking related lung pathologies.* IEEE Transactions on Medical Imaging, 2006. **25**(4): p. 464-475.
2. Zavaletta, V.A., B.J. Bartholmai, and R.A. Robb, *High Resolution Multidetector CT-Aided Tissue Analysis and Quantification of Lung Fibrosis.* Academic Radiology, 2007. **14**(7): p. 772-787.
3. Zhao, G. and M. Pietikäinen, *Dynamic Texture Recognition Using Local Binary Patterns with an Application to Facial Expressions.* IEEE Transactions on Pattern Analysis and Machine Intelligence, 2007: p. 915-928.