Learning techniques for Computer Aided Diagnosis in Chest Radiography

Course CS229: Final Write-up Arpit Aggarwal (SU # 05375755)

Introduction

Lung nodule detection in chest radiographs is an important area for Computer Aided Diagnosis (CAD) primarily because lung cancer still claims over one million deaths every year. Early detection of the disease is critical but the fact remains that only 20% [1] of cases are detected in the first phase.

Radiologists can miss up to 30% of lung nodules (which may develop into cancer) in chest radiographs due to the background anatomy of the lungs which can camouflage the nodules. CAD helps radiologists by performing preprocessing of the images and suggesting the most likely locations for nodules. Detection of lung nodules proceeds through techniques for suppressing the background structures in lungs which include the blood vessels, ribs and the bronchi. The resulting images provide chest structures which make good candidates for nodules and can be further classified on the basis of features like size, contrast and shapes. Simple rule based classifications on such features tend to produce a lot of false positives. In this project I investigate a supervised learning technique to come up with non-trivial secondary classifiers for separating nodules from vessels, bifurcations, crossings and other background structures.

Input Data

The input data is a series of axial radiographs from different patients. The radiographs are 512 x 512 grayscale images with intensities in Hounsfield units (HU). The data from these slices can be concatenated in an array to get a 3D array of voxels. I have used a set of 304 radiographs from a single patient for training. The size of the voxels in the dataset is 0.683594 x 0.683594 x 1.0 mm^3.

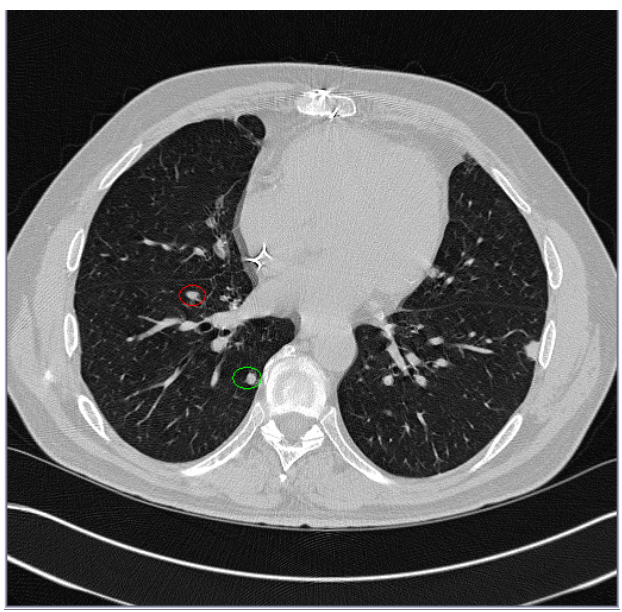


Figure 1 – Shows a single input axial radiograph of a lung with an actual nodule (labeled true positive - green circle) and one false positive (red circle).

Also a set of candidate locations for nodules (rough center points of the structures) found by the pre-processing (using the first classifier) of the images is provided. These have been correctly labeled as either a nodule (true positive of the first classifier) or not (a false positive of the same classifier). They are used to train the secondary classifier. The single patient data-set had 1000 such locations, out of which 82 were labeled as true positive and the rest as false positives.

Features

Given the candidate location, a trivial set of features would be to take the intensities of the voxels or the average intensity of a block of voxels about the point of interest. But this would result in a high dimensional feature space and also not exploit the knowledge that one already possesses about the nodules. At the same time such features would be susceptible to rotation and scale variance.

One important heuristic which can be used to distinguish nodules from other structures is that they are essentially spherical where as other structures like vessels, ribs or bronchi are tubular. Following are the set of features that I analyzed to model this distinction –

Marching Cubes with visibility test

Figure 2: Surface rendering of a Bifurcation using Marching Cubes.

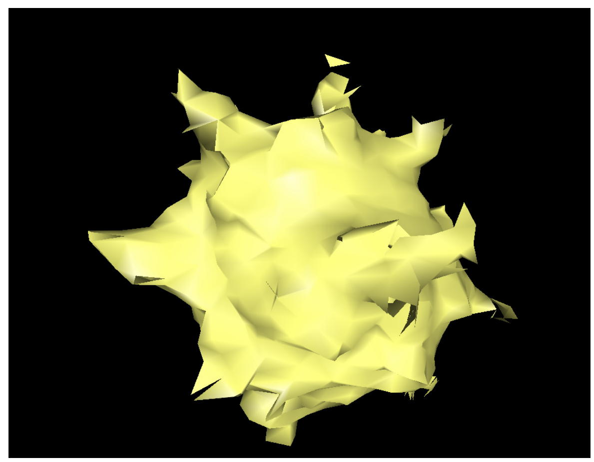
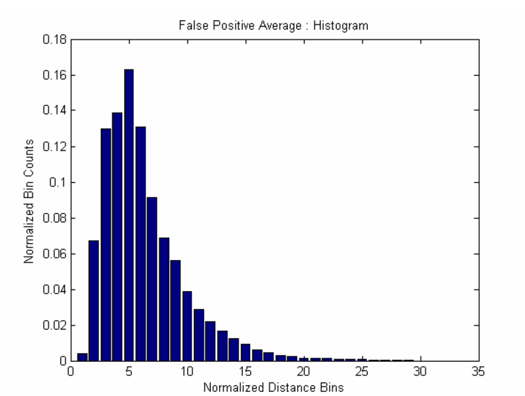
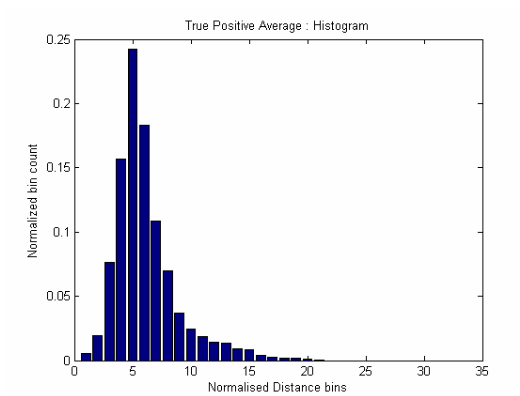


Figure 3: Surface rendering of a Lung Nodule using Marching Cubes.





To represent the shape of the structure through the features, I used the distances of points on the surface of the structures from their center points (provided as input). To find the surface I ran the Marching cubes algorithm to determine the triangles making up iso-surface. Details of the algorithm can be found in [2].

For each triangle visible from the center point the distance to its centroid is calculated. A histogram of all the distances normalized by their mean is found out. Each bin in the histogram is taken as a feature for the classifier. Different numbers of features were tried out as a trade off between high bias and computational complexity. Finally 30 such features were used to define the classifier.

This set of features resulted in average performance on the training set (high bias) indicating that a better set of features was required.

Lantern Transform with rays in a single direction

The marching cubes algorithm had the problem of carrying over into surrounding structures while calculating the surface of a particular structure. This skewed the distance histograms. Thus, I decided to sample the surface directly. For this 3D rays were cast out from the center point over equally spaced directions (calculated by distributing a set of points uniformly over a unit sphere) and their length when they hit a surface (sudden drop in intensities) was calculated. A histogram of the various distances was then plotted. Each bin in this histogram represented a feature.

Lantern Transform with a pair of anti-parallel rays

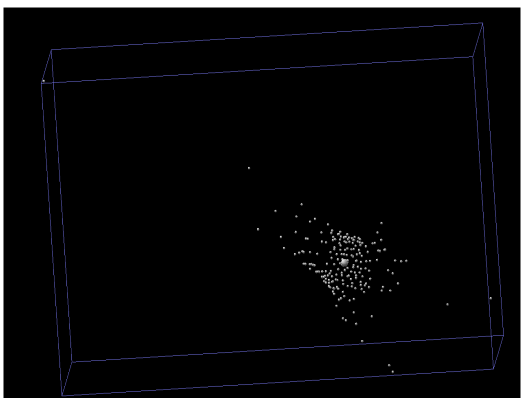


Figure 4: Sampled surface of a Bifurcation

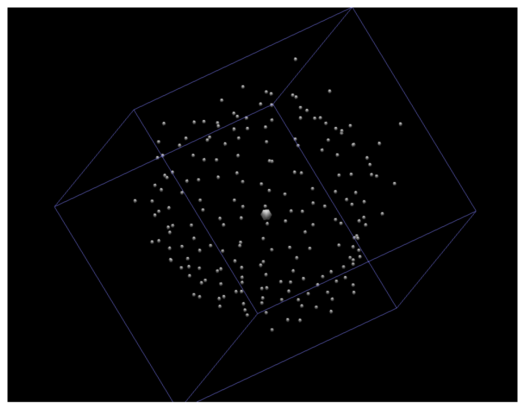
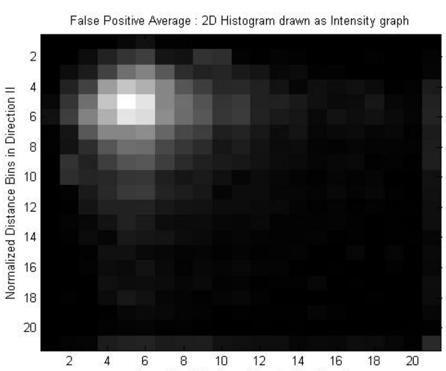
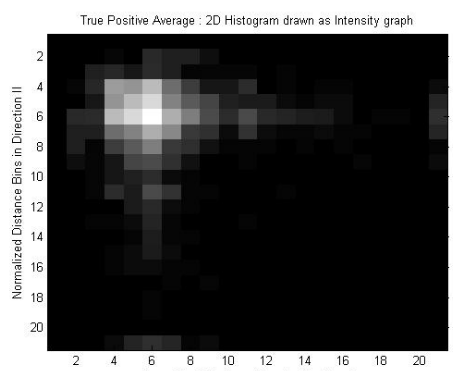


Figure 5: Sampled surface of a Nodule





Further, I employed the correlation between rays that are anti-parallel to each other. So rays were cast out in opposite directions and the pairs of distances were used for the histograms. These gave a better indication of the symmetric nature of the structure. The resulting 2D histogram had n^2 bins (n bins corresponding to distance in 1 direction) and since the histograms formed a symmetric matrix only half the bin counts ($n^2/2$) were used as features. I chose a set of 200 features for classification.

The above features were made scale invariant by normalizing the distances. Normalization was done with the mean of the distances for each histogram.

Classifier & Results

Since this is a binary classification problem, Logistic linear regression was used to learn a linear classifier with a bias. The classifier had 31 parameters for the first feature set and 201 when training using the second. Batch gradient ascent was used to maximize the log-likelihood and the learning rate was kept at a constant value of .01 because convergence can be achieved for a fixed value for alpha.

Following are the classification results achieved on the test and training set for both classifiers.

Lantern Transform

	Training Set	Test Set
Sensitivity (True Positive Ratio)	**48.15**	**23.08**
Specificity (True Negative Ratio)	**99.89**	**97.78**
Overall Accuracy	**95.67**	**88.35**

Marching Cubes

	Training Set	Test Set
Sensitivity (True Positive Ratio)	**37.18**	**23.08**
Specificity (True Negative Ratio)	**96.55**	**93.25**
Overall Accuracy	**91.17**	**84.31**

Even though the overall accuracy was found to be pretty good, binary classifiers which have a varied distribution in the two classes of the training data are better evaluated using ROC curves.

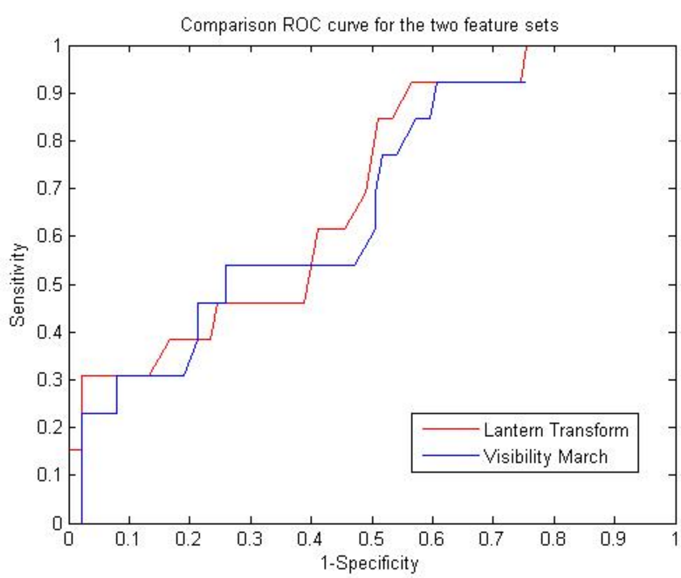


Figure 6: Shows the ROC curves for the two feature sets plotted by varying the threshold value of the classifier

From the figure, one realizes that Lantern Transform results in a better classifier as it has a larger area under the ROC. Depending on the context and the cost associated with the two classes (nodule and non-nodule) one can choose a threshold value along the curve. For certain values the Visibility March outperforms the Lantern Transform and is thus used in that context. The combined performance is then the convex hull of the two curves. In our case since high sensitivity is critical as we don’t want to drop any actual nodules, the lantern transform with a threshold value of -5.3 can be chosen. This ensures a sensitivity value of 1 and at the same time cuts down 25% of the false positives.

Bayesian Logistic regression with MAP estimate was also tried as a classifier but expectedly did not offer significantly better performance as the earlier classifier did not suffer from the problem of over fitting.

Testing

Testing was done on a set of 100 candidate locations which had 12 nodules and 88 other structures. These radiographs belonged to a patient different from the one on which the training was performed. Since the real test of the classifier is to correctly identify nodules in different patients, cross validation was not used while learning/evaluating the classifier on the two sets of radiographs.

Future work

Features could be further improved to get better classification. In particular in the marching cubes feature, the distances could be weighted by the area of the triangles. Also the angle at which the 3D ray is incident on to the surface of the structure could be considered to form a feature.

A better classifier could be built by using a combination of some of the above features instead of using them alone. Also ensemble learning could be used to get better results.

Acknowledgement

I want to acknowledge the valuable inputs given by Asst. Prof. David Paik, Radiology Department, Stanford School of Medicine, during the course of this project. He also provided the training and test radiographs that were essential for the viability of the project.

References:

1. Ellis J. R. C., Gleeson F. V., Lung cancer screening, The British Journal of Radiology, 74 (2001), 478–485
2. Lorensen W. E., Cline H. E, Marching Cubes: A High Resolution 3D Surface construction algorithm, Computer Graphics, Volume 21, Number 4, July 1987.
3. Doi K., et al., Computer-aided diagnosis in radiology: potential and pitfalls, European Journal of Radiology 31 (1997) 97–109