Unsupervised classification of cirrhotic livers using MRI data

Gobert Lee[a], Masayuki Kanematsu[b], Hiroki Kato[b], Hiroshi Kondo[b], Xiangrong Zhou[a], Takeshi Hara[a], Hiroshi Fujita[a] and Hiroaki Hoshi[b]

[a]Department of Intelligent Image Information, Division of Regeneration and Advanced Medical Sciences, Graduate School of Medicine, Gifu University, Gifu 501-1194, Japan
[b]Department of Radiology, Gifu University School of Medicine, Gifu 501-1194, Japan

ABSTRACT

Cirrhosis of the liver is a chronic disease. It is characterized by the presence of widespread nodules and fibrosis in the liver which results in characteristic texture patterns. Computerized analysis of hepatic texture patterns is usually based on regions-of-interest (ROIs). However, not all ROIs are typical representatives of the disease stage of the liver from which the ROIs originated. This leads to uncertainties in the ROI labels (diseased or non-diseased). On the other hand, supervised classifiers are commonly used in determining the assignment rule. This presents a problem as the training of a supervised classifier requires the *correct* labels of the ROIs. The main purpose of this paper is to investigate the use of an unsupervised classifier, the k-means clustering, in classifying ROI based data. In addition, a procedure for generating a receiver operating characteristic (ROC) curve depicting the classification performance of *k*-means clustering is also reported. Hepatic MRI images of 44 patients (16 cirrhotic; 28 non-cirrhotic) are used in this study. The MRI data are derived from gadolinium-enhanced equilibrium phase images. For each patient, 10 ROIs selected by an experienced radiologist and 7 texture features measured on each ROI are included in the MRI data. Results of the *k*-means classifier are depicted using an ROC curve. The area under the curve (AUC) has a value of 0.704. This is slightly lower than but comparable to that of LDA and ANN classifiers which have values 0.781 and 0.801, respectively. Methods in constructing ROC curve in relation to *k*-means clustering have not been previously reported in the literature.

Keywords: Computer-aided diagnosis (CAD), classification, k-means clustering, ROC analysis, magnetic resonance (MR) images, cirrhosis

1. INTRODUCTION

Cirrhosis of the liver is one of the leading causes of death by disease, killing more than 20,000 people in the United States each year. Cirrhosis of the liver is characterized by the presence of widespread fibrosis and regenerative nodules in the liver. The fibrosis and nodules formation causes distortion of the normal liver architecture, resulting in characteristic texture patterns. Though presented with great potential, interpretation of texture patterns is a challenge for human observers. This calls for the interest in machine analysis of hepatic texture patterns.

Texture analysis for computer-aided diagnosis (CAD) in medical images has been studied in many disciplines including the diagnosis of breast cancer in mammograms [1-5], lung nodules in chest radiographs [6-8], osteoporosis in bone x-ray images [9-11], and abnormalities in kidney and liver [12-14]. Analysis is typically based on regions-of-interest (ROIs). A classifier is then subsequently used to classify the data (feature vectors) into a number of pre-defined classes. Linear discriminant analysis (LDA), artificial neural network (ANN) and support vector machine (SVM) are among the popular choices of classifiers. The above named classifiers are in the category of supervised classifier which employs a training data set (with known class labels) for training the classification rule.

Medical Imaging 2008: Computer-Aided Diagnosis, edited by Maryellen L. Giger, Nico Karssemeijer
Proc. of SPIE Vol. 6915, 691514, (2008) · 1605-7422/08/$18 · doi: 10.1117/12.770862

Problem in diagnosing abnormalities using a scheme involving ROIs and supervised classifier is that a sampled ROI may or may not contain the features that typically represent the disease stage of the organ/structure from which the ROI originated. Yet, disregard of its disposition, the ROI will be given the same label as any other ROI from that organ/structure. If a supervised classifier is subsequently used, the uncharacteristic ROI will constitute an outlier and will have an influence in the training of the classifier.

In this paper, classification of livers as cirrhotic or non-cirrhotic based on texture features measured on ROIs is studied with the use of a different type of classifier, the unsupervised classifier. An unsupervised classifier differs from a supervised classifier in two ways: (1) it works with unlabeled data, and (2) training of a classification rule with the use of a training data set is not required. K-means clustering is a simple and popular unsupervised classifier and is chosen for this study. The ROI texture feature data employed in this study is a subset of the data set reported in Kato et al. [15].

In addition, a method for generating an ROC curve in depicting the classification performance of k-means clustering is also described in this paper. Receiver operating characteristic (ROC) analysis [16] is widely used for evaluating performance of classification/CAD schemes, however, schemes employing k-means clustering typically attain only a single measure of sensitivity and specificity. In order to enable full comparisons between classification schemes employing k-means clustering and those employing other classifiers, an ROC curve generation method for k-means clustering is developed.

Finally, with the use of the above ROC curve generating method, the classification results using k-means clustering are compared with the classification results using a classical LDA classifier and an ANN classifier, with the former computed in this study and the latter obtained from Kato et. al. [15].

2. METHODOLGY

This section describes a *k*-means clustering algorithm and the proposed method in generating ROC curve associated with k-means clustering. The k-means clustering algorithm partitions the data of interest, hepatic ROI data in this study, into a number of k clusters. The determination of the number of clusters, *k*, is based on the silhouette index.

2.1. K-means clustering

Among a number of unsupervised learning algorithms, K-means clustering is simple and is one of the most popular algorithms. As mentioned before, the main differences between a supervised classifier and an unsupervised classifier is that a training data set with *known* class labels is required for the former to train the classification rule, whereas such a training data set and the knowledge of the class labels in the data set are not required for the latter. This makes it most suitable when there are uncertainties in the class labels.

2.1.1. The algorithm

The k-means clustering algorithm partitions a given data set into *k* mutually exclusive clusters such that the sum of the distances between data and the corresponding cluster centroid is minimized. The above distance measure between two data points is taken as a measure of similarity. A number of distance measures can be used depending on the data. The Minkowski distance, the Euclidean distance and the Mahalanobis distance are some typical examples of distance measures. In this study, the standard Euclidean distance was used as the distance measure.

Mathematically, given a set of data vectors $X = [x_1, ..., x_n]$ where n is the number of observations, the k-means clustering algorithm groups the data into k clusters with the aim at minimizing an objective function, a squared error function. As the Euclidean distance was adopted as the distance measure in this study, the objective function J is then explicitly given as

$$J = \sum_{j=1}^{k} \sum_{i \in c_j} \| x_i - \mu_j \|^2 , \quad (1)$$

where c_j is the *j*-th cluster and μ_j is the centroid of the cluster c_j . Therefore, the k-means clustering algorithm is an iterative algorithm that finds a suitable partition which minimizes the sum squared error. The algorithm begins with the initialization of *k* cluster centroids. Different approaches in initialization have been suggested. A simple method is to initialize the problem by randomly select *k* data points from the given data. The remaining data points are classified into the k clusters by distance. The centroids are then updated by computing the centroids in the *k* clusters. As in other optimization problems with a random component, the results of *k*-means clustering are initialization dependent. This is usually dealt with by running the algorithms several times, each with a different initialization. The best solution from the multiple runs is then taken as the final solution.

2.1.2. Determining the number of clusters

The *k*-means clustering algorithm requires the knowledge of *k*, the number of clusters. Unfortunately, this is usually not known *a prior*. Furthermore, real data are often corrupted with noise and the clusters in real data sets are often not well separated. As a result, the data set can be interpreted with several possible numbers of clusters. This means determining the number of clusters is an ill-posed problem. Nevertheless, a number of approaches have been suggested in finding the number of clusters. Here, the number of clusters is determined by assessing the 'quality' of the clustering results over a range of *k*.

In this study, the silhouette index was employed as a cluster validity index. A silhouette index was defined on each and every data point (feature vector) in the feature space. The silhouette index of the *i*-th data point is given by

$$S_i = \frac{b_i - a_i}{\max(a_i, b_i)} , \quad (2)$$

where a_i is the average distance of the *i-th* point to other points in the same cluster and b_i is the average distance of the *i-th* point to points in its nearest neighbour cluster. The silhouette index indicates how well a data point is clustered. In particular, the silhouette index S_i is a normalized difference measure of the average distance of the *i*-th point to all other points in the same cluster and that of the *i*-th point to all other points in the closest cluster with distance being the Euclidean distance. The value of the silhouette index S_i ranges from +1 to -1. A value of unity indicates that the point is very distant from its neighbor clusters while a value of -1 indicates that the point is closer to points in its nearest neighbor cluster than to those in its own cluster.

An average silhouette is obtained by averaging the silhouette values over the number of data points. The average silhouette

$$S(k) = \frac{1}{N} \sum_{i=1 \ldots N} S_i \quad (3)$$

is an index measuring the overall 'quality' of the partition where S_i is the silhouette index defined in Equation 2 and *N* is the number of observations. Similar to the point silhouette index, a high value in the average silhouette indicates the 'quality' of the partition is good. Furthermore, the number of clusters, k, is a parameter of the average silhouette. The number of clusters can be determined by maximizing $S(k)$ over a range of *k*.

2.2. ROC curve

ROC analysis [16] is widely used in evaluating CAD schemes or classification results. An ROC curve is a plot of the true positive fraction (tpf) verses the false positive fraction (fpf). The true positive fraction is sometimes referred

to as the hit rate, detection rate or probability of detection while the false positive fraction is sometimes referred to as the false alarm rate. In other words, the ROC curve depicts the trade-off between sensitivity and specificity, where sensitivity is, yet, another term for true positive fraction and specificity is the unit complement of false positive fraction (1-fpf).

2.2.1. Typical ROC curve

In this paper, the term typical ROC curve refers to the ROC curve constructed with the assumption that all observations can be represented by a single univariate random variable, X. It is the ROC curve that readers are familiar with [16]. Consider two groups, G_0 and G_1 where G_0 is the non-diseased group occupying the lower end of the univariate axis and G_1 is the diseased group occupying the higher end of the axis, the probability distribution of X conditional on group G_i is denoted by f_i and their corresponding cumulative distribution by F_i where $i = 1, 2$. For continuous random variable X, the true positive fraction and the false positive fraction are given by

$$\text{tpf} = \int_t^{\infty} F_1(x)dx, \qquad \text{fpf} = \int_t^{\infty} F_0(x)dx\,,$$

and the ROC curve, C, is given by

$$C(x) = 1 - F_1(F_0^{-1}(1-x)),$$

where F_0^{-1} is defined by $F_0^{-1}\,F_0 = 1$. As the distributions of the two groups are usually not known *a priori*, the ROC curve C(x) is commonly estimated by estimating F_0 and F_1 with appropriate distribution models (parametric estimator) or their respective empirical distributions (non-parametric estimator).

For multivariate observations, commonly used classifier such as a linear discriminant function or an artificial neural network will transform the multivariate observations into univariates. Distribution of the transformed univariate can then be estimated as described in the above. The probability density functions of the transformed univariates are the underlying curves from which the ROC curve is generated. In particular, if a linear discriminant function is used in the transformation, the transformed value is often referred to as the discriminant score. Furthermore, this univariate random variable is also called the decision-variable as the decision of an observation belonging to one group or the other is based upon a threshold value of this variable (Figure 1).

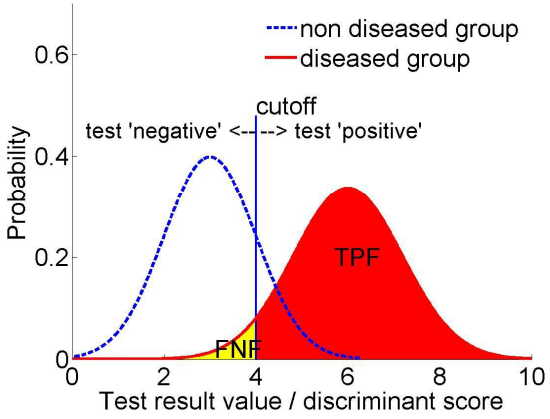

Figure 1. The underlying probability density functions that generate an ROC curve.

Outputs of classifiers such as a linear discriminant analysis (LDA) and artificial neural network (ANN) are univariates, hence, ROC curve construction is straightforward. However, it should be noted that in the above two classifiers, the support for the LDA output is [-inf, inf] while that for the ANN output is [0, 1]. Consequently, assumption of a binormal ROC curve would not be suitable for the ANN classifier.

2.2.2. ROC curve for k-means clustering

Contrary to commonly used classifiers such as LDA and ANN, k-means clustering does not transform the multivariate feature input into a univariate. This makes the construction of an ROC curve non-trivial. In the remaining of this Section, a method in generating an ROC curve depicting the k-means clustering performance is described. The method is a non-parametric method. It takes on any k-means clustering output without making any assumptions of the distributions or density functions of the clusters. An iterative procedure is employed in order to obtain a number of (tpf, fpf) ROC curve data points. The procedure is described as follows.

##
1. Apply k-means clustering with $k = 2$.
2. Compute the fractions T_1 and T_2 of cirrhotic livers detected in clusters C_1 and C_2, respectively.
3. Determine the true positive cluster by comparing T_1 and T_2, that is,

$$\begin{cases} C_1 = \text{true positive cluster} & \text{if} \quad T_1 > T_2 \\ C_2 = \text{true positive cluster} & \text{if} \quad T_2 > T_1 \end{cases}.$$

4. Compute (tpf, fpf).
5. Relocate a point x_i from C_1 to C_2 such that the change in J, as shown in Equation 1, is minimum.
6. Compute (tpf, fpf).
7. Repeat Step 5 - 6 until C_1 becomes an empty cluster.
8. Reinstate the k-means clustering results.
9. Relocate a point x_i from C_2 to C_1 such that the change in J, as shown in Equation 1, is minimum.
10. Compute (tpf, fpf).
11. Repeat Step 9 - 10 until C_2 becomes an empty cluster.
##

Once all the (tpf, fpf) data points are obtained, the ROC curve can be generated by simply plotting the true positive fraction verses the false positive fraction.

3. EXPERIMENTAL DATA

MRI data of 44 patient cases (16 cirrhotic; 28 non-cirrhotic) were used in this study. The data is a subset of a database reported in a study by Kato et al. [15]. The 44 patients were identified from the archive of Gifu University Hospital, Japan in a retrospective study. All the 44 patients had partial hepatectomy for malignant hepatic lesions in the period between February 2000 and January 2002 with preoperative evaluations of hepatic tumors using gadolinium-enhanced MRI performed within 2 weeks prior to surgery. Diagnoses among the 44 patients were hepatocellular carcinoma in cirrhosis, hepatocellular carcinoma in chronic viral hepatitis type C or type B, cholangioccllular carcinoma and liver metastasis. Surgical specimens were retrospectively examined by a pathologist who was blinded to patient histories and radiology and surgery reports. The degree of hepatic fibrosis (Figure 2) in non-tumorous liver parenchyma was evaluated using the 5-point Desmet scale [17] with F0 indicates no fibrosis; F1, mild fibrosis; F2, moderate fibrosis; F3, severe fibrosis; and F4, cirrhosis. For the evaluation of a

two-class CAD scheme employed in Kato et al. [15], livers with pathologic grades of F3 and F4 were taken as true positives (cirrhotic livers) and livers with pathologic grades F0-F2 were taken as true negatives (non-cirrhotic livers), resulting in a total of 16 true positives and 28 true negatives.

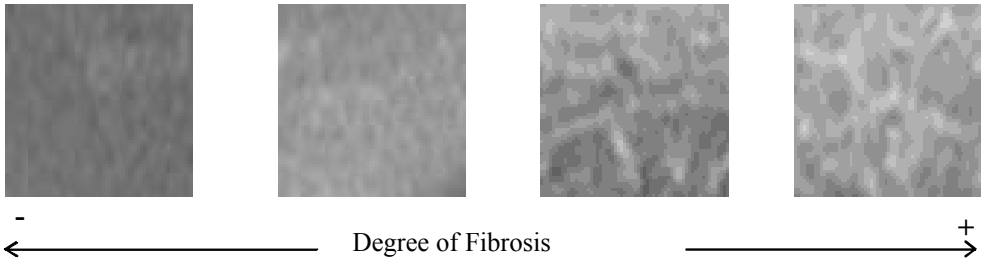

Figure 2. Different degree of fibrosis ranging from the very mild (left) to the very severe (right).

The preoperative hepatic MRI images were acquired using a 1.5T MR scanner (Signa Horizon; GE Medical Systems, Milwaukee, Wis.). MRI images of several sequences were obtained including T1- and T2-weighted MRI as well as gadolinium-enhanced pre-contrast, hepatic arterial, portal venous and equilibrium phases MRI. All MRI images were acquired in the axial plane. The depth resolution of the images was 16 bits (65536 gray scale levels) and the size of the axial images was 512 × 512 pixels with a section thickness of 8-10 mm and a 2- to 3-mm intersection gap. Depending on the size of the coverage, a MRI sequence of a patient typically consisted of a stack of 18-22 (axial) slice images.

Only the gadolinium-enhanced equilibrium phase images were of interest in this study. This is due to the findings in Kato et al. [15] that performance of the 7 texture features extracted from the gadolinium-enhanced equilibrium phase image is superior to that extracted from other MRI sequence or combination of sequences. From the equilibrium phase images of each patient, 10 ROIs were selected by a radiologist who was blinded to the patient information and the pathologic diagnosis. The ROIs were 32 × 32 pixels in size and were placed in the liver parenchyma. Care was taken to avoid large blood vessels, focal hepatic lesions and prominent hepatic artifacts. The locations of the ROIs were planned to cover the liver strategically with eight in the right liver lobe (typically, two in each of Couinaud's liver segments V, VI, VII and VIII [18]) and two in the left lobe (typically, one in each of Couinaud's liver segments II and III). The seven texture features were contrast, angular second moment, entropy, mean and inverse difference moment derived from co-occurrence matrix [19] as well as the mean and standard deviation derived from the image intensity histogram.

4. RESULTS

4.1. Number of clusters *k*

A range of values were investigated in finding the most appropriate k value. For each candidate k value, the k-means clustering algorithm was run 50 times. The best solution with the least sum of distances (Equation 1) was taken as the final solution for that particular k. Finally, the average silhouette index $S(k)$ was used in assessing the cluster validity with different k. The average silhouette index with a range of k is shown in Figure 3. It can be seen that $S(k)$ increases monotonically in the range $k = 2, ...8$. Consider the small size of the data set, dividing the data into as many as 8 clusters would mean that each cluster would contain a few data. The monotonic increase in $S(k)$ in this scenario was thought to be due to the classes in the data are not well separated. Figure 3 also shows that there are two regions where the increase of $S(k)$ is small. The two regions are from $k = 2$ to $k = 3$, and from $k = 6$ to $k = 7$. The small increase in $S(k)$ indicates that the clusters are fairly stable. Due to the small number of samples in this study, the region with a smaller k, that is, $k = 2$ was taken as the optimal number of clusters.

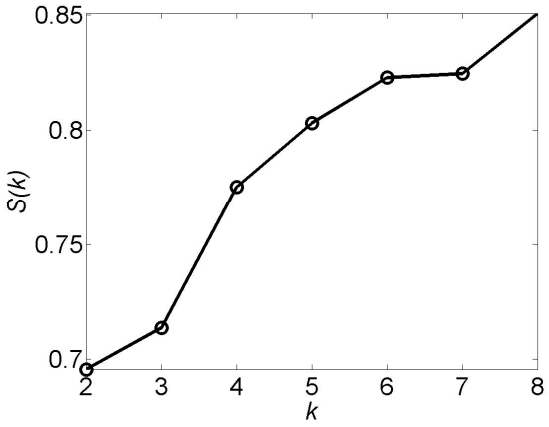

Figure 3. Plot of the average silhouette index $S(k)$ for $k = 2, 3, \ldots 8$.

4.2. Classification using k-means clustering

An ROC curve depicting the performance of the unsupervised k-means clustering algorithm is shown in Figure 4. Area under the ROC curve (AUC) was found to be 0.704. The ROC curve was constructed using the method proposed in this paper. For comparison, classification performances using two other classifiers, LDA and ANN, were also investigated. The AUC was found to be 0.781 for a LDA classifier using all cases for testing and training. For an ANN classifier, an AUC of 0.801 was claimed in Kato et al. [15]. The results was achieved using a second dataset of 8 patient cases (not available to this study) for training the ANN and the same dataset of 44 patient cases employed in this study for testing the classifier.

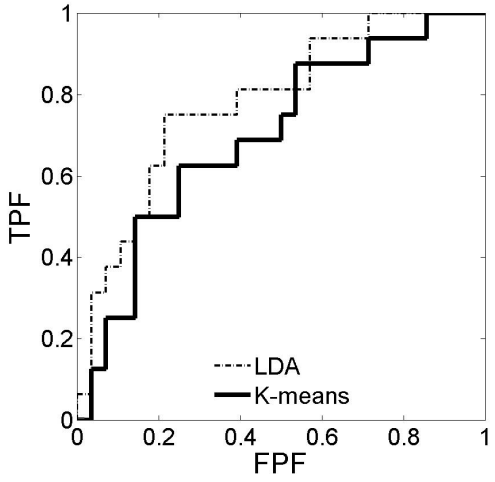

Figure 4. ROC curves obtained using k-means clustering and linear discriminant analysis (LDA). All cases were used for training and testing in obtaining the two ROC curves. The area under the ROC curve (AUC) has a value of 0.781 for LDA and 0.704 for k-means clustering.

5. DISCUSSION AND CONCLUSION

The classification results obtained using a k-means classifier was depicted in an ROC curve. The performance of the classification was compared to that of two commonly used supervised classifiers, namely, the LDA and the ANN by comparing the AUCs. The AUC obtained using the unsupervised classifier was found to be slightly lower than that of the LDA and ANN. This could be due to the reduced amount of information presented to the k-means classifier. The use of an unsupervised classifier, on the one hand, eliminates the dependency of the ROI labels which may be subjected to error due to sampling. On the other hand, ignoring the ROI label could mean loss of information and could affect the classification results had the labels be correct.

Quantitative comparison of the three ROC curve obtained using k-means classifier, LDA and ANN should be performed in order to assess if the apparent difference in the three AUC values is statistically significant. Method in comparing ROC curves is well-known [Hanley]. However, the method assumes that the ROC curve is generated based on two univariate probability distributions which is not the case for the ROC curve associated with k-means clustering. Methods in comparing the k-means clustering ROC curve with other ROC curves are being investigated.

Only a single sensitivity value and a single specificity (i.e. one data point on an ROC curve) are usually available in indicating the performance of a scheme/algorithm associated with a k-means classifier. In order to have a full analysis of the performance and be able to compare with other schemes/algorithms, the generation of ROC curves is necessary. Methods in generating ROC curves associated with k-means clustering have not been previously reported in the literature. A procedure for generating such ROC curves was described in this paper. The use of ROC analysis for evaluation is becoming a trend for medical diagnosis/CAD schemes. The proposed ROC generating procedure enables k-means clustering to be evaluated against other classifier on a common ground.

ACKNOWLEDGMENTS

This research was supported in part by a grant for the Knowledge Cluster Gifu-Ogaki (KCGO), referred to as the “Robotics Advanced Medical Cluster,” from the Ministry of Education, Culture, Sports, Science and Technology, Japan.

REFERENCES

1. S. Sahiner, H.-P. Chan, and N. Petrick et al., "Computerized characterization of masses on mammograms: the rubber band straightening transform and texture analysis", *Med. Phy.* Vol.25(4), 1998, pp.516-526.
2. N.R. Mudigonda, R.M. Rangayyan and J.E.L. Desautels, "Gradient and texture analysis for the classification of mammographic masses", *IEEE Trans. Med. Imag*., Vol.19(10), 2000, pp.1032-1043.
3. R. Gupta and P.E. Undrill, "The use of texture analysis to delineate suspicious masses in mammography", *Phys. Med. Biol.* Vol.40, 1995, pp.835-855.
4. H-P. Chan, B. Sahiner, and N. Petrick et al. "Computerized classification of malignant and benign microcalcifications on mammograms: texture analysis using an artifical neural network", *Phys. Med. Biol.*, Vol. 42, 1997, pp.549-567.
5. M. Zhang, M.L. Giger, and K. Doi, "Mammographic texture analysis for the detection of spiculated lesions", In: *3rd Int'l Workshop on Digital Mammography* (eds. K. Doi, M. L. Giger, R. M. Nishikawa, R. A. Schmidt), Elsevier Science, Amsterdam, pp. 347-350, 1996.
6. X. Chen, K. Doi, and S. Katsuragawa et al., "Automated selection of regions of interest for quantitative analysis of lung textures in digital chest radiographs", *Med. Phys*., Vol.20, 1993, pp.975-982.

7. B. van Ginneken, S. Katsuragawa, and B. ter Harr Romeny B et al., "Automatic detection of abnormalitites in chest radiographs using local texture analysis", *IEEE Trans Med Imag.,* Vol.21(2), 2002, pp.139-149.
8. E.A. Hoffman, J.M. Reinhardt, and M. Sonka et. al, "Characterization of the interstitial lung diseases via density-based and texture-based analysis of computed tomography images of lung structure and function", *Acad. Radiol*. Vol.10(10), 2003, pp.1104-1118.
9. L. Lan, M.L. Giger, and J. R. Wilkie et al., "Automated selection of region of interest for radiographic texture analysis", *Proc SPIE Medical Imaging*, Vol. 6514, 2007, pp.651436.
10. M.R. Chinander, M.L. Giger, and R.D. Shah et al., "Investigation of using bone texture analysis on bone densitometry images", *Proc SPIE*, Vol. 4684, 2002, pp.860-863, 2002.
11. J.R. Wilkie, M.L. Giger, and M.R. Chinander et al., "Comparison of radiographic texture analysis from computed radiography and bone densitometry systems", *Med. Phys*. Vol.31, 2004, pp.882-891.
12. G.N. Lee, X. Zhang, and M. Kanematsu et al., "Classification of cirrhotic liver on MR images using texture analysis", *Int'l J. CARS,* Vol.1, Supp. 1, 2006, pp.379-381.
13. H. Yoshida, D. Casalino, and B. Keserci et al., "Wavelet packet based texture analysis for differentiation between benign and malignant liver tumors in ultrasound images", *Phys. Med. Biol*. Vol. 48, 2003, pp.3735-3753.
14. D.-Y. Kim and J.-W. Park, "Computer-aided detection of kidney tumor on abdominal computed tomography scans", *Acta Radiol*. Vol.45, 2004, pp.791-795.
15. H. Kato, M. Kanematsu, and X. Zhang et al., "Computer-aided diagnosis of hepatic fibrosis: preliminary evaluation of MRI texture analysis using the finite difference method and an artificial neural network", *AJR*, Vol.189, 2007, pp.117-122.
16. C.E. Metz, "Basic principles of ROC analysis", *Seminars in Nuclear Medicine,* Vol.8, 1978, pp.283-298.
17. V.J. Desmet, M. Gerber, and J.H. Hoofnagle et al., "Classification of chronic hepatitis: diagnosis, grading and staging", *Hepatology*, Vol. 19, 1994, pp.1513-1520.
18. H. Gray, *Gray's Anatomy* 36th ed. Churchill Livingston. 1980.
19. R.M. Haralick, K. Shanmugam, and I. Dinstein, "Texture features for image classification", *IEEE Trans. Sys. Man Cybern*. Vol. SMC-3-6, 1973, pp.610-621.