Artificial Life Feature Selection Techniques for Prostrate Cancer Diagnosis Using TRUS Images

S.S. Mohamed[1], A.M. Youssef[2], E.F. El-Saadany[1], and M.M.A. Salama[1]

[1] Department of Electrical and Computer Engineering,
University of Waterloo, Ontario, Canada
(smohamed, ehab, msalama)@hivolt.uwaterloo.ca
[2] Concordia Institute for Information Systems Engineering,
Concordia University, Montréal, Quebec, Canada
youssef@ciise.concordia.ca

Abstract. This paper presents two novel feature selection techniques for the purpose of prostate tissue characterization based on Trans-rectal Ultrasound (TRUS) images. First, suspected cancerous regions of interest (ROIs) are identified from the segmented TRUS images using Gabor filters. Next, second and higher order statistical texture features are constructed for these ROIs. Furthermore, a representative feature subset with the best discriminatory power among the constructed features is selected using two artificial life techniques: the Particle Swarm Optimization (PSO) and the Ant Colony Optimization (ACO). Both the PSO and ACO are tailored to fit the binary nature of the feature selection problem. The results are compared to the results obtained using the Genetic Algorithm (GA) feature selection approach. When Support Vector Machine (SVM) classifier is applied for the purpose of tissue characterization, the features obtained using the PSO and ACO outperforms the features obtained using the GA, i.e., they are capable of discriminating between suspicious cancerous and non-cancerous in a better accuracy. The obtained results demonstrate excellent tissue characterization with 83.3% sensitivity, 100% specificity and 94% overall accuracy.

1 Introduction

Prostate cancer is the highest-incidence cancer and the second leading cancer killer in men. It is only curable at an early stage; therefore, early detection is highly recommended [1]. Different types of diagnosis such as the Prostatic Specific Antigen (PSA) value, family history, age, race, prostate volume, as well as Digital Rectal Examination (DRE) lack reliability and therefore, are not sufficient for accurate diagnosis. Depending on these results, the doctor usually refers the patient to the next common diagnostic stage which is using Trans-rectal Ultrasound (TRUS) imaging system. Based on the radiologist scrutiny to the TRUS image a biopsy operation might follow [2]. The radiologist experience plays an important role in identifying the biopsy locations.

TRUS provides information about the size and shape of the prostate; it is also used for identifying different gland zones. TRUS is considered the dominant imaging modality for diagnosis of Prostatism as well as detection and staging of prostate

M. Kamel and A. Campilho (Eds.): ICIAR 2005, LNCS 3656, pp. 903 – 913, 2005.

cancer. However, TRUS is still deficient in early and accurate detection of tumours due to the low quality and noise characteristics of the TRUS images. The only way for recognizing the suspicious zones is with the aid of expert radiologist, which makes this process time consuming and operator dependent. Therefore, mimicking the expert radiologist decision is recommended as an assistive method for newer radiologists. This is achieved by the means of Computer Aided Diagnosis (CAD).

CAD of prostate cancer is a developing field where several aspects can be tackled. A typical CAD system consists of a segmentation stage, region of interest identification, feature construction, feature selection and classification. The work in this paper is focuses on ROI identification, feature extraction and selection where the selected features are then tested by a classifier that is capable of dealing with the noisy and distorted features of the TRUS images.

The first phase in CAD diagnosis from TRUS images is *segmentation* where the prostate boundaries are being detected from the TRUS image, a step that leads to determining the volume of the gland. Lots of research has been done in this area using different methods such as statistical shape model [3], super ellipses [4] and wavelet analysis [5] which makes it well established.

The second important phase is *ROI identification*, which deals with highlighting the most probable cancerous regions in the gland, a step that is usually achieved by an expert radiologist. This step is crucial as studying the whole image is computationally demanding. Moreover, choosing the incorrect ROIs results in misleading features that might lead to inaccurate medical decisions. A promising ROI identification method using Gabor multi-resolution analysis was proposed and applied to TRUS images in [6].

The next stage is *feature construction* where different statistical and spectral features are constructed from the identified ROIs. The features used in this paper include second order as well as higher order statistical features. The obtained features might have some redundancy or correlation; therefore, selection among features is necessary in order obtain a highly representative feature set [7]. Several methods can be used for feature selection. These methods can be categorized into classifier independent and classifier dependent feature selection. For the classifier independent feature selection, features are ranked according to the information content of each feature. While in the classifier dependent, a pre-specified classifier is used and the features that lead to the best classification results are selected.

Mutual Information Feature Selection (MIFS), as a classifier independent algorithm, was applied to TRUS images features in [8] and leads to excellent results, however, on the expense of the computational effort. This paper focuses on classifier dependant feature selection algorithms. These methods are Particle Swarm Optimization (PSO) and Ant Colony Optimization (ACO). The results of these two artificial life techniques are then compared to the Genetic Algorithm (GA) based feature selection.

The ultimate goal of this work is to mimic the radiologist's decision which is achieved using several consecutive stages that are highlighted in figure 1. The first stage is ROI identification using Gabor filters; the second stage is extracting statistical features from the identified ROIs. The third phase is a classifier based feature selection based on GA, PSO or ACO. The results obtained are compared to the doctor's diagnosis.

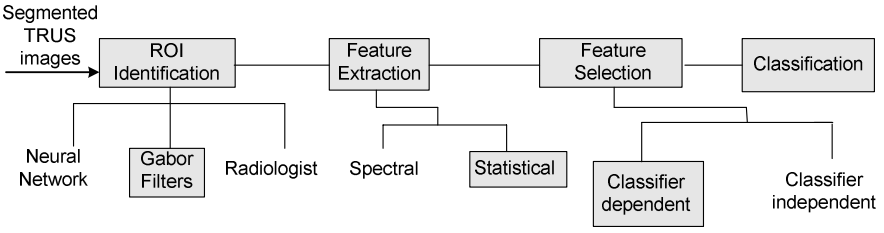

Fig. 1. Prostate Cancer Diagnosis from TRUS images

2 ROI Identification

ROI segmentation is a vital stage for TRUS image feature extraction for the purpose of prostate cancer diagnosis and it is usually performed by the aid of an expert radiologist. With the goal of assisting radiologist's decision and getting accurate rapid results, there is a great need for an automated ROI segmentation algorithm. Multi-resolution filtering is an excellent method for texture investigation, which is achieved by using Gabor multi-resolution analysis that is able to segment the image according to the frequency response of the pixels. The pixels that have similar response will be assigned to the same cluster. The Gabor function was chosen for this application for its high

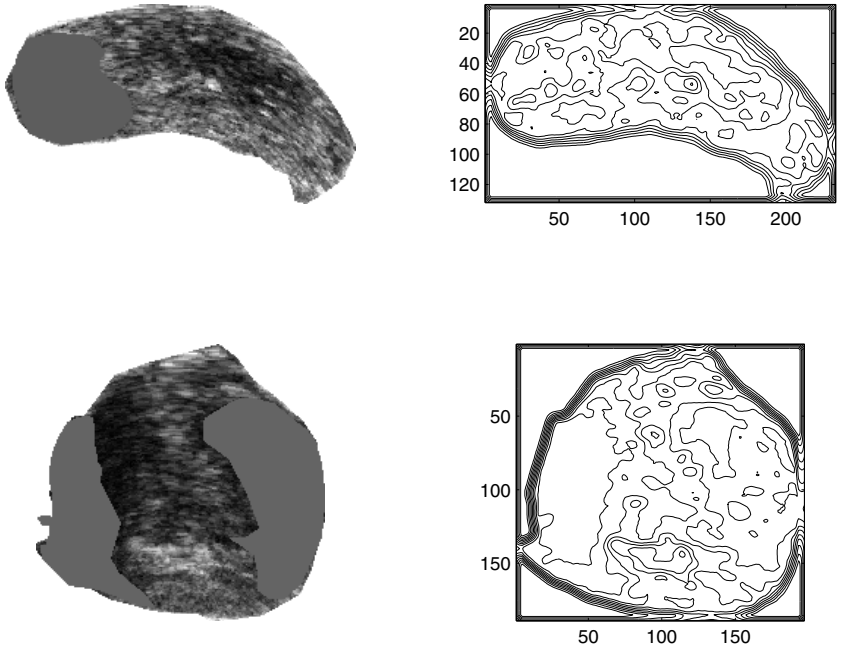

Fig. 2. Two different TRUS professionally segmented images and the corresponding ROIs

localization in both the spatial frequency domain as well as the spatial domain. The Gabor function in the spatial domain is a Gaussian modulated sinusoid, while in the spatial-frequency domain the Gabor function becomes two shifted Gaussians at the location of the modulating frequency. This method is applied in this paper and a sample of the segmented images and the original TRUS images is shown in figure 2.

3 Texture Feature Construction

Texture is used to describe the local spatial variations in image brightness which is related to image properties such as coarseness, and regularity. This is achieved by performing numerical manipulation of digitized images to get quantitative measurements. Texture analysis can potentially expand the visual skills of the expert eye by extracting image features that are relevant to diagnostic problem and not necessarily visually extractable. Statistical approaches represent texture with features that depend on relationships between the grey levels of the image. It is very helpful to know that different tissues have different textures. Benign tumors are described as regular masses with homogenous internal echoes, while carcinomas are masses with fuzzy borders and heterogenous internal echoes. In this work statistical texture features of identified ROIs are constructed. Statistical texture features proved its high recognition ability in ultrasound images. These features has been used for fetal lung maturity [9], liver tissue characterization [10], prostate cancer recognition [8] as well as some other applications. Second and higher order statistical features are used in this work where different texture features are constructed from the identified regions of interest of the TRUS images.

3.1 Second Order Statistics

The second order statistical features are considered crucial as the human visual system is capable of identifying different textures only if their second order statistics are different. However, textures which differ in higher-order statistics but have the same first- and second-order statistics cannot be recognized spontaneously by the human visual system. These features were used earlier for fetal lung maturity, liver tissue characterization, as well as prostate cancer diagnosis [8, 9, 10]. The two sets of second order statistical features used in this work are the Grey Level Difference Matrix (GLDM) and the Grey Level Difference Vector (GLDV). A set of nine features are constructed using GLDM and GLDV and the details of both methods are explained in [8].

3.2 Higher Order Statistics

With the relative immaturity of CAD applied to TRUS images, it would be unsuitable to think about the image analysis only using second-order statistical techniques. Moreover, as textures differing in third or higher-order statistics escape the capabilities of the human perceptual system, it is expected that considering higher-order analysis of TRUS images will give good results. Two higher order statistical algorithms are adopted and applied for the first time to the TRUS images in this work which are the Neighborhood Grey Tone Difference Matrix (NGTDM) and the Grey Tone Run Length Matrix

GTRLM where nine features are extracted from both matrices. The details and mathematical formulation of these methods are explained in [11,12] respectively.

- **Neighborhood Grey Tone Difference Matrix NGTDM**

NGTDM features are the properties, that might be used to discriminate between different textural patterns, include coarseness, contrast, complexity, busyness (fineness), shape, directionality and texture strength. In this approach the i^{th} entry in NGTDM is a summation of the differences between all pixels with grey-tone i and the average value of their surrounding neighbors. NGTDM applied in this work used a square region of five pixels as the neighborhood size. If i is the grey-level at (x,y) then the average grey-level over the square neighborhood centered at (x,y) is given by:

$$A_i = A(x, y) = \frac{1}{W - 1} \sum_{m=-d}^{d} \sum_{n=-d}^{d} i(x + n, y + m), \tag{1}$$

Where $(m,n) \neq (0,0)$, d specifies the neighborhood size and $W = (2d+1)^2$ it follows that the i^{th} entry in the NGTDM is given as:

$$s(i) = \begin{cases} \sum |i - A_i|, & for\ i \in N_i\ if\ N_i \neq 0 \\ 0, & Otherwise. \end{cases} \tag{2}$$

Five texture features are constructed from the NGTDM which are: coarseness, contrast, busyness and complexity. The details of these features is explained in [11].

- **Grey Tone Run Length Matrix GTRLM**

GTRLM calculates the run length of a specific image grey-tone in a direction α within a textured ROI. A run isa set of successive pixels having the same or similar intensities along a specified direction. The considered pixels have to be linearly adjacent in the direction α. The run length is the number of pixels enclosed within the run. The number of runs with grey-tone i of run length j in some α direction is denoted by $R(\alpha) = \left[r'(i, j|\alpha)\right]$. Five features are computed from the GTRLM which are: short run emphasis, long run emphasis, grey tone distribution, run length distribution and run percentage. The mathematical formulation of these features is explained in [12].

4 Feature Selection

The output of the previous stage is a feature vector that is composed of GLDV, GLDM, NGTDM and GTRLM features. This feature vector is applied to a classifier that should identify some classes (e.g., cancerous and non-cancerous). This feature vector might have some redundant and correlated features (curse of dimensionality) which is the main motivation for using the feature selection techniques.

The principle of feature selection decides on a subset of features, which preserve most of the information needed for pattern classification [7]. An optimal set is a subset of features, which forfeit none of the information needed for classification and results in no increase in the minimum probability of error, when a decision rule is applied in both the observation and the subset space. The Feature Selection algorithms used in this paper are classifier dependant FS methods. This means that the possible feature subsets

are obtained and the classifier performance is tested for each subset and finally the best discriminatory feature subset is chosen.

4.1 Particle Swarm Optimization (PSO)

PSO [13, 14, 15] is a new population based stochastic optimization technique inspired by social behavior of bird flocking or fish schooling. Each particle is treated as a point in a 2^n dimensional space where n denote the total number of features. Let $\boldsymbol{v_{ij}}$ and $\boldsymbol{p_{ij}}$ denote the j^{th} component of the i^{th} particle velocity $\boldsymbol{v_i}$ and position $\boldsymbol{p_i}$ respectively.

We define the fitness of each particle as the recognition accuracy corresponding to the features selected by this particle using a pre-specified classifier (in this work, we use Support Vector Machine classifier.)

The algorithm described in here is a slightly modified version of the PSO algorithm to fit the binary nature of the feature selection problem.

The system is initialized with a group of random particles with $0 \le p_{i,j} \le 1$, and then searches for optima by updating generations. In each iteration, each particle position is updated by following two "best" particles. The first one (denoted by *pbest*) is the best fitness it has achieved so far. The fitness value is also stored. Another "best" value that is tracked by the particle swarm optimizer is the best value (denoted by *gbest*), obtained so far by any particle in the population.

After finding the two best values, the particle updates its velocity and positions with following equations

$$v_{i,j} = v_{i,j} + c_1 r_1 (pbest_{i,j} - p_{i,j}) + c_2 r_2 (gbest_{i,j} - p_{i,j}), \tag{3}$$

$$v_{i,j} = 2\sigma(v_{i,j}) - 1, \tag{4}$$

$$p_{i,j} = f_i(p_{i,j} + v_{i,j}), \tag{5}$$

Where $1 \le r_1, r_2 \le 0$ are uniformly distributed random variables, c_1, c_2 are learning factors and $\sigma(x)$ is the sigmoid function given by

$$\sigma(x) = \frac{1}{1 + \exp(-\alpha x)} \tag{6}$$

$f_i = (T(p_{i,0}), T(p_{i,1}), \ldots T(p_{i,2^n-1}))$ where

$$T(x) = \begin{cases} 1, & x \ge 0.5, \\ 0, & x < 0.5. \end{cases} \tag{7}$$

The output of the above adopted algorithm is a vector composed of ones that correspond to the selected features and zeros that correspond to the rejected features.

4.2 ANT Colony Optimization

ANT Colony Optimization [16] is another heuristic optimization method for solving optimization problems which borrows ideas from biological ants. Experiments with real ants showed that ants go from the nest to the food source and backwards then, after a while, the ants prefer the shortest path from the nest to the food source. The ants communicate indirectly by laying pheromone trails and following trails with higher pheromone. Naturally, larger amount of pheromone will accumulate on the shorter paths to good food sources because larger number of ants will cross it back and forth per unit time as compared to longer paths.

The originally proposed ANT colony optimization algorithm fits naturally in optimization problems corresponding to the selection of optimum permutation (such as the traveling sales person problem), i.e., we can apply it for the feature selection problem if we fixed the size of the subset of features to be chosen.

In here, we present a simple algorithm that borrows ideas from the ANT colony but doesn't have the above constraint, i.e., we don't have to pre-determine the size of the optimal feature subset.

The system is initialized with a group of ants moving across a full binary tree of depth n and 2^n leaves. Each leaf corresponds to one of the possible 2^n feature subsets. The root of the tree corresponds to the nest of the ants and the accuracy of the classifier based on the feature subset associated with each leaf corresponds to the amount of food found at the food source.

The algorithm proceeds by iterating through the following three basic steps:

- Construct a solution for all ants: At each node, each ant has to make a (statistical) decision whether to follow the right path or the left path. At the first iteration, all the ants will move randomly. However, on subsequent iterations, the ants' choices will be influenced by the intensity of the pheromone trails left by preceding ants. A higher level of pheromone on the right path gives an ant a stronger stimulus and thus a higher probability to turn right and vice versa. Let $Pher_l(R)$ and $Pher_l(L)$ denote the value of the pheromone accumulated at the right edge and the left edge of a given node at the l^{th} level of the tree. Then the ants' behavior equivalent to having each ant choosing a uniformly distributed random variable $0 \le r \le 1$ and choosing to follow the right edge at the l^{th} level of the tree if $r \ge \frac{Pher_l(L)}{Pher_l(R) + Pher_l(L)}$, and to follow the left edge otherwise.
- Do a global pheromone update: For our problem, this step is also different than the one proposed in the original ANT colony optimization algorithm. Instead of updating the pheromone along the visited arcs only, we update all the corresponding 2^{l-1} arcs at l^{th} level of the tree. The amount of pheromone laid by each ant corresponds to the amount of food (i.e., the classifier accuracy) that the ant finds at the leaf of the tree at the end of the path followed by this ant.
- Evaporate pheromone: After each iteration, a portion of the pheromone of the edge is evaporated according to a local updating rule, such that the probability of the selection of that edge by other ants decreases. This

prevents construction of similar paths by the set of ants and increases the diversity of the system. The rate of evaporation provides a compromise between the rate of convergence and reliability of convergence. Fast evaporation causes the search algorithm to be stuck at local optima, while slow evaporation lowers the rate of convergence. After enough iteration of the algorithm, the pheromone of the good edges which are used in constructing of low-cost paths will increase and the pheromone of the other edges will evaporate. Thus, in the higher iterations the probability of constructing low-cost paths increases.

As in the PSO method the accuracy criterion is based on SVM classifier.

5 Classification

Support Vector Machines are known to be a leading method for solving non-linear classification problems [8]. SVM depends mainly on pre-processing the data to represent patterns in a higher dimensionality space, usually much higher than the original feature space. This is achieved with a suitable non-linear mapping $\Phi(.)$ to a sufficiently high dimension [8]. Data from two classes are always separated by a hyper-plane. In binary classification, the task is to find a function $\Phi(.)$ that separates the two classes by learning from a set of samples. The hyperplane is selected so that it maximizes the margin between the two classes. The vectors (samples) defining the hyperplane are the most difficult patterns to classify and are called Support Vectors.

Each of the four feature sets as well as the selected feature subsets using the feature selection algorithms above are examined using the Support Vector Machine classifier.

6 Results and Discussion

The constructed feature sets are separately tested using the SVM classifier. The results are shown in table 1 where cancer and non-cancer attributes represents the suspicious cancerous cases and the non-suspicious cancerous cases according to the radiologist opinion. A set of 96 regions were used in this study where 80 regions were used as training set and a set of 16 regions were used as the test set. The following parameters are used to evaluate these feature:

Accuracy: 1 - probability of misclassification
False Negative: probability of classifying a cancerous prostate as a normal prostate.
False Positive: probability of classifying a cancerous prostate as a normal prostate.
Sensitivity: 1 - false negative rate
Specificity: 1 - false positive rate

Specificity, sensitivity and accuracy are the measures used to test different feature sets applied in this work. The table shows that the classification accuracy obtained using second order statistics is much better than that obtained accuracy using the higher order statistics. It is expected to obtain this kind of result due to the fact that the main target is to mimic the expert radiologist, whose classification accuracy is bounded by the human vision capabilities that is limited to the second order statistics. It is also

clear that the GLDV features obtained better results than the GLDM features which show that the GLDM features have redundant information that is confusing the classifier.

Table 1. Classification results using the constructed feature sets

GLDM		Cancer	Non-cancer
	cancer	5	1
	Non-cancer	1	9
83.33% Sensitivity; 90% Specificity; 87.75% Accuracy			
GLDV		Cancer	Non-cancer
	cancer	5	1
	Non-cancer	0	10
83.33% Sensitivity; 100% Specificity; 93.75% Accuracy			
GTRLM		Cancer	Non-cancer
	cancer	4	2
	Non-cancer	2	8
Sensitivity = 66.67%; Specificity 90%; 81.25% Accuracy			
NGTDM		Cancer	Non-cancer
	cancer	4	2
	Non-cancer	2	8
Sensitivity = 66.67%; Specificity 80%; 75% Accuracy			

The constructed feature sets are combined together to form a feature set that includes GLDM, GLDV, NGTDM and GTRLM features. The proposed feature selection algorithms are then applied to the obtained feature set. Genetic algorithms [16] have been used for the past decade for feature selection applications. A basic binary GA feature selection is used in this work for the purpose of comparison with the results obtained using the above two techniques. The classification results show that the GA obtained the least classification accuracy with 83.33% sensitivity; 90% specificity and 87.75% accuracy where the PSO and ACO obtained a better classification results with 83.33% sensitivity; 100% specificity and 93.75% accuracy.

Table 2. Classification results using the selected feature sets

GA		Cancer	Non-cancer
	Cancer	5	1
	Non-cancer	1	9
83.33% Sensitivity; 90% Specificity; 87.75% Accuracy			
PSO		cancer	Non-cancer
	cancer	5	1
	Non-cancer	0	10
83.33% Sensitivity; 100% Specificity; 93.75% Accuracy			
ACO		cancer	Non-cancer
	cancer	5	1
	Non-cancer	0	10
83.33% Sensitivity; 100% Specificity; 93.75% Accuracy			

7 Conclusions

Two novel feature selection algorithms were applied for accurate feature selection for the purpose of prostate cancer diagnosis using TRUS images. ROIs were identified from the segmented prostate TRUS images using Gabor multi-resolution analysis leading to accurate identified regions. Second and higher order statistical texture features such as GLDM, GLDV, NGTDM and GTRLM were constructed from these automatically segmented regions. Moreover, a feature subset representing the most salient and uncorrelated features was generated utilizing three different artificial life techniques where the well established GA is compared to both the PSO and the ACO. Finally these features were used for tissue characterization using SVM algorithm. The obtained results revealed the out performance of the ACO and the PSO compared to the basic binary GA. The sensitivity was 83.3% for all three feature selection methods and the specificity was 90% for GA and 100 % for the PSO and ACO. Moreover, it is observed that the selected features using both the PSO and the ACO were from the second order statistical features which prove that the system is limited with the human visual system and cannot go beyond what the HVS can recognize. This shows a great radiologist mimicking capability of the proposed system.

References

1. Scardino, PT., "Early detection of prostate cancer". ; Urol Clin North Am. 1989 Nov; 16(4) Pages: 635-55.
2. Scheipers, U., Lorenz, A., Pesavento, A., Ermert, H., Sommerfeld, H., Garcia-Schurmann, M., Kuhne, K., Senge, T. and Philippou, S "Ultrasonic multifeature tissue characterization for the early detection of prostate cancer", IEEE Ultrasonics Symposium, Volume: 2 , 7-10 Oct. 2001 Pages: 1265 -1268 vol.2
3. Dinggang S., Yiqiang Z. and Davatzikos, C."Segmentation of prostate boundaries from ultrasound images using statistical shape model" IEEE Transactions on Medical Imaging, Volume: 22 , Issue: 4 , April 2003 Pages:539 – 551
4. Lixin G., Pathak, S.D., Haynor, D.R., Cho, P.S. and Yongmin K, **"**Parametric shape modeling using deformable super ellipses for prostate segmentation", IEEE Transactions on Medical Imaging, Volume: 23, Issue: 3, March 2004 Pages:340 – 349
5. Chiu, B., Freeman, G., Salama, M.M.A. and Fenster, A., **"**Prostate segmentation algorithm using dyadic wavelet transform and discrete dynamic contour", Physics in Medicine and Biology, Vol. 49, No.20, pp. 4943-4960.
6. Mohamed, S.S., Salama, M.M.A., Kamel, M. and Rizkalla, K. "Region of interest based prostate tissue characterization using least square support vector machine LS-VSM", Proceedings of ICIAR, Lecture Notes in Computer Science, Vol. 3212, pp. 51-58,
7. R. Duda, P. Hart, D. Stork, "Pattern Classification", John Wiley and Sons.2001
8. S.S. Mohamed, M.M. Salama "Computer Aided diagnosis for Prostate cancer using Support Vector Machine" medical imaging conference 2005, California, SPIE.
9. Bhanu, K.N.; Ramakrishnan, A.G.; Suresh, S. and Chow, T.W.P.; "Fetal lung maturity analysis using ultrasound image features" Information Technology in Biomedicine, IEEE Transactions on , Volume: 6 , Issue: 1 , March 2002 Pages:38 – 45

10. Chung-Ming W.; Yung-Chang C. and Hsieh, K; "Texture features for classification of ultrasonic liver images". Medical Imaging, IEEE Transactions on, Volume: 11 Issue: 2, June 1992 Page(s): 141 –152
11. Amadasun, M. and King, R., "Textural features corresponding to textural properties", IEEE Transactions on Systems, Man, and Cybernetics, vol. 19, no. 5, pp. 1264-1274, September 1989.
12. Galloway, M. M. "Texture analysis using gray-level run lengths", Computer Graphics and Image Processing, vol. 4, pp. 172-179, June 1975.
13. Kennedy J. and Eberhart, R. " Particle Swarm Optimization", Proc. IEEE, Int Conf. Neural Network, pp 1942-1948, 1995
14. "Special Issue on Particle Swarm Optimization", IEEE Transactions on Evolutionary Computation publication information, Volume: 8, Issue: 3, June 2004.
15. Eberhart R.C. and Shi, Y. "Comparison between genetic algorithms and particle swarm optimization" Proc. IEEE international conference on evolutionary Comp. pp 611-616, May 1998
16. Dorigo, M. Maniezzo, V. and Colorni, A. "The ant system: optimization by a colony of cooperating agents", IEEE Transactions on Systems, Man, and Cybernetics-Part B 26(1) (1996), pp. 29–41.
17. Goldberg, D. "Genetic algorithms in search, optimization and machine learning", Reading, MA: Addison, Wesley.