Image Features Extraction for Masses Classification in Mammograms

Ramzi CHAIEB[1], *Amira BACHA*[1]
[1]SAGE research unit, National Engineering School of Sousse, University of Sousse
Sousse, Tunisia
e-mails: ramzi.chaieb@hotmail.com, amira_bacha@live.com

Karim KALTI[1,2]
[2]Computer Science Department, Faculty of Science of Monastir, University of Monastir
Monastir, Tunisia
e-mail: karim.kalti@gmail.com

Fradj BEN LAMINE[3]
[3]Department of Electronics and Informatics, High School of Sciences and Technologies of Hammam Sousse
University of Sousse, Sousse, Tunisia
e-mail: benlamine.fradj@gmail.com

***Abstract*—Computer aided diagnosis of breast cancer is becoming increasingly a necessity given the exponential growth of performed mammograms. In particular, the breast mass diagnosis and classification arouse nowadays a great interest. Indeed, the complexity of processed mass shapes and the difficulty to distinguish between them require the use of appropriate descriptors. In this paper, suitable characterization methods for breast pathologies are proposed and the study of different classification methods is addressed. In order to analyze the mass shapes, a segmentation is performed manually. Once the images are segmented, a study of various descriptors proposed in the literature is conducted. In order to compare different approaches of characterization, a comparative study is performed. The descriptors commonly used in the breast cancer field are compared to test their ability to characterize the breast lesions. Obtained results show that statistical approaches of texture provides the best classification result.**

Keywords-Breast cancer; Computer Aided Diagnosis (CAD); Characterization; Selection; Classification; Evaluation

I. INTRODUCTION

Breast cancer is the leading cause of cancer deaths among the female population [1]. The only way today to reduce this mortality is the early detection of breast cancer using imaging techniques of the breast. Mammography is one of the most effective tools for prevention and early detection of breast cancer. It is a screening tool used to localize the suspicious tissues in the breast such as micocalcifications, masses, architectural distortion and bilateral asymmetry [2]. A mass is defined as a space-occupying lesion seen in at least two different projections [3]. Mass density can be high, isodense, low, and fat containing. Moreover, mass margin can be circumscribed, micro-lobulated, indistinct, and spiculated. Finally, mass shape can be round, oval, lobular, and irregular [4]. In recent years, screening campaigns are organized in several countries. These campaigns generate a huge stream of mammograms and it is still difficult for expert radiologists to provide accurate and consistent analysis. Hence the importance of the Computer Aided Diagnosis (CAD) systems which have been used widely in order to help the radiologists in detecting lesions and in making diagnostic decisions [4]. The radiologist is responsible for making the final decision.

A generic CAD system includes segmentation, feature extraction, and classification stages. In this paper, the objective is to develop a novel system for the diagnosis of breast masses focusing on the feature extraction stage. The use of automated image analysis techniques for screening of mammography images helps radiologists in the diagnosis of breast cancer.

In the process of pattern recognition, the goal is to achieve the best classification rate using required features. The extraction of features from regions of the image is one of the important phases in pattern recognition and many researchers are interested in this research domain. Features are defined as the input variables which are used for class prediction. The quality of a feature is related to its ability to discriminate observations from different classes. The description task often generates a large number of features and the obtained feature space may include a large number of irrelevant features. This will induce greater computational cost, occupy a lot of storage space and decrease the classification performance. Thus, a feature selection phase is needed to avoid these problems. In this current study, we propose an automated computer scheme in order to select an optimal subset of features for masses classification in digital mammography.

In the literature, various numbers of techniques are studied to describe breast abnormalities in digital

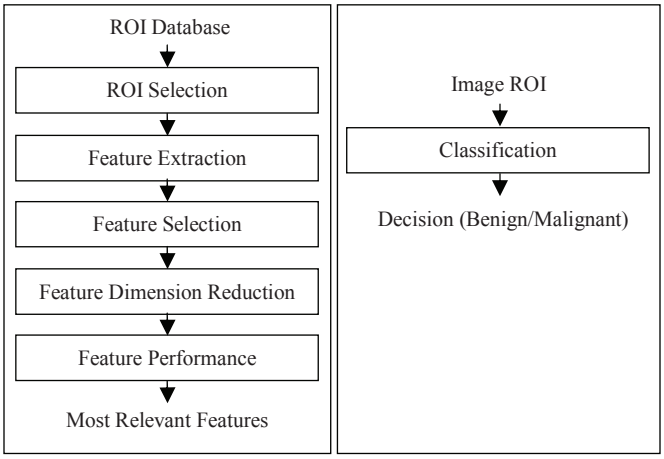
978-1-4799-5934-1/14/$31.00 ©2014 IEEE

mammograms. A lot of research has been done on the textural and shape analysis on mammographic images [1, 4, 7].

The remaining of this paper is organized as follows. In the next section, an overview of the proposed approach is illustrated. Sections 3 to 5 describe the process of selecting features. Section 6 presents the combination of the three classifiers. The experimental results are evaluated and discussed in section 7. Finally, concluding remarks are given in the last section.

II. METHODOLOGY

The proposed approach is composed of two main stages: characterization and classification stages. Each of these introduced stages is explained in detail by the flowchart given by figure 1.



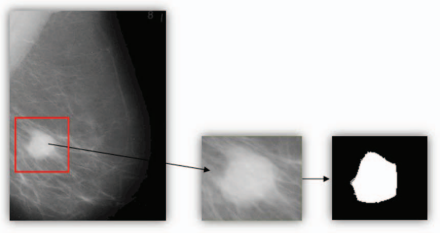
(a) Characterization stage (b) Classification stage

Figure 1. Flowchart of the proposed approach performed in this study

A. ROI Selection

The first step in the system is to perform ROI selection. This involves separating the suspected areas they may contain abnormalities from the image. The ROIs are usually very small and limited to areas being determined as suspicious regions of masses. The suspicious area is an area that is brighter than its surroundings, has almost uniform density, has a regular shape with varying size, and has fuzzy boundaries. In this paper, manual segmentation has been utilized. In fact, each ROI has been selected by an image editing tool. For instance, figure 2 depicts the output ROI obtained after executing this operation.

After running this step, we obtain for each mammogram a binary mask. Figure 2.c shows one of the output acquired from this operation by using figure 2.a as input. In figure 2.c, the white region indicates the mass and the black region indicates the background. The resulting binary image produced will be useful in shape description.



(a) (b) (c)

Figure 2. (a) Manual selection of ROI specified by an expert radiologist, (b) ROI containing only one mass, (c) Mass shape in binary format

B. Feature Extraction

Various features have been proposed in literature for the characterization of masses. These features are organized into families according to their nature (Gabor, shape, Tamura,...) [19]. The majority of studies choses one family and analyses its performance. In this work, we propose to study the performance of a set of feature families. Then, we make a comparison between these families in order to select the best feature set. Finally, the most discriminant features are selected from the obtained feature set. The process is described in figure 3.

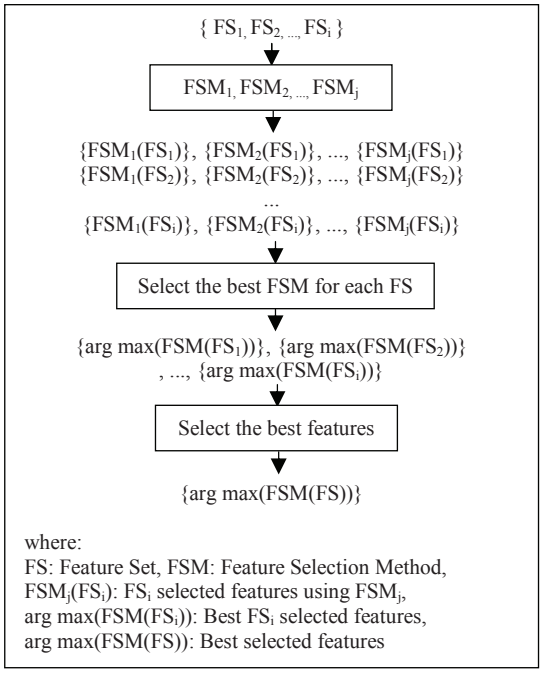


Figure 3. Feature selection steps

Two kinds of features set can be used for the discrimination between benign and malignant masses: Texture features and shape features. The features used in this study are listed in table 1.

TABLE I. FEATURES USED IN THIS STUDY

	Features	
Statistical textural features	First Order Statistics (FOS) [6]	Mean value of gray levels, Mean square value of gray levels, Standard Deviation, Variance, Skewness and Kurtosis.
	Gray Level Co-occurrence Matrices (GLCM) [7]	Mean, Variance, Dissimilarity, Correlation, Diagonal Moment, Energy, Homogeneity, Entropy, Contrast, Sum Average, Sum Entropy, Sum Variance, Correlation information 1, Correlation information 2, Difference Entropy, Difference Mean, Difference Variance, Promenance and Shade. Four values were obtained for each feature corresponding to the four directions (θ=0°, 45°, 90° and 135°)
	Gray Level Difference Matrices (GLDM) [8]	Mean, Contrast, Angular Second Moment, Entropy and Inverse Difference Moment.
	Gray Level Run Length Matrices (GLRLM) [8]	Short Runs Emphasis (SRE), Long Runs Emphasis (LRE), Gray Level Non-Uniformity (GLNU), Run Length Non-Uniformity (RLNU), Run Percentage (RP), Low Gray Level Run Emphasis (LGRE), High Gray Level Run Emphasis (HGRE), Short Run Low Gray Level Emphasis (SRLGLE), Short Run High Gray Level Emphasis (SRHGLE), Long Run Low Gray Level Emphasis (LRLGLE) and Long Run High Gray Level Emphasis (LRHGLE). Four values were computed for each feature, corresponding to the angles of 0°, 45°, 90° and 135°.
	Tamura features [9]	coarseness, contrast, direction, linearity, regularity and roughness
Frequential textural features	Gabor transform [10]	Five scales and six orientations are used. the features vector f is created using μ_{mn} and σ_{mn} as the feature components.
	Two-dimensional wavelet transform	Energy measures at different resolution levels
Shape features	Seven Hu's invariant moments [15, 16]	
	Other features [17]	Area, Perimeter, Compactness, Aspect ratio, Major Axis Length, Minor Axis Length, Eccentricity, Orientation, Convex polygon area, Euler Number, Equivalent diameter, Solidity, Extent and Mean Intensity.

III. FEATURE SELECTION

At the stage of feature analysis, 159 features are generated for each ROI. The feature space is very large and complex due to the wide diversity of the normal tissues and the variety of the abnormalities. Only some of them are significant. With a large number of features the computational cost will increase. Irrelevant and redundant features may affect the training process and consequently minimize the classification accuracy. The main goal of feature selection is to reduce the dimensionality by eliminating irrelevant features and selecting the best discriminative ones. Many search methods are proposed for feature selection. These methods could be categorized into sequential or randomized feature selection. Sequential methods are simple and fast but they could not backtrack, which means that they are candidate to fall into local minima. The problem of local minima is solved in randomized feature selection methods, but with randomized methods it is difficult to choose proper parameters.

To avoid problems of local minima and choosing proper parameters, we have opted to use five sequential and randomized feature selection methods. Then, using an evaluation criterion, we retain the method that gives the best results among them. The feature selection methods that we have used are: Tabu Search (TS), Genetic Algorithm (GA), ReliefF Algorithm (RA), Sequential Forward Selection (SFS) and Sequential Backward Selection (SBS). We used all extracted features as input for selection methods individually.

A. Tabu Search

The Tabu Search is a meta-heuristic approach that can be used to solve combinatorial optimization problem. TS is conceptually simple and elegant. It has recently received widespread attention [12]. It is a form of local neighbourhood search. It differs from the local search techniques in the sense that tabu search allows moving to a new solution which makes the objective function worse in the hope that it will not trap in local optimal solutions. Tabu search uses a short-term memory, called tabu list, to record and guide the process of the search. To avoid cycling, solutions that were recently explored are declared forbidden or Tabu for a number of iterations. The Tabu status of a solution is overridden when certain criteria (aspiration criteria) are satisfied. In this study, the size of the tabu list is set to be equal to 2. This size is reasonable to ensure diversity. In addition to the tabu list, we can also use a long-term memories and other prior information about the solutions to improve the intensification and/or diversification of the search.

B. Genetic Algorithm

Genetic algorithm is based in randomness in its search procedure to escape falling in local minima. By GA we create a population of solutions based on the chromosomes and evolve the solutions by applying genetic operators such as mutation and crossover to find best solution based on the predefined fitness function.

The initial population of GA is created using the following formula:

$$P=round\,((L-1)\times rand\,(DF,200\times DF))+1 \quad (1)$$

where L and DF represent, respectively, the number of input features and the desired number of selected features (In this work, $DF=L/2$).

In GA, we use a fitness function based on the principle of Max-Relevance and Min-Redundancy (mRMR) [13]. The idea of mRMR is to select the set S with m features $\{x_i\}$ that satisfies the maximization problem:

$$\max_i \Phi_i(D,R),\ \Phi(D,R)=D-R \quad (2)$$

where D and R represent the max-relevance and min-redundancy, respectively, and are computed by the following formula:

$$D = \frac{1}{|S|}\sum_{x_i} I(x_i, y) \quad (3)$$

$$R = \frac{1}{|S|^2}\sum_{x_i,x_j \in S} I(x_i, x_j) \quad (4)$$

where $I(x_i,y)$ and $I(x_i,x_j)$ represent the mutual information, which is the quantity that measures the mutual dependence of the two random variables and is computed using the following formula:

$$I(x,y)=H(x)+H(y)-H(x,y) \quad (5)$$

where H(.) is the entropy.

We applied TS and GA ten times, each time the results are slightly changed from previous. Then, the results are ordered with respect to the score obtained by TS algorithm and fitness function, respectively. The best solution corresponds to the highest score for both TS and GA.

C. *ReliefF Algorithm*

The key idea of the ReliefF Algorithm is to estimate the quality of features according to how well their values distinguish between instances that are near to each other [14]. ReliefF assigns a grade of relevance to each feature, and those features valued over a user given threshold are selected.

D. *Sequential Forward Selection and Sequential Backward Selection*

In SFS, we start with empty list of selected feature, and successively we add one useful feature to the list until no useful feature remains in the extracted input list. The selection of useful feature is based on a criteria function. SBS instead begins with all features and repeatedly removes a feature whose removal yields to the maximal performance improvement.

IV. FEATURE DIMENSION REDUCTION

After selecting the most relevant descriptors, the next step is to reduce the dimensionality of the feature set in order to minimize the computational complexity. The dimension reduction is carried out using the Principal Component Analysis (PCA) method [18].

V. FEATURE PERFORMANCE

This stage has two main purposes: choosing the input parameters giving the best results and comparing types of features. Features are evaluated according to their discriminatory power. Five measurements derived from Rodrigues approch [11] can be used as criteria for this purpose: Class Classifier (CC) measurement, Class Variance (CV) measurement, Total Class Classifier (TCC) measurement which corresponds to the sum of CCs, Weighted Average of Class Classifiers (WACC) measurement which corresponds to the weighted average of CCs and Weighted Average of Class Variances (WACV) measurement which corresponds to the weighted average of CVs.

VI. CLASSIFICATION

Once the features relating to masses have been extracted and selected, they can then be used as inputs to the classifier to classify the regions of interest into benign and malignant. In the literature, various classifiers are described to distinguish between normal and abnormal masses. Each classifier has its advantages and disadvantages. In this work, we propose to study the performance of three classifiers: Multi Layer Perceptron (MLP), Support Vector Machines (SVM) and K-Nearest Neighbors (K-NN) which have performed well in mass classification [19]. Then, we make a combination of these classifiers in order to exploit the advantages of each one of them and to improve the accuracy and efficiency of the classification system. The multiclassifier system is described in figure 4.

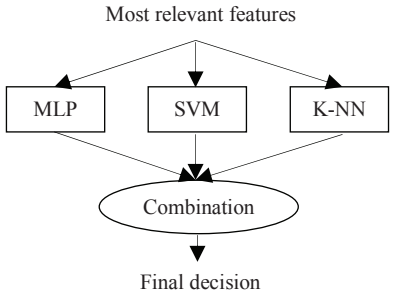


Figure 4. The multiclassifier system

A. *Multi Layer Perceptron*

MLP is regarded as one prominent tool for training a system as a classifier [19]. It uses simple connection of the artificial neurons to imitate the neurons in human. It also has a huge capacity of parallel computing and powerful remembrance ability. MLP is organized in successive layers: an input layer, one or more hidden layers and output layer.

B. *Support Vector Machines*

SVM is an example of supervised learning methods used for classification [20]. It is a binary classifier that for each given input data it predicts which of two possible classes comprises the input. It is based on the idea to look for the hyperplane that maximizes the margin between two classes.

C. *K-Nearest Neighbors*

The K-nearest neighbors of an unknown sample are selected from the training set in order to predict the class label as the most frequent one occurring in the K-neighbors. K is a parameter which can be adjusted, it is usually an integer. It specifies the number of nearest neighbors. In this

work, K is equal to 3 and Euclidean distance is used as a distance metric.

The K-NN classifier is well explored in the literature and has been proved to have good classification performance on a wide range of data sets.

In order to accumulate the classifiers advantages and to exploit the complementarity between the three classifiers, we applied a fusion method: Majority Vote (MV) which is based on voting algorithm. Then, the accuracy obtained after this combination is used as final result of classification.

VII. Experimental Results

The dataset used for this experiment is composed of 322 mammograms from the MIAS database [5] which includes 208 normal breasts, 63 ROIs with benign lesions and 51 ROIs with malignant lesions, 80% set of abnormal images are used for training and 20% used for testing. A digitised mammogram is initially fed into the system. Any areas representing suspected part of a mammogram are cropped from the image and further analysis is carried out on these. Each of the regions of interest (ROIs) contains only one mass. Textural and shape features are calculated and extracted from the cropped ROI's. In order to achieve optimum discrimination, a normalization procedure can be performed by assigning a weight to all features in order to measure their similarity on the same basis. The technique used was to project each feature onto a unit sphere. Features values are normalized between 0 and 1 according to the following formula:

$$Y=(X-Min)/(Max-Min) \quad (6)$$

where X is the initial feature value and Y is the feature value after normalization.

Our experiment involved a binary classification, in which the classifier should recognize, whether a given ROI contains benign or malignant tissue. The goal of the experiments is to test which descriptor is best for describing mammographic images i.e. to find out which combination of descriptors will give the best results. The effectiveness of the different feature sets is listed in table 3.

TABLE II. Performance Obtained for GLRLM Features

		Benign Class		Malignant Class		Total		
Select. Method	Dir	CC	CV	CC	CV	TCC	WACC	WACV
TS	H	0.019	0.175	0.911	0.487	0.931	0.402	0.309
	V	0.039	0.1662	0.955	0.338	0.995	0.432	0.239
GA	H	0.019	0.158	0.955	0.417	0.975	0.420	0.269
	V	0.039	0.088	0.985	0.324	1.024	0.444	0.189
ReliefF	H	0.039	0.160	0.941	0.513	0.980	0.425	0.311
	V	0.039	0.127	0.926	0.341	0.965	0.419	0.219
SFS	H	0.019	0.167	0.897	0.525	0.916	0.395	0.321
	V	0	0.205	0.897	0.335	0.897	0.384	0.260
SBS	H	0.019	0.184	0.897	0.515	0.916	0.395	0.326
	V	0.039	0.138	0.970	0.341	1.009	0.438	0.225

TABLE III. Features Performance

Features	Input Parameters	TCC		WACC		WACV
		Rank	Value	Rank	Value	
GLRLM (using GA)	Vertical	1	1.024	1	0.444	0.189
All Texture Features (using SFS)		2	0.995	3	0.432	0.303
GLCM (using SFS)	d=1,θ=45°	3	0.990	2	0.435	0.176
All Texture frequential Features (using TS)		4	0.990	4	0.427	0.306
FOS (using TS)	None	5	0.985	5	0.422	0.118
Wavelets (using ReliefF)	None	6	0.975	6	0.420	0.295
All Features (using SBS)		7	0.970	7	0.416	0.366
Gabor Filters (using SFS)	None	8	0.970	8	0.416	0.428
Tamura (using SBS)	None	9	0.960	9	0.414	0.311
All Texture Statistical Features (using GA)		10	0.955	10	0.409	0.269
Hu's Invariant Moments (using SFS)	None	11	0.955	11	0.409	0.354
GLDM (using SFS)	d=1	12	0.955	12	0.409	0.426
Shape (using SFS)	None	13	0.941	13	0.403	0.207
All Shape Features (using ReliefF)		14	0.911	14	0.390	0.350

The best performances obtained for the texture and shape features corresponds to GLRLM features using GA and Hu's invariant moments using SFS, respectively. The GLRLM performance is more than performance obtained for GLRLM and Hu's invariant moments features as shown in table 4.

TABLE IV. Performance Obtained for GLRLM and Hu's Invariant Moments Features

	Benign Class		Malignant Class		Total		
Select. Method	CC	CV	CC	CV	TCC	WACC	WACV
TS	0.019	0.290	0.970	0.365	0.990	0.427	0.322
GA	0.039	0.234	0.955	0.342	0.995	0.432	0.281
ReliefF	0.019	0.274	0.970	0.367	0.990	0.427	0.314
SFS	0	0.406	0.941	0.404	0.941	0.403	0.405
SBS	0.019	0.275	0.955	0.368	0.975	0.420	0.315

Most relevant GLRLM and Hu's invariant moments features are GLNU, RLNU, LGRE, SRLGLE, LRLGLE and Second moment.

A number of different measures are commonly used to evaluate the classification performance. These measures defined in table 6 are derived from the confusion matrix which describes actual and predicted classes as shown in table 5.

TABLE V. Confusion Matrix

Actual	**Predicted**	
	Positive	Negative
Positive	TP (True Positive)	FP (False Positive)
Negative	FN (False Negative)	TN (True Negative)

where:

TP: predicts abnormal as abnormal, FP: predicts abnormal as normal, TN: predicts normal as normal and FN: predicts normal as abnormal.

TABLE VI. PERFORMANCE CRITERIA

Criterion	Formulas
Rate of Positive Predictions	RPP=(TP+FP)/(TP+FN+FP+TN)
Rate of Negative Predictions	RNP=(TN+FN)/(TP+FN+FP+TN)
True Positive Rate (Sensitivity)	TPR=TP/(TP+FN)
False Negative Rate	FNR=FN/(TP+FN)
False Positive Rate	FPR= FP/(TN+FP)
True Negative Rate (Specificity)	TNR=TN/(TN+FP)
Positive Predictive Value	PPV=TP/(TP+FP)
Negative Predictive Value	NPV=TN/(TN+FN)
Accuracy	AC=(TP+TN)/(TP+FP+TN+FN)
Mathews Correlation Coefficient	MCC=(TP×TN-FP×FN)/ √((TP+FP)×(TP+FN)×(TN+FP)×(TN+FN))
Fmeasure	F=(2×PPV×TPR)/(PPV+TPR)
Gmean	G=√(TNR×TPR)

A Receiver Operating Characteristic (ROC) curve is also used for this stage. It is a plotting of true positive as a function of false positive. Higher ROC, approaching the perfection at the upper left hand corner, would indicate greater discrimination capacity.

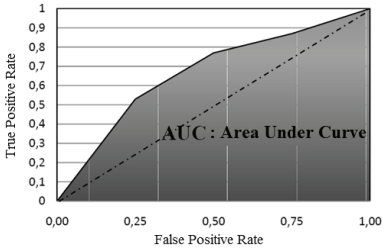


Figure 5. The ROC curve

VIII. CONCLUSION

Digital mammography is the most common method for early breast cancer detection. Automated analysis of these images is very important, since manual analysis of these images is slow, costly and inconsistent. In this paper, we made an analysis using nine different techniques for feature extraction. According to the provided examination, we can conclude that the best feature performance was achieved in the case of GLRLM descriptor.

REFERENCES

[1] A.R. Razavi, H. Gill, H. Ahlfeldt, and N. Shahsavar, Predicting metastasis in breast cancer: Comparing a decision tree with domain experts,. *J. Medical Systems,* vol. 31, pp.263-273, 2007.

[2] American Cancer Society, "Cancer prevention and Early Detection Facts and Figures", Atlanta, GA: American Cancer Society, 2003.

[3] American College of Radiology, *ACR BI-RADS. Mammography, Ultrasound & Magnetic Resonance Imaging, 4th ed*., American College of Radiology, Reston, VA, 2003.

[4] J. Bozek, M. Mustra, K.Delac, and M. Grgic, .A survey of image processing algorithms in digital mammography,. *J. Recent Advances in Multimedia Signal Processing and Communications*, vol. 231, pp. 631. 657, 2009.

[5] J. Suckling *et al*, .The Mammographic Image Analysis Society digital mammogram database,. *Exerpta Medica, Int. Congress Series 1069,* 1994, pp. 375-378.

[6] R. C. Gonzalez, and R. E. Woods, *Digital Image Processing*, Prentice-Hall, Inc., New Jersey, 2002, pp. 76-142.

[7] R. M. Haralick, K. Shanmugam and I. Dinstein, "Textural features for image classification", *IEEE Trans. System Man. Cybernetics*, vol. SMC-3, pp. 610–621, 1973.

[8] M.M. Galloway, Texture classification using gray level run length, Comput. Graphics Image Process. 4 (1975) 172–179.

[9] H. Tamura, S. Mori, and T. Yamawaki, "Texture features corresponding to visual perception", IEEE Trans. On Systems, Man, and Cybernetics, vol. Smc-8, No. 6, Juin 1978.

[10] B. S. Manjunath and W. Y. Ma. "Texture features for browsing and retrieval of large image data" IEEE Transactions on Pattern Analysis and Machine Intelligence, (Special Issue on Digital Libraries), Vol. 18 (8), August 1996, pp. 837-842.

[11] José Fernando Rodrigues Jr, Agma J.M.Traina et Caetano Traina Jr. "Enhanced Visual Evaluation of Feature Extractors For Image Mining". The 3rd ACS/IEEE International Conference on Computer Systems and Applications. 2005.

[12] Muhammad Atif Tahir, Ahmed Bouridane, Fatih Kurugollu, "Simultaneous feature selection and feature weighting using Hybrid Tabu Search/K-nearest neighbor classifier", Pattern Recognition Letters 28 (2007) 438–446.

[13] Hanchuan Peng, Fuhui Long, and Chris Ding, "Feature selection based on mutual information: criteria of max-dependency, max-relevance, and min-redundancy," IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 27, no. 8, pp. 1226-1238, 2005.

[14] Yuxuan SUN, Xiaojun LOU, Bisai BAOA, "Novel Relief Feature Selection Algorithm Based on Mean-Variance Model", Journal of Information & Computational Science 8: 16 (2011) 3921–3929.

[15] Hu, M. K. (1962), "Visual Pattern Recognition by Moment Invariants", IEEE Transactions on Information Theory, Vol IT-8.

[16] Chaur-Chin Chien (1993), "Improved Moment Invariants for Shape Discrimination", Pattern Recognition Vol. 26, No. 5, pp. 683-686.

[17] Mingqiang, Y., Kidiyo, K., and Joseph, R. (2008), "A survey of shape feature extraction techniques", Pattern Recognition, pages 43–90.

[18] Principal Components Analysis. 36-490. Spring 2010.

[19] H.D. Cheng, X.J. Shi, R. Min, L.M. Hu, X.P. Cai, H.N. Du, "Approaches for automated detection and classification of masses in mammograms", Pattern Recognition 39 (2006) 646 – 668.

[20] V. Vapnik, "The Nature of Statistical Learning Theory", Springer-Verlag, 1995.