A Computer-Aided Diagnosis System for Breast Cancer Combining Features Complementarily and New Scheme of SVM Classifiers Fusion

Nabiha Azizi, Yamina Tlili-Guiassa and Nawel Zemmal

Computer Science Department, Badji Mokhtar University, Bp n 12, Annaba, 23000, Algeria

nabiha.azizi@univ-annaba.net

Abstract

Breast cancer is reported as the second most deadly cancer in the world and the main of mortality among the women, on which public awareness has been increasing during the last few decades. This is why several works are made to develop help tools for disease diagnosis. Computer-Assisted Diagnosis (CAD) is based on 3 main steps: segmentation, feature extraction and classification in order to generate a final decision. Classification phase is the key step in this process; for that, many research have been accentuated in this domain and many techniques were be proposed. Kernel combination is a current active topic in the field of machine learning. It takes benefit of classifier algorithms. it allows to choose the kernel functions according to the features vectors. The combination of Kernel-based classifiers was proposed as a research way allowing reliability recognition by using the complementarily which can exist between classifiers. This study investigated a computer-aided diagnosis system for breast cancer by developing a novel classifier fusion scheme based on fusion of three support vector machine classifier. Each one is associated with an homogenous family of features (Hu moments; central moments, Haralick moment) as efficient learning algorithm and diversity between features family as fusion criteria to ensure best performance. Our experiments demonstrated that developed system using Database for Screening Mammography (DDSM) database achieve very encouraging results when compared with past works using the same information.

Keywords: *Support Vector Machine classifier, Computer-aided diagnosi; mammography, Hu moments; central moments, GLCM (Grey Level Co-occurrence Matrix); fusion classifier, majority voting.*

1. Introduction

The breast cancer constitutes, in world, the most frequent cause of cancer death at the woman. In Algeria, the breast cancer represents nearly 50% of gynecological cancer at the woman, and during these fifteen (15) last years, the incidence of the breast cancer was multiplied by three (3). Because of it late diagnosis, it often results from it a heavy, mutilating and expensive treatment (processing) which is accompanied by a high mortality rate[1].Various studies confirmed that detection in early stage of the infra-clinical cancers can improve the forecast and that the mammography constitutes in that case the best diagnostic technique. All the radiologists recognize the difficulty of the mammographic examination which still increases by the tissues type of the examined breast, the conditions of realization, the number of available stereotype (pictures), *etc.* [2] (Figure 1). The uniformization of the screening, the decrease of experts number and the quality requirements regarding public health make indispensable recourse to technologies able to help in the diagnosis. For this reason, several researches were led these last years to develop help tools for the diagnosis (CAD Computer-Assisted Detection) of this disease. These tools present a rather general scheme: After the segmentation of the mammography, the following step is the extraction of the characteristics of the image, then interpretation in order to identify the anomalies. Because the number of training and test examples corresponds to the number of patients, the number of such examples is typically lower than for most other application areas.

Figure 1. Breast Images Sample: (a) Normal Fatty Tissue Breast (b) Normal Dense Glandular Tissue Breast (c) Fatty Glandular Tissue with Malignant tumor (d) Fatty Glandular Tissue with Benign Tumor

The examples also tend to be relatively imbalanced; the number of examples of anomalies or patients with diseases is, fortunately, much lower than the number of examples of normal patients. However, this can pose difficulties for machine learning algorithms especially classification algorithms. With so few positive (*e.g.*, disease present) examples, classification algorithms will tend to be strongly biased toward predicting that new examples are negative. However, incorrectly predicting that an example is negative (false negative) is typically much worse than incorrectly predicting that an example is positive. The number of attributes also tends to be much higher than the number of examples in problems such as human genome analysis and image analysis. This mandates the use of feature selection and/or feature extraction methods, and the selection of the appropriate such method(s) add significantly to the challenge of using machine learning for these problems.

Most CAD research has been performed using relatively homogeneous data sets collected at one institution, acquired using one type of digitizer or digital detector, or using features drawn from one source such as human-interpreted findings versus computer extracted features [3]. Increasingly however, there is a trend toward boosting diagnostic performance by combining data from many different sources to create heterogeneous data.

We defined heterogeneous data as comprising multiple, distinct groups. Specifically, for this study, we considered as heterogeneous various types of features extracted from the same image, especially computer-extracted and human-extracted features.

Combining heterogeneous data types for classification is a difficult machine-learning problem, but one that has shown promise in bioinformatics applications. To meet the challenge of combining heterogeneous data types, we turned to ensemble of classifiers techniques that operates by the following two steps: (1) select the appropriate kernel function of SVM Classifiers use feature subsets to generate pool of classifiers and (2) these individual selected classifiers are then optimally combined by using multiple classifiers tools.

In fact, MCS is a set of individual classifiers whose decisions are combined when classifying new patterns. There are many different reasons for combining multiple classifiers to solve a given learning problem [4, 5]. First, MCSs try to exploit the local different behaviour of the individual classifiers to improve the accuracy of the overall system.

The principal diagnostic technique for breast cancer detection is X-ray mammography. Screening programs have been introduced in many European countries to invite women to a periodic radiological breast examination.

In such screenings, radiologists are often required to examine large numbers of mammograms with a double reading, that is, two radiologists examine the images independently and then compare their results. In this way an increment in sensitivity (the rate of correctly identified images with a lesion) of up to 15% is obtained [6, 7]. Since a double reading allowed by the availability of two skilled radiologists for the same reports is not common to realize in almost all radiological centres. In recent years, different computer aided detection (CAD) systems have been developed as a support to radiologists working in mammography: one may hope that the ‘‘second opinion’’ provided by CAD might represent a lower cost alternative to improve the diagnosis.

The role of our system is to provide to doctors symbolic information on the image contents. The interpretation of the medical images represents an effective tool for the processing of visual information; she allows inevitably not only to detect and to locate the tumor but also to specify or rather to envisage the gravity of the tumor, in term of benign / malignant, by exploiting the characteristics extracted from the image. However, it is difficult to find a unique feature representation or distance function to compare images accurately for all types of queries. In other words, different feature representations might be complementary in nature and will have their own limitations.

In this work, we present an approach of analysis and help to the diagnosis based on the combination of classifiers because of his crucial interest and proved systems reliability using this technique. Proposed approach begins with a manual segmentation, translated by the extraction of the edge of the mass. We begin afterward the phase of characteristics extraction, indeed we used three heterogeneous families of characteristics separated between three vectors of characteristics and which are: the matrix of co-occurrence (which aims at extracting texture characteristics), the moments of Hu and central moments (these last two families are used to describe image shape). For classification stage, a supervised classification tool is used. We are focus on support vector machines (SVM) because they limit the risk of over-training, their capacities of regularization (this risk being particularly important when the number of characteristic is big) and their good results in practice. Then, we are going to combine final decisions of every classifier by applying combination functions of all results given by every SVM to generate a definitive decision of type of tumor (benign or malign).

The rest of the article appears as follows: in the nest section, we illustrate combination of classifiers technique. Section 3 presents the general architecture of our system as well as the description of modules composing of the proposed approach. Experimental part which includes the database used for the validity, the results of each of the classifiers and the general system combining these classifiers is described in Section 4. Finally we will end with a conclusion and some perspectives of future extensions.

2. Features Extraction, Selection and complementarity in Classifier combination paradigm

Individual classification models are recently challenged by combined pattern recognition systems, which often show better performance. Multiple classifier systems (MCSs) were shown to outperform single classifiers for a wide range of classification problems [5, 8]. The reason is that a combination of classifiers reduces risks associated with picking an inadequate single classifier, choosing a space of classifiers not containing the optimal classifier, and falling into local error minima during training [9].

Classifier fusion assumes that all individual classifiers are competitive, instead of complementary. For this reason, each component takes part in the decision of classifying an input test pattern.

The problem arouse naturally as a need of improvement of classification rates obtained from individual classifiers. Fusion of data/information can be carried out on three levels of abstraction closely connected with the flow of the classification process: *data level* *fusion, feature level fusion, and classifier fusion* [10]. There is little theory about the first two levels of information fusion. However, there have been successful attempts to transform the numerical, interval and linguistic data into a single space of symmetric trapezoidal fuzzy numbers and some heuristic methods have been successfully used for feature level fusion.

Two strategies are accepted for combining classifier decisions: fusion and selection. In classifier fusion, each member of the ensemble is supposed to have knowledge of the whole feature space and thus, they are either complementary or competitive [11]. That condition not always is fulfilled. In some cases, the simple voting might perform even worse than any of the members of the ensemble. On the other hand, studies pointed out that the selection between different classifier could be better accomplished if the classifiers are "specialized" on the recognition of different partitions of the data set. Each classifier will be thus required to exhibit a high value of classification accuracy only for the patterns that belong to a particular "region" of the feature space [12].

In this approach, called classifier selection, only one classifier decides the class label for the test sample. In the simple voting (by majority), the final decision is taken according to the number of votes given by the individual classifiers to each one of the classes, thus assigning the test pattern to the class that has obtained a majority of votes.

3. New Scheme of SVM Classifier Fusion based on Kernel Function Adaptation and Features Diversity

For mass classification (as the most effective stage) is feature extraction. Texture and shape are the commonly used features in the analysis and interpretation of images. Here, we suggest to distinguish underlying three families if features based on textures and shape in mammography based on classifier fusion approach [4] in the three following stages:

- feature extraction of texture and shape based image which are: Hu moments; central moments, Haralick moment;

- individual training of the three SVM classifiers, each one with an homogenous features- In experimentation, following feature extraction, an appropriate kernel function of SVM classifier is utilized in breast mass classification. For this, we have tested four (4) kernel functions to select the appropriate one for each feature family. Once the kernel function was be determined for each features family, the associated SVM will be used to construct the MCS which their output classes were be combined.

- Using majority voting as fusion function to generate final decision of the mass.

The proposed approach can be summarized by the following scheme (Figure 2).

3.1. The Learning Base

The process of classification always requires a base of learning as entry. To create a base of learning, it is to have individuals (here images medical) which we know with certainty the membership class.

In our study we used 1100 images from the base Digital Database for Screening Mammography (DDSM). It is the Marathon database of South Florida University.

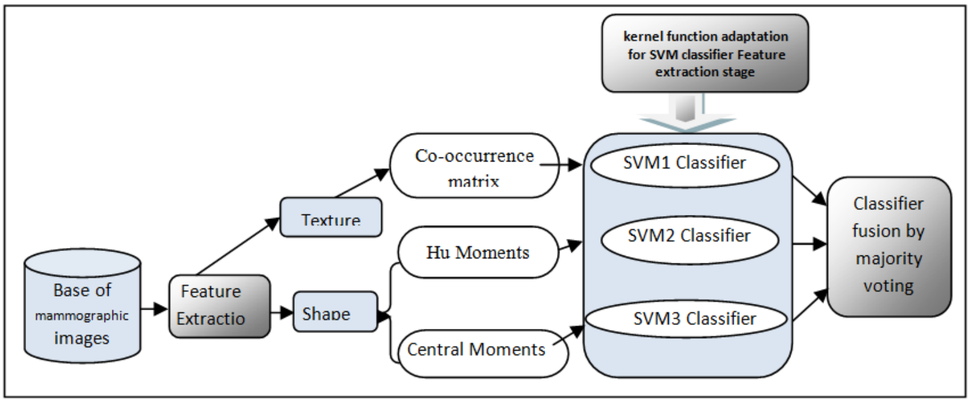

Figure 2. Classification Process of Mammographic Images

A description of this base was made by "American college of Radiology" in the lexicon of BI RADS (Breast Imaging Reporting and Data System). The corresponding base contains 2620 patients' files classified in three groups: normal, benign and malignant. Every case contains the extern oblique views(MLO) and external face (CC) of each breast, and metadatas associated with every case. They are contained in the file`` *ics*`` and Overlay files when they exist. The lexicon of the BI RADS contains a guide allowing standardizing the language used in the mammographic reports, which are protected in Overlay.7files [13].

3.1.1. Extraction of mass contour: The extraction of the masses, in our approach, is manual because our base of learning contains mass images surrounded by a red circle. In the segmentation, we are going to extract the outline of the shape to be analyzed with the help of "Image J " as image processing tool. Our image will have the following shape (Figure 3).

Figure 3. Extraction of the Mass Outline by *ImageJ* Tool

3.2. Features Extraction

After segmentation of the mammographic image, the next step is the characteristics extractions that describe the regions of the image. The methods of image analysis are variable according to the types of the feature extracted from the image as the morphological characteristics, texture characteristics and shape characteristics. Here, we used three families of features, one of them is based on the texture and the two others are based on the shape, and which are described as follows:

3.2.1. Co-occurrence Matrix: Texture is one of the significant characteristics used in identifying objects of interest or regions in an image. Upon given a query region, similar regions can be retrieved in CAD applications by using texture features [14].

In [15], Haralick described texture as one of three fundamental types of feature used by humans to distinguish regions in a grayscale image.

In 1973 Haralick introduced the co-occurrence matrix and texture features for automated classification of rocks into six categories. Today, these features are widely used for different kinds of images. Some of these features include angular second moment, contrast, correlation, as well as a variety of entropy measures (Eq. 1).

To analyse the benefice of combining these futures with others, our first feature vector is based on image texture and it is calculated from the co-occurrence matrix

We can resume method objective as a statistical method which consists in constructing of co-occurrence matrix to represent the relations between the pixels of an image.The matrix is the joint probability between two gray levels i, j which are given in a spatial relationship. This relation is defined in terms of the distance and angle between these two pixels. The angle is used to evaluate the direction of texture and the application of several values of distance can give a meaningful description of the size of the texture periodicity.

"Energy" $= f_1 = \sum_{i,j} g(i,j)^2$

"Energy" $= f_1 = \sum_{i,j} g(i,j)^2$

"Entropy" $= f_2 = -\sum_{i,j} g(i,j) \log_2 g(i,j)$, or 0 if $g(i,j) = 0$

"Correlation" $= f_3 = \sum_{i,j} \frac{(i-\mu)(j-\mu)g(i,j)}{\sigma^2}$

"Difference Moment" $= f_4 = \sum_{i,j} \frac{1}{1+(i-j)^2} g(i,j)$

"Inertia" $= f_5 = \sum_{i,j} (i-j)^2 g(i,j)$ (sometimes called "contrast")

"Cluster Shade" $= f_6 = \sum_{i,j} ((i-\mu) + (j-\mu))^3 g(i,j)$ (1)

"Cluster Prominence" $= f_7 = \sum_{i,j} ((i-\mu) + (j-\mu))^4 g(i,j)$

"Haralick's Correlation" $= f_8 = \frac{\sum_{i,j} (i,j) g(i,j) - \mu_t^2}{\sigma_t^2}$ where μ_t and σ_t are the mean and standard deviation of the row (or column, due to symmetry) sums.

Above, $\mu =$(weighted pixel average) $= \sum_{i,j} i \cdot g(i,j) = \sum_{i,j} j \cdot g(i,j)$ (due to matrix summetry), and

$\sigma =$ (weighted pixel variance) $= \sum_{i,j} (i-\mu)^2 \cdot g(i,j) = \sum_{i,j} (j-\mu)^2 \cdot g(i,j)$ (due to matrix summetry)

3.2.2. Hu Moments: The moments are an essential issue in the field of pattern analysis is the objects recognition regardless of their position, size and orientation.

Moments and the related invariants have been extensively analyzed to characterize the patterns in images in a variety of applications. The well-known moments include geometric moments, zernike moments, rotational moments, and complex moments.

Moment invariants are firstly introduced by Hu. In [16], Hu derived six absolute orthogonal invariants and one skew orthogonal invariant based upon algebraic invariants, which are not only independent of position, size and orientation but also independent of parallel projection.

The idea of using moments in shape recognition gained prominence when Hu (1962), derived a set of invariants using algebraic invariants, which can be resumed by the following formula (Eq. 2).

$$\begin{aligned}
M_1 &= (\eta_{20} + \eta_{02}),\\
M_2 &= (\eta_{20} - \eta_{02})^2 + 4\eta_{11}^2,\\
M_3 &= (\eta_{30} - 3\eta_{12})^2 + (3\eta_{21} - \eta_{03})^2,\\
M_4 &= (\eta_{30} + \eta_{12})^2 + (\eta_{21} + \eta_{03})^2,\\
M_5 &= (\eta_{30} - 3\eta_{12})(\eta_{30} + \eta_{12})[(\eta_{30} + \eta_{12})^2 - 3(\eta_{21} + \eta_{03})^2]\\
&\quad + (3\eta_{21} - \eta_{03})(\eta_{21} + \eta_{03})[3(\eta_{30} + \eta_{12})^2 - (\eta_{21} + \eta_{03})^2],\\
M_6 &= (\eta_{20} - \eta_{02})[(\eta_{30} + \eta_{12})^2 - (\eta_{21} + \eta_{03})^2]\\
&\quad + 4\eta_{11}(\eta_{30} + \eta_{12})(\eta_{21} + \eta_{03}),\\
M_7 &= (3\eta_{21} - \eta_{03})(\eta_{30} + \eta_{12})[(\eta_{30} + \eta_{12})^2 - 3(\eta_{21} + \eta_{03})^2]\\
&\quad - (\eta_{30} + 3\eta_{12})(\eta_{21} + \eta_{03})[3(\eta_{30} + \eta_{12})^2 - (\eta_{21} + \eta_{03})^2].
\end{aligned} \tag{2}$$

3.2.3. Central moments: As their name indicates it, the central moments are calculated from shape center. Next formula describes general spatial moments of the object.

From the spatial moments, the central moments can be derived by reducing the spatial moments with the center of gravity (x_c, y_c) of the object, so all the central moments refer to the center of gravity of the object. Expressed as formula the central moments are calculated as follows:

$$\mu_{p,q} = \int\int (x - x_c)^p (y - y_c)^q f(x,y)dxdy \tag{3}$$

Central moments with an order equal to 3 are:

$$\mu_{0,0} = m_{0,0} \tag{4}$$

$$\mu_{1,0} = \mu_{0,1} = 0 \tag{5}$$

The central moments of first or higher order can directly be derived from the spatial moments by:

$$\mu_{p,q} = \frac{m_{p,q}}{m_{0,0}} - \left(\frac{m_{1,0}}{m_{0,0}}\right)^p * \left(\frac{m_{0,1}}{m_{0,0}}\right)^q \tag{6}$$

Using above formula, the central moments of first and second order can be derived from spatial moments as follows (Eq.7):

$$\begin{aligned}
\mu_{1,0} &= \frac{m_{1,0}}{m_{0,0}} - \left(\frac{m_{1,0}}{m_{0,0}}\right) = 0\\
\mu_{0,1} &= \frac{m_{0,1}}{m_{0,0}} - \left(\frac{m_{0,1}}{m_{0,0}}\right) = 0\\
\mu_{2,0} &= \frac{m_{2,0}}{m_{0,0}} - \left(\frac{m_{1,0}}{m_{0,0}}\right)^2 = \frac{m_{2,0}}{m_{0,0}} - x_c^2\\
\mu_{0,2} &= \frac{m_{0,2}}{m_{0,0}} - \left(\frac{m_{0,1}}{m_{0,0}}\right)^2 = \frac{m_{0,2}}{m_{0,0}} - y_c^2\\
\mu_{1,1} &= \frac{m_{1,1}}{m_{0,0}} - \left(\frac{m_{1,0}}{m_{0,0}}\right) * \left(\frac{m_{0,1}}{m_{0,0}}\right) = \frac{m_{1,1}}{m_{0,0}} - x_c * y_c
\end{aligned} \tag{7}$$

Hu moments and central moments must be calculated from a binary image; for this, we have applied OTSU method as robust binarization method.

- ***OTSU Method:*** In computer vision and image processing, *Otsu's method* is used to automatically perform histogram shape-based image thresholding, or, the reduction of a graylevel image to a binary image. The algorithm assumes that the image to be thresholded contains two classes of pixels or bi-modal histogram (*e.g.* foreground and background) then calculates the optimum threshold separating those two classes so that their combined spread (intra-class variance) is minimal.[2] The extension of the original method to multi-level thresholding is referred to as the Multi Otsu method.[3] Otsu's method is named after Nobuyuki Otsu.

In fact, the histogram of the image is realized at first, then, image pixels are separated in two classes, the first one having a maximal level (255) and the second one a minimal level(0). Automatic thresholding method for separation of the histogram consists as its name indicates it to determining a threshold value for which the accumulated histogram attains 50 % of its maximal value. Within the framework of the binarization by Otsu method, the separation is made from the moments of the first second order know the average and the standard deviation. Furthermore, in the optics to ensure process independency from the number of points in the image, we have normalized calculated histogram.

3.3. Classification

Classification is the final stage of any image-processing system where each unknown pattern is assigned to a category. The degree of difficulty of the classification problem depends on the variability in feature values for objects in the same category, relative to the difference between feature values for objects in different categories Duda *et al.*, (2000).

The performance of the process is very much dependent on the capability of the classifier. Some classifiers that have been applied for mass classification include a statistical bayesian model [7], a three layered multi-layer perceptron (MLP) [19, 2], Adaptive Nero Fuzzy Inference System (ANFIS) classifier [16], Gaussian kernel based Radial Basis Function network [2], *K*-Nearest Neighbors (KNN) algorithm, support vector machine (SVM) [20, 9]. A detection scheme using successive enhancement learning SVM was proposed in [6]. In this paper kernel function adaptation are applied are applied to the to outperform mammograms classification using complenetarity between features.

Our objective is to develop an automated imaging system for mass detection of digital mammograms. Techniques such as neural networks (RN), fuzzy logic (FL) and support vector machine (SVM) are most commonly used. In this study, we use a new scheme of SVM (Support Vector Machine) fusion based on kernel function adaptation ensuring diversity between features.

3.4. SVM (Support Vector Machine) Classifier

For many years Neural Networks was the ultimate champion, it was the most effective learning algorithm.SVM became popular because of its success in handwritten digit recognition (in NIST (1998)). It gave accuracy that is comparable to sophisticated and carefully constructed neural networks with elaborated features in a handwriting recognition task. Much more effective “off the shelf ” algorithm than Neural Networks: It generalize good on unseen data and is easier to train and doesn’t have any local optima in contrast to neural networks that may have many local optima and takes a lot of time to converge. SVM has successful applications in many complex, real-world problems such as text and image classification, hand-writing recognition, data mining, bioinformatics, medicine and biosequence analysis and even stock market.

SVM is an emerging machine learning technology that has already been successfully used for image classification in both general and medical domain [5, 11, 17, 23]. It performs the classification between two classes by finding a decision surface that is based on the most informative points of the training set [22].

Figure 4 shows the simplest case in which the data vectors (marked by 'X's and'O's) can be separated by a hyperplane. In such a case, there may exist many separating hyperplanes. Among them, the SVM classifier seeks a separating hyperplane that produces the largest separation margin [22]. Many learning techniques attempt to minimize the classification error in the training phase, only. Those methods, however, do not guarantee alow error rate in the testing phase. In statisticall earning theory, the SVM is claimed to efficiently address this issue [23]. In a more general case in which the data points are not linearly separable in the input space, an on linear transformation isused to map the data vector X into a higher-dimensional space(called feature space) prior to applying the linear maximum-margin classifier. To avoid the potential pit fall of over-fitting in this higher-dimensional space, the SVM uses a kernel function in which then on linear mapping is implicitly embedded. According to Cover's theorem, a function can be considered as a kernel provided that it satisfies Mercer's conditions [22]. To optimize the SVM classifier boundary the following relation should be maximized (Eq. 8):

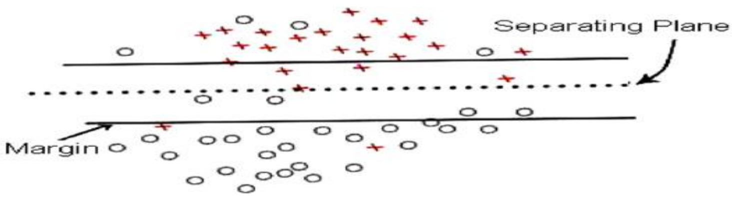

Figure 4. Support Vector Machine Classification with a Linear Hyperplane that Maximizes the Separating Margin between the Two Classes

$$L(\underline{\alpha}) = \sum_{i=1}^{v} \alpha_i - \frac{1}{2} \sum_{i,j=1}^{v} y_i y_j \alpha_i \alpha_j K(x_i, x_j), \quad 0 \le \alpha_i \le C \tag{8}$$

While

$$\sum_{i=1}^{v} y_i \alpha_i = 0, \quad w = \sum_{i=1}^{N} \alpha_i y_i x_i, \quad \alpha_i [y_i (w^T x_i + b) - 1 + \xi_i] = 0 \tag{9}$$

where K(xi, xj) is the SVM kernel, *v* shows number of total samples, and C is a user-specified positive parameter to control trade off between the SVM complexity and the number of non-separable points. This quadratic optimization problem is solved and a solution to $\alpha = (\alpha_1, \alpha_2, ..., \alpha_{nsv})$ is obtained, where each α_i is a Lagrange coefficient, and ns*v* denotes number of support vectors. The slack variables ξ_i are used to relax the constraints of the canonical hyperplane equation. Note that in a typical SVM learning problem only a small fraction of the training examples will typically qualify as support vectors. The slack variables xi are used to relax the constraints of the canonical hyperplane equation [6]. More details about SVM process can be found in [vapni]

3.5. Combination by Majority Vote

Kernel combination is a current active topic in the field of machine learning. It takes benefit of Kernel-based classifier algorithms. Advantages of merging modalities at kernel level are numerous. First, it allows to choose the kernel functions according to the modalities.

Voting strategies can be applied to a multiple classifier system assuming that each classifier gives a single class label as an output and no training data are available. There are a number of approaches to combination of such uncertain information units in order to obtain the best final decision. However, they all lead to the generalised voting definition.

More than an approach of fusion, the principle of the vote is a method of combination particularly adapted to the decisions. Let us note $S_j(x) = i$ the fact that the source S_j decides on d_i, *i.e*. for example attribute(award) the class C_i to the observation x. We suppose here that the decisions d_i is exclusive. To every source we associate the indicator function:

$$M_i^j(x) = \begin{cases} 1 & \text{si} \quad S_j(x) = i, \\ 0 & \text{sinon.} \end{cases} \tag{10}$$

The combination of sources is written by:

$$M_k^E(x) = \sum_{j=1}^{m} M_k^j(x), \tag{11}$$

For any K, the operator of combination is thus associative and commutative. The rule of the majority vote consists in choosing decision taken by maximum of sources.

4. Experimental Results and Discussion

4.1. Used Database

For train and test our SVM and as we detailed in the Section 3, our database is built from DDSM (Digital Database for Screening Mammography) mammographic images database. 300 images were used for training and 100 for testing.

4.2. Features Extraction

Our work consists in representing mammography image by three families of characteristics which are: the matrix of co-occurrence, Hu moments and central moments detailed in the Section 3. Figure 4 is an example of binary image of extracted mass using ImageJ tool.

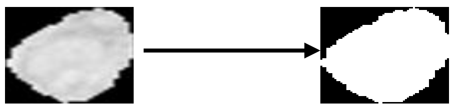

Figure 5. Example of an Image Binarised with OTSU Method

After application of the extraction module, mammography image will be indexed by 3 descriptors vectors representing the three families of characteristics (The moments of Hu, central moments and co-occurrence matrix) as shown in Figure 6 for Haralick features.

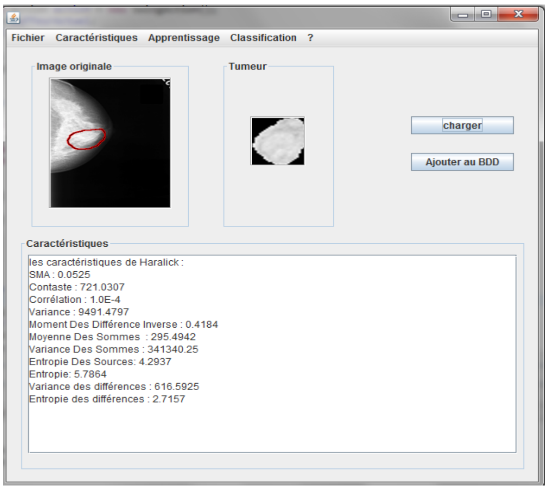

Figure 6. Haralick Features Extracted from Image

4.3. Classification

Each digital mammography image will be represented by three family of features. In our study, to choose the appropriate kernel function for each feature ensemble, we have tested four kernel function which are; Sigmoid function, polynomial function, Rbf function and tangent function.

Our objective is to assign the best SVM classifier to each features ensemble. These classifiers are trained in learning stage independently. Each SVM generates its own response due to its local features. But to select the best kernel function of each SVM classifier, we have tested four kernel functions for all features family. Table 1 shows obtained results (Table 1):

To have final decision about an unknown mammography image, combination method will be used which is majority voting technique defined above. In this stage, only best Svm classifier (associating with best kernel function) of each features family is chosen for fusion.

Table 2 summarizes the classification rate of each classifier (SVM) taken independently, MCS using three SVMs and Svm results using as entry the hole vectors of three families of characteristics.

Table 1. Obtained Results of SVM Classifiers with different Kernel Function and Independent Features Ensemble

	Used Kernel function	SVM with co-occurrence Matrix	SVM with Hu Moments	SVM with central Moments	Svm with all features
Acc rate for benign mass	Sigmoid	88,94%	**87,21%**	82,65%	87,86
	Tangent	88,64%	86,05%	82,88%	88,59
	polynomial	87,94%	83,97%	82,67%	86,30
	rbf	**90,07%**	87,04%	**83,42%**	**89,30**
Acc rate for Malign mass	Sigmoid	**91,56%**	88,23%	84,17%	**90,25**
	Tangent	87,42%	86,34%	83,78%	87,19
	polynomial	89,92%	**88,72%**	84,31%	89,68
	rbf	90,12%	87,31%	**84,74%**	89,79

Table 2. Best Accuracy of Single Classifier and MCS Results

	SVM with co-occurrence Matrix	SVM with Hu Moments	SVM with central Moments	MCS with voting	SVM With al features and best kernel function
Acc rate for benign mass	90,07% Rbf	87,21% sigmoid	83,42% Rbf	**91,27%**	89,30% Rbf
Acc rate for benign mass	91,56% Sigmoid	88,72% Poly	84,74% Rbf	**92,18%**	90,25% Sigmoid

We made our tests on a sample of 100 images. Fir this, the obtained performance of our MCS was evaluated with the cross-validation technique [23] and according to the results, we noticed that every method of extraction features trained by Svm classifier gave different rates of classification compared to others, for example the matrix of co-occurrence gave a high rate of recognition compared with two other methods. By combining the results of three families of characteristics, we obtained higher rates what improves more the recognition and thus have a good decision. From these results we found that the combination of classifiers can give good results and improves the performance of the classification system better than using fusion of all features in one features vector.

We can also conclude that choosing appropriate kernel function for each features ensemble improve significantly the classification rate of single classifier and also Multiple classifier system with majority voting fusion method.

5. Conclusion

Automatic detection of the breast cancer has been studied during more than two decades. This study was led to establish the right way to plan the future evolution of image processing in medicine and the health. In a MCS, performance mainly depends on the accuracy of the individual classifiers and on the specific way of combining the individual decisions. Correspondingly, it results crucial to appropriately handle the combination of decisions in order to attain the most accurate system. In the present work we proposed and evaluated a new system for detection of breast masses that is based on classifier fusion with features cooperation. This approach represents a new help tool for the diagnosis which is based on classifiers combination seen that it is an effective concept to have a high performance without increasing the complexity of the existing classification techniques. We represented our images by three feature vectors each is calculated from a different method, then the image were classified by the selected support vector machines (SVM) by associating the best kernel function. The results of these SVM are then combined, and we noticed according to the results combination classifier, feature fusion and kernel function adaptation of each SVM classifier improved the quality of decision significantly .The conclusions of our study are the following: The proposed CAD system performs well when tested on mammograms. Also, we can see that the sensitivity of our CAD system is better for malignant masses than for benign masses for all kernel function.

References

[1] M. D. E. Júlia de Oliveira and A. A. de Araújo, "Breast Lesions Classification applied to a reference database", 2nd International Conference: E-Medical Systems, **(2008)** October, pp. 29-31.

[2] J. L. Jesneck, L. W. Nolte, J. A. Baker, C. E. Floyd and J. Y. Lo, "An optimized approach to decision fusion of heterogeneous data for Breast Cancer Diagnosis", Medical Physics, vol. 33, no. 8, **(2006)**, pp. 2945-2954.

[3] M. X. Rodríguez-Álvarez, P. G. Tahoces, C. Cadarso-Suárez and M. J. Lado, "Comparative study of ROC regression techniques_Applications for the computer-aided diagnostic system in breast cancer detection", Computational Statistics and Data Analysis, vol. 55, **(2011)**, pp. 888-902**.**

[4] L. Xu, A. Krzyzak and C. Y. Suen, "Methods of combining multiple classifiers and their application to handwriting recognition", IEEE transactions on systems, man and cybernetics, vol. 22, no. 3, **(1992)**, pp. 418-435.

[5] F. Roli, G. Giacinto and G. Vernazza, "Methods for designing multiple classifier systems", MCS'2001, 2nd International Workshop on multiple classifier systems, Cambridge, UK, **(2001),** pp. 78-87.

[6] F. Moayedi, Z. Azimifar, R. Boostani and S. Katebi, "Contourlet-based mammography mass classification using the SVM family", Computers in Biology and Medicine, vol. 40, pp. 373-383, **(2010)**.

[7] C. Varela, P. G. Tahoces, A. J. Méndez, M. Souto and J. J. Vidal, "Computerized detection of breast masses in digitized mammograms", Computers in Biology and Medicine, vol. 37, **(2007)**, pp. 214-226.

[8] D. Ruta and B. Gabrys, "An Overview of Classifier Fusion Methods", Computing and Information systems, vol. 7, **(2000)**, pp. 1-10.

[9] N. Azizi, N. Farah, M. Sellami and A. Ennaji, ''Using diversity in Classifiers set selection for Arabic handwritten recognition", Lecture notes in Computer science, Springer, LNCS 5997, **(2010)**, pp. 235-244, Egypt.

[10] N. Azizi, N. Farah and A. Ennaji, "New Dynamic Classifiers Selection Approach for Handwritten Recognition", Int. Conference on Artificial Neural Networks ICANN, Lecture Notes in Computer Science 7553, **(2012)**, pp. 189-196.

[11] T. Woloszynski and M. Kurzynski, A probabilistic model of classifier competence for dynamic ensemble selection", Pattern Recognition, vol. 44, **(2011)**, pp. 2656-2668.

[12] L. I. Kuncheva, C. S. C. Whitaker and R. P. W. Duin, "Is independence good for combining classifiers", Proceedings of the 15th International Conference on Pattern Recognition, Barcelona, Spain, **(2000)**, pp. 168-171.

[13] http://marathon.csee.usf.edu/Mammography/Database.html, DDSM: Digital Database for Screening Mammography,University of South Florida Digital Mammography Home Page.

[14] A. Jalalian, B. T. M. Syamsiah, H. R. Mahmud, M. B. Saripan, A. B. Ramli and B. Karasfi, "Computer-aided detection/diagnosis of breast cancer in mammography and ultrasound: a review", Clinical Imaging, press, **(2012)**.

[15] R. Haralick, K. Shanmugan and I. Distein, "Textural Features for image classification", IEEE Trans. system Man and Cybern, vol. 3, **(1973)**, pp. 610-621.

[16] Z. Huang, J. Leng, "Analysis of Hu's Moment Invariants on Image Scaling and Rotation", ECU Publications Pre., vol. 7, **(1973)**, pp. 476-480.

[17] Y. Li, A. Poulos, D. McLean and M. Rickard, "A review of methods of clinical image quality evaluation in mammography", European Journal of Radiology, vol. 74, **(2010)**, pp. 122-131.

[18] J. Zhang and J. Hu, "Image Segmentation Based on 2D Otsu Method with Histogram Analysis", IEEE International Conference on Computer Science and Software Engineering, **(2008)**, pp. 105-109.

[19] C. J. C. Burges, "A Tutorial on Support Vector Machine**s** for Pattern Recognition", Kluwer Academic Publishers, Boston, Manufactured in The Netherlands, **(1998)**, pp. 1-43.

[20] Z. Al-Ameen, G. Sulong and M. G. Johar, "Fast Deblurring Method for Computed Tomography Medical Images Using a Novel Kernels Set", International Journal of Bio-Science and Bio-Technology, vol. 4.3, **(2012)**, pp. 9-20.

[21] F. Moayedi, R. Boostani, Z. Azimifar and S. Katebi, "A support vector based fuzzy neural network approach for mass classification in mammography", International Conference on Digital Signal Processing, **(2007)**, Britain.

[22] rsb.info.nih.*gov/ij/*, ImageJ for image processing in Java.

[23] M. Stone Cross-validatory choice and assessment of statistical predictions. J Roy Stat Soc B, vol. 36, no. 1, **(1974)**, pp. 111e47.

[24] V. Vapnik and S. Kotz, "Estimation of Dependences Based on Empirical Data", Springer Series in Statistics, 1982, ISBN: 978-0387907338, **(1982).**

Author

Nabiha AZIZI was born in Algeria in 1975. After succeeding in the baccalaureat, majoring in Maths and Technology in 1993, she graduated from the University of Badji Mokhtar Annaba, Algeria, with a state Engineering degree in Computer science majoring in materials and software, in 1998. She has her master degree in pattern recognition in 2003 and her Ph.D. degree in Computer Science from the University of Annaba, Algeria in 2010. Since 2003 until now, she is senior lecturer at department of computer science in Annaba University (Algeria).

Dr. Nabiha AZIZI is an IEEE Member and her research interests include medical Imaging, computer aided diagnostic, machine learning, and ensemble of classifier construction, fusion and selection.