

Breast density classification to reduce false positives in CADe systems

Noelia Vállez^a, Gloria Bueno^a, Oscar Déniz^a, Julián Dorado^b, José Antonio Seoane^b, Alejandro Pazos^b, Carlos Pastor^c

^a *VISILAB, Engineering School, Universidad de Castilla-La Mancha, Spain*

^b *RNASA-IMEDIR Group, Computer School, Universidade da Coruña, Spain*

^c *Department of Radiology, Hospital General Universitario de Ciudad Real, Spain*

Abstract

This paper describes a novel weighted voting tree classification scheme for breast density classification. Breast parenchymal density is an important risk factor in breast cancer. Moreover, it is known that mammogram interpretation is more difficult when dense tissue is involved. Therefore, automated breast density classification may aid in breast lesion detection and analysis. Several classification methods have been compared and a novel hierarchical classification procedure of combined classifiers with linear discriminant analysis (LDA) is proposed as the best solution to classify the mammograms into the four BIRADS tissue classes. The classification scheme is based on 298 texture features. Statistical analysis to test the normality and homoscedasticity of the data was carried out for feature selection. Thus, only features that are influenced by the tissue type were considered. The novel classification techniques have been incorporated into a CADe system to drive the detection algorithms and tested with 1459 images. The results obtained on the 322 screen-film mammograms (SFM) of the mini-MIAS dataset show that 99.75% of samples were correctly classified. On the 1137 full-field digital mammograms (FFDM) dataset results show 91.58% agreement. The results of the lesion detection algorithms were obtained from modules integrated within the CADe system developed by the authors and show that using breast tissue classification prior to lesion detection leads to an improvement of the detection results. The tools enhance the detectability of lesions and they are able to distinguish their local attenuation without local tissue density constraints.

Keywords

Breast tissue classification; Weighted voting tree classifier; Texture analysis; CADe system; False positive reduction

1. Introduction

Breast cancer continues to be an important health problem. Early detection can potentially improve breast cancer prognosis and significantly reduce female mortality. Computer-aided detection/diagnosis systems (CAD) can be of tremendous help to radiologists in the detection and classification of breast lesions, [1], [2], [3] and [4]. Computer-aided detection systems are abbreviated as CADe systems and computer-aided diagnosis system as CADx systems. The development of reliable CAD systems is an important and challenging task because the automated interpretation of mammogram lesions remains very difficult. Moreover, the presence of dense breast tissue is one of the potential problems. Dense tissue may cause suspicious areas to be almost invisible

and may be easily misinterpreted as calcifications or masses [5], [6], [7] and [8]. Since the discovery by Wolfe [9] of the relation between mammographic parenchymal patterns and the risk of developing breast cancer in 1976, there has been a heightened interest in investigating breast tissue density [10], [11] and [12]. There are several research articles that describe epidemiological studies including the estimation of breast cancer risk [13], [14], [15], [16], [17], [18], [19] and [20], and diagnosis support by means of content-based image retrieval [21] and [23] based on breast tissue density information.

Our research has been prompted by this need to classify breast tissue and drive the development of CAD algorithms for the automated analysis of breast lesions. Recent studies have shown that the performance of the CAD system is improved if breast density information is considered [23], [24], [25] and [26]. These studies showed an overall sensitivity of CAD system of 88.5% with an accuracy of 78% [23] and [26]. The CAD's sensitivity was usually low in density types 3 and 4. In masses with density 3 the specificity was 79% and in those with density 4 were 45%. The specificity was up to 80% both for microcalcifications present in the four types of densities and for masses with densities 1 and 2. Thus, it is necessary to adjust the input parameters of a lesion detection algorithm to control its sensitivity depending on the tissue type in order to reduce false positive detections, especially in dense tissue, at almost the same detection rate.

The rest of this paper is organized as follows. Section 2 shows an overview of the problem in the literature. Section 3 describes the methods and materials used in this work. These include the feature extraction procedure, the statistical analysis, the classifiers tested, the training and testing procedures and the experimental database used. Section 4 explains the integration of the classifiers into the CADe and the system implementation. Mode of availability and system requirements are explained in Section 5. Section 6 describes the results obtained with the proposed methods. Finally, in Section 7, the main conclusions are drawn.

2. Background

Several studies dealing with the breast tissue classification problem have been described in the literature [27], [28], [29], [30], [31], [32], [33], [34], [35], [36], [37], [38], [39], [40] and [41]. The American College of Radiology BIRADS [42] is the standard guide on the assessment of mammographic images, where the images are classified into four categories: (T.I) fatty, (T.II) fatty-glandular or fibroglandular, (T.III) heterogeneously dense and (T.IV) extremely dense. The first type of breast density, T.I, is almost entirely fat, that is, glandular tissue is less than 25%. In T.II breast density, there are scattered fibroglandular tissues ranging from [25–50%) of the breast. Dense tissue in T.III ranges [50–75%) of the breast tissue. The last category of breast density, T.IV, means that the breast contains greater than 75% glandular and fibrous tissue. The features used for the classification process are based on: (a) grey-level histograms or morphological analysis based on 1st-order statistic, (b) texture information based on 2nd-order statistic as well as textons and (c) space-frequency properties. The most effective features used for characterization are extracted from the gray level histogram and texture patterns [28] and [21]. Some studies have indicated that histogram information alone might not be sufficient to classify mammograms according to BIRADS categories [5] and [36]. Features are extracted from different areas: (a) the whole breast area, (b) the breast area without the pectoral muscle or (c) segmented areas according to their tissue appearance. The main drawback of the former techniques (b) and (c) is their reliance on an initial segmentation of the breast. Wrong segmentation may cause errors on the classification [31]. Moreover, the pectoral muscle may contain suspicious areas to analyze.

A variety of classification methods have been used based on neural networks (NN), support vector machines (SVM), linear Bayes normal classifier (LBN) classification trees (Trees) and feature histogram comparison against χ^2 and Student's *t*-test distributions. Most of these methods have been demonstrated on screen-film mammogram (SFM) databases, including the public mini-MIAS (Mammographic Image Analysis Society) dataset, and only two methods have shown results on a full-

field digital mammogram (FFDM) dataset [41] and [43]. Liu et al. [41] report a correct classification percentage (CCP) of 86.4% on a database composed of 88 FFDM, which is not statistically representative. Tortajada et al. [43] report a CCP of 92% on a database composed of 236 FFDM but only 72% on a dataset composed of 831 SFM. The sensitivity of this method depends on the mammogram types and it has the additional drawback of the segmentation process prior to classification.

Furthermore, none of the research studies statistically assessed the influence of the tissue types for each calculated feature. The classification methods are summarized in Table 1 in comparison with the proposed method. The table includes the number and type of features, the database used, the cross-validation method used to train and test the classifiers, i.e., 10-fold (10FCV) or leave-one-out (LOOCV), the number of BIRADS density categories classified, if there is pectoral muscle segmentation previous to the classification and the global results obtained. The classification results are given as a correct classification percentage (CCP), that is, the overall sensitivity of the classification method. Notice that the overall accuracy for the 2-class classification problem is equivalent to the overall sensitivity but that is not the case for the 3- and 4-class classification problem (accuracy being defined in terms of sensitivity and specificity) [44].

Table 1. Methods for breast parenchymal density classification from the literature in comparison with the proposed method. Columns show: classification methods (NN: neural networks, SVM: support vector machines, LBN: linear Bayes normal, Trees, χ^2 and Student's *T*-test), the number and type of features, the dataset, i.e., number of mammograms and type (SFM: screen-film mammograms, FFDM: full-field digital mammograms and public mini-MIAS: Mammographic Image Analysis Society), the cross-validation method used (10FCV: 10-fold, LOOCV: leave-one-out), the number of BIRADS density categories classified, if there is pectoral muscle segmentation prior to classification and the correct classification percentage (CCP) obtained.

Reference	Classifiers	Features	Dataset	Validation method	Tissue types	Segmentation	CCP
Bovis and Singh [27]	NN	180 texture and 136 frequency	377 SFM	10FCV	4	No	71.40%
Wang et al. [28]	NN	4 grey-level histogram	195 SFM	10FCV	4	No	71.00%
Petroudi et al. [29] and [30]	χ^2	40 textons	132 SFM	χ^2	4	Yes	75.76%
Bosch et al. [31]	SVM	Semantic textons	322 mini-MIAS	LOOCV	4	Yes	93.40%
Wang et al. [39]	SVM	Iterative textons	322 mini-MIAS	LOOCV	4	Yes	89.00%
Castella et al. [34]	LBN	36 texture	352 SFM	LOOCV	4	No	76.00%
Boehm et al. [35]	<i>t</i> -Test	3 texture	100 SFM	LOOCV	3	No	89.00%
Oliver et al. [36]	LBN	216 texture & 16 morphological	322 mini-MIAS	LOOCV	4	Yes	82.75%
Oliver et al. [32]	LBN	216 texture & 112 LBP	322 mini-MIAS	LOOCV	4	No	79.25%
Oliver et al. [24]	Tree	216 texture & 16 morphological	200 FFDM	LOOCV	2	Yes	90.00%
Cheng et al. [45]	SVM	textons in bag-of-words	23 SFM	LOOCV	4	No	80.70%
Subashini et al. [38]	SVM	9 texture	43 mini-MIAS	3FCV	3	Yes	95.44%
Oliveira et al. [21] and [23]	SVM	CBIR	5024 SFM	10FCV	4	No	80.00%
Tzikopoulos et al. [20]	SVM	fractal	322 mini-MIAS	LOOCV	3	Yes	85.70%
Liu et al. [41]	SVM	Moments	88 FFDM	10FCV	4	Yes	86.40%
Tortajada et al. [43]	LBN	Texture and morphological	236 FFDM	LOOCV	4	Yes	92.00%

Table 1. Methods for breast parenchymal density classification from the literature in comparison with the proposed method. Columns show: classification methods (NN: neural networks, SVM: support vector machines, LBN: linear Bayes normal, Trees, χ^2 and Student's *T*-test), the number and type of features, the dataset, i.e., number of mammograms and type (SFM: screen-film mammograms, FFDM: full-field digital mammograms and public mini-MIAS: Mammographic Image Analysis Society), the cross-validation method used (10FCV: 10-fold, LOOCV: leave-one-out), the number of BIRADS density categories classified, if there is pectoral muscle segmentation prior to classification and the correct classification percentage (CCP) obtained.

Reference	Classifiers	Features	Dataset	Validation method	Tissue types	Segmentation	CCP
Tortajada et al. [43]	LBN	Texture and morphological	831 SFM	LOOCV	4	Yes	72.00%
Proposed method	Voting tree	277 texture	322 mini-MIAS	LOOCV	3	No	99.68%
Proposed method	Voting tree	277 texture	322 SFM (mini-MIAS)	LOOCV	4	No	99.75%
Proposed method	Voting tree	218 texture	1137 FFDM	LOOCV	2	No	96.76%
Proposed method	Voting tree	218 texture	1137 FFDM	LOOCV	4	No	91.52%

There is still a need to carry out the statistical analysis of the classification features, improve the classification, test it with larger databases that also include FFDM and show how classification influences lesion detection.

In this paper we extend our previous work from [46] and [47]. Additional classification methods have been compared here using with principal component analysis (PCA), Linear Discriminant Analysis (LDA), feature ranking (FR) and feature selection (FS). A novel weighted voting tree classification scheme is proposed as the best solution. Moreover, a statistical analysis of the data, including normality tests, homoscedasticity tests and the analysis of variance, is carried out to assess the influence of the tissue types on each of the 298 calculated features. Thus, only features that are significantly influenced by tissue type are considered. Experimental results with the four BIRADS classes have been obtained on 1459 mammograms. These mammograms are from two datasets, one composed by 322 screen-film mammograms (mini-MIAS) and other composed by 1137 full-field digital mammograms (FFDM), both with a range of densities and abnormalities. The results obtained with the proposed method improve over existing techniques using the same type of dataset (mini-MIAS or FFDM in this case), the same number of tissue classes (2–4), the same validation methods (10-fold cross-validation, 10FCV, and leave-one-out cross-validation, LOOCV) and without previous segmentation (see Table 1).

3. Methods and materials

Our proposal is to apply texture analysis to the whole breast. Thus all mammograms are pre-processed to identify the breast region and remove background and possible labels. This process consists on finding a point of the breast region boundary and then, obtaining the complete one using the 8-directional chain code [48] and the Otsu's threshold as the chain code threshold. The output of this process is illustrated in Fig. 1 for the different tissue types considered. The pectoral muscle is not removed since some of the malignant and abnormal lesions may appear in this area and it has also some textural information. Fig. 2 illustrates this area with some mammograms from the used FFDM dataset. The boundary obtained in the pre-processing step may not be so well defined but the definition of the breast region boundary does not affect the next steps because all texture features are calculated inside of the tissue breast region previously identified.

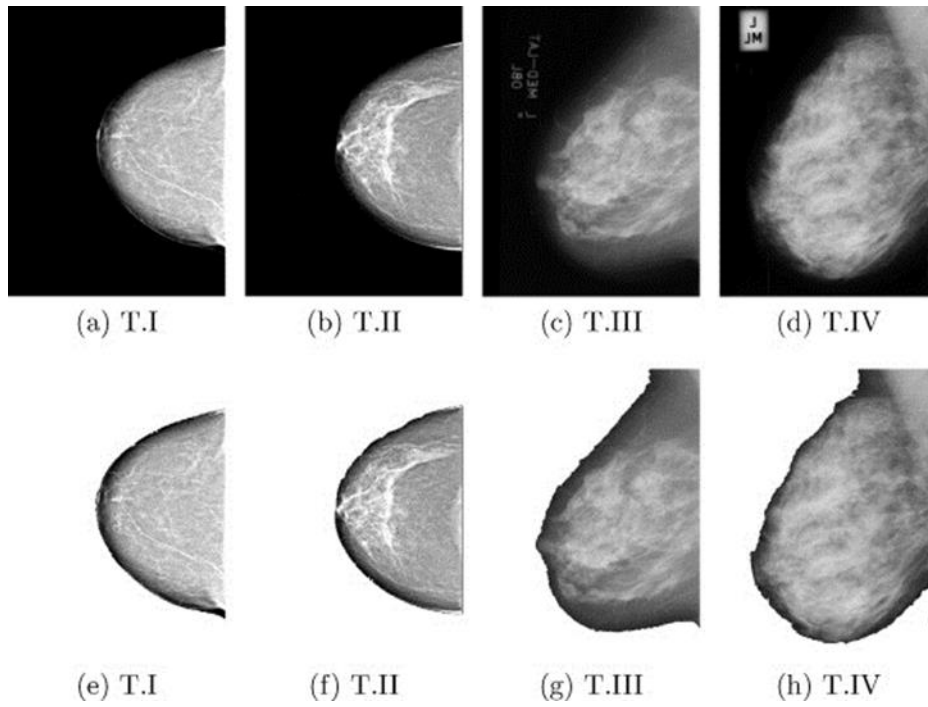


Fig. 1. Mammography examples of the four BIRADS breast density classification types illustrated with the FFDM (a and b) and SFM datasets (c and d) and their preprocessed results (e–h): (T.I) fatty, (T.II) fatty-glandular or fibroglandular, (T.III) heterogeneously dense and (T.IV) extremely dense.

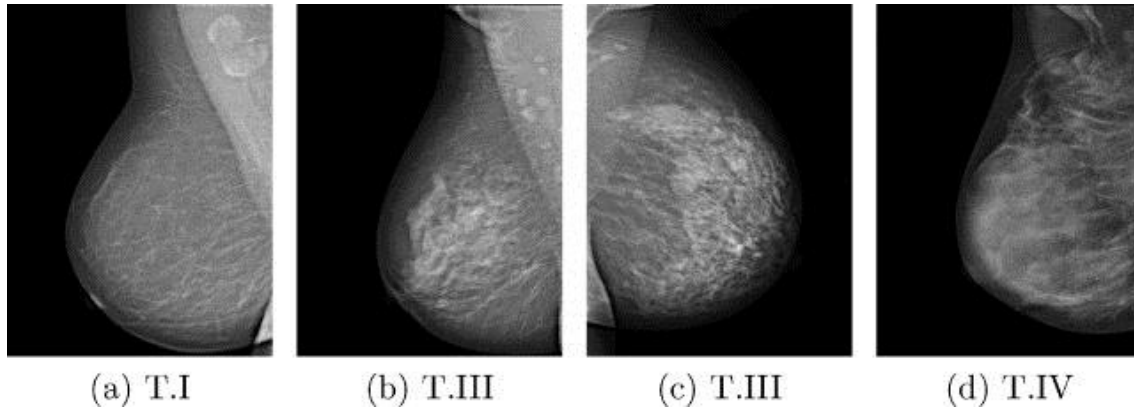


Fig. 2. Examples of mammograms from the FFDM dataset with abnormal lesions in the pectoral muscle.

3.1. Experimental database

Two datasets were considered. One composed of 322 SFM obtained from the mini-MIAS public database and another one composed of 1137 FFDM provided by local Hospitals. We focus our attention on the use of a FFDM dataset. Both datasets were labelled according to the BIRADS categories by four expert radiologists from the Hospital General Universitario de Ciudad Real, using the majority vote opinion. The number of images of each type is shown in Table 2 and Table 3. The mini-MIAS dataset was used also to compare our results with those of other authors and we test the proposed method with the mini-MIAS original classification. The image sizes are 1024×1024 and 3328×4084 respectively for the SFM and FFDM datasets.

Table 2. Number of images of each type for SFM and FFDM databases.

Dense tissue	[0–25%)	[25–50%)	[50–75%)	[75–100%]
Type	T.I	T.II	T.III	T.IV
SFM	84	102	92	44
FFDM	288	289	289	271

Table 3. Number of images of each type for mini-MIAS database.

Dense tissue	[0–33%)	[33–66%)	[66–100%]
Type	T.I	T.II	T.III
mini-MIAS	106	104	112

The SFM database contains images from right and left mediolateral oblique projections (RMLO, LMLO) whereas the FFDM database contains also images from right and left craniocaudal projections (RCC, LCC).

3.2. Feature extraction and statistical analysis

Most studies on texture classification are based on statistical and morphological features obtained from the image (see Section 2). Here we analyze 298 features. All features have been calculated using only the pixel values inside the extracted breast area of the preprocessed images. 241 features have been drawn from the histograms and the co-occurrence matrices by means of the Haralick's coefficients, [49]. The co-occurrence matrices have been calculated for a distance d equal to 1, 3 and 5 and an angle a equal to 0° , 45° , 90° and 135° . These features are listed in detail in [46]. The other 57 features are: mean, variance, kurtosis and asymmetry of the local binary patterns (LBP); Chebyshev moments and mean and variance of 24 Gabor filters (6 orientations, 4 frequencies) [50], [51] and [52]. These 57 features are complementary to the previous ones. They cope with non-stationary grey texture images which may be the case for mammograms with breast density T.II and T.III. Whereas features defined from the co-occurrence matrix may cope with stationary grey texture images, which is mainly the case of breast density T.I and T.IV. A stationary signal (image) is a signal where there is no change in its properties versus a non-stationary signal where there is change in the properties. By definition T.I and T.IV have homogeneous density, non-dense and dense respectively, while T.II and T.III have heterogeneous density.

Once all features have been calculated, a statistical analysis is carried out to include only features influenced by tissue type. In the literature, several sets of different features have been applied to automatic breast tissue classification without determining if these features really contain discriminant information (Table 1). Performing a statistical analysis on the above data is necessary to know if the texture features have significant differences between the four different tissue types. In other words, if they constitute a set of discriminant features of the problem or not.

Feature analysis entails the analysis of variance. In this case, the classification criterion (tissue type) is an independent variable called factor. Then, each feature variable is divided into four groups by the independent variable. Thus, the feature set contains the dependent variables of the problem. This analysis has been carried out to check whether or not the means of several groups are all equal for each feature and then, if we can conclude that the factor has a significant influence on the results.

The most common procedure to test if the means of two populations are equal is the Student's t -test. However, it is known that using the two sample t -test to evaluate more than two groups of data and performing all the possible pairwise comparisons, the likelihood of making a Type I error in at least one of our comparisons increases [53]. This makes the analysis of variance (ANOVA) more adequate in this case [54]. However, there are two main restrictions to apply ANOVA, the normality and the homoscedasticity of the data. The lack of the normality in some of the features is not a problem for large sets of data due to the Central Limit Theorem (more than 30 samples). In cases where features present significant differences between their variances it is possible to apply ANOVA if the sizes of the groups are nearly equal. Unfortunately, this is not our case and we can only apply the ANOVA analysis over the features that have homogeneity of variances. With the features that do not fulfill ANOVA conditions, it is possible to apply the non-parametric alternative, the Kruskal-Wallis test (KWt) [55]. The KWt tries to perform an analysis similar to ANOVA though with relaxed conditions. Then, we use this technique with the features that have significant differences between the group variances.

In order to test the normality of the data the Kolmogorov–Smirnov test (KS) was used [56]. It considers a null hypothesis H_0 that data comes from populations with normal distributions and an alternative hypothesis H_1 that data comes from non-normal populations. Setting the significance level to $\alpha = 0.05$, from the results of the KS test applied to the SFM database 102 of the 298 features obtained a p -value $>\alpha$ for each tissue type so we can assume that these features have a normal distribution with a confidence of 95%. For the FFDM database only 32 features obtained a p -value $>\alpha$. The next step of the analysis is to test homoscedasticity. To check it we run the Levene's test [57]. One advantage of this test is that it does not require normality of the underlying data. Levene's test has as null hypothesis, H_0 , that the variances are equal between all types. 134 and 74 features have equal variances between all types for SFM and FFDM datasets respectively.

Finally, ANOVA has been applied over the features that follow the normality and homoscedasticity conditions and the KWt has been applied with the others. The results of applying ANOVA show that in case of using the SFM dataset only in 11 of the features the tissue type is not influential. With the FFDM dataset there are 52 features with no tissue type influence. On the other hand, the results of the KWt with the non-homoscedastic features show that all these features are influenced by the tissue type in the SFM dataset, but in case of using the FFDM dataset we obtain another 19 features without tissue influence. Summarizing, we conclude that all the calculated features except for 11 features for the SFM dataset and 71 features for the FFDM have significant differences influenced by the tissue type. Table 4 shows these features which correspond to those that do not fulfill ANOVA and the non-homoscedastic. Table 4 also indicates the feature category, that is, 54 statistical and 22 space-frequency descriptors are not influenced by the tissue type. 1st-order statistical descriptors measure typical statistics in image histogram. They are sensitive to global variation of gray pixel levels, although they ignore their local correlation. The *mode* and the *minimum* are those descriptors non-influenced by the tissue type. 2nd-order statistical descriptors measure statistics in co-occurrence matrix defined as the distribution of co-occurring neighbor gray values. The 2nd-order statistical descriptors non-influenced by the tissue type are: *difference variance*, *energy*, *difference entropy*, *measure of correlation*, *maximum probability* and *homogeneity* at different distances, d , and angles, a . Space-frequency descriptors do not really constitute descriptors by themselves but transformations where features, somehow hidden, arise with higher visibility. Here 24 Gabor filters (Gbf _{i}) formed calculating the energy at every scaled level have been used. The mean and variance of 3 and 11 Gbf for SFM and FFDM dataset respectively were disregarded as non-influenced by the tissue type. They correspond to the first and the last scales, which contain mostly noise.

Table 4. Features non-influenced by the tissue type. These features do not fulfill ANOVA and the non-homoscedastic.

Dataset	Variable	Co-occurrence matrices	Category
SFM	Minimum		1st Statistical
FFDM	Mode		1st Statistical
SFM	Measure correlation	$d = 5 \forall a$	2nd Statistical
FFDM	Energy	$\forall d \forall a$	2nd Statistical
FFDM	Difference variance	$d = 3a = 0^\circ, 90^\circ$	2nd Statistical
FFDM	Difference entropy	$d = 3a = 0^\circ, 45^\circ, 135^\circ$	2nd statistical
FFDM	Homogeneity 1, 2	$d = 3, 5 \forall a$	2nd statistical
FFDM	Homogeneity 2	$d = 1a = 0^\circ, 45^\circ, 135^\circ$	2nd statistical
FFDM	Max. Probability	$\forall d \forall a$	2nd statistical
SFM, FFDM	Gbf _{i} , $i = 1, 2, 24$		Space-frequency
FFDM	Gbf _{i} , $i = 3 \dots 6, 20 \dots 23$		Space-frequency

Considering the results above, most of the calculated features are influenced by the tissue type and it therefore makes sense to solve this classification problem by using these features.

3.3. Dimension reduction

Although some calculated features have been discarded as a result of the statistical feature analysis (Section 3.2) the total number of features still remains high (287 and 227 for the SFM and the FFDM datasets respectively). Large numbers of features could

reduce classifier accuracy (Hughes' phenomenon) and increase computational time [58]. Then, to reduce and select the feature space, LDA, PCA and FR and FS of individual performance were applied for each classification method. PCA transforms a number of variables that can be correlated into a smaller set of uncorrelated variables called principal components and tries to find a subspace whose basis vectors correspond to the maximum-variance directions in the original space [59]. On the other hand, LDA searches for those vectors in the underlying space that best discriminate among classes. Fig. 3 depicts the feature values of the two first components of the LDA reduced space for the SFM, FFDM and mini-MIAS image datasets.

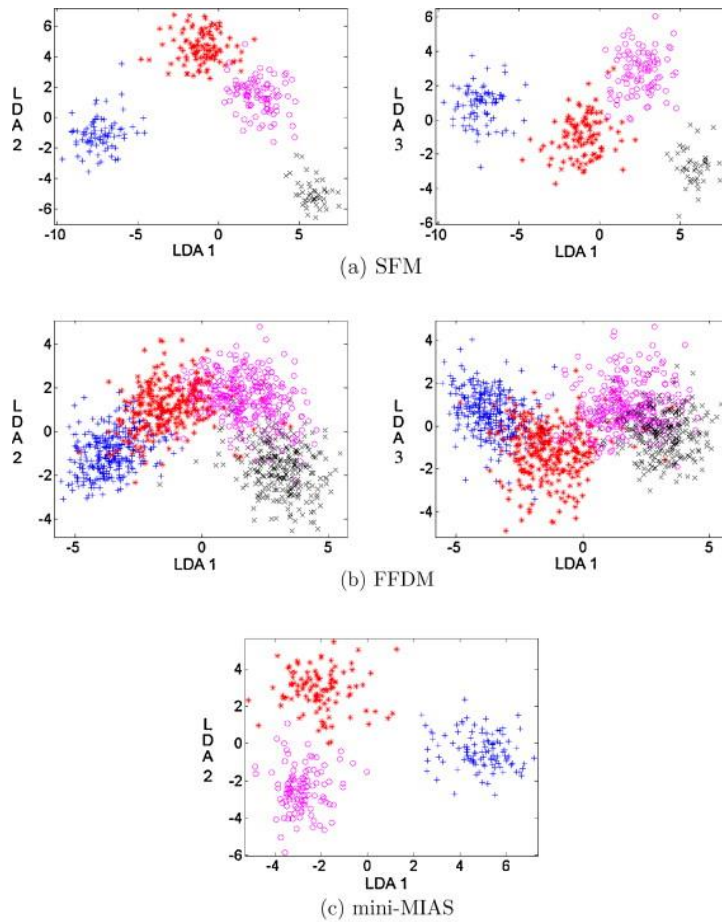


Fig. 3. Values projected on the new axes of the LDA reduced space for the SFM, FFDM and mini-MIAS image datasets. Blue-T.I, red-T.II, pink-T.III and black-T.IV. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of the article.)

FR is based on the results from intra-cluster and inter-cluster distances between the tissue types. These distances measure the variability within and between different groups. Finally, FS is a forward selection with the nearest-neighbor criterion to obtain the optimal feature reduced set.

Different tests were performed by varying the number of components from the space reduced by PCA and FR. This number of components varies between 10 and 270 or 210 (for SFM or FFDM datasets) at intervals of 10. LDA transforms the feature dataset into a new dataset with the number of groups minus 1 and FS has been used with the optimal features returned by this technique. We select all of these features from the training partitions. The average errors for all tested classifiers were measured and the

reduction/selection techniques show improvements of 8%, 11%, 20% and 52% with PCA, FR, FS and LDA respectively.

Fig. 4 shows the analysis of the feature selection and reduction methods, as well as the effect of the classification with the selected features. Fig. 4(a-c) shows the errors with all features (ALL) and with the selected ones, that is, before and after applying the selection/reduction methods for the different classification methods tested on the selected features. These are: support vector machines (SVM), linear Bayes normal classifier (LBN), naive Bayes classifier (NAIVE), k-nearest neighbor (k-NN) with k equal to 1, quadratic classifier (QD), nearest mean classifier (NMC), Fisher classifier (FISH), and parzen classifier (PARZ) [60] and [61]. Fig. 4(d) shows the errors when using different set of descriptors under the same classifier (Fisher). It is shown how the combination of different types of descriptors can strengthen the capacity of the classifiers to discriminate. Thus, from this chart, the use of the statistical features (STAT) plus Gabor (GB) complements their efficacy. Note that adding LBP and Chebyshev moments (ChM) barely affects performance.

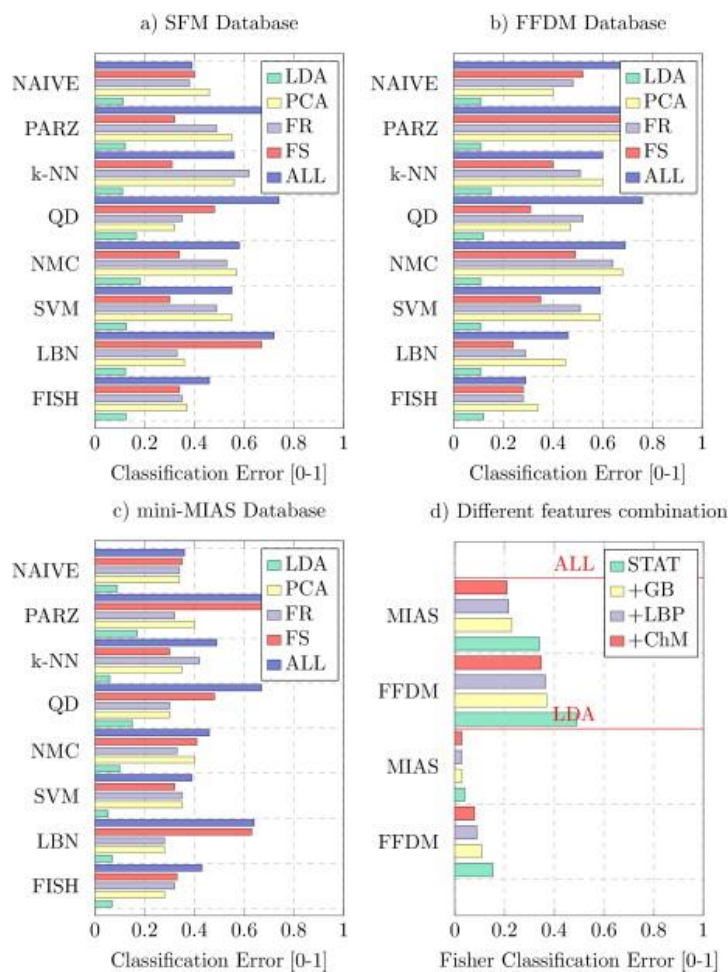


Fig. 4. Analysis of features. (a-c) Comparison of feature reduction and feature selection methods. (d) Comparing descriptor types under the same Fisher classifier. The leave-one-out cross-validation method is used to train and test the classifiers.

3.4. Classifiers, data training and testing

To train and test classifiers the 10FCV and the LOOCV methods were used. The 10FCV method consists of randomly dividing the data into 10 different groups containing approximately the same number of samples. One of these groups is selected to test the classifier and training is performed on the rest of the groups. The process is then repeated with the other groups of the dataset as the test set, and the average classification error is obtained. The LOOCV method works like 10FCV but dividing the data into N groups, being N the number of samples. A sample case is composed of all mammogram projections of a woman. Then, RMLO and LMLO projections of a woman from the SFM database are a sample case and RMLO, LMLO, RCC and LCC projections from the FFDM database are another sample case. At the classification stage, the projections of each woman are analyzed by the classifier trained using the mammograms of all other women in the database. Thus, a leave-one-woman-out methodology is used to avoid bias. The performance of the proposed classification scheme with both methods is shown and discussed in Section 6.

Examining the results [46], it was observed that the classifiers with the best global CCP do not necessarily have the best CCP for each tissue type because their training algorithms try to minimize the global error. In order to improve the individual results of the classifiers we propose to combine them in a weighted voting classification scheme. The weighted voting scheme is based on the idea that not all voters are equal. Instead, if a classifier has better performance at one class, a larger weight is assigned to it for detecting instances of this class. The process begins by training and testing the classifiers individually. Then, the best classifier, j , for each tissue type, i , is selected. After selecting the classifiers, five weights are assigned to each classifier, one of them corresponds to the global sensitivity of the classifier, w_{g_i} , and the other four correspond to the sensitivity for each tissue type, $w_{1...4i}$. Finally, the classifiers are applied individually over the test set obtaining four decisions. Each decision has a weight as a result of multiplying the global classifier sensitivity times the individual class sensitivity of the decision. Adding all the weights, the tissue type that has the largest value becomes the final decision ($\max_i \sum_{j=1}^4 w_{gj} * w_{ij}$). Fig. 5 depicts the proposed weighted voting combination.

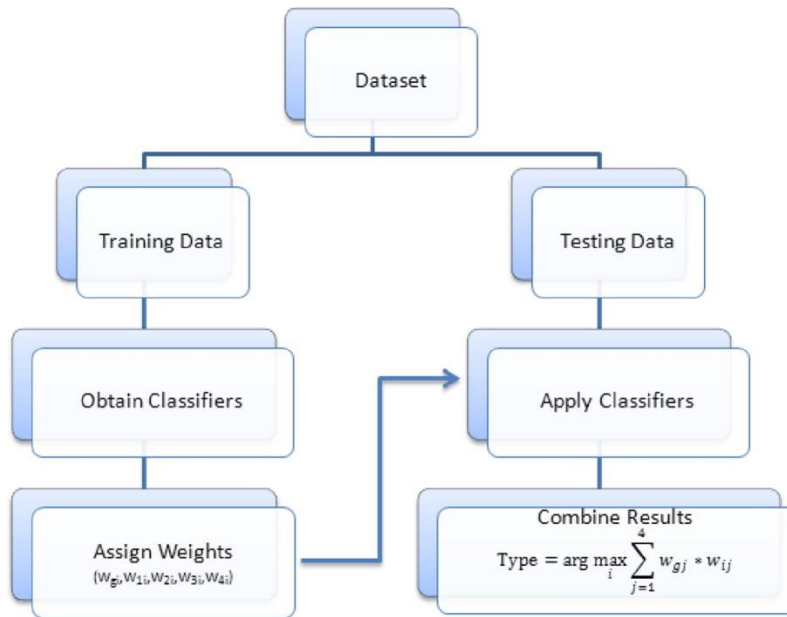


Fig. 5. Proposed weighted voting combination scheme for the 4-class classification problem.

Mammographic density can be evaluated and reported by radiologists on the basis of visual analysis of mammograms by following two different radiologic patterns of dense and non-dense tissue in the breast. Thus, the first evaluation by radiologists follows a two-class problem, differentiating between fatty and dense breast types. Afterwards, percentage density is visually analysed to discriminate between subtle glandular and dense tissue types. Therefore, a tree structure looks the natural way to mimic the classification of mammographic parenchyma patterns. Moreover, the higher classification errors are given by densities Type II and III. Then, in order to make the classification problem easier the four tissue types could be merged into two main groups, fatty and dense, according to the percentage of dense tissue. Then, it is possible to get a two level classification problem with a tree structure in which the first level separates fatty mammograms, $\{T.I \cup T.II\}$, from dense ones, $\{T.III \cup T.IV\}$ (see Fig. 6). The second level separates each of the two main tissue types in the original two grouped tissues. Then, it separates T.I from T.II and T.III from T.IV, leading to the four classification groups.

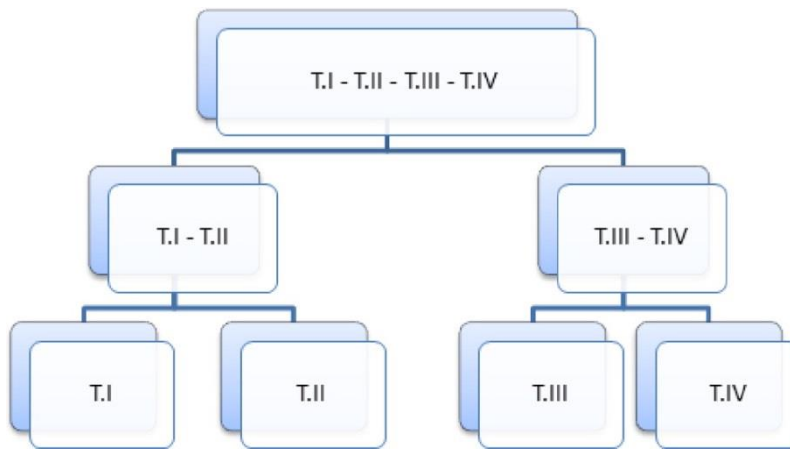


Fig. 6. Proposed tree classifier structure.

Finally, we propose to use the above tree classifier structure combined with the weighted voting classification scheme resulting in a weighted voting tree classification scheme. Therefore, only two individual classifiers are needed in each node of the tree and three weights are calculated for each individual classifier (Fig. 7).

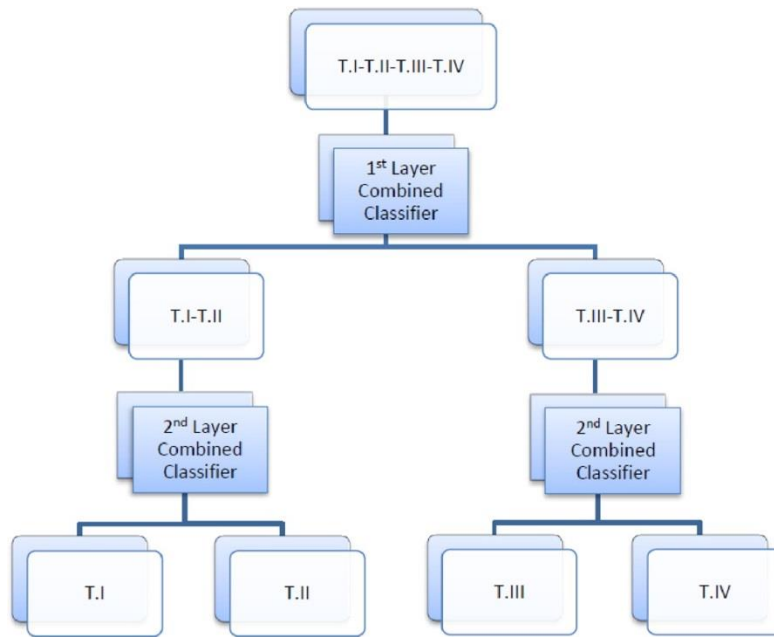


Fig. 7. Proposed weighted voting tree combination scheme for the two stage classification problem.

4. Systems and program

A CADe system for mammography with different detection methods has been developed. These are β -splines, wavelet, adaptive filtering and fuzzy k -means [4] and [62]. The methods may be potentially applied to all lesions and tissues. However, after comprehensive tests with these methods we concluded that their accuracy, in terms of sensitivity and specificity, was tissue and lesion type dependent [26] and [63]. It is necessary to adjust the input parameters to control the sensitivity of the algorithm depending on the tissue type, especially in areas of high density, in order to reduce the false positive rate at almost the same true positive detection rate. Then, prior to the detection algorithm, tissue classification is applied. Fig. 8 shows a scheme of the system. The system is divided into three stages: (i) mammographic image selection and preprocessing, (ii) mammographic density classification and (iii) lesion detection. The training and testing process described in previous sections are done for the feature selection and the voting tree classification. Once the tissue classification is done, lesion detection is carried out with the defined set of input parameters.

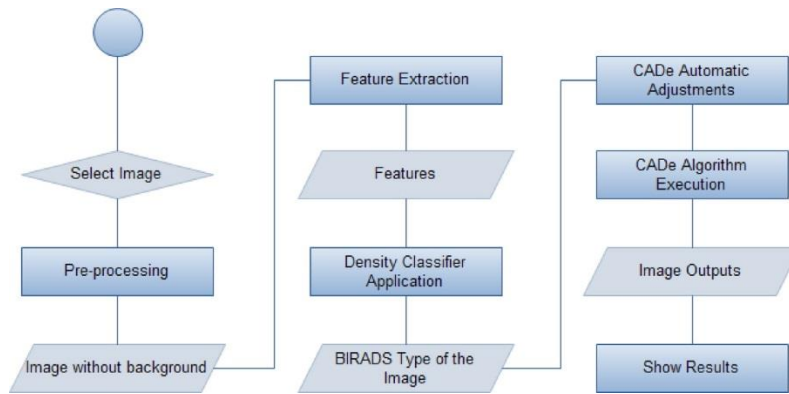


Fig. 8. Scheme of the CADe algorithm execution steps.

The input parameters of the detection methods are the number of clusters in the fuzzy k -means, the number of iterations in the wavelet method and the angular rate in the adaptive filtering. These parameters should be increased when processing T.IV and proportionally decreased with the other types. Higher order wavelets are smoother and better able to distinguish between the various frequencies. Thus, if the noise is hardly noticeable or there are various frequencies, we may need more levels in order to get the fine details of the image. Usually, the detail information of level 7 was considered for T.IV, as the frequencies covered by this level were similar than the frequency content of the dense mammogram. Also in the case of T.III and T.IV, wavelets and β -spline algorithms are applied in conjunction with the fuzzy and the adaptive filtering, respectively. Wavelets have the advantage of being able to separate the fine details in an image and deblurred. Moreover, wavelets and β -splines are able to enhance image contrast by adjusting its gray-level probability density function. This processing makes lesion detection easier in T.III and T.IV mammograms by fuzzy k -means and the adaptive filtering algorithms. In terms of lesions, adaptive filtering and β -splines are better used for microcalcifications, wavelets for distortions, and fuzzy k -means for mass lesions.

β -spline filtering uses the first derivative of a cubic spline model. It is applied on the images over both the X - and Y -axis directions obtaining new coefficients that are rescaled from 0 to 255 for visualization purposes. The results are shown in Fig. 9 for both axes. It is possible to see how the output simulates a raised relief of the image. This is due to the intensity changes produced on the image when converting from discrete to continuous coefficients with the β -spline transform.

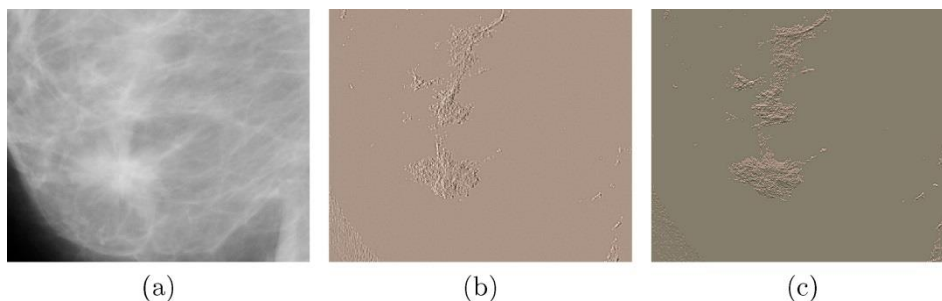


Fig. 9. β -spline filtering: (a) Selected region obtained from the original images corresponding to a mammography with T.II, (b) Highlighted contrast by β -splines applied to X -axis and (c) Highlighted contrast by β -splines applied to Y -axis.

The method has also been compared to wavelet analysis. Several wavelet transforms were tested, and the Debauchies transform was found to give the best results, with 20 coefficients (DB20) over high frequencies, after 3 iterations for T.I and T.II, and 7 iterations for T.III and T.IV. Section 6.2 illustrates these results.

4.1. Implementation

In order to make the CADe system accessible it has been developed based on a client-server architecture. The CADe can be accessed through a web browser from any computer with network and user privilege access. The server is responsible for carrying out the appropriate operations and facilitating the communication with the image database to provide the necessary information to clients. We selected Java as the programming language for the CADe, more specifically, Java Applets for the application interface and JavaScript for the CADe – web communication.

The software developed follows a three layer architecture. This architecture performs a code division according to the responsibilities that has every part of the application code. The classical division of the code divides it into a presentation layer, a domain layer, and a data layer. The presentation layer handles the code of the user interface, the domain layer contains all the business logic required to perform all the application functions and the persistence layer is responsible for managing data persistence in databases or files.

Finally, the software development process is selected according to the software structure that has to be developed. The quality of the developed software often depends on the methodology used for its construction. Currently, one of the most used methods to develop a complete system is the Unified Process [64]. The Unified Process is a methodology for developing object-oriented software using the Unified Modelling Language (UML) for representing system models. This development methodology is perfectly compatible with the three layer architecture.

5. Mode of availability

A demo version of the CADe and the FFDM database is available upon request at the VISILAB website [65]. The CADe application can be accessed by a web browser with Java Applets enabled. It runs on a typical modern PC and has no specific hardware requirements. The FFDM database provided is composed of 1137 jpg images of size 3328×4084 pixels.

6. Results and discussion

Results have been divided in two main categories, one evaluating density classification and other evaluating the CADe results with and without tissue classification. The tests have been performed on a PC with 2 Intel Xeon CPU E5-2690 (2.9 GHz) processors, 64GB of RAM and Windows 7 Professional 64-bit with SP1.

6.1. Density classification results

The different classifiers were analyzed using 10FCV and LOOCV. Experiments were carried out with all features (287 features for SFM and 227 for FFDM databases) and reducing the feature space with PCA, LDA, FR and FS. Weighted with respect to the number of mammograms of each type, classification with LDA provides better results. The best classifiers for all tested databases and both validation methods are shown in Table 5 and Table 6. Results show the 99.21% of the SFM mammograms correctly classified using 10FCV and LOOCV. For the FFDM dataset, on average 89.34% of the images are correctly classified. The CCP for the mini-MIAS dataset is 99.06% for both 10FCV and LOOCV with SVM classifier and LDA.

Table 5. Best correct classification percentage (CCP) for SFM and FFDM databases using the classifiers with the feature reduction techniques.

Dataset	CV	Classifier	Red	TI	TII	TIII	TIV	CCP
SFM	10FCV	<i>k</i> -NN	LDA	100%	99.01%	97.82%	100%	99.21%
	LOOCV	LBN	LDA	100%	99.01%	97.82%	100%	99.21%
FFDM	10FCV	LBN	LDA	91.66%	83.73%	88.23%	94.09%	89.43%
	LOOCV	LBN	LDA	91.66%	83.73%	87.88%	93.72%	89.25%

Table 6. Best correct classification percentage (CCP) for mini-MIAS database using the classifiers with the feature reduction techniques.

Dataset	CV	Classifier	Red	TI	TII	TIII	CCP
mini-MIAS	10FCV	SVM	LDA	99.10%	99.05%	99.03%	99.06%
	LOOCV	SVM	LDA	99.10%	99.05%	99.03%	99.06%

After that, we combine the best classifiers for each tissue type in order to obtain better results with the weighted voting combination scheme explained in Section 3.4. Thus, we have four individual classifiers with their corresponding feature selection techniques applied to both, SFM and FFDM datasets and three individual classifiers for the mini-MIAS database classified in three tissue types. The results of this scheme for 10FCV and LOOCV are shown in Table 7 and Table 8. The classifiers which have been used in the combination are also shown together with the final CCP for each tissue type.

Table 7. Results of the weighted voting combination scheme. Rows contain the initial classifiers that have been combined and their feature selection techniques. The percentage shown is the final correct classification percentage of the combined classifier by each tissue type. Columns contain the database and the cross-validation used.

Dataset	SFM		FFDM	
	10FCV	LOOCV	10FCV	LOOCV
Classifiers	NMC + LDA	LBN + LDA	FISH + LDA	FISH+LDA
	FISH + LDA	FISH + LDA	LBN + LDA	NMC + LDA
	LBN + LDA	LBN + LDA	PARZ + LDA	SVM + LDA
	LBN + LDA	LBN + LDA	FISH + LDA	PARZ + LDA
T.I CCP	100%	100%	93.05%	92.70%
T.II CCP	98.03%	97.05%	81.31%	81.66%
T.III CCP	91.30%	89.13%	87.88%	86.85%
T.IV CCP	100%	100%	94.83%	93.72%
Final CCP	97.33%	96.54%	89.27%	88.73%

Table 8. Results of the weighted voting classification scheme for the mini-MIAS classification. Rows contain the initial classifiers that have been combined and their feature selection techniques. The percentages shown are the final correct classification percentage of the combined classifier by each tissue type. Columns contain the cross-validation used.

Dataset	mini-MIAS	
	10FCV	LOOCV
Classifiers	LBN + LDA SVM + LDA FISH + LDA	LBN + LDA NAIVE + LDA FISH + LDA
T.I CCP	100%	100%
T.II CCP	98.11%	99.05%
T.III CCP	99.03%	99.03%
Final CCP	99.05%	99.36%

Finally, the weighted voting tree combination scheme, also explained in Section 3.4, was applied. Now, we have 6 individual classifiers with their corresponding feature selection techniques for this scheme for the SFM and FFDM datasets (2 for the first layer or separation node and 4 for the second layer or the last two separation nodes) and 4 individual classifiers for the mini-MIAS dataset. The results of this scheme for 10FCV and LOOCV are shown in Table 9 for the SFM and FFDM datasets, together with the classifiers obtained in the combination and the CCP for each tissue type. The weighted voting tree classification scheme improves upon previous results. On average, and weighted with respect to the number of mammograms of each type, the results reflect up to 99% and 97% of samples correctly classified in the 1st layer, i.e., when the number of classes is reduced to fatty and dense densities only, for SFM and FFDM datasets. In the 2nd layer, classifying into four BIRADS categories, for the SFM dataset we obtain up to 99.75% by means of both validation methods, 10FCV and LOOCV. For the FFDM dataset the results reflect up to 91.63% of mammograms correctly classified. The best results for the mini-MIAS dataset were obtained separating T.I from {T.II,T.III} in the first layer and then T.II from T.III in the second layer. In both layers the CCP obtained was 99.68% with FISH classifier and LDA.

Table 9. Results of the weighted voting tree combination scheme. Results are classified according to the separation level. On each level, rows contain the initial classifiers that have been combined and their feature selection technique. The percentage shown are the final correct classification percentages of the combined classifier by each type. Columns contain the database and the cross-validation used.

Dataset Criterion	SFM		FFDM	
	10FCV	LOOCV	10FCV	LOOCV
1st Layer	PARZ + LDA	FISH + LDA	NAIVE + LDA	NAIVE + LDA
Classifiers	FISH + LDA	FISH + LDA	SVM + LDA	SVM + LDA
{T.I, T.II} CCP	99.46%	98.92%	96.01%	95.32%
{T.III, T.IV} CCP	98.52%	99.26%	98.21%	98.21%
1st Layer CCP	98.99%	99.04%	97.11%	96.76%
2nd Layer	FISH + LDA	FISH + LDA	FISH + LDA	SVM + LDA
Classifiers	FISH + LDA	FISH + LDA	NAIVE + LDA	NMC + LDA
	FISH + LDA	FISH + LDA	PARZ + LDA	SVM + LDA
	FISH + LDA	FISH + LDA	FISH + LDA	FISH + LDA
T.I CCP	100%	100%	93.75%	93.05%
T.II CCP	99.01%	99.01%	83.04%	82.00%
T.III CCP	100%	100%	93.41%	96.19%
T.IV CCP	100%	100%	96.30%	94.83%
Final CCP	99.75%	99.75%	91.63%	91.52%

The confusion matrices for the weighted voting tree combination scheme are shown in Table 10 and Table 11 for SFM and FFDM datasets respectively and using both LOOCV and 10FCV methods. The Tables also shown the False Positive Rate (FPR) and the True Positive Rate (TPR) or sensitivity, similar to the CCP. The specificity and the accuracy of the classification method may be obtained from these measures. The specificity is equal to $(1 - \text{FPR})$ and the accuracy is equal to $(\text{sensitivity} * \text{positive cases} + \text{specificity} * \text{negative cases})$. The results show an overall specificity of 99.9% for the SFM and 97.2% for the FFDM dataset and an overall accuracy of 99.8% for the SFM and 94.3% for the FFDM dataset.

Table 10. SFM. Weighted voting tree classification scheme confusion matrices together with the true positive rate or sensitivity and the false positive rate per tissue type.

Types	Estimated				True total	TPR	FPR
	T.I	T.II	T.III	T.IV			
T.I	84	0	0	0	84	100%	0.0%
T.II	0	101	1	0	102	99%	0.4%
T.III	0	0	92	0	92	100%	0.0%
T.IV	0	0	0	44	44	100%	0.0%

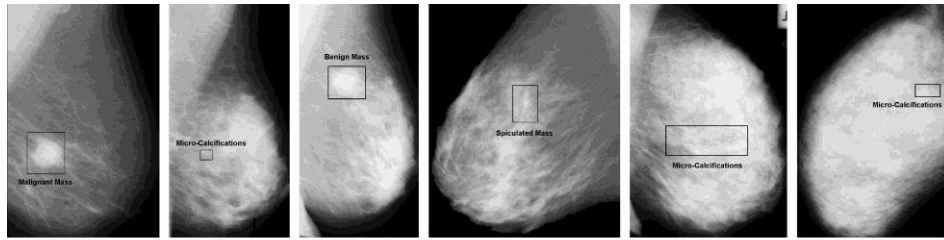
Table 11. FFDM. Weighted voting tree classification scheme confusion matrices together with the true positive rate or sensitivity and the false positive rate per tissue type.

Types	Estimated				True total	TPR	FPR
	T.I	T.II	T.III	T.IV			
<i>(a) Final results – 2nd layer using 10FCV</i>							
T.I	270	16	2	0	288	93.75%	3.2%
T.II	25	240	18	6	289	83.04%	2.9%
T.III	2	8	270	9	289	93.42%	3.4%
T.IV	0	1	9	261	271	96.30%	1.7%
<i>(b) Final results – 2nd layer using LOOCV</i>							
T.I	268	18	2	0	288	93.05%	3.3%
T.II	28	237	18	6	289	82.00%	3.2%
T.III	0	8	278	3	289	96.19%	3.8%
T.IV	0	1	13	257	271	94.83%	1.0%

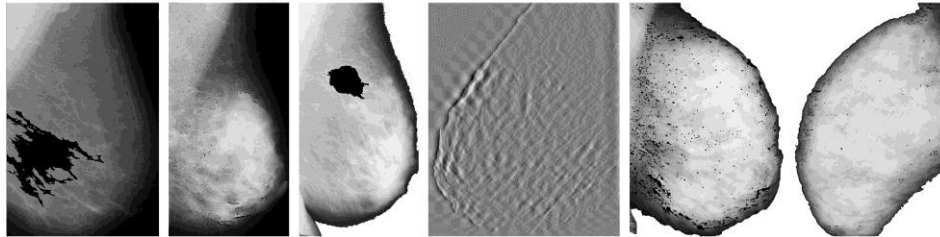
Summarizing, the proposed approach reflects up to 99% of samples correctly classified into 4 BIRADS classes by means of the weighted voting tree classification scheme for the SFM dataset. For the FFDM dataset on average 91.57% of samples were correctly classified using 10FCV and LOOCV. When the number of classes is reduced to fatty and dense densities only, the results of CCP are 99% and 96.93% for SFM and FFDM datasets respectively. Therefore, our approach improves upon previous results reported in the literature for breast tissue classification (see Table 1).

6.2. CADe results

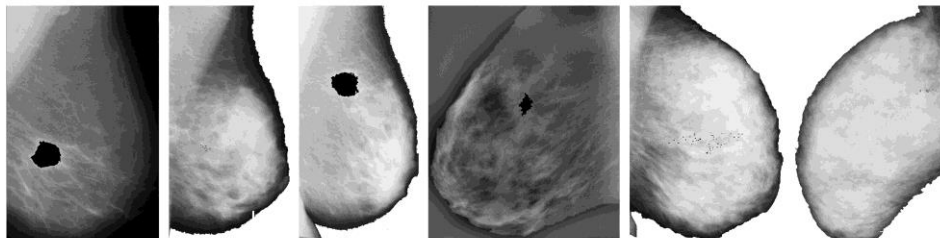
These classification methods have been integrated into a CADe system and applied prior to the detection algorithms as explained in Section 4. Fig. 10 illustrates the result of the detection algorithms after tissue type classification. The figures show the original image with the spatial location of the lesion, the lesion and tissue type, the detected lesion marked in black, and the parameters used for each algorithm, those are, the number of clusters, c , in the fuzzy k -means, the number of iterations, $iter$ in the DB20 wavelet method and the angular rate γ in the adaptive filtering. A wrong selection of γ in the adaptive filtering can lead to a high number of FPR, and a wrong number of iterations in the wavelet algorithm does not allow visualizing the lesion. This is illustrated in Fig. 10 2nd row with the results of the detection algorithms without previous classification of the breast tissue.



(a) Original images with manual lesion detection



(b) Automatic lesion detection without breast tissue classification



(c) Automatic lesion detection with breast tissue classification

Fig. 10. Lesion detection for different tissue types. 1st column TI where Fuzzy k -means algorithm has been applied, 2nd column TII where adaptive filter has been applied, 3rd column TIII where Fuzzy k -means algorithm has been applied, 4th column TIII with wavelet processing, 5th and 6th columns TIV where adaptive filtering has been applied.

The sensitivity and the specificity of the CADe were analyzed with FFDM 724 mammograms with and without breast tissue classification. All density classes were balanced, that is, they were equally represented by 181 mammograms from each type. The mammograms contain two lesion types, that is, masses and microcalcifications. All lesions of the FFDM dataset were manually located by the four expert radiologists, similarly as in Fig. 10(a). The lesions and their location in the SFM dataset were obtained from the information provided by the mini-MIAS database. Then, the automatic detection provided by the CADe system, similarly as in Fig. 10(c), was compared with the manual one. The results of this analysis is shown in Table 12 for masses and microcalcifications present in the four types of densities. On average TPR increases 13% and the FPR decreases 14% with tissue type classification. The main differences are in the adaptive filtering and wavelet processing applied to microcalcifications and masses for T.III and T.IV. This conclusion and performance of the CAD system is similar to that reported in the literature [23], [24] and [26]. The overall sensitivity obtained previous tissue type classification is 77.5% and after 90.5%. The overall specificity obtained previous tissue type classification is 77.87% and after 91.6%. That is, an overall accuracy of 91% is obtained with our CADe if the breast densities are taken into account.

Table 12. Sensitivity and specificity for the CADe algorithms without and with breast tissue classification

Lesions	Without breast density				With breast density			
	T.I	T.II	T.III	T.IV	T.I	T.II	T.III	T.IV
<i>Masses</i>								
Sensitivity	88%	79%	50%	48%	90%	91%	89%	90%
Specificity	92%	94%	93%	90%	92%	93%	92%	94%
<i>Microcalcifications</i>								
Sensitivity	87%	85%	93%	90%	91%	90%	92%	91%
Specificity	90%	86%	38%	40%	91%	90%	90%	91%

It is worth mentioning the comments made by the clinicians: The tools improve the resolution, in terms of the detectability of lesions, and additionally, they are able to distinguish their degrees of attenuation. The ability of wavelets to homogenize the background and of β -spline filtering to provide contrast and relief was judged to be quite useful. Both wavelets and β -spline work well in analyzing the resolution, which means that they properly characterize the border of the region of interest. They project the image onto a gray background which highlights the spicules, distortions and parenchyma. The β -spline transform keeps the original size of the calcium nodes. The filtering presented has shown to be successful at highlighting breast lesions on different tissue types.

7. Conclusions

In this work a novel hierarchical procedure based on weighted classifiers on texture features has been proposed for breast tissue classification.

Our approach reflects up to 99% of samples are correctly classified into 4 BIRADS classes by means of the weighted voting tree classification scheme for the SFM dataset. For the FFDM dataset on average 91.57% of samples were correctly classified using 10FCV and LOOCV. When the number of classes is reduced to fatty and dense densities only, the results of CCP are 99% and 96.93% for SFM and FFDM datasets respectively. This improves upon previous results reported in the literature.

Moreover, a large database of full-field digital mammograms classified into the four BIRADS categories by radiologists has been used. The classification information has been incorporated into a CADe system to show how classification influences lesion detection. Breast density classification improves CADe results not only for masses, as shown in [25], but also for microcalcifications.

A statistical analysis of the 298 calculated texture features has been carried out to include only features significantly influenced by the tissue type. In this step we concluded that the tissue type affects the values of most features. Furthermore, the discriminatory power of the features was analyzed using the relation between the inter/intra cluster distances, PCA and LDA. The best results were obtained with LDA. Future work may additionally consider other textural features and the use of the bag-of-words methodology.

The processing tools implemented into the CADe system have been tested and qualitatively validated by expert clinicians at Hospital General Universitario de Ciudad Real. The results of the lesion detection algorithms obtained from the CADe system show that using breast tissue classification prior to lesion detection leads to an improvement of the detection results. Therefore, the ability to detect suspicious lesions on dense and heterogeneous tissue has been tested.

Finally, the breast parenchymal procedure presented helps to adjust correctly the parameters of the CADe algorithms, which improves the detection results of the whole system increasing the true positive rate and decreasing the false positive rate.

Acknowledgment

The authors acknowledge partial financial support from the Spanish Research Ministry through projects RETIC COMBIOMED and TIN2011-24367.

References

- [1]. R. Birdwell. The preponderance of evidence supports computer-aided detection for screening mammography. *Radiology*, 253 (2009), pp. 9–16.
- [2]. J. Tang, R. Rangayyan, J. Xu, I. El Naqa, Y. Yang. Computer-aided detection and diagnosis of breast cancer with mammography: Recent advances. *IEEE Transactions on Information Technology in Biomedicine*, 13 (2) (2009), pp. 236–251.
- [3]. N. Karssemeijer. Computer aided detection in breast imaging: more than perception aid. Proc. of the 2010 IEEE Intern. conference on Biomedical Imaging: From Nano to Macro, ISBI'10, IEEE Press, Piscataway, NJ, USA (2010) pp. 273–273.
- [4]. G. Bueno. Fuzzy systems and deformable models, series in medical physics and biomedical engineering. *Intelligent and Adaptive Systems in Medicine*, Taylor & Francis Group, London (2008) Ch. 10, pp. 305–329.
- [5]. C. Zhou, H. Chan, N. Petrick, M. Helvie, M. Goodsitt, B. Sahiner, L. Hadjiiski. Computerized image analysis: estimation of breast density on mammograms. *Medical Physics*, 28 (2001), pp. 1056–1069.
- [6]. N. Boyd, G. Dite, J. Stone, A. Gunasekara, D. English, M. McCredie, G. Giles, D. Tritchler, A. Chiarelli, M. Yaffe, J. Hopper. Heritability of mammographic density, a risk factor for breast cancer. *New England Journal of Medicine*, 347 (12) (2002), pp. 886–894.
- [7]. G. Ursin, L. Hovanessian-Larsen, Y. Parisky, M. Pike, A. Wu. Greatly increased occurrence of breast cancers in areas of mammographically dense tissue. *Breast Cancer Research*, 7 (5) (2005), pp. 605–608.
- [8]. R. Brem, J. Hoffmeister, J. Rapelyea, *et al.* Impact of breast density on computer-aided detection for breast cancer. *American Journal of Roentgenology*, 184 (2005), pp. 439–444.
- [9]. J. Wolfe. Risk for breast cancer development determined by mammographic parenchymal pattern. *Cancer*, 37 (1976), pp. 2486–2492.
- [10]. N.F. Boyd, J.W. Byng, R.A. Jong, E.K. Fishell, L.E. Little, A.B. Miller, G.A. Lockwood, D.L. Tritchler, M.J. Yaffe. Quantitative classification of mammographic densities and breast cancer risk: Results from the canadian national breast screening study. *Journal of the National Cancer Institute*, 87 (1995), pp. 670–675.
- [11]. J. Li, L. Szekely, L. Eriksson, B. Heddsen, A. Sundbom, K. Czene, P. Hall, K. Humphreys. High-throughput mammographic-density measurement: a tool for risk prediction of breast cancer. *Breast Cancer Research*, 14 (4) (2012), p. R114 <http://dx.doi.org/10.1186/bcr3238> URL <http://breast-cancer-research.com/content/14/4/R114>.
- [12]. M. Sandberg, J. Li, P. Hall, M. Hartman, I. dos Santos-Silva, K. Humphreys, K. Czene. Change of mammographic density predicts the risk of contralateral breast cancer: a case–control study. *Breast Cancer Research*, 15 (4) (2013), p. R57 <http://dx.doi.org/10.1186/bcr3451> URL <http://breast-cancer-research.com/content/15/4/R57>.
- [13]. P. Carney, D. Miglioretti, B. Yankaskas, K. Kerlikowske, R. Rosenberg, C. Rutter, B. Geller, L. Abraham, S. Taplin, M. Dignan, G. Cutter, R. Ballard-Barbash. Individual and combined effects of age, breast density, and hormone replacement therapy use on the accuracy of screening mammography. *Annals of Internal Medicine*, 138 (3) (2003), pp. 168–175.
- [14]. J.A. Harvey, V.E. Bovbjerg. Quantitative assessment of mammographic breast density: relationship with breast cancer risk. *Radiology*, 230 (2004), pp. 29–41.
- [15]. M. Yafee, N. Boyd. Mammographic breast density and cancer risk: the radiological view. *Gynecological Endocrinology*, 21 (2005), pp. 6–11.
- [16]. R. Highnam, M. Jeffreys, V. McCormack, R. Warren, G. Davey Smith, M. Brady. Comparing measurements of breast density. *Physics in Medicine and Biology*, 52 (2007), pp. 5881–5895.
- [17]. D. Kontos, P. Bakic, A. Troxel, E. Conant, A.D. Maidment. Digital breast tomosynthesis parenchymal texture analysis for breast cancer risk estimation: a preliminary study. Proc. of the 9th Intern. workshop on Digital Mammography, IWDM '08, Springer-Verlag, Berlin, Heidelberg (2008), pp. 681–688.
- [18]. Z. Lao, Z. Huo. Quantitative assessment of breast dense tissue on mammograms. Proc. of the 16th IEEE Intern. conference on Image processing, ICIP'09, IEEE Press, Piscataway, NJ, USA (2009), pp. 2577–2580.
- [19]. G. Ursin, S. Qureshi. Mammographic density a useful biomarker for breast cancer risk in epidemiologic studies. *Norsk Epidemiologi*, 19 (1) (2009), pp. 59–68.
- [20]. S.D. Tzikopoulos, M.E. Mavroforakis, H.V. Georgiou, N. Dimitropoulos, S. Theodoridis. A fully automated scheme for mammographic segmentation and classification based on breast

- density and asymmetry. *Computer Methods and Programs in Biomedicine*, 102 (2011), pp. 47–63.
- [21]. J.E. de Oliveira, A. Machado, G.C. Chavez, A. Lopes, T.M. Deserno, A. Araújo. Mammosys: A content-based image retrieval system using breast density patterns. *Computer Methods and Programs in Biomedicine*, 99 (2010), pp. 289–297.
- [23]. T.M. Deserno, M. Soiron, J.E. de Oliveira, A. Araújo. Towards computer-aided diagnostics of screening mammography using content-based image retrieval. T. Lewiner, R. Torres (Eds.), *Sibgrapi 2011 (XXIV Conference on Graphics, Patterns and Images)*, IEEE Computer Society, Maceió – AL, Brazil (2011).
- [23]. A. Malich, D.R. Fischer, M. Facius, A. Petrovitch, J. Boettcher, C. Marx, A. Hansch, W.A. Kaiser. Effect of breast density on computer aided detection. *Journal of Digital Imaging*, 18 (3) (2005), pp. 227–233.
- [24]. A. Oliver, X. Lladó, E. Pérez, J. Pont, E. Denton, J. Freixenet, J. Martí. A statistical approach for breast density segmentation. *Journal of Digital Imaging*, 23 (2010), pp. 527–537.
- [25]. A. Oliver, X. Lladó, J. Freixenet, R. Martí, E. Prez, J. Pont, R. Zwigelaar. Influence of Using Manual or Automatic Breast Density Information in a Mass Detection CAD System. *Academic Radiology*, 17 (7) (2010), pp. 877–883.
- [26]. C. Romero, C. Varela, R. Cuenca, A. Almenar, J. Pinto, M. Botella. Impact of mammographic breast density on computer-assisted detection (CAD) in a breast imaging department. *Radiologia*, 54 (5) (2011), pp. 456–461.
- [27]. K. Bovis, S. Singh. Classification of mammographic breast density using a combined classifier paradigm. *4th Intern. Workshop on Digital Mammography (2002)*, pp. 177–180.
- [28]. X.H. Wang, W.F. Good, B.E. Chapman, Y.-H. Chang, W.R. Poller, T.S. Chang, L.A. Hardesty. Automated assessment of the composition of breast tissue revealed on tissue-thickness-corrected mammography. *American Journal of Roentgenology*, 180 (2003), pp. 227–262.
- [29]. S. Petroudi, T. Kadir, M. Brady. Automatic classification of mammographic parenchymal patterns: a statistical approach. *Proc. IEEE Conf. Eng. Med. Biol. Soc.*, Vol. 1 (2003), pp. 798–801.
- [30]. S. Petroudi, M. Brady. Breast density segmentation using texture. *Lecture Notes in Computer Science*, 4046 (2006), pp. 609–615.
- [31]. A. Bosch, X. Munoz, A. Oliver, J. Martí. Modeling and classifying breast tissue density in mammograms. *Proc. of the 2006 IEEE Computer Society Conference on Computer Vision and Pattern Recognition - Volume 2, CVPR '06*, IEEE Computer Society, Washington, DC, USA (2006), pp. 1552–1558.
- [32]. A. Oliver, X. Lladó, R. Martí, J. Freixenet, R. Zwigelaar. Classifying Mammograms Using Texture Information. *Proc. Medical Image Understanding and Analysis (2007)*, pp. 223–227.
- [33]. H.S. Sheshadri, A. Kandaswamy. Experimental investigation on breast tissue classification based on statistical feature extraction of mammograms. *Computerized Medical Imaging and Graphics*, 31 (1) (2007), pp. 46–48.
- [34]. C. Castella, K. Kinkel, M.P. Eckstein, P.E. Sottas, F.R. Verdun, F.O. Bochud. Semiautomatic mammographic parenchymal patterns classification using multiple statistical features. *Academic Radiology*, 14 (12) (2007), pp. 1486–1499.
- [35]. H.F. Boehm, T. Schneider, S.M. Buhmann-Kirchhoff, T. Schlossbauer, D. Rjosk-Dendorfer, S. Britsch, M. Reiser. Automated classification of breast parenchymal density: Topologic analysis of X-ray attenuation patterns depicted with digital mammography. *American Journal of Roentgenology*, 191 (2008), pp. 275–282.
- [36]. A. Oliver, J. Freixenet, R. Martí, J. Pont, E. Pérez, E. Denton, R. Zwigelaar. A Novel Breast Tissue Density Classification Methodology. *IEEE Trans. on Inform. Techn. in Biomed.*, 12 (2008), pp. 55–65.
- [37]. A. Tagliafico, G. Tagliafico, S. Tosto, F. Chiesa, C. Martinoli, L.E. Derchi, M. Calabrese. Mammographic density estimation: Comparison among BI-RADS categories, a semi-automated software and a fully automated one. *The Breast*, 18 (1) (2009), pp. 35–40.
- [38]. T.S. Subashini, V. Ramalingam, S. Palanivel. Automated assessment of breast tissue density in digital mammograms. *Computer Vision and Image Understanding*, 114 (2010), pp. 33–43.
- [39]. J. Wang, Y. Li, Y. Zhang, H. Xie, C. Wang. Bag-of-features based classification of breast parenchymal tissue in the mammogram via jointly selecting and weighting visual words. *Proc. of the 2011 Sixth Intern. Conference on Image and Graphics, ICIG '11*, IEEE Computer Society, Washington, DC, USA (2011), pp. 622–627.
- [40]. B. Keller, D. Nathan, Y. Wang, Y. Zheng, J. Gee, E. Conant, D. Kontos. Adaptive multi-cluster fuzzy c-means segmentation of breast parenchymal tissue in digital mammography. *Proc. of the 14th Intern. Conference on Medical Image Computing and Computer-assisted Intervention, MICCAI'11*, Springer-Verlag, Berlin, Heidelberg (2011), pp. 562–569.
- [41]. Q. Liu, L. Liu, Y. Tan, J. Wang, X. Ma, H. Ni. Mammogram density estimation using sub-region classification. *4th International Conference on Biomedical Engineering and Informatics (2011)*.
- [42]. *Breast Imaging Reporting and Data System Atlas (BIRADS)*. ACR, Reston, VA (2003).

- [43]. M. Tortajada, A. Oliver, R. Marty, M. Vilagran, S. Ganau, L. Tortajada, M. Sentys, J. Freixenet. Adapting breast density classification from digitized to FFD, Lecture Notes in Computer Science, Vol. 7361. Breast Imaging: 11th International Workshop, IWDM., Vol. 7361 of Lecture Notes in Computer Science (2012), pp. 561–568.
- [44]. D. Powers. Evaluation From precision, recall and f-measure to roc, informedness, markedness and correlation. *Journal of Machine Learning Technologies*, 2 (1) (2011), pp. 37–63.
- [45]. E. Cheng, N. Xie, H. Lin, P.R. Bakic, A.D.A. Maidment, V. Megalooikonomou. Mammographic image classification using histogram intersection. Proc. of the 2010 IEEE Intern. conference on Biomedical imaging: from nano to Macro, ISBI'10, IEEE Press, Piscataway, NJ, USA (2010), pp. 197–200.
- [46]. G. Bueno, N. Vázquez, O. Déniz, P. Esteve, M.A. Rienda, M. Arias, C. Pastor. Automatic breast parenchymal density classification integrated into a CADe system. *International Journal of Computer Assisted Radiology and Surgery* (2011), pp. 309–318.
- [47]. N. Vázquez, G. Bueno, O. Déniz, P. Esteve, M.A. Rienda, C. Pastor. Automatic breast tissue classification based on BIRADS categories, IWDM'10 (2010), pp. 259–266.
- [48]. H. Freeman. On the encoding of arbitrary geometric configurations. *IRE Transactions on Electronic Computers*, EC-10 (2) (1961), pp. 260–268.
- [49]. R. Haralick, S. Sternberg, X. Zhuang. Image Analysis Using Mathematical Morphology. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 9 (4) (1987), pp. 532–550.
- [50]. T. Maenpaa, M. Pietikainen. Texture analysis with local binary patterns. *Handbook of Pattern Recognition and Computer Vision*, Word Scientific (2005) pp. 197–216.
- [51]. R. Mukundan, S.H. Ong, P.A. Lee. Image analysis by Tchebichef moments. *IEEE Transactions on Image Processing*, 10 (9) (2001), pp. 1357–1364.
- [52]. J.D. Buf. Gabor phase in texture discrimination. *Signal Processing*, 21 (1990), pp. 221–240.
- [53]. R.E. Walpole, R.H. Myers, S.L. Myers. *Probability & Statistics for Engineers & Scientists*. Pearson Education (2006).
- [54]. J. Stevens. *Applied Multivariate Statistics for the Social Sciences*. Mahwah (2002).
- [55]. W. Kruskal, W. Wallis. Use of ranks in one-criterion variance analysis. *Journal of the American Statistical Association*, 47 (260) (1952), pp. 583–621.
- [56]. M. Stephens. EDF statistics for goodness of fit and some comparisons. *Journal of the American Statistical Association*, 69 (347) (1974), pp. 730–737.
- [57]. H. Levene. Robust Tests for the Equality of Variance. I. Olkin (Ed.), *Contributions to Probability and Statistics*, Stanford University Press, Palo Alto, CA (1960), pp. 278–292.
- [58]. S. Theodoridis, K. Koutroumbas. *Pattern Recognition*. Elsevier Inc. (2009).
- [59]. A.M. Martínez, A.C. Kak. PCA versus LDA. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 23 (2) (2001), pp. 228–233.
- [60]. L.I. Kuncheva. *Combining Pattern Classifiers*. John Wiley & Sons, Inc. (2004).
- [61]. R.O. Duda, P.E. Hart, D.G. Stork. *Pattern Classification*. John Wiley & Sons, Inc. (2001).
- [62]. G. Bueno, M. Ruiz, S. Sánchez. B-Spline Filtering for Automatic Detection of Calcification Lesions in Mammograms. Proc. of the Intern. Conference on Information Optics, WIO'06 (2006), pp. 60–70.
- [63]. N. Vázquez, G. Bueno, O. Déniz, M. Fernández-Carrobles, C. Pastor, M. Rienda, P. Esteve, M. Arias. CADe system integrated within the electronic health record. *Journal of Biomedicine and Biotechnology (BioMed Research International)*, 2013 (2013), p. 14 pages <http://dx.doi.org/10.1155/2013/219407> Article ID 219407.
- [64]. I. Jacobson, G. Booch, J. Rumbaugh. *The Unified Software Development Process*. Addison-Wesley Professional (1999).
- [65]. VISILAB Group. URL <http://visilab.etsii.uclm.es>.