HHS Public Access

Author manuscript

Int J Comput Assist Radiol Surg. Author manuscript; available in PMC 2015 September 25.

Published in final edited form as:

Int J Comput Assist Radiol Surg. 2014 November ; 9(6): 1005–1020. doi:10.1007/s11548-014-0992-1.

Optimization of breast mass classification using sequential forward floating selection (SFFS) and a support vector machine (SVM) model

Maxine Tan[1], **Jiantao Pu**[2], and **Bin Zheng**[1,2]

[1]School of Electrical and Computer Engineering, University of Oklahoma, Norman, OK 73019

[2] Department of Radiology, University of Pittsburgh, Pittsburgh, PA 15213

Abstract

Purpose: Improving radiologists' performance in classification between malignant and benign breast lesions is important to increase cancer detection sensitivity and reduce false-positive recalls. For this purpose, developing computer-aided diagnosis (CAD) schemes has been attracting research interest in recent years. In this study, we investigated a new feature selection method for the task of breast mass classification.

Methods: We initially computed 181 image features based on mass shape, spiculation, contrast, presence of fat or calcifications, texture, isodensity, and other morphological features. From this large image feature pool, we used a sequential forward floating selection (SFFS)-based feature selection method to select relevant features, and analyzed their performance using a support vector machine (SVM) model trained for the classification task. On a database of 600 benign and 600 malignant mass regions of interest (ROIs), we performed the study using a ten-fold cross-validation method. Feature selection and optimization of the SVM parameters were conducted on the training subsets only.

Results: The area under the receiver operating characteristic curve (AUC) = 0.805±0.012 was obtained for the classification task. The results also showed that the most frequently-selected features by the SFFS-based algorithm in 10-fold iterations were those related to mass shape, isodensity and presence of fat, which are consistent with the image features frequently used by radiologists in the clinical environment for mass classification. The study also indicated that accurately computing mass spiculation features from the projection mammograms was difficult, and failed to perform well for the mass classification task due to tissue overlap within the benign mass regions.

Conclusions: In conclusion, this comprehensive feature analysis study provided new and valuable information for optimizing computerized mass classification schemes that may have potential to be useful as a "second reader" in future clinical practice.

Corresponding author, Maxine Tan: Maxine.Y.Tan-1@ou.edu.

Conflict of interest

Maxine Tan, Jiantao Pu and Bin Zheng declare they have no conflict of interest.

Keywords

Computer-aided diagnosis of mammograms; Breast cancer; Pattern classification; Feature selection

1. Introduction

Computer-aided detection and diagnosis schemes are currently being developed for a wide range of medical image modalities and applications, such as computed tomography (CT), mammography, magnetic resonance imaging (MRI), and virtual colonoscopy. In this study, we focus on mass classification in mammography. Most recent statistics report that breast cancer is the second most prevalent cancer in women after skin cancer, and ranks second as a cause of cancer death in women after lung cancer [1]. It is also a very heterogeneous disease [2]. Studies show that the majority of breast cancers are detected among women with no known cancer risk factors [3,4]. Thus, a uniform, population-based mammography screening protocol is currently recommended for all women who are older than a qualifying age. The task of interpreting mammograms is, however, difficult due to the large variability of breast abnormalities, overlapping dense fibro-glandular tissues, and the low cancer detection rate (i.e., three to five cancers in 1000 non-baseline screening examinations) [5]. The result of these issues generates high false positive (FP) recall rates that substantially reduce the efficacy of screening mammography.

Although commercialized computer-aided detection schemes have been routinely used in clinical practice in many medical institutions to date, these schemes only focus on detecting the suspicious mass regions and also produce relatively higher FP rates. Thus, using the current computer schemes is unable to help radiologists reduce FP recalls. In order to help reduce high FP rates that generate a large number of unnecessary biopsies and/or additional expensive imaging examination procedures, developing computer-aided diagnosis (CAD) schemes that enable to classify between suspicious breast lesions (e.g., mass-like lesions) identified by radiologists into malignant and benign ones has also been attracting research interest in recent years [6-9]. Finding optimal features that can be computed from the images by the CAD schemes is one of the most important steps to achieve optimal performance in classifying between malignant and benign lesions identified from the mammograms. The purpose of this study is to investigate an effective artificial intelligent method to search for and select optimal or relevant features from a large image feature pool to classify between benign and malignant mass regions, and then analyze their performance using a support vector machine (SVM) model for the classification task.

To date, many different features have been proposed for mass classification in the literature. For example, in [10], Varela et al. investigated a variety of features that incorporated mass border information for the classification of mammographic masses. The study compared the classification accuracy obtained using different feature space combinations (border, interior and outer) on trained 3-layer feed-forward neural networks (with 8 hidden nodes). The authors concluded that the border and outer areas contained the most valuable information for mass classification. Te Brake et al. [11] presented features related to intensity, contrast,

isodensity, location, linear texture, etc. to discriminate malignant masses from normal tissue in mammography.

Shi et al. [12] presented a novel mass characterization/classification method based on level set segmentation with new image features and patient information. The used image features were related to the presence of microcalcifications within or surrounding the mass, abruptness of mass margin and patient age. The authors used a linear discriminant analysis classifier and stepwise feature selection to merge extracted features into a classification score. Nandi et al. presented a novel mass classification method using genetic programing and feature selection in [13]. In their experiments, the authors extracted 22 features for each region of interest (ROI), namely four edge-sharpness measures, four shape factors and 14 statistical texture features. They also used five feature selection methods: sequential forward selection, sequential backward selection and three statistical measures (Student's *t* test, Kolmogorov-Smirnov test and Kullback-Leibler divergence) to refine the pool of features available to a genetic programming classifier.

Rangayyan et al. [14] proposed acutance and shape measures to discriminate between benign and malignant mammographic tumors using 39 images from the Mammographic Image Analysis Society (MIAS) database [15] and an additional set of 15 local cases. The features were computed on manually-input contours of the tumor ROIs, and the results indicated that the computed acutance and shape measures were important for the classification task. Retico et al. [16] computed 16 shape, size and intensity features for each mass ROI, and used a feed-forward neural network to classify the masses as benign or malignant. The experiments were conducted on 226 mass-like lesions (109 malignant and 117 benign) extracted from a mammographic database.

In [17], Mavroforakis et al. analyzed a list of clinical features for tumor benign/malignant classification, and performed statistical analysis on those features including the *t*-test and multivariate analysis of variance. The authors also assessed the discriminative power of each of the features using four different classifiers, namely artificial neural networks, SVMs, *k*-nearest-neighbor, and linear discriminant analysis. In [18], Kilday et al. introduced a set of features based on the normalized radial length, defined as the Euclidean distance of each pixel on the object contour to the object's centroid, which was subsequently applied by other groups, also for cancer detection and classification [19-21].

In other earlier work [22,23], Huo et al. examined the relevance of automatic computerized schemes to differentiate malignant from benign masses using two features that measure the amount of mass spiculation (based on an analysis of the radial gradient of its border/contour), and with different classifiers. Sahiner et al. [24] presented the rubber band straightening transform (RBST) for mass benign/malignant characterization. The RBST transforms a band of mass border pixels onto the Cartesian plane. In the RBST-transformed image, spiculations approximately resemble vertical lines. The results showed that the features extracted from the RBST-transformed image yielded better classification results than the same features extracted from the untransformed images. Other studies that subsequently incorporated the RBST transform can be found in [25,26].

Although many image features have been tested independently by different research groups, there is no consensus on which features are the optimal features or how to select the optimal features for the computer-aided mass classification tasks. Finding the highly-performing and robust image features remains a difficult and challenging issue. In this study, we investigated a new approach to compare a large number of features that have been previously used, and then search for the optimal features for the breast mass classification task. The structure of this paper is as follows: In Section 2, the datasets and materials used in this study are explained. In Section 3, the methods of the experiments are described. The first subsection of Section 3, namely Section 3.1 describes the features computed in our experiments. We computed and analyzed altogether 181 shape, texture, contrast, spiculation, and isodensity based features as well as features to determine the presence of fat and calcifications in this study. Section 3.2 describes the feature selection algorithm (sequential floating forward selection with the Bayes classifier [27,28]) to extract relevant features from the initial feature set, and Section 3.3 explains the classifier used for the classification task (SVM) and the classification (training and validation) methodology. The Results are reported in Section 4, and finally, the Discussion and Conclusions are presented in Section 5.

2. Materials

We have assembled a large image database by selecting regions of interest (ROIs) from the digitized screen-film-based mammograms. Each ROI depicts one mass region detected by radiologists during the original mammogram reading and later verified by pathology examinations from the biopsy specimens. Detailed descriptions of the image database and its characteristics have been reported in previous studies [19,29,30]. In brief, the original screen-film mammograms were collected from several medical institutions, and were digitized using several film digitizers (with a pixel size of around $50\times50\mu m$ and 12-bit gray level resolution).

A ROI with a fixed size of 512×512 pixels was extracted from the center of each identified suspicious mass region. Using the ROI center as the seed for a region growing algorithm, the mass region was segmented by our CAD scheme [31,32] with a defined boundary contour. For each segmented mass region, the automated segmentation result was visually examined: if there was noticeable segmentation error, the mass boundary was manually corrected (re-drawn) as reported in our previous study [33]. This process helps yield the high feature relevance. Of all the analyzed features, the shape and novel location-based isodensity and contrast features would likely be the most affected by the accuracy of the segmentations, as an accurate segmentation is a prerequisite for the reliable extraction of these features.

In this study, we used 1200 ROIs (600 malignant masses and 600 benign) randomly selected from a local clinical database and the Digital Database for Screening Mammography (DDSM) [34]; the ratio of benign/malignant ROIs selected from the local database and from the DDSM database was about 50% to 50%. Each ROI image was first reduced (subsampled) by a pixel averaging method using a kernel of 8×8 pixels in both the x and y directions, thus, increasing the pixel size from $50\times50\mu m$ in the original digitized image to $400\times400\mu m$ in the downsampled image. Figures 1 and 2 show ROI examples of a benign

and a malignant mass from our image database, and their corresponding segmentations automatically generated by our mass segmentation algorithm.

3. Methods

In this section, we describe the three stages of the classification task, namely feature computation (Section 3.1), feature selection (Section 3.2) and classification and experimental methodology (Section 3.3).

3.1. Feature computation

In order to extract and select useful features that are fully representative of the mass characteristics, they should be extracted from different regions of the mass [10,11,19,12,24,32]. We have preliminarily analyzed various features in [35], using independent training and testing set experiments. In this study, we extracted and analyzed various features based on mass shape, spiculation, contrast, texture, isodensity, as well as the presence and location of fat or calcifications. A summary of the computed features according to their feature grouping/type and number is given in Table 1. We also combined 27 previously-computed features [19,32] along with the newly-computed features in Table 1 for each of the ROIs in our image database. A total of 181 features were thus extracted from each of the ROIs to build an initial feature pool. Descriptions of each of the feature groups and their computation methodology will be given in turn, in the following subsections.

3.1.1. Shape features—Many different shape or geometric features have been proposed in the literature. These include features related to margin sharpness and spiculation, circularity, convexity, acutance measure, Fourier descriptors, etc. For a more detailed description, the reader is referred to [36]. The shape-based features are computed on all the pixels that make up the mass ROI (segmented mass). In this section, we describe all the shape features listed in Table 1; there is a mixture of novel and established features in the list.

The eccentricity is defined as the ratio of the distance between the foci of the ellipse that has the same second-moments as the segmented mass, and its major axis length. Its value varies from 0 (a circle) to 1 (a line segment). The equivalent diameter is a scalar that specifies the diameter of a circle with the same area as the ROI; it is computed as $\sqrt{4 * Area / \pi}$ whereby *Area* is defined as the number of pixels in the mass ROI. The extent is a scalar that specifies the ratio of the pixels in the ROI to the pixels in the total bounding box whereby the bounding box is defined as the smallest rectangle containing all the pixels located in the region of the mass ROI.

The convex area of the segmented mass specifies the number of pixels in the convex hull of the ROI, namely the smallest convex polygon that can contain the ROI. We also computed the major and minor axis lengths of the ellipse that has the same normalized second central moments as the segmented mass. The orientation of the ellipse that has the same second-moments as the segmented mass is computed, and it is the angle (ranging from -90 to 90 degrees) between the x-axis and the major axis of the ellipse. The solidity feature is also

computed, and it is defined as the proportion of the pixels in the convex hull that are also in the ROI.

The shape factor ratio is ($Perimeter^2$)/ $Area$ as defined in [18,37,36]. The ratio of the major axis length to the minor axis length of the ellipse was also computed as a feature, and is self-explanatory. Finally, we computed the modified compactness that was proposed by Rangayyan et al. [14,38] in order to restrict the range of the shape factor ratio to (0, 1), and obtain increasing values with higher shape complexity or roughness. The definition of modified compactness is given by: $C=1-\frac{4^*\pi^*Area}{Perimeter^2}$.

3.1.2. Spiculation features—Unlike human eyes that are quite sensitive to the presence of mass boundary spiculation, segmenting spiculation patterns both manually and automatically is often difficult (as shown in figure 1). In this section, we introduce a new method to detect the presence of spiculation within a mass based on the divergence and curl of the normalized gradient of the image. The presence of spiculation within the mass region has frequently been associated with malignant masses whereas a more homogeneous appearance is frequently associated with benign masses [39,40], although benign lesions can also be possibly surrounded by the spiculated patterns due to the tissue overlapping on the two-dimensional (2D) projection images [41]. Thus, many computerized methods have been applied to compute and assess the degree of spiculation as a feature for mass detection and/or classification [42-47].

We hereby present novel features to measure the spiculation of the mass regions based on the divergence of the normalized gradient (DNG) and the curl of the normalized gradient (CNG). The divergence is an operator that measures the magnitude of a vector field's source or sink at a given point. It is a signed scalar that represents the density of the outward flux of a vector field from an infinitesimal region around that point. The DNG feature has been applied successfully for the detection of lung nodules in computed tomography (CT) images in [48,49].

In the context of our application, a benign mass can be modeled as a circular region with homogenous intensity against a darker background. The DNG operator applied on the normalized image intensity gradient vector field requires the computation of many image partial derivatives. For the reduction of noise sensitivity, the original image should be blurred with a Gaussian smoothing operator prior to the application of the DNG operator.

The Gaussian smoothing operator is defined in 2D as $G(x,y)=\frac{1}{2\pi\sigma^2}e^{\frac{x^2+y^2}{2\sigma^2}}$ whereby σ determines the width or standard deviation of the Gaussian smoothing operator. If we consider the Gaussian as an aperture function of some observation, it is also referred to as *scale* [50].

Figure 3 shows the gradient vector field of a 2D circular region with a decreasing intensity along its radial axis. The computation of the DNG of this region will produce a maximum value at the location of its center point. The DNG is given by $k=div(\vec{w})$, whereby $\vec{w}=\vec{\nabla}L/\parallel\vec{\nabla}L\parallel$ and L is the image intensity.

On the other hand, if a region is spicular (and non-homogeneous), its gradient vector field will look similar to that shown in figure 4. The curl is an operator that measures the degree of rotation of a vector field. The direction of the curl is the axis of rotation, and can be determined by the right-hand rule. The magnitude of the curl is the magnitude of the rotation of the field.

Similar to the computation of the DNG operator, the original image should be blurred with a Gaussian smoothing operator prior to the application of the CNG operator, for the reduction of noise sensitivity. The computation of the CNG is given by $\nabla \times \vec{w}$ whereby $\vec{w}$ is as previously defined for the computation of the DNG. From figure 4, it is obvious that the field is rotating; the CNG feature will give a high value at the center point of this region. The computation of the CNG will produce a high result at the location of the center points of spicular regions whereas the DNG will give a high result at the center points of homogeneous regions. In the context of our application, we expect a higher result of the CNG feature for malignant masses, and a higher result of the DNG feature for benign masses.

Also, in order to compute the CNG and DNG features effectively for different-sized masses, they have to be computed at multiple scales or standard deviations, σ of the Gaussian smoothing operator. We expect the result that maximum responses of the DNG and CNG features will be obtained when the width of the Gaussian smoothing operator matches the diameter of the mass region. To compute the range and values of σ, we used the method proposed by Li et al. [51,52]. Assuming that the diameters of the mass regions are in a range of [d_0, d_1], and the number of scales to be used is N, then we can calculate the scales used for a series of Gaussian smoothing filters as follows [51]:

$$\sigma = d_0/4, \sigma_2 = r\sigma_1, \sigma_3 = r^2\sigma_1, \ldots, \sigma_N = r^{N-1}\sigma_1 = d_1/4, \quad (1)$$

whereby $r = (d_1/d_0)^{1/(N-1)}$.

In our experiments, after measuring the sizes of the mass regions, we selected the values of $d_0 = 12$, $d_1 = 100$ and $N = 9$. For each pixel and for both the CNG and DNG features, the final feature value is computed as the maximum value from the output of all the individual filters at N scales. As the CNG and DNG features should give maximum values at the location of the centers of a malignant and benign mass, respectively, we also located the maxima points of the CNG and DNG features within each of the segmented mass regions.

The following features were computed over the *whole* mass region, namely the regions within the segmented masks (refer to figures 1 and 2): mean, maximum, minimum, standard deviation, and median of the computed CNG and DNG features. The same features (mean, maximum, minimum, standard deviation, and median) were also computed over the *maxima points* of the CNG and DNG located within the segmented regions (due to the irregularity of the intensities and texture within the mass region, frequently, more than one maxima point is obtained per mass region). The total number of computed features to measure spiculation within the mass regions is, thus, 20 altogether.

3.1.3. Features to detect the presence and location of fat within the mass regions—Although in the clinical practice radiologists routinely detect and/or examine whether there is presence of fat within a mass, which is recognized as a good indicator to distinguish between benign and malignant, this feature has rarely been computed and used in computerized schemes for classifying masses. We hereby proposed and computed four features in this study to determine the location and presence of fat within the mass regions. In order to extract the fat regions from the mass ROIs, we applied an empirically determined threshold of 2600 to extract the fat regions defined by the locations within the mass boundary whose pixel values fell below this threshold. For the first fat-related feature, we counted the number of segmented pixels. The second feature is defined as the ratio of the value of the first feature to the total pixels of the segmented mass region. In order to compute the third feature, we performed connected-component labeling (8-connectivity) on the regions extracted from the first (thresholding) step, and counted the number of connected regions that were obtained.

The fourth feature is computed to estimate the location of the fat regions within the mass, in order to detect their distributions along the border or within the center of the mass region. To compute this feature, we first calculated the distances between the centroids (center of mass) of all the fat regions (extracted from the third feature) to the centroid of the whole mass region. We then computed the average of the distances, and divided this value by the mean radial length [18] of the mass region in order to compensate for mass size (i.e. normalization for different-sized masses).

3.1.4. Features to detect the presence of calcifications within the mass regions—We computed three features to quantify the presence of calcifications within the mass regions. First, we extracted the calcified regions within the mass regions using our scheme. The first feature to assess the presence of calcifications within the mass is given by the area or number of pixels of the calcified regions. The second feature is computed by calculating the ratio of this value to the area of the whole segmented mass region. To compute the third and final feature, we performed connected-component labeling (8-connectivity) on the extracted regions, and counted the number of regions obtained.

3.1.5. Texture-based features—Many previous studies have analyzed the usefulness of texture features to discriminate masses [24,10,53,54]. An exhaustive list is provided here [8]. In this work, we computed 22 gray level run length based features, and four gray level co-occurrence matrix based features.

We computed the gray level run length based texture features [55] using the toolbox provided by Wei [56]. The gray level run length based features consist of: short run emphasis, long run emphasis, low gray level run emphasis, high gray level run emphasis, short run low gray level emphasis, short run high gray level emphasis, long run low gray level emphasis, long run high gray level emphasis, gray level non-uniformity, run length non-uniformity, and run percentage. In order to compute the run length texture features, we first reduced the gray level range of the images from 4096 to 256 gray levels resulting in an 8 bit depth in computing the gray level run length matrix. The scheme computed four run length matrices along 0, 45, 90, and 135 degrees. The average and maximum values of the

four run length features computed along the four directions were then computed to obtain the final feature values.

We also computed four gray level co-occurrence matrix based features, namely, contrast, correlation, energy, and homogeneity using a standard computation method [57]. These statistics were computed from a gray level co-occurrence matrix (created by calculating how often a pixel with gray level value i occurs horizontally adjacent to a pixel with gray level value j).

3.1.6. Contrast-based features—Many contrast-based features are proposed in the literature for mass detection or classification [58,22,59]; for an exhaustive list, the reader is referred to [36,6,8]. In this work, we use four contrast measures previously proposed in [11,10,37] and computed over different regions of the mass and background. The contrast features were computed between predefined regions between inner and outer segments of the mass. In [11], an area outside the mass contour defined as $0.6R$ where R is the effective radius of the segmented region was used as the definition of the outer region; in [10], the background region was defined as the region outside the contour and within 2 cm from it. Both methods defined the interior segment of the mass as the region inside its contour.

This work differs from the previous methods in that we examined the effects of computing the contrast features over different regions of the mass and over different sizes of the background. We proposed this new approach of computing the contrast features as different regions of the mass and background have different intensity and structural appearance. The region at the center of the mass might have different intensities and structure compared to the region around its contour. Similarly, the background region immediately outside the mass contour might have a different (higher) intensity and structural appearance than a background region that is further away from the mass contour. In our computation, we first segmented the background or outer region denoted by O, by dilating the segmented mass with a flat, disk-shaped ('disk') structuring element (SE). We used three different sizes of the disk radius to segment the outer region: the mean radial length of the mass, ½ of the mean radial length, and ¼ of the mean radial length.

We also defined the inner region denoted by I, by the interior segment of the mass within its contour. By performing an erosion operation on region I, we obtained two separate segments of the mass: the eroded image after applying the erosion operator, I_1 and the resultant image obtained by subtracting I_1 from I, denoted by I_2. In order to perform the erosion operation, we also used a 'disk' SE with three disk radius sizes: mean radial length of the mass, ½ of the mean radial length, and ¼ of the mean radial length. The contrast-based features were computed between the inner and outer regions of the mass (between O and I, O and I_1, and O and I_2). The different regions defined for the computation of the contrast-based features are depicted more clearly in figure 5. In figure 5(c), I_1 is defined as the innermost region of the mass (lighter gray: erosion of I by SE = ¼ of the mean radial length), I_2 as the mass region adjacent to its contour (white: $I - I_1$), and the outer region O is shown in dark gray (dilation of I by SE = ½ of the mean radial length).

The four computed contrast-based features were previously used in [11,10,37]. The first contrast feature is defined in [10] as:

$$C1=\frac{E_1 - E_0}{m_{dense} - m_{fat}} \quad (2)$$

where E_I and E_O represent the average values of the inner and outer regions, respectively, m_{dense} is the average value of the dense tissue within the mass; we had previously defined dense tissue in [60] as the region that encompasses the pixel values above the median value of the segmented mass (i.e. 50% of the mass region). m_{fat} is defined as the average value of the fat tissue within the mass whereby the fat tissue was earlier defined in Section 3.1.3.

The second contrast feature is also defined in [10], and is computed as:

$$C2=\frac{(E_I - E_O)^2}{\sigma_1+\sigma_O} \quad (3)$$

where σ_I and σ_O are the pixel value standard deviations of the inner and outer regions, respectively. We computed the third contrast feature [11], representing the distance between the normalized histograms of the inner and outer regions:

$$C3=\sum_y |H_I(y) - H_O(y)| \quad (4)$$

where y denotes the pixel value, and $H_X(y)$ denotes the value of the normalized histogram of region X.

The fourth contrast feature is the simplest and most commonly-computed contrast feature [19,11], and is just the difference between the average values of the regions I and O:

$$C4=E_I - E_O \quad (5)$$

The four features (C1 to C4) were also computed between the regions I_1 and O as well as between I_2 and O. Taking into account the four contrast features computed over the different sizes of the inner and outer regions, namely three sizes of O and five defined inner regions, we obtain altogether 5×3 = 15 different regions to compute the contrast-based features. As we compute four contrast measures over each region, C1-C4, we obtain altogether 15×4 = 60 contrast features to be given to the classifier at the classification stage.

3.1.7. Isodensity-based features—Some regions that look suspicious in mammograms might actually just be structures that look suspicious due to projection. The region of a mass should usually be dense, and the presence of many holes or low-intensity subregions within the mass is indicative of structure due to projection [11]. Te Brake et al. [11] presented two isodensity-based features to quantify the amount of pixels within the mass region that have a lower intensity than its background. In [11], the authors incorporated the use of the segmented mass region for the inner region, I and $0.6R$ for the outer region, O. The authors also mentioned that one of the limitations of their method was that the location of the low

intensity regions within the mass (or their topology) could not be determined using their method. To overcome this limitation, we proposed the use of incorporating different sizes and locations of the inner mass region with different sizes of the outer mass region. Thus, the isodensity-based features were also computed on inner and outer regions defined on the mass segments. The definitions of the two computed isodensity features is given in [11], and the reader is referred to that reference for the detailed explanation.

As performed in Section 3.1.6 for the computation of the contrast-based features, we compute the isodensity features over three sizes of O and five defined inner regions (5×3 = 15 regions). Altogether, 15×2 = 30 isodensity features were then computed using our proposed method.

3.1.8. Previously-computed features—In addition, we had previously computed 27 ROI or local morphological features as shown in Table 2. A detailed description of these features has been provided in our previous publications [19,32,61]. The features in Table 2 consist of different intensity, contrast, shape, border segment, and topology based features.

3.2. Feature selection – Fast and accurate SFFS algorithm

We have altogether 181 computed features for our image database of 1200 mass ROIs. Due to the *curse of dimensionality* [62], and to avoid the risk of "over-fitting" on the training set, we employed a feature selection method proposed by Ververidis and Kotropoulos [27,28].

In general, feature selection methods can be broadly divided into *wrappers* and *filters* [63,64]. *Filters* perform feature selection independently of the classifier whereas *wrappers* select features that optimize the correct classification rate (CCR). The method proposed by Ververidis and Kotropoulos [27,28] is an improvement of the well-known Sequential Forward Floating Selection (SFFS) feature selection method proposed by Pudil et al. [65]. It uses the CCR of the Bayes classifier as the criterion employed in SFFS in a *wrapper*-based framework. A popular technique of estimating the CCR is to use *N-fold cross-validation*. In this technique, the dataset is split into N parts. N-1 parts are used as the training data, and the remaining part is used as the testing data to estimate the CCR. Each of the N parts is used in turn as the testing data. The resulting CCRs of each part are finally averaged to obtain the mean CCR (MCCR). In [27], the authors implement *repeated N-fold cross-validation*, which is simply the *N-fold cross-validation* repeated many times [66] as the variance of the MCCR estimated by *repeated N-fold cross-validation* varies less than that measured by *N-fold cross-validation.*

In [27], the authors used statistical tests employing a proved lemma to improve the speed and accuracy of SFFS, when the criterion for feature selection is the CCR of the classifier. The authors showed that using this method in the context of speech emotion recognition, SFFS computational time was reduced, and the CCR for the selected feature subset varied less than the CCR of the conventional SFFS method. Although their approach was tested on the Bayes classifier and for speech emotion recognition, it can also be applied to other applications. However, we chose to re-use the Bayes classifier due to its speed in training and validation compared to other classifiers, such as artificial neural networks.

We reused all the standard parameters that were recommended in [27], except two parameters, M, the total number of feature insertions and exclusions, and γ, the difference between the cross-validated CCR required to add another feature. First, instead of $M = 25$ recommended in [27], we chose $M = 35$ as initial experiments showed that the CCR graph frequently peaked beyond $M = 25$ and between the value of 25 and 35. Secondly, although $\gamma = 0.0125$ was used in the previous experiment [27], the study also indicated that γ should be selected based on the computational load that one can afford (a lower value of γ means a higher consistency of features selected by the algorithm, but a higher computational time is required). As a trade-off between computational time requirement and consistency of selected features, we selected $\gamma = 0.0075$ for our task.

Specifically, the subset of selected features begins with an empty set, as was also implemented in [27], and proceeds with M feature insertions and exclusions. At each insertion step, the feature that maximizes the MCCR is added to the current feature subset. After the insertion of a feature, a conditional deletion step is examined. At a deletion step, a feature that maximizes the MCCR on the new feature subset is sought. If the deletion of this feature improves the highest MCCR of the original/previous feature subset, it is deleted from the feature subset. Otherwise, an inclusion step follows. After deleting a particular feature, another feature that satisfies the conditions of the deletion step is searched for another deletion. The feature selection algorithm can also start with a randomly-selected feature subset, however, initialization with a random feature subset might include features that are irrelevant or redundant that might be unbeneficial to the classifier in the long run.

3.3. Classification and experimental methodology

Many classifiers have been proposed for mass classification in the literature including artificial neural networks [10,67,68], linear discriminant analysis [24,69] and dynamic fuzzy neural networks [70]. In this study, we trained and optimized a SVM classifier using the LIBSVM package [71].

In order to train and validate our SVM classifier, we applied a 10-fold cross-validation scheme whereby the 600 malignant ROIs and 600 benign ROIs were randomly and individually segmented into 10 exclusive partitions or subgroups. In each training and testing cycle, 9 malignant and 9 benign subgroups were used to train the model. After training, the model was applied to each ROI in the remaining one malignant and one benign subgroup to generate a likelihood score of the ROI being malignant. The scores range from 0 to 1 with a higher score indicating a higher probability of the ROI being cancerous. This procedure was iteratively executed 10 times using the different combinations of 9 positive and 9 negative subgroups. In this way, each of the 1200 malignant and benign cases will be tested once with a model-generated probability score. Feature selection using SFFS and the optimization of the SVM parameters were performed only on the training partitions, in order to minimize the possibility of introducing bias during the SVM optimization process.

We designed a LIBSVM classifier with the radial basis function (RBF) kernel, defined as $K(x_1, x_j) = exp\left(-\alpha \parallel x_i - x_j \parallel^2\right)$, $\alpha > 0$, on the training set of instance-label pairs $(\mathbf{x}_i, y_i)$, $i = 1,...,l$ where $\mathbf{x}_i \in R^n$ and $y_1 \in \{1, -1\}\ \forall i =, 1 \ldots, l$. A recommended five-fold cross-

validation method with a parallel “grid-search” [72] was used to determine the penalty parameter of the error term and c. We performed linear normalization of all input features to normalize the feature values between 0-1. For input feature vector $\mathbf{x_i}$, the corresponding linearly-normalized feature vector $\mathbf{x_i'}$, is expressed as :

$$\mathbf{x_i}' = \frac{\mathbf{x_i} - \mathbf{min}(\mathbf{x_i})}{\mathbf{max}(\mathbf{x_i}) - \mathbf{min}(\mathbf{x_i})} \quad (6)$$

In order to evaluate the accuracy of our system, we calculated the area under a receiver operating characteristic (ROC) curve (AUC). AUC was computed using a ROC curve fitting program that applies an expanded bi-normal model and the maximum likelihood estimation method (ROCKIT http://www.-radiology.uchicago.edu/krl/, University of Chicago, USA, 1998). We computed AUC along with the corresponding 95% confidence interval. Since we used a 10-fold cross-validation method, 10 SVMs with 10 different combinations of feature sets were independently built and tested. The detailed features selected and implemented in the 10 SVMs are not exactly identical. We thus also performed an analysis of the features that were most frequently selected by the SFFS-based feature selection method for our task of mass classification.

Furthermore, to analyze the benefit/usefulness of the modified SFFS feature selection algorithm, we compared its performance with the SVM classifier trained with all 181 features. We also compared modified SFFS’s performance with a different feature selection method, namely sequential forward selection (SFS) using an evaluation criterion based on achieving a small within-class distance and a large between-class distance [73,74]. The performances of the different methods were analyzed by comparing their AUC results, and statistical significance tests at the 5% significance level were performed on the average AUC results computed per fold.

4. Results

Figure 6 shows the ROC curves using the SVM classifier-generated probability scores for all 1200 images in our dataset. The area under the ROC curve, AUC = 0.805 with a standard deviation of 0.012 and 95% confidence interval of [0.779, 0.828] was obtained for the modified SFFS feature selection algorithm. For SFS feature selection combined with SVM, AUC = 0.749 with a standard deviation of 0.014 and 95% confidence interval of [0.721, 0.775]. For SVM trained with all features, AUC = 0.807 with a standard deviation of 0.012 and 95% confidence interval of [0.782, 0.831]. None of the three methods had differences in AUC results that were statistically significant at the 5% significance level, although the computed p-values were much lower for modified SFFS compared with SFS ($p = 0.0801$), and for SVM trained with all features compared with SFS ($p = 0.0503$). For modified SFFS compared with SVM trained with all features, the computed p-value was much higher ($p = 0.7420$).

We also obtain the confusion matrix in Table 3 using the classifier prediction score of 0.5 as a threshold on the probability scores generated by the features selected by modified SFFS and trained with the SVM classifier. Table 3 shows that 74.4% (893/1200) of the ROIs are

correctly classified by our system whereas 25.6% (307/1200) of ROIs are misclassified, at this prediction threshold. In the cancer image subgroup, 75.5% (453/600) are correctly predicted whereas in the benign image subgroup, the prediction accuracy is 73.3% (440/600). These results indicate that malignant images are more accurately classified or predicted by our classification scheme than benign images at this particular threshold.

Table 4 displays the frequency of the selection of features by the SFFS-based feature selection algorithm according to their feature groups or type. In the far-left column of Table 4, the feature type or group is shown. The middle column of Table 4 displays the fraction of that feature group out of all available features (181); for example, in the first row of Table 4, there are 11 shape features altogether out of the total of 181 computed features. Finally, the far-right column displays the percentage of features that were selected in each group, namely the average percentage value of selected features within a group, and the standard deviation intervals computed over the 10-fold cross-validation experiment. For example, in the first row, an average of 60% of 11 shape features (60% × 11 = 6.6 features) were selected over the 10-fold cross-validation experiment with standard deviation intervals of 15% or 1.65 features.

5. Discussion

Improving classification accuracy of breast lesions has high clinical impact to help increase cancer detection sensitivity while also reducing the FP recall rates and/or the number of unnecessary benign biopsies. For this purpose, selecting optimal image features and accurately assessing their values play a critical role in lesion classification. Although radiologists routinely assess three image features (namely lesion density in relationship to its background, lesion boundary shape, and the presence of fat inside the lesion) in interpreting mammograms, previously-developed CAD schemes typically used many different image features. Since the features used by different schemes also vary substantially, there is no agreement or accepted reference on what are the most effective features for classifying between malignant and benign masses. In this work, we conducted a comprehensive feature analysis that covered the majority of feature categories used in the previous CAD schemes. From these feature categories, we extracted and computed 181 features to build an initial feature pool. We examined and applied a novel wrapper-based SFFS feature selection algorithm. Although this new algorithm has shown to perform faster and more accurately than the conventional SFFS algorithm in a speech emotion recognition problem [27,28], to the best of our knowledge, it has never been previously applied to the field of CAD of medical images.

In this study we demonstrated that using this modified SFFS algorithm to select mass image features, a SVM classifier yielded an AUC value of 0.805 for image-based evaluation on our dataset. This value is very comparable with the AUC result achieved by training the SVM using all 181 features (AUC = 0.807). The SFFS method also selects only 25.8±6.4 features on average per fold, thus requiring a much lower feature computation time requirement compared with SVM trained on all features. To compute all 181 features on the whole image database of 1200 ROIs, the total processing time using a regular Dell precision T3600 workstation was around 23 minutes whereas the processing time required to compute 26

features selected by SFFS on the same workstation was around 2 minutes. This represents a speed-up of around 11 fold (i.e. around 11 times faster) in the feature computation time requirement with SFFS feature selection. The classifier training and validation time requirement will also likely be less for the 26 features; however, the improvement gained at this stage is likely to be negligible compared to the speed-up obtained by SFFS at the feature computation stage.

Although the image-based classification results are typically lower than the case-based classification in which the information of two mass regions depicting on craniocaudal (CC) and mediolateral oblique (MLO) is combined, the classification performance of our study is higher or very comparable with respect to other methods reported in the literature [53,75,10] as references (shown in Table 5). Furthermore, the performance of our mass classification system is also comparable to the average performance of radiologists (e.g. AUC = 0.81 in [22]).

Although previous CAD schemes have used many complicated texture features that may not be recognizable by the human eye, aiming to improve classification of breast masses, radiologists actually use and access much simpler image features including mass shape, presence of fat within the mass region and mass density compared to its background in their decision making. From the results in Table 4, it can be observed that the shape features were selected most frequently by the SFFS-based algorithm. This was followed by the features to detect the presence of fat and the isodensity features. These results correlate well with our observation of the visual preference of the feature selection made by the radiologists in reading mammograms. Hence, the results of this study indicated that although there may be quite a big difference between human vision and computer vision, a computer-aided scheme that uses a small number of simple features mimicking the common features used by radiologists may also perform well in the mass classification task. We believe that a computer-aided scheme that uses the same types of features with the radiologists can have a number of advantages. For example, the computed feature values can be more robust due to the elimination of the inter-reader variability during the subjective assessment of these features. The combination of these features can also be made by more powerful machine learning or statistical models (e.g., SVM), so that the classification results may not be duplicated with radiologists' decisions. Finally, due to the simplicity of the features used in the scheme, the scheme classification results can be more explainable to the radiologists and thus increase their confidence to consider the scheme classification results. This is also very important when applying an automated scheme in clinical practice. Therefore, this result, combined with the similar AUC results obtained by our SVM classification scheme compared to the average performance of radiologists, indicates that our proposed mass classification scheme has potential to function as a "second reader" in assisting radiologists to make better decisions about whether to recall the women and/or recommend lesion biopsy in clinical practice.

From Table 4 we also found that the spiculation features were not selected frequently by the SFFS algorithm, and implemented in the SVM classifier indicating that these features did not perform well in the mass classification task when applied to our dataset. To understand this unexpected result, we examined the spiculation feature values computed on some of the

malignant and benign mass ROIs in our image database. Figure 7 and figure 8 display an example of a malignant mass and a benign mass, respectively. In both figures, the DNG and the CNG features were computed, and their maximum values corresponding to the location of the maxima points within the segmented mass were located and computed across several scales, σ. Note that in figure 8, there are two maxima points of the CNG feature within the image computed at the same scale; the maximum (returned) value is the blue 'plus' marker located close to the center of the mass region.

From figures 7 and 8, we observed that the maximum values of the DNG and the CNG features are close to the mass center. Some other examples of the maximum DNG and CNG features obtained on some of the benign and malignant masses in our database are displayed in figure 9; the top row of figure 9 displays the malignant masses whereas the benign masses are shown in the bottom row.

From our experiments and the results displayed in figure 9, we found that the DNG feature yields slightly higher values for the malignant masses. We had initially expected the benign masses to yield a higher value of the DNG due to their assumed homogeneous intensity; however, the results indicate that the region within the benign masses, in general, might not be homogeneous. Another surprising result that is linked to the first result is that the CNG feature is also high for the benign masses. Our initial hypothesis was that the CNG feature will yield a high value for malignant masses due to its spicular interior and produce a low value for benign masses. However, the results show that both the malignant and benign masses yield a high value of the CNG feature. These unexpected results could be due to tissue overlap within the benign mass regions. The results indicate that it may not be feasible to measure the amount of spiculation in 2D mammographic images using the DNG and CNG features; however, these features may be useful for measuring spiculation in higher dimensions (3D). This hypothesis needs to be analyzed in our future work/studies.

We also recognized that this preliminary study has a number of limitations. For example, this is only a technology development study. We only assess the mass classification performance of a standalone computerized scheme. No observer performance study was involved. As a result, a number of empirically-determined thresholds or parameters (e.g., a fixed threshold of 2600 to detect fat inside the mass region) may not be optimal. In order to overcome these limitations, the adaptive thresholds or image processing methods need to be developed and tested with the feedback from the observers (e.g., radiologists) in the future studies. In addition, although a number of previous studies have shown that the use of computer-aided diagnosis schemes could be beneficial in the clinical practice, evaluating the impact of our scheme on radiologists' performance is also an important avenue for future studies.

In summary, we developed and tested a new computer-aided breast mass classification scheme through a comprehensive image feature selection process. From a large pool of 181 image features computed from many different feature categories, we demonstrated that using an efficient and accurate SFFS-based feature selection algorithm, relevant features were selected and provided to a SVM model in a 10-fold cross-validation experiment, whereby feature selection and parameter tuning were performed only on the training subsets in order

to minimize bias due to parameter optimization/tuning of the SVM model. In this limited study performed on 1200 ROIs, our CAD system performs competitively with an image-based AUC of 0.805 in this challenging task of mass classification in mammography. It is in the agenda to examine the performance of our scheme on a more extensive and diverse image database in future work to further evaluate the feasibility or potential of using CAD schemes as a "second reader" to assist radiologists in breast mass classification in future clinical practice.

Acknowledgements

This work was supported in part by the National Cancer Institute, National Institutes of Health, under Grant R01CA160205.

References

1. American Cancer Society. Cancer Facts & Figures 2013. 2013. http://wwwcancerorg/research/cancerfactsstatistics/cancerfactsfigures2013/index
2. Siegel R, Naishadham D, Jemal A. Cancer statistics, 2013. CA Cancer J Clin. 2013; 1; 63:11–30. [PubMed: 23335087]
3. Madigan MP, Ziegler RG, Benichou J, Byrne C, Hoover RN. Proportion of breast cancer cases in the United States explained by well-established risk factors. J Natl Cancer Inst. 1995; 22; 87:1681–1685. [PubMed: 7473816]
4. Amir E, Freedman OC, Seruga B, Evans DG. Assessing women at high risk of breast cancer: a review of risk assessment models. J Natl Cancer Inst. 2010; 10; 102:680–691. [PubMed: 20427433]
5. Sickles EA, Wolverton DE, Dee KE. Performance parameters for screening and diagnostic mammography: specialist and general radiologists. Radiology. 2002; 3; 224:861–869. [PubMed: 12202726]
6. Rangayyan RM, Ayres FJ, Desautels JEL. A review of computer-aided diagnosis of breast cancer: Toward the detection of subtle signs. Journal Of The Franklin Institute-Engineering And Applied Mathematics. 2007; 344(3-4):312–348.
7. Oliver A, Freixenet J, Marti J, Perez E, Pont J, Denton ERE, Zwiggelaar R. A review of automatic mass detection and segmentation in mammographic images. Med Image Anal. 2010; 2; 14:87–110. [PubMed: 20071209]
8. Elter M, Horsch A. CADx of mammographic masses and clustered microcalcifications: A review. Med Phys. 2009; 6; 36:2052–2068. [PubMed: 19610294]
9. Horsch AH, Elter M. Needs assessment for next generation computer-aided mammography reference image databases and evaluation studies. Int J CARS. 2011; 6; 6:749–767. A.
10. Varela C, Timp S, Karssemeijer N. Use of border information in the classification of mammographic masses. Phys Med Biol. 2006; 51:425–441. [PubMed: 16394348]
11. Brake, GMt; Karssemeijer, N.; Hendriks, JH. An automatic method to discriminate malignant masses from normal tissue in digital mammograms. Phys Med Biol. 2000; 10; 45:2843–2857. [PubMed: 11049175]
12. Shi J, Sahiner B, Chan HP, Ge J, Hadjiiski L, Helvie MA, Nees A, Wu YT, Wei J, Zhou C, Zhang Y, Cui J. Characterization of mammographic masses based on level set segmentation with new image features and patient information. Med Phys. 2008; 1; 35:280–290. [PubMed: 18293583]
13. Nandi RJ, Nandi AK, Rangayyan RM, Scutt D. Classification of breast masses in mammograms using genetic programming and feature selection. Med Bio Eng Comput. 2006; 8; 44:683–694. [PubMed: 16937210]
14. Rangayyan RM, El-Faramawy NM, Desautels JE, Alim OA. Measures of acutance and shape for classification of breast tumors. IEEE Trans Med Imaging. 1997; 6; 16:799–810. doi: 10.1109/42.650876. [PubMed: 9533580]

15. Suckling, J.; Parker, J.; Dance, DR.; Astley, S.; Hutt, I.; Boggis, C.; Ricketts, I.; Stamatakis, E.; Cerneaz, N.; Kok, SL.; Taylor, PM.; Betal, D.; Savage, J. The Mammographic Image Analysis Society digital mammogram database; Proc 2nd International Workshop on Digital Mammography; 1994. p. 375-378.
16. Retico A, Delogu P, Fantacci ME, Kasae P. An automatic system to discriminate malignant from benign massive lesions on mammograms. Nuclear Instruments and Methods in Physics Research Section A. 2006; 569:596–600.
17. Mavroforakis M, Georgiou H, Dimitropoulos N, Cavouras D, Theodoridis S. Significance analysis of qualitative mammographic features, using linear classifiers, neural networks and support vector machines. Eur J Radiol. 2005; 1; 54:80–89. [PubMed: 15797296]
18. Kilday J, Palmieri F, Fox MD. Classifying mammographic lesions using computerized image analysis. IEEE Trans Med Imaging. 1993; 4; 12:664–669. [PubMed: 18218460]
19. Zheng B, Lu A, Hardesty LA, H. SJ, Hakim CM, Ganott MA, Gur D. A method to improve visual similarity of breast masses for an interactive computer-aided diagnosis environment. Med Phys. 2006; 1; 33:111–117. [PubMed: 16485416]
20. Sahiner B, Petrick N, Chan HP, Hadjiiski LM, Paramagul C, Helvie MA, Gurcan MN. Computer-aided characterization of mammographic masses: accuracy of mass segmentation and its effects on characterization. IEEE Trans Med Imaging. 2001; 12; 20:1275–1284. doi:10.1109/42.974922. [PubMed: 11811827]
21. Bruce LM, Adhami RR. Classifying mammographic mass shapes using the wavelet transform modulus-maxima method. IEEE Trans Med Imaging. 1999; 12; 18:1170–1177. doi: 10.1109/42.819326. [PubMed: 10695529]
22. Huo Z, Giger ML, Vyborny CJ, Wolverton DE, Schmidt RA, Doi K. Automated computerized classification of malignant and benign masses on digitized mammograms. Acad Radiol. 1998; 3; 5:155–168. [PubMed: 9522881]
23. Huo Z, Giger ML, Vyborny CJ, Wolverton DE, Metz CE. Computerized classification of benign and malignant masses on digitized mammograms: a study of robustness. Acad Radiol. 2000; 12; 7:1077–1084. [PubMed: 11131052]
24. Sahiner B, Chan HP, Petrick N, Helvie MA, Goodsitt MM. Computerized characterization of masses on mammograms: the rubber band straightening transform and texture analysis. Med Phys. 1998; 4; 25:516–526. [PubMed: 9571620]
25. Sahiner B, Chan HP, Petrick N, Helvie MA, Goodsitt MM. Design of a high-sensitivity classifier based on a genetic algorithm: application to computer-aided diagnosis. Phys Med Biol. 1998; 10; 43:2853–2871. [PubMed: 9814523]
26. Hadjiiski L, Sahiner B, Chan HP, Petrick N, Helvie M. Classification of malignant and benign masses based on hybrid ART2LDA approach. IEEE Trans Med Imaging. 1999; 12; 18:1178–1187. doi:10.1109/42.819327. [PubMed: 10695530]
27. Ververidis D, Kotropoulos C. Fast and accurate sequential floating forward feature selection with the Bayes classifier applied to speech emotion recognition. Signal Process. 2008; 12; 88:2956–2970. doi:10.1016/j.sigpro.2008.07.001.
28. Ververidis D, Kotropoulos C. Information loss of the mahalanobis distance in high dimensions: application to feature selection. IEEE Trans Pattern Anal Mach Intell. 2009; 12; 31:2275–2281. doi:10.1109/tpami.2009.84. [PubMed: 19834146]
29. Park SC, Wang XH, Zheng B. Assessment of performance improvement in content-based medical image retrieval schemes using fractal dimension. Acad Radiol. 2009; 10; 16:1171–1178. [PubMed: 19524455]
30. Zheng B, Leader JK, Abrams G, Shindel B, Catullo V, Good WF, Gur D. Computer-aided detection schemes: the effect of limiting the number of cued regions in each case. Am J Roentgenol. 2004; 3; 182:579–583. doi:10.2214/ajr.182.3.1820579. [PubMed: 14975949]
31. Gur D, Stalder JS, Hardesty LA, Zheng B, Sumkin JH, Chough DM, Shindel BE, Rockette HE. Computer-aided detection performance in mammographic examination of masses: assessment. Radiology. 2004; 2; 233:418–423. doi:10.1148/radiol.2332040277. [PubMed: 15358846]

32. Zheng B, Chang Y-H, Gur D. Computerized detection of masses in digitized mammograms using single-image segmentation and a multilayer topographic feature analysis. Acad Radiol. 1995; 11; 2:959–966. [PubMed: 9419667]
33. Wang XH, Park SC, Zheng B. Improving performance of content-based image retrieval schemes in searching for similar breast mass regions: an assessment. Phys Med Biol. 2009; 4; 54:949–961. doi:10.1088/0031-9155/54/4/009. [PubMed: 19147902]
34. Heath M, Bowyer K, Kopans D, Kegelmeyer P Jr. Moore R, Chang K, Munishkumaran S. Current Status of the Digital Database for Screening Mammography. Digital Mammography. 1998; 13:457–460. doi:10.1007/978-94-011-5318-8_75.
35. Tan, M.; Pu, J.; Zheng, B. A new mass classification system derived from multiple features and a trained MLP model; Proc SPIE (Medical Imaging 2014: Computer-Aided Diagnosis); 2014. Accepted:
36. Cheng HD, Shi XJ, Min R, Hu LM, Cai XP, Du HN. Approaches for automated detection and classification of masses in mammograms. Pattern Recognit. 2006; 4; 39:646–668. doi:http://dx.doi.org/10.1016/j.patcog.2005.07.006.
37. Zheng B, Leader JK, Abrams GS, Lu AH, Wallace LP, Maitz GS, Gur D. Multiview-based computer-aided detection scheme for breast masses. Med Phys. 2006; 9; 33:3135–3143. [PubMed: 17022205]
38. Shen L, Rangayyan RM, Desautels JEL. Detection and classification of mammographic calcifications. Int J Pattern Recognit Artif Intell. 1993; 6; 7:1403–1416.
39. Sickles EA. Breast masses: mammographic evaluation. Radiology. 1989; 2; 173:297–303. [PubMed: 2678242]
40. D'Orsi CJ, Kopans DB. Mammographic feature analysis. Semin Roentgenol. 1993; 3; 28:204–230. [PubMed: 8211230]
41. Franquet T, De Miguel C, Cozcolluela R, Donoso L. Spiculated lesions of the breast: mammographic-pathologic correlation. Radiographics. 1993; 4; 13:841–852. [PubMed: 8356272]
42. Zwiggelaar R, Astley SM, Boggis CRM, Taylor CJ. Linear structures in mammographic images: detection and classification. IEEE Trans Med Imaging. 2004; 9; 23:1077–1086. doi:10.1109/tmi.2004.828675. [PubMed: 15377116]
43. Sampat, MP.; Whitman, GJ.; Markey, MK.; Bovik, AC. Evidence based detection of spiculated masses and architectural distortions; Medical Imaging 2005: Image Processing; 2005. p. 26-37.
44. Vyborny CJ, Doi T, O'Shaughnessy KF, Romsdahl HM, Schneider AC, Stein AA. Breast cancer: importance of spiculation in computer-aided detection. Radiology. 2000; 3; 215:703–707. [PubMed: 10831688]
45. Jiang L, Song E, Xu X, Ma G, Zheng B. Automated detection of breast mass spiculation levels and evaluation of scheme performance. Acad Radiol. 2008; 12; 15:1534–1544. doi:http://dx.doi.org/10.1016/j.acra.2008.07.015. [PubMed: 19000870]
46. Sampat MP, Bovik AC, Whitman GJ, Markey MK. A model-based framework for the detection of spiculated masses on mammography. Med Phys. 2008; 5; 35:2110–2123. [PubMed: 18561687]
47. Kegelmeyer WP Jr. Pruneda JM, Bourland PD, Hillis A, Riggs MW, Nipper ML. Computer-aided mammographic screening for spiculated lesions. Radiology. 1994; 2; 191:331–337. [PubMed: 8153302]
48. Tan M, Deklerck R, Cornelis J, Jansen B. Phased searching with NEAT in a Time-Scaled Framework: Experiments on a computer-aided detection system for lung nodules. Artificial Intelligence in Medicine. 2013; 3; 59:157–167. doi:http://dx.doi.org/10.1016/j.artmed.2013.07.002. [PubMed: 24028824]
49. Tan M, Deklerck R, Jansen B, Bister M, Cornelis J. A novel computer-aided lung nodule detection system for CT images. Med Phys. 2011; 10; 38:5630–5645. [PubMed: 21992380]
50. Romeny, BMtH. Front-End Vision & Multi-Scale Image Analysis. Springer Verlag; Berlin: 2003.
51. Li Q, Sone S, Doi K. Selective enhancement filters for nodules, vessels, and airway walls in two- and three-dimensional CT scans. Med Phys. 2003; 8; 30:2040–2051. [PubMed: 12945970]
52. Li Q, Arimura H, Doi K. Selective enhancement filters for lung nodules, intracranial aneurysms, and breast microcalcifications. Proc CARS. 2004; 1268:929–934.

53. Mudigonda NR, Rangayyan RM, Desautels JE. Gradient and texture analysis for the classification of mammographic masses. IEEE Trans Med Imaging. 2000; 10; 19:1032–1043. doi: 10.1109/42.887618. [PubMed: 11131493]
54. Gupta S, Markey MK. Correspondence in texture features between two mammographic views. Med Phys. 2005; 6; 32:1598–1606. [PubMed: 16013719]
55. Tang X. Texture information in run-length matrices. IEEE Trans Image Proc. 1998; 11; 7:1602–1609.
56. Wei, X. Gray Level Run Length Matrix Toolbox v1.0. Beijing Aeronautical Technology Research Center; 2007. http://wwwmathworkscom/matlabcentral/fileexchange/17482-gray-level-run-length-matrix-toolbox Last accessed: 12 December 2013
57. Haralick RM, Shanmugam K, Dinstein I. Texture features for image classification. IEEE Trans Syst Man, Cybern. 1973; 3(6):610–621.
58. Li H, Wang Y, Liu KJ, Lo SC, Freedman MT. Computerized radiographic mass detection--part I: Lesion site selection by morphological enhancement and contextual segmentation. IEEE Trans Med Imaging. 2001; 4; 20:289–301. [PubMed: 11370896]
59. Petrick N, Chan HP, Sahiner B, Helvie MA. Combined adaptive enhancement and region-growing segmentation of breast masses on digitized mammograms. Med Phys. 1999; 8; 26:1642–1654. [PubMed: 10501064]
60. Chang Y-H, Wang X-H, Hardesty LA, Chang TS, Poller WR, Good WF, Gur D. Computerized assessment of tissue composition on digitized mammograms. Acad Radiol. 2002; 8; 9:899–905. [PubMed: 12186438]
61. Zheng B, Wang X, Lederman D, Tan J, Gur D. Computer-aided detection; the effect of training databases on detection of subtle breast masses. Acad Radiol. 2010; 20650667; 17:1401–1408. [PubMed: 20650667]
62. Powell, WB. Approximate Dynamic Programming: Solving the Curses of Dimensionality. 1st. Wiley-Interscience; 2007.
63. Langley, P. AAAI fall symposium on relevance. AAAI Press; New Orleans: 1994. Selection of relevant features in machine learning; p. 140-144.
64. Kohavi R, John GH. Wrappers for feature subset selection. Artif Intell. 1997; 97(1-2):273–324.
65. Pudil P, Novovičová J, Kittler J. Floating search methods in feature selection. Pattern Recognit Lett. 1994; 11; 15:1119–1125.
66. Burman P. A comparative study of ordinary cross-validation, v-fold cross-validation and the repeated learning-testing methods. Biometrika. 1989; 3; 76:503–514. doi:10.2307/2336116.
67. Huo Z, Giger ML, Vyborny CJ, Bick U, Lu P, Wolverton DE, Schmidt RA. Analysis of spiculation in the computerized classification of mammographic masses. Med Phys. 1995; 10; 22:1569–1579. [PubMed: 8551981]
68. Mavroforakis ME, Georgiou HV, Dimitropoulos N, Cavouras D, Theodoridis S. Mammographic masses characterization based on localized texture and dataset fractal analysis using linear, neural and support vector machine classifiers. Artif Intell Med. 2006; 2; 37:145–162. doi:10.1016/j.artmed.2006.03.002. [PubMed: 16716579]
69. Gupta S, Chyn PF, Markey MK. Breast cancer CADx based on BI-RADS descriptors from two mammographic views. Med Phys. 2006; 6; 33:1810–1817. [PubMed: 16872088]
70. Lim WK, Er MJ. Classification of mammographic masses using generalized dynamic fuzzy neural networks. Med Phys. 2004; 5; 31:1288–1295. [PubMed: 15191321]
71. Chang C-C, Lin C-J. LIBSVM: A library for support vector machines. ACM TIST. 2011; 2(3):27.
72. Hsu, C-W.; Chang, C-C.; Lin, C-J. A practical guide to support vector classification Technical report. National Taiwan University; Taipei 106, Taiwan: 2009.
73. Theodoridis, S.; Koutroumbas, K. Pattern Recognition. 4th. Academic Press; San Diego, CA, USA: 2008.
74. Theodoridis, S.; Koutroumbas, K. Introduction to Pattern Recognition: A MATLAB Approach. Academic Press; 2010.

75. Rangayyan RM, Mudigonda NR, Desautels JE. Boundary modelling and shape analysis methods for classification of mammographic masses. Med Biol Eng Comput. 2000; 5; 38:487–496. [PubMed: 11094803]

76. Stone M. Cross-validatory choice and assessment of statistical predictions. J Royal Statistical Society B. 1974; 36:111–147.

77. Land JWH, McKee DW, Anderson FR, Lo JY. Breast cancer classification improvements using a new kernel function with evolutionary-programming-configured support vector machines. Proc SPIE. 2004; 5370:880–887.

78. Heidt, S-R.; Elter, M.; Wittenberg, T.; Paulus, D. Model-Based Characterization of Mammographic Masses. In: Meinzer, H-P.; Deserno, T.; Handels, H.; Tolxdorff, T., editors. Bildverarbeitung für die Medizin 2009. Informatik aktuell. Springer; Berlin Heidelberg: 2009. p. >287-291.doi: 10.1007/978-3-540-93860-6_58

79. Eltonsy, NH.; Elmaghraby, AS.; Tourassi, GD. Bilateral Breast Volume Asymmetry in Screening Mammograms as a Potential Marker of Breast Cancer: Preliminary Experience; Proc. ICIP 2007; 2007. p. V - 5-V - 8.Sept. 16 2007-Oct. 19 2007

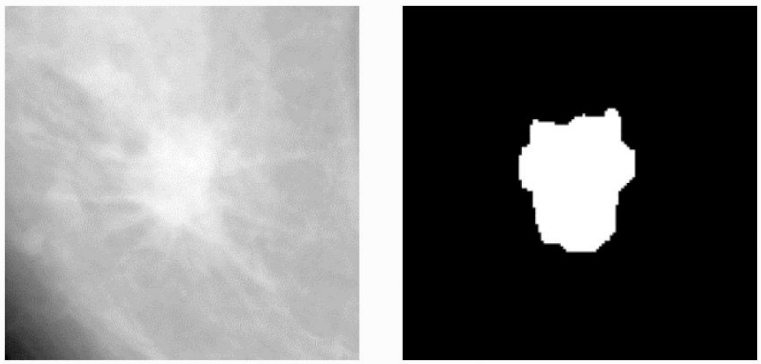

Fig. 1.
Example of a malignant mass ROI and its corresponding segmentation mask

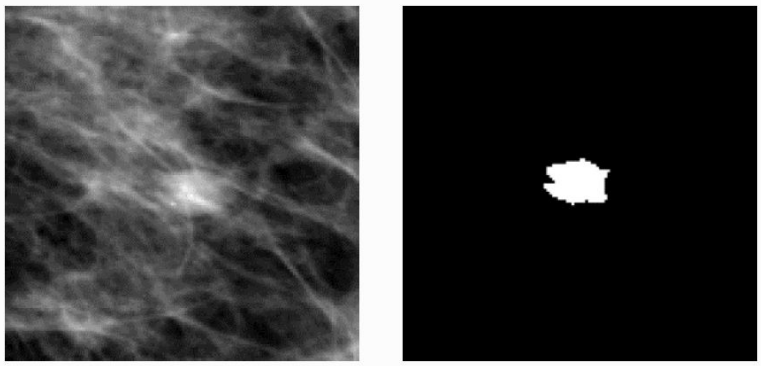

Fig. 2.
Example of a benign mass ROI and its corresponding segmentation mask

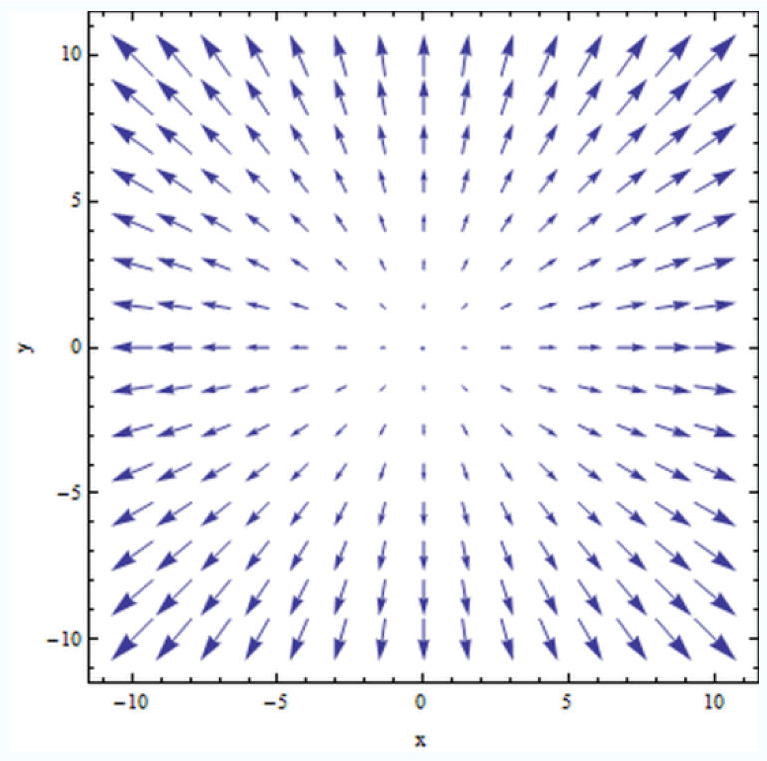

Fig. 3.

An example of the vector field of the gradient of a circular region with a decreasing intensity along its radial axis. The computation of the divergence of the normalized gradient (DNG) of this region will produce a maximum value at the location of its center point

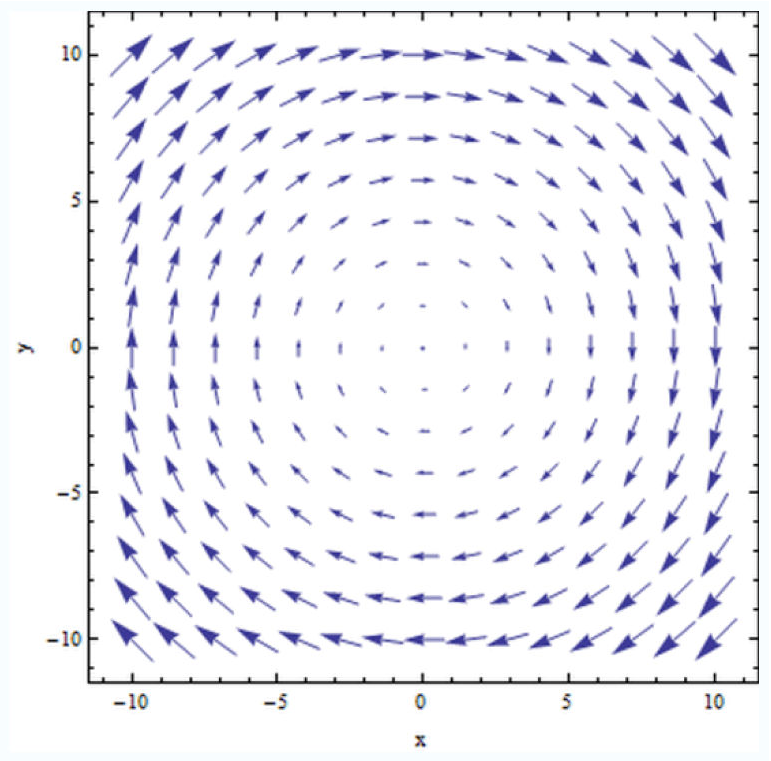

Fig. 4.

An example of the vector field of the gradient of a circular region with spicular texture or intensities. The computation of the curl of the normalized gradient (CNG) of this region will produce a maximum value at the location of its center point

Fig. 5.

A (a) malignant mass, (b) its original segmentation mask and (c) different regions defined for computation of the contrast-based features: I_1 (innermost lighter gray region of the mass), I_2 (white region of the mass adjacent to its contour), and O (darker gray background region). Three sets of contrast features are computed from (1) between O and I whereby I is the interior segment of the mass within its contour ($I_1 + I^2$), (2) between O and I_1, and (3) between O and I_2

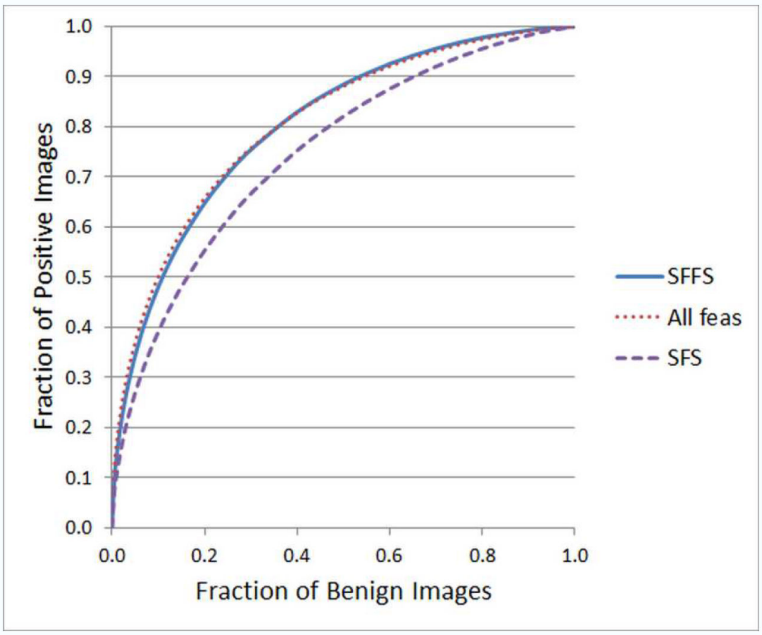

Fig. 6.

ROC curves of the fast and accurate SFFS feature selection algorithm (SFFS), SVM trained with all features (All feas), and SFS based on a small within-class distance and a large between-class distance (SFS) of applying our image-based mass classification system to correctly classify the mass ROIs in our image database as benign or malignant. The computed AUC results and 95 confidence intervals are as follows – SFFS: 0.805 [0.779, 0.828]; All feas: 0.807 [0.782, 0.831] and SFS: 0.749 [0.721, 0.775].

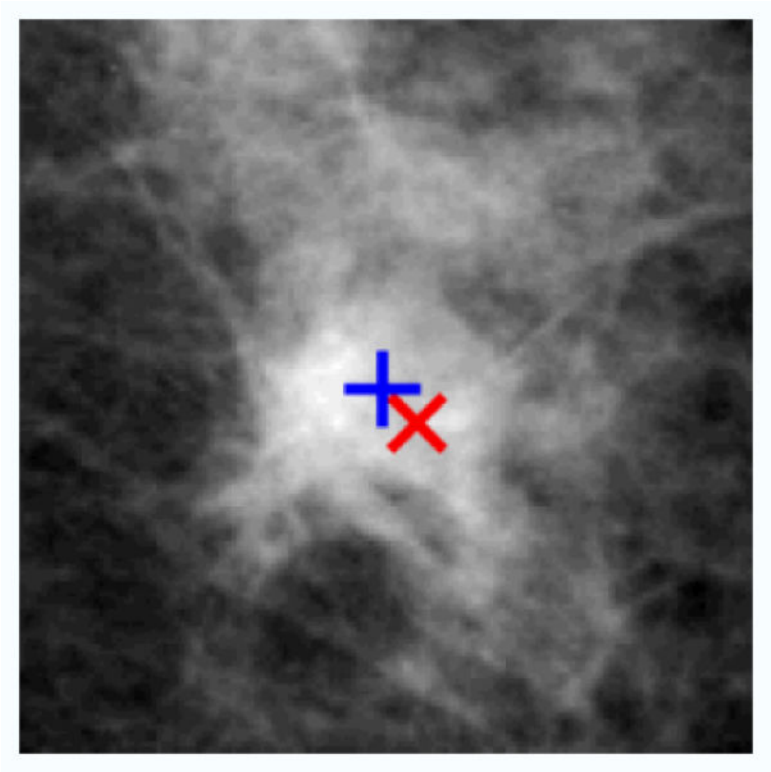

Fig. 7.

A malignant mass, the maxima point where the DNG feature is maximal (red ‘cross’ marker; DNG = 3.5, computed at $\sigma = 3.9$), and the maxima point where the CNG feature is maximal (blue ‘plus’ marker; CNG = 6.8, computed at $\sigma = 14.7$)

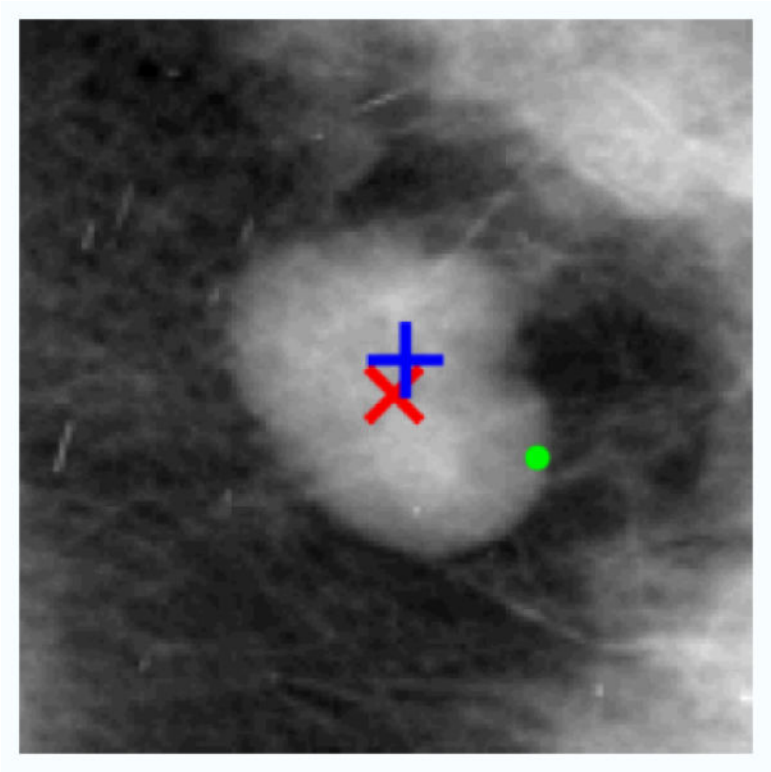

Fig. 8.

A benign mass, the maxima point where the DNG feature is maximal (red ‘cross’ marker; DNG = 4.5, computed at $\sigma = 6.6$), and the maxima point where the CNG feature is maximal (blue ‘plus’ marker; CNG = 31.8, computed at $\sigma = 31.7$). The green ‘dot’ marker shows the location of another lower maxima point of the CNG, computed at the same scale as the other (blue) maxima point

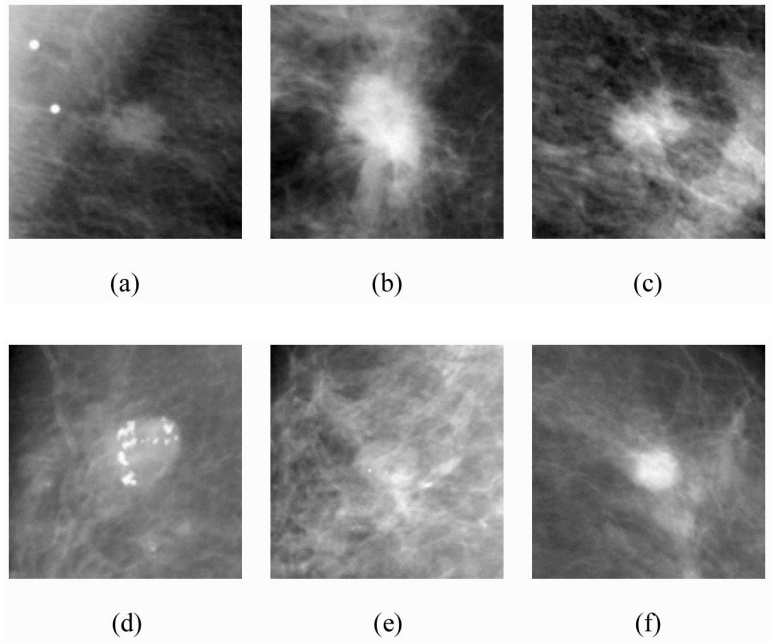

Fig. 9.

Examples of (a) first malignant mass (DNG = 10.9; CNG = 39.8), (b) second malignant mass (DNG = 4.6; CNG = 9.5), (c) third malignant mass (DNG = 4.4; CNG = 5.3), (d) first benign mass (DNG = 12.6; CNG = 10.2), (e) second benign mass (DNG = 1.6; CNG = 9.9), (f) third benign mass (DNG = 0; CNG = 8.8)

Table 1

Summary of newly-computed CAD image features according to feature type, number and description

Feature type	Feature number	Description
Shape	1-11	Eccentricity, equivalent diameter, extent, convex area, major axis length, minor axis length, orientation, solidity, shape factor ratio, ratio of major to minor axis length, modified compactness
Fat	12-15	Size (pixel number), size factor ratio (size/mass area), region number, average distance to the mass center (average distance/mean radial length of mass region)
Presence of calcifications	16-18	Size (pixel number), size factor ratio (size/mass area), region number
Texture	19-44	4 gray level co-occurrence matrix based features, 22 average and maximum values of gray level run length based texture features (computed along 0, 45, 90, and 135 degrees)
Spiculation	45-64	Features computed on the maxima points and on the whole image of the divergence of the normalized gradient and the curl of the normalized gradient of the segmented mass regions
Contrast	65-124	Contrast-based features (previously defined in [11,32,10]) computed for different-sized regions and locations of the segmented mass and background
Isodensity	125-154	Isodensity-based features (previously defined in [11]) computed for different-sized regions and locations of the segmented mass and background

Table 2

List of local image features previously proposed in [19,32,61]

No.	Feature
1	Region conspicuity
2	Region size
3	Region contrast
4	Average local noise (pixel value fluctuation)
5	Standard deviation of local noise
6	Skewness of local noise
7	Mean radial length / region size
8	Standard deviation of radial lengths
9	Skewness of radial lengths
10	Ratio between the maximum and minimal radial length
11	Shape factor ratio (perimeter of boundary^2 / region size)
12	Region circularity
13	Standard deviation of pixel values inside growth region
14	Skewness of pixel values inside growth region
15	Kurtosis of pixel values inside growth region
16	Average value of gradient of boundary contour pixels
17	Standard deviation of gradient of boundary contour pixels
18	Skewness of gradient of boundary contour pixels
19	Standard deviation of pixel values in the surrounding background
20	Skewness of pixel values in the surrounding background
21	Kurtosis of pixel values in the surrounding background
22	Average noise of pixel values in the surrounding background
23	Standard deviation of pixel value noise in the surrounding background
24	Skewness of pixel value noise in the surrounding background
25	Ratio of number of “minimum” pixels inside growth region
26	Average pixel value depth of “minimum” pixels inside growth region
27	Center position shift (distance between gravity center and local minimum) / region size

Table 3

A confusion matrix of prediction results using the fast and accurate SFFS feature selection algorithm, and obtained by applying a threshold of 0.5 on the classifier-generated probability scores

Prediction ->	**Benign images**	**Malignant images**
Benign images	440	160
Malignant images	147	453

Table 4

Features selected by the SFFS-based feature selection algorithm in the ten-fold cross-validation experiments. The 181 proposed features can be divided into 8 feature groups or types listed in the far-left column. The fraction of the features represented in each group is shown in the middle column. The last column displays the average percentages of the features selected by the SFFS-based algorithm along with their standard deviation intervals

Feature type	Fraction	Average percentage and std. dev. intervals
Shape	11/181	60.0±15.0%
Fat	4/181	40.0±24.2%
Presence of calcifications	3/181	6.7±14.1%
Texture	26/181	18.8±6.4%
Spiculation	20/181	14.0±7.0%
Contrast	60/181	7.7±3.9%
Previously-computed morphological features	27/181	17.4±4.3%
Isodensity	30/181	22.0±5.5%

Table 5

Performance level (AUC values) and database description of several previously reported studies applying CAD schemes for mammographic mass classification

CAD system	Year	AUC	Comments on number of masses and on reported AUC values
Varela et al. [10]	2006	0.81	1076 mass ROIs trained and tested using a leave-one-case-out method
Shi et al. [12]	2008	0.84	329 ROIs (132 benign and 197 malignant) evaluated as an independent test set
Retico et al. [16]	2006	0.80	226 massive lesions (117 benign and 109 malignant) trained and tested using 5 replications of the two-fold cross-validation method [76]
Mudigonda et al. [53]	2000	0.85; 0.67 (jack-knife classification)	54 images (28 benign and 26 malignant) from a local database and the MIAS database
Land [77]	2004	0.78	1979 cases (994 benign and 985 malignant) trained and tested using a five-fold cross-validation method
Huo [23]	2000	0.82 or 0.81	110 cases (50 malignant and 60 benign) evaluated as an independent test set. AUC value depends on the dataset and on whether the hybrid or ANN classifier was used
Heidt [78]	2009	0.60	206 ROIs (92 benign and 114 malignant) trained and tested using a ten-fold cross-validation method
Eltonsy [79]	2007	0.80	350 cases (82 normal and 268 malignant) trained and tested using a five-fold cross-validation method