
ORIGINAL ARTICLE

Identification of architectural distortions in mammograms using local binary patterns and radial lengths through an exhaustive evaluation framework

Sevastianos E. Chatzistergos¹ | Ioannis I. Andreadis¹ |
Konstantina S. Nikita¹

¹School of Electrical and Computer Engineering, National Technical University of Athens, Attiki, 15772, Greece

Correspondence

Sevastianos E. Chatzistergos, School of Electrical and Computer Engineering, National Technical University of Athens, Attiki, 15772, Greece
Email: schatzist@biosim.ntua.gr

Present address

[†]School of Electrical and Computer Engineering, National Technical University of Athens, Attiki, 15772, Greece

Funding information

A method based on the combination of Local Binary Pattern (*LBP*) operator and radial lengths (*RL*) is presented aiming at the identification of Architectural Distortions (*ADs*) in mammograms. *LBP* operator, a number of its variants and *RL* are combined together producing a high-dimensional feature space. A process, based on the combination of Principal Component Analysis (*PCA*) and *t* – *test*, is used to effectively transform feature space and reveal the most descriptive features. The classification step is performed using a Support Vector Machine (*SVM*) classifier. Open access databases (Mammographic Image Analysis Society-*MIAS* and Digital Database for Screening Mammography-*DDSM*) are used through an exhaustive evaluation framework that aims at eliminating both mammogram selection bias and limited subtlety variation, thus enabling a fair and complete comparison procedure. Furthermore, in order to provide a testbed for future comparisons, a dataset is constructed from all the available *AD* Regions Of Interest (*ROIs*) in *DDSM* (163 *AD* vs 375 *ROIs* from specific normal cases) and is used to further evaluate the performance of the pro-

conflicts of interest: none

posed method. The method performed flawlessly and classified correctly all cases.

KEYWORDS

Local Binary Pattern (*LBP*), Principal Component Analysis (*PCA*), t-test, texture classification, architectural distortion (*AD*), mammogram

1 | INTRODUCTION

Mammography is the main imaging technique for the detection and diagnosis of breast cancer, however about 10% of all cancerous lesions are missed by radiologists due to false interpretation of mammographic images (Sundaram et al., 2011). Reading mammograms is known to be a very demanding job for radiologists since judgments depend on training, experience, and subjective criteria (Cheng et al., 2006).

Architectural distortion (*AD*) is defined as a focal defect of fibroglandular distribution which results in a modification of the architecture of breast parenchyma (radiating-out spicules) without being accompanied by an increased density or mass (Lavoue et al., 2016). Unfortunately, due to its subtlety and variability in presentation, it is often missed during screening (Burrell et al., 1996). Even modern imaging techniques, like digital mammography, have failed to alter this fact. More specifically, a recent study (Suleiman et al., 2016) has shown that despite the availability of improved post-processing tools for digital mammography, *AD* detection remains a difficult task for readers.

The basic characteristic of *ADs* that most methods try to exploit is the existence of radiating spicules. A method based on line extraction has been presented (Nemoto et al., 2009), where a likelihood of *AD* association is based on density and geometric measures on the lines extracted. Furthermore, a number of methods based on the analysis of oriented structures, derived mainly through Gabor filters, have been reported (Karssemeijer and te Brake, 1996; Ayres and Rangayyan, 2007; Banik et al., 2013; Rangayyan et al., 2013). More specifically, approaches based on the combination of orientation fields with phase portraits (Ayres and Rangayyan, 2007), (Banik et al., 2013) or the deviation of oriented structures from the expected orientation of breast tissue (Rangayyan et al., 2013) have been proposed.

However, methods based on multi-scale analysis suffer from the fact that *AD* spicules have variable size and contrast. Additionally, phase portrait analysis often fails since spicules usually present an incomplete "star" shape (Biswas and Mukherjee, 2011). As a result, different approaches have been proposed that ignore the specific shape properties of *ADs* and focus mainly on texture analysis. One such approach has been based on fractal dimension (Tourassi et al., 2006; Beheshti et al., 2016; Guo et al., 2009). In another case, a number of texture characteristics have been defined using a set of multi-scale oriented filter bank responses. Gaussian Mixture Models have then been used in order to define a set of texture primitives (textons) (Biswas and Mukherjee, 2011). Recently, a method to discriminate between various breast lesion types based on local energy-based shape histograms has been presented (Wajid and Hussain, 2015). The authors reported 100% accuracy in discriminating *ADs* (22 *AD* cases) from other lesion types but no evaluation was presented against normal *ROI*s. A different approach is based on the empirical mode decomposition of the image followed by a model-based approach (Zyout and Togneri, 2015). The method uses a large number of image *ROI*s from *DDSM* database but performance analysis is limited only to the Area Under the receiver operating characteristic Curve (*AUC*). Although, appropriate performance metrics are reported in most studies, their direct comparison is not feasible, since, in the majority of cases, either restricted access datasets are used or, even in the case of public databases, the exact image subset considered is not adequately defined. Thus, the reported performance is

highly dependent on the dataset used.

The methods described so far, can be categorized as either performing *AD* detection (Tourassi et al., 2006; Guo et al., 2009; Wajid and Hussain, 2015) or *AD* identification (Karssemeijer and te Brake, 1996; Nemoto et al., 2009; Ayres and Rangayyan, 2007; Banik et al., 2013; Rangayyan et al., 2013; Biswas and Mukherjee, 2011; Beheshti et al., 2016). In the detection case, methods receive whole mammograms as inputs and attempt to define the existence of *ADs*. In the identification case, methods receive various *ROI*s as input and try to categorize them either as normal tissue or *AD*. In the current work, we attempt to perform *AD* identification.

Local Binary Pattern (*LBP*) has become quite popular lately as one very efficient texture operator (Ojala et al., 2002). The main idea of *LBP* is to represent local patterns by calculating the difference of a central image pixel from its surrounding pixels in a small neighborhood around it. *LBP*, despite its simplicity, has been proven to be very effective in describing local image patterns, and has achieved impressive classification results on representative texture databases (Hong et al., 2014). *LBP* has recently been used for mass detection on mammographic images. Texture features from gray level co-occurrence matrix (*GLCM*) and a variation of *LBP* called completed local binary pattern (*CLBP*) along with geometry features have been combined in order to discriminate masses from normal breast tissue (Guo et al., 2010). Furthermore, non-uniform patterns, that in normal *LBP* procedures are ignored, have been used towards the discrimination between benign and malignant mass lesions (Liu and Zeng, 2015).

To the best of authors' knowledge, *LBP* has not been previously used for the identification of *ADs* in mammographic images. Moreover, in an earlier work, we have highlighted the potential of radial lengths (*RL*) to enhance the contrast on mammograms, achieving promising results (Chatzistergos et al., 2014). Motivated by the simplicity and efficacy of *LBP* in texture analysis and the ability of *RL* to reveal diagnostically critical information, we propose the combination of these two methodologies, in the process of discrimination between *AD* and normal tissue or other lesion types (masses, calcifications). Furthermore, in an attempt to reduce the dimensionality of the feature set, a feature selection process based on the combination of Principal Component Analysis (*PCA*) and *t* – *test* is proposed. The final classification step is performed using Support Vector Machines (*SVM*).

In order to enable fair comparisons with the results reported in previous studies, as well as with future studies, we exploit two well established open access mammogram databases (*MIAS*, *DDSM*) and we propose a complete evaluation framework based on the creation of multiple, randomly assembled datasets. Thus, we attempt to avoid image selection bias and limited subtlety variation, which are inherent in most published studies reporting results based on fixed, manually selected datasets. We also provide performance metrics on a dataset constructed from all the available *AD* Regions of Interest (*ROI*s) in the *DDSM* (163 *AD* vs 375 *ROI*s from specific normal cases), which may be used as a representative testbed for the task of *AD* identification, given its large size.

The rest of this work is organized as follows: Section 2 briefly reviews *LBP* and its variations. The classification process and the proposed evaluation framework are presented in Section 3, followed by information regarding the construction of a benchmarking dataset from public databases in Section 4. In Section 5, the classification performance of the proposed method is evaluated and discussed.

2 | LBP BASED SCHEMES

2.1 | Classical LBP

Given an image I and a point g_c at (x_c, y_c) , *LBP* is computed by first subtracting $I(x_c, y_c)$ from the gray values of image points in a narrow neighborhood around the original point g_c . If the result of the subtraction is greater or equal to zero, the certain neighborhood point is assigned the value 1, otherwise 0 is assigned. The assigned values from all

neighborhood locations are then summed up

$$LBP_{P,R} = \sum_{p=0}^{P-1} s(g_p - g_c) \cdot 2^p \quad (1)$$

where, $s(x) = \begin{cases} 1, & x \geq 0 \\ 0, & x < 0 \end{cases}$, g_c is the gray value of the central pixel (x_c, y_c) , g_p is the gray value of its neighbors and P is the total number of involved neighbors.

Another important parameter for LBP is the size of the local neighborhood which is defined through the radius R . Given that a central pixel is located at (x_c, y_c) , the coordinates of g_p are $(x_c + R \cdot \cos(2 \cdot \pi \cdot p/P), y_c + R \cdot \sin(2 \cdot \pi \cdot p/P))$. If the coordinates are different from the exact pixel locations, interpolation is used to infer the gray value of the particular points.

LBP as defined in (1) is not rotation invariant. To make it rotation invariant the number of spatial transitions (bitwise changes 1/0) of LBP needs to be defined

$$U(LBP_{P,R}) = |s(g_{P-1} - g_c) - s(g_0 - g_c)| + \sum_{p=1}^{P-1} |s(g_p - g_c) - s(g_{p-1} - g_c)|. \quad (2)$$

LBP patterns whose U value is 0 or 2 ($U \leq 2$) are called uniform, since they have limited transitions or else discontinuities in their binary representation. By definition, exactly $P + 1$ uniform binary patterns can occur in a circularly symmetric neighborhood set of P pixels. LBP can now be expressed in a rotation invariant way ($LBP_{P,R}^{riu2}$) if only uniform patterns are considered

$$LBP_{P,R}^{riu2} = \begin{cases} \sum_{p=0}^{P-1} s(g_p - g_c), & \text{if } U(LBP_{P,R}) \leq 2 \\ P + 1, & \text{otherwise} \end{cases}. \quad (3)$$

$LBP_{P,R}^{riu2}$, which has $P + 2$ distinct output values can now be used to create an image histogram to represent image texture

$$H(k) = \sum_{i=1}^H \sum_{j=1}^W f(LBP_{P,R}^{riu2}(i, j), k), k \in [0, P + 1],$$

$$f(x, y) = \begin{cases} 1, & x = y \\ 0, & \text{otherwise} \end{cases} \quad (4)$$

where (H,W) refer to image height and width, respectively.

2.2 | Completed LBP

As presented in the previous section, *LBP* is exclusively based on the sign of the calculated differences, discarding all information found in the magnitude of the differences. However, this has the disadvantage that *LBP* is unaware of how “strong” or else important a local structure is. To address this issue and integrate magnitude information into classic *LBP* a modification of *LBP* called Completed *LBP* (*CLBP*) is proposed (Guo et al., 2010).

In *CLBP* a local region is represented by its central pixel gray value and a local difference sign-magnitude transform (*LDSMT*)

$$d_p = g_p - g_c = s_p \cdot m_p, \quad (5)$$

$$s_p = \text{sign}(d_p), \quad (6)$$

$$m_p = |d_p|. \quad (7)$$

LDSMT decomposes the image local structure into two complementary components, difference signs and difference magnitudes, which are then coded in a binary format using the following operators:

$$CLBP_S_{P,R} = \sum_{p=0}^{P-1} t(s_p, 0) \cdot 2^p, \quad (8)$$

$$CLBP_M_{P,R} = \sum_{p=0}^{P-1} t(m_p, c) \cdot 2^p, \quad (9)$$

$$t(x, c) = \begin{cases} 1, & x \geq c \\ 0, & x < c \end{cases}. \quad (10)$$

where c is a threshold that is defined adaptively and usually set equal to the mean value of m_p from the whole image.

The central pixel is simply coded by a binary code after global thresholding

$$CLBP_C_{P,R} = t(g_c, c_I) \quad (11)$$

where c_I is the mean gray level of the whole image.

2.3 | Thresholded CLBP (TCLBP)

Another disadvantage of *LBP* is that it tends to be sensitive to noise, particularly in near-uniform image regions and smooth weak illumination gradients. To address exactly this issue, a work presented in (Tan and Triggs, 2010) incorporated a threshold around the value of the central pixel introducing the notion of thresholded *LBP*.

Specifically, given a local neighborhood defined around a central pixel (x_c, y_c) with gray value g_c , a thresholded version of *CLBP* can be defined using a parameter E_c which is either added (12) or subtracted (13) from the central pixel value (g_c)

$$g_c^+ = g_c + E_c, \quad (12)$$

$$g_c^- = g_c - E_c. \quad (13)$$

Using the new values for the central pixel, (5), (8), (9) are transformed to

$$d_p^+ = g_p - g_c^+ = s_p^+ \cdot m_p^+, \quad (14)$$

$$TCLBP_{S_{p,R}^+} = \sum_{p=0}^{P-1} t(s_p^+, 0) \cdot 2^p, \quad (15)$$

$$TCLBP_{M_{p,R}^+} = \sum_{p=0}^{P-1} t(m_p^+, c) \cdot 2^p, \quad (16)$$

$$d_p^- = g_p - g_c^- = s_p^- \cdot m_p^-, \quad (17)$$

$$TCLBP_{S_{p,R}^-} = \sum_{p=0}^{P-1} t(s_p^-, 0) \cdot 2^p, \quad (18)$$

$$TCLBP_{M_{p,R}^-} = \sum_{p=0}^{P-1} t(m_p^-, c) \cdot 2^p. \quad (19)$$

For $E_c = 0$, *TCLBP* becomes identical to *CLBP*.

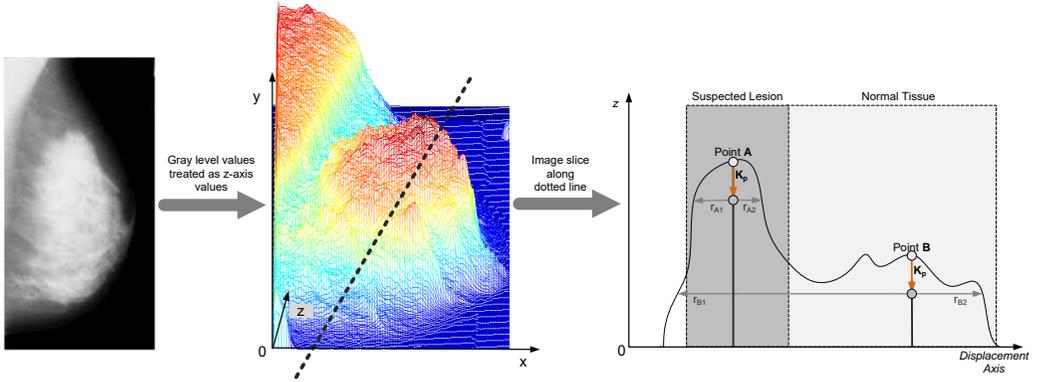


FIGURE 1 Original image (left), 3D image representation after treating gray level values as z-axis values (center), demonstration of Radial Lengths Method on an image slice (right). Radial lines start propagating at each image point after a value K_p is subtracted from the image gray value at the corresponding point. Propagation will stop if the image boundaries are met. Points located at a suspected lesion region with high gray level values and close to strong edges (Point A) will present smaller radial length values (r_{A1}, r_{A2}). On the other hand, points at low intensity regions away from strong edges (Point B), which are commonly considered as normal tissue regions, will present larger radial length values (r_{B1}, r_{B2}).

Two ways to define E_c are evaluated. At first, all differences between g_c and g_p are calculated ($DIFF_i$) for the entire image ROI and then E_c^{mean} and E_c^{std} are defined using (20) and (21) respectively

$$E_c^{mean} = \frac{1}{T_N} \sum_i |DIFF_i|, \quad (20)$$

$$E_c^{std} = \sqrt{\frac{1}{T_N} \sum_i (DIFF_i - E_c^{mean})^2}, \quad (21)$$

where T_N is the total number of differences for the particular image ROI .

2.4 | Radial Lengths and CLBP

In our previous work (Chatzistergos et al., 2014) we demonstrated the ability to reveal diagnostically critical information in mammograms by improving the local contrast, using a series of equally rotated radials that propagate until they meet a certain limit. The basic idea is to treat gray level values at each image point (x_0, y_0) as values in the z-axis (z_0) and create this way a 3D surface from the original image (Fig.1). Given the surface point (x_0, y_0, z_0) , a new point $(x_0, y_0, z_0 - K_p)$ is defined by subtracting a certain value K_p from z_0 . From this point, a fixed number of lines can start propagate until they meet the surface limit. The lines must be equally rotated between each other and remain parallel to the surface specified by axes x, y . The line lengths produced this way, comprise the radial lengths of image I at point (x_0, y_0) . The method described, is termed Radial Lengths Method (RLM).

In this work, we investigate the potential of *RLM* to reveal strong local edges, which are of great importance in most image classification tasks. More specifically, diagnostically critical information usually lies in high intensity image regions (e.g. masses, calcifications) and therefore these regions are expected to contain a large number of strong edges (Fig.1). On the other hand, regions with normal tissue are expected to present low intensity values and contain large number of weak edges. The radial lengths from an image point located at a high intensity region and close to a strong edge (Fig.1, Point A) will present lower values (r_{A1}, r_{A2}) as compared to the radial length values (r_{B1}, r_{B2}) from an image point (Fig.1, Point B) located at a low intensity region away from strong edges.

At each image point a total number of Q equally rotated radial lengths are defined. Furthermore, parameter K_P is defined using the maximum gray level value of each image using

$$K_P = a \cdot \max(I). \quad (22)$$

Parameter a is user defined and takes values in $(0, 1)$. Smaller values of a force radial lengths to propagate closer to the image surface and therefore small gray level fluctuations have strong impact on the radial lengths. In other words, smaller values of a introduce noise, while larger values reduce the sensitivity to important local edges. It was visually determined that clinically important information is better revealed when a values lie between 0.05 and 0.15. Therefore, in the current work a is set equal to $\{0.05, 0.10, 0.15\}$.

Once radial lengths (r_q , where $q \in [0, Q - 1]$) at each image point (x, y) are calculated, the radial differences (R_q) are determined

$$R_q = r_q - c, \quad (23)$$

where

$$c = \frac{1}{H \cdot W \cdot Q} \sum_{x=1}^H \sum_{y=1}^W \sum_{q=0}^{Q-1} r_q(x, y). \quad (24)$$

This subtraction of c from r_q is performed as an equivalent to the subtraction of g_c from g_p used in classical *LBP*.

Furthermore, based on the notion of *CLBP* equations (5), (8), (9) are transformed to

$$R_q = s_q \cdot m_q, \quad (25)$$

$$RL-LBP_{S_{Q,a}} = \sum_{q=0}^{Q-1} t(s_q, 0) \cdot 2^q, \quad (26)$$

$$RL-LBP_{M_{Q,a}} = \sum_{q=0}^{Q-1} t(m_q, c_M) \cdot 2^q, \quad (27)$$

where

$$c_M = \frac{1}{H \cdot W \cdot Q} \sum_{x=1}^H \sum_{y=1}^W \sum_{q=0}^{Q-1} R_q(x, y). \quad (28)$$

A large number of radial lengths is desired so as not to omit useful information from the image. However, as Q becomes larger, so does the computational cost. Q values, multiples of 2, between 4 and 32 have been evaluated. Preliminary results have indicated small correlation between Q value and performance, for $Q \geq 8$. In the following steps, Q is set equal to 16, so as to ensure good performance without overly increasing computational cost.

One of the reasons why LBP has been proven to be so effective in characterizing various texture types is the fact that by imposing the use of uniform patterns it actually distinguishes between strong and weak local edges, rejecting the latter as noise. This characteristic of LBP is what we want to further enhance by defining $CLBP$ on r_q instead of the actual image values. The ability of RLM to reveal important local edges is demonstrated in Fig. 2 (e) to (h), where the mean radial length value, at each image point, is provided. One can notice that the various lesions are becoming more obvious. Fig. 2 (i) to (l) and (m) to (p) present respectively, the $LBP_{S_{8,1}}$ and $RL-LBP_{S_{16,0.05}}$ histograms of the corresponding lesion types. It can be seen that both methods provide differentiating histograms for different lesion types, supporting the hypothesis that the combination of those two methods can lead to improved classification performance. More specifically and as far as $RL-LBP_{S_{16,0.05}}$ is concerned, it can be seen that in Fig. 2 (r) at bin 9, ARCH, CALC and the pair CIRC, NORM have clearly separable values. Furthermore, at bins 1 and 2, CIRC clearly differs from the rest.

2.5 | Feature histograms assembly

For each image ROI a number of binary codes using the operators determined so far, are defined. In all cases the rotation invariant version of the operators is used. For the sign and magnitude components, the relating binary codes are defined using a series of variations. More specifically, seven thresholds ($E^{Thres} = +E_c^{std}, -E_c^{std}, +E_c^{mean}, -E_c^{mean}, 0$) as well as two $\{P, R\}$ pairs ($\{P, R\} = \{1, 8\}, \{2, 16\}, \{3, 24\}$) are used. The computations are performed on average values of block subregions rather than individual pixels (Jia et al., 2013), with block sizes (W_L) ranging from 1 to 20 pixels ($W_L = 0, 5, 10, 15, 20$). Furthermore, sign and magnitude components are calculated for the radial lengths defined for three a parameters ($a = 0.05, 0.10, 0.15$). Following the notion of $CLBP$, a binary code for the central grey level is also defined. The produced codes are then used to produce one dimensional histograms or combined together to produce two dimensional (2D) or three dimensional (3D) histograms. The multidimensional histograms are created by using different binary code sequences at each dimension. These multidimensional histograms are then converted back to 1D by rearranging their values. This conversion is performed for the sake of simplicity and uniformity in the classification step. The overview of the histogram (feature vector) creation process is given in Fig. 3.

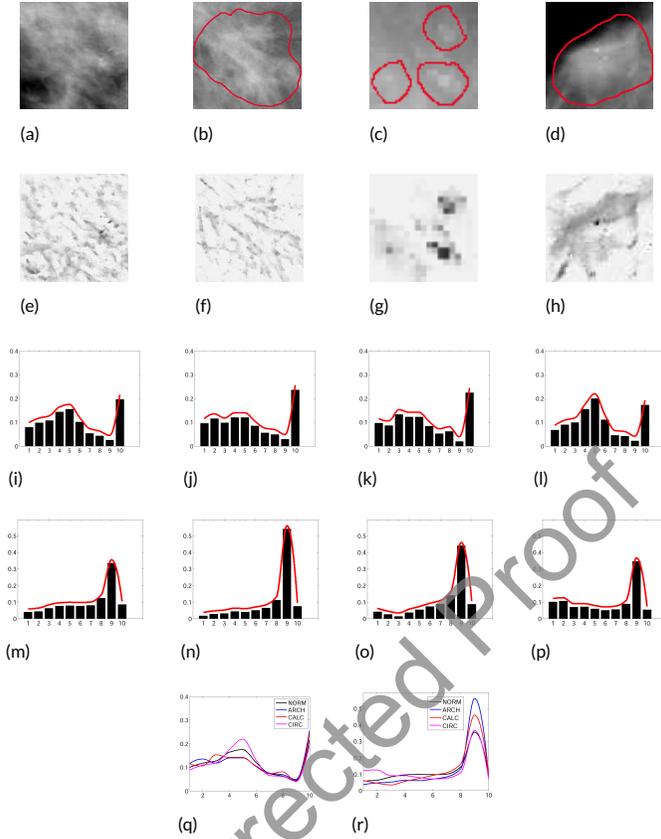


FIGURE 2 Various image $ROIs$ ((a) to (d)) received from $MIAS$ database, along with the mean value of radial lengths at each ROI point ((e) to (h)). Plots at (i) to (l) and (m) to (p) demonstrate LBP_{S_8} and $RL - LBP_{S_{16,0.05}}$ histograms, respectively. An envelop line is included for each histogram to better highlight its fluctuations. A comparison of the various envelop lines for LBP_{S_8} and $RL - LBP_{S_{16,0.05}}$ histograms is given at (q) and (r) respectively. $ROIs$ include: (a) normal tissue (NORM), (b) AD (ARCH), (c) calcifications (CALC) and (d) circumscribed mass (CIRC).

3 | CLASSIFICATION

3.1 | Feature selection

The number of the produced histograms for each image ROI is very high (1320 histograms), therefore there is a need to reduce the dimensionality of the feature set and reveal features with strong discrimination ability. To this end, a framework based on the combination of PCA and $t - test$ is used (Fig. 4). From the original set of $ROIs$, a specific dataset is randomly assembled. Then, based on the 10-fold cross validation procedure a training and a testing set is properly defined. Test and training sets in this stage (Fig. 4.A) consist of 1320 histograms each of which contains from 10 until up to 1,000 bins.

Next, PCA on both training and testing sets is used to transform histograms into linearly uncorrelated variables called principal components (Martinez and Kak, 2001). Each component is associated with a value V_p that expresses

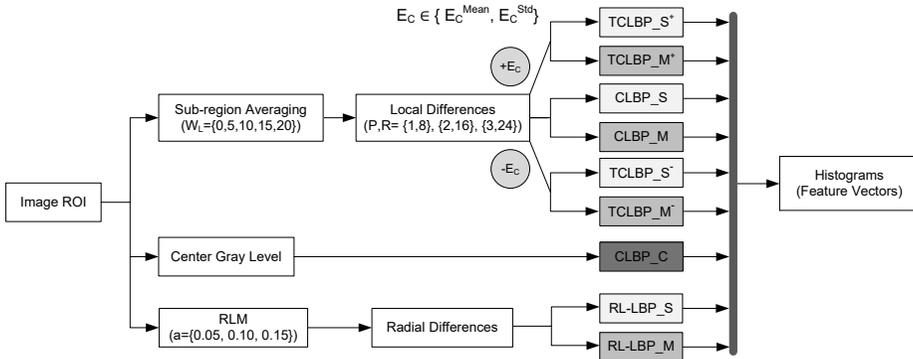


FIGURE 3 Overview of feature histograms creation process. Feature histograms are created following routes A,B and C. Route A: Subregion averaging for different window lengths (W_L) is followed by Local Differences calculation for different pairs of neighborhood size (R) and number of involving neighbors (P). Either a positive (+) or negative (-) threshold E_c is then used and results are given as inputs to appropriate sign ($_S$) or magnitude ($_M$) $CLBP$ or thresholded $CLBP$ ($TCLBP$) operators. Route B: Center gray level is given as input to $CLBP_C$ operator. Route C: Radial length method (RLM) for different a values is performed, followed by Radial Differences calculation and given as input to appropriate sign ($_S$) or magnitude ($_M$) $RL-LBP$ operator

the percentage of the total variance for the particular component. Components with $V_p < 1$ are rejected. Principal components from all different histograms of a single case (ROI) are aligned together and form a new feature vector (Fig. 4.B). Feature vectors from the training set are then used to perform a statistical test ($t-test$). The purpose is to define the features for which there is significant statistical difference between cases that belong to different categories (like AD or normal). All other features from training and testing sets are rejected (Fig. 4.C). The statistical test used is the two-sided $t-test$ with significance level 5%. In order to further reduce the number of features, PCA is used once more. Again, components with $V_p < 1$ are rejected. The resulting features from the original training and testing sets (Fig. 4.D) are used for the final classification step.

3.2 | Evaluation framework

In order to perform the final classification step, a SVM classifier with linear kernel is used. This classifier is selected mainly due to the fact that it is a well-established and effective classifier, used successfully in many classification tasks (Wei et al., 2005; Liu and Tang, 2014; Andreadis et al., 2011). The feature selection framework, described so far, effectively defines a reduced number of features but experimentation showed that there is strong relation between the actual number of features used and the classification performance.

Therefore, a procedure based on 10-fold cross validation and misclassification rate is implemented in order to define the number of features that maximize classification performance. Specifically, from each training set, a certain number of features is considered and the set is segmented into ten partitions, nine of which are used to train the classifier. The trained classifier is used to classify the remaining partition. This step is repeated ten times in total, until all partitions are used for testing. By calculating the percentage of misclassified cases for each test segment the misclassification rate for the particular number of features is determined. Fig. 5 presents an example plot of misclassification rate in relation to the number of features used. The final number of features to be used is the one that minimizes the misclassification

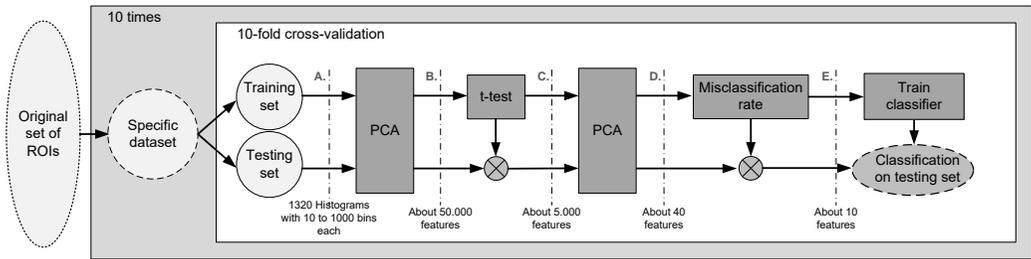


FIGURE 4 Feature selection and evaluation framework. From the original set of *ROIs* a specific dataset is defined and separated into training and testing sets in order to perform 10-fold-validation (point A). Principal component analysis (*PCA*) is then performed on the entire set (point B), followed by *t - test* on just the train set. After this step all statistical insignificant features from both training and testing set are rejected (point C). *PCA* is performed again (point D) and followed by misclassification rate process on train set to define the optimal features number (point E). These features are then used to train the classifier and perform the final classification step on the testing set.

rate (Fig. 4.E). The order of features is not considered since they are actually sorted in descending order, based on the percentage of the total variance defined in the lastly performed *PCA*.

4 | DATASETS

In order to enable proper comparisons with the results reported in previous studies of the literature, the data selection process used in (Biswas and Mukherjee, 2011) is utilized. More specifically, four datasets (*M1*, *M2*, *SM1*, *SM2*) are constructed from *MIAS* Suckling et al. (1994) database and two (*H*, *L*) from *DDSM* (Heath et al., 2001). Additionally, a larger dataset (*F*) is constructed and proposed, including all the available *AD ROIs* from *DDSM* database, in an attempt to provide a representative testbed for the task of *AD* identification

4.1 | MIAS

The first dataset from *MIAS* (*M1*) consists of 19 *AD ROIs* and 45 normal *ROIs*, the second (*M2*) consists of 19 *AD ROIs* and 28 *ROIs* with other abnormalities like calcifications and circumscribed masses. The third dataset (*SM1*) contains 18 spiculated masses with 45 normal *ROIs* and the fourth (*SM2*) contains 18 spiculated *ROIs* and 28 *ROIs* with other abnormalities. In images with lesions, *ROIs* are defined using the smallest bounding rectangle to the supplied ground truth, while in normal cases, *ROIs* are randomly selected. More specifically, on each case a point belonging to the actual breast region is randomly selected and used as the *ROI*'s center, while an edge is randomly selected from a pool consisted of all *AD ROI* edges. If the defined rectangle region is found to contain less than 90% of breast tissue it is rejected and the procedure is repeated. The reason we use edge values from the pool of lesion *ROI* edges is to avoid any possible bias of either fixed or completely random edge values.

Table 1 shows the distribution of the various lesion types in the *MIAS* database. It can be seen that there are 209 normal cases of which 45 need to be selected, furthermore there are 46 cases with calcifications or circumscribed masses from which 28 need to be selected. In order to avoid a possible bias in the case images are manually selected, we randomly select the required number of images from each lesion category. Therefore, all available *ROIs* are used to

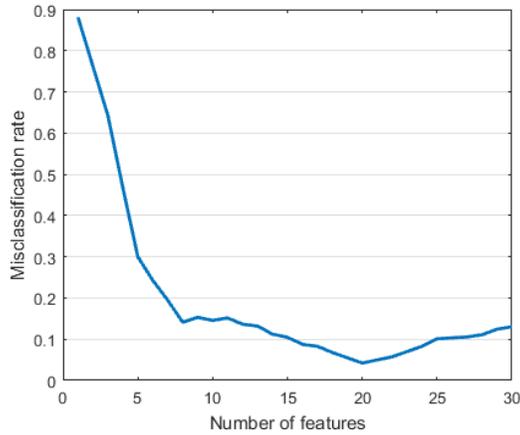


FIGURE 5 Misclassification rate (max. 1) in relation to the number of features used.

TABLE 1 MIAS lesion distribution.

Lesion type	Total number of cases
Circumscribed masses	25
Spiculated masses	19
Calcification	21
Architectural distortion	19
Normal	209
Other	29

from the *ROI* set and a special, random based, procedure is used to assemble each specific dataset. It should be noted that special care is taken for the selected images to be evenly distributed in the various density categories.

4.2 | DDSM

The first dataset from *DDSM* (*H*) consists of 40 *AD ROI*s and 40 normal *ROI*s from mammograms digitized using the *Howtek* scanner while *L* dataset consists of 40 *AD ROI*s and 40 normal *ROI*s from mammograms digitized using the *Lumisys* scanner (Biswas and Mukherjee, 2011). Given the large number of available images in *DDSM*, an approach similar to the one described for *MIAS* datasets is used. More specifically, the required images for each dataset are randomly selected from a larger dataset, consisted of 54 *AD ROI*s from *MLO* views, along with 80 normal cases digitized using *Howtek* scanner and 80 normal cases digitized using *Lumisys*. *DDSM* database provides breast density categorization into 4 categories, based on Breast Imaging Reporting and Data System (*BI-RADS*) (Sickles et al., 2013). For the normal cases, special care is taken once more so as to ensure that equal number of images belong to the two lower density categories and the two higher ones. More specifically since 40 cases from each scanner have to be selected, 3 cases are selected from breast density category 1, 17 cases from category 2, 17 cases form category 3 and 3

TABLE 2 Dataset composition

Dataset	Regions of interest (ROIs)
M1	19 Architectural distortions vs 45 Normal
M2	19 Architectural distortions vs 28 Other
SM1	18 Spiculated masses vs 45 Normal
SM2	18 Spiculated masses vs 28 Other
H	40 Architectural distortions vs 40 Normal (Howtek)
L	40 Architectural distortions vs 40 Normal (Lumisys)
F	163 Architectural distortions vs 375 Normal

from category 4. We select more images from density ratings 2 and 3 since these are the most common ratings.

Moreover, a larger dataset is constructed and used in an attempt to further evaluate the performance of the proposed method. The dataset (F) consists of all *AD ROIs* available in *DDSM* and thus, any bias caused by manually selecting cases is removed. Both benign (29 cases) and malignant (134 cases) *AD ROIs* are included, along with 375 normal *ROIs* from *MLO* and *CC* images in *DDSM*. A total number of 211 normal *ROIs* digitized using Howtek scanner (subfolders normal-7 and normal-8), and 164 normal *ROIs* digitized using Lumisys scanner (normal-9 and normal-10) have been randomly selected from the corresponding *DDSM* subfolders.

The various datasets are shown in Table 2.

5 | RESULTS AND DISCUSSION

The performance of the proposed method is evaluated following the process presented in Section 3.2 for each of the considered datasets enumerated in Table 2. The classification process is repeated for 10 times for each different dataset. In each repetition the dataset is randomly assembled using different images from the original *ROI* set. The classification process on *F* dataset is performed just once due to the fact that it contains all *DDSM AD ROIs* and *ROIs* from all normal cases in certain *DDSM* sub folders.

Table 3 presents evaluation results (accuracy) when *CLBP* is considered (*D*), when *CLBP* is combined with thresholding and region averaging (*DE*) and when *DE* is further combined with radial lengths (*DRLE*).

It can be seen that for the majority of datasets, *DRLE* feature set performs better than *DE* and *DE* performs better than *D*. The only exceptions are *M2* and *SM2* datasets where the performance between the three feature sets is almost the same. It should be noted though, that the specific datasets contain the smallest number of cases, 47 and 46 respectively.

A performance comparison of the proposed method with other state-of-the-art methods is provided in Table 4. For the proposed method, the performance metrics are actually averaged values since the whole classification process is performed ten times. The mean values are accompanied by their relating standard deviations. Furthermore, the metrics for the repetition with the highest accuracy are also provided.

It can be seen that the averaged metric values for the proposed method are generally inferior to the ones presented in the literature. The only exception is *SM1*. If the best performing single repetition is considered, the proposed method clearly outperforms comparing methods for *SM1* and can be considered almost similar, though inferior, for *M1* and *H* datasets. We believe that the low classification accuracy obtained is mainly due to the statistical nature of *PCA* and

TABLE 3 Performance (mean accuracy %) on S1, S2, SM1, SM2, H and L datasets for *DRLE*, *DE* and *D* feature sets.

Dataset	Feature Set	Accuracy	Sensitivity	Specificity	Area Under the Curve (AUC)
M1	<i>DRLE</i>	72.55	54.14	80.32	0.67
	<i>DE</i>	68.97	49.63	77.14	0.63
	<i>D</i>	66.97	6.77	92.38	0.50
M2	<i>DRLE</i>	64.58	61.19	66.81	0.64
	<i>DE</i>	64.32	59.21	67.67	0.63
	<i>D</i>	53.13	42.11	60.35	0.51
SM1	<i>DRLE</i>	79.37	48.15	91.85	0.70
	<i>DE</i>	75.4	43.52	88.15	0.66
	<i>D</i>	70.11	17.59	91.11	0.54
SM2	<i>DRLE</i>	51.37	42.06	57.14	0.50
	<i>DE</i>	50.15	38.89	57.14	0.48
	<i>D</i>	50.15	43.65	54.19	0.49
H	<i>DRLE</i>	76.04	76.25	75.83	0.76
	<i>DE</i>	75.21	72.92	77.5	0.75
	<i>D</i>	65.83	72.08	59.58	0.66
L	<i>DRLE</i>	73.04	71.43	74.64	0.73
	<i>DE</i>	70.36	68.57	72.14	0.70
	<i>D</i>	68.75	68.93	68.57	0.69

D: only *CLBP* considered.

DE: *CLBP* combined with thresholding and region averaging.

DRLE: *DE* combined with radial lengths.

t – *test*. It seems that these procedures fail to correctly estimate the statistical properties of feature distributions when limited samples are used, influencing strongly the overall classification performance.

If *F* is considered, the performance for *D* is almost perfect while for *DE* and *DRLE* all cases are correctly classified (Table 5). The increased performance when *F* dataset is used can be attributed to the large number of cases within the set, which allows *PCA* and *t* – *test* based feature selection process to perform properly.

6 | CONCLUSIONS

A method to perform AD and other mammogram lesions identification has been presented. The method is based on a number of modifications of the classical *LBP* and their combination with radial lengths. *PCA* and *t* – *test* are adopted to perform feature selection while the final lesion classification is performed using *SVM*.

Special care has been taken to evaluate the proposed method's performance in terms of metrics directly comparable with other studies, which usually report performance on either private databases or on arbitrarily selected data subsets

TABLE 4 Comparison of proposed approach with state-of-the-art

Dataset	Performance Metric	proposed (averaged value $\mu \pm \sigma$)	proposed (single repetition with higher accuracy)	method A	method B
M1	Accuracy (%)	72.54 \pm 4.22	76.56	81.60	83.60
	Sensitivity (%)	54.14 \pm 8.44	52.63	84.20	82.40
	Specificity (%)	80.32 \pm 5.04	86.67	79.10	71.90
	AUC (%)	0.67 \pm 0.05	0.70	0.83 ^a	0.77 ^a
M2	Accuracy (%)	64.58 \pm 4.45	68.75	82.50	76.40
	Sensitivity (%)	61.18 \pm 7.41	57.89	85.50	76.40
	Specificity (%)	66.81 \pm 5.51	75.86	81.00	76.50
	AUC (%)	0.64 \pm 0.05	0.67	0.83 ^a	0.77 ^a
SM1	Accuracy (%)	79.37 \pm 3.01	82.54	76.30	-
	Sensitivity (%)	48.15 \pm 7.59	50	78.20	-
	Specificity (%)	91.85 \pm 5	95.56	74.30	-
	AUC (%)	0.70 \pm 0.03	0.73	-	-
SM2	Accuracy (%)	51.37 \pm 8.89	61.7	74.40	-
	Sensitivity (%)	42.06 \pm 13.17	61.11	79.00	-
	Specificity (%)	57.14 \pm 8.17	62.07	72.10	-
	AUC (%)	0.50 \pm 0.09	0.62	-	-
H	Accuracy (%)	76.04 \pm 5.27	82.5	86.50	-
	Sensitivity (%)	76.25 \pm 5.65	80	87.40	-
	Specificity (%)	75.83 \pm 9.83	85	84.10	-
	AUC (%)	0.76 \pm 0.05	0.83	0.87 ^b	-
L	Accuracy (%)	73.04 \pm 3.45	76.25	88.30	-
	Sensitivity (%)	71.43 \pm 10.49	77.50	89.20	-
	Specificity (%)	74.64 \pm 10.15	75	86.70	-
	AUC (%)	0.73 \pm 0.03	0.76	0.87 ^b	-

^a averaged M1 and M2.

^b averaged H and L.

method A, proposed by Biswas and Mukherjee (2011) ; method B, proposed by Ayres and Rangayyan (2005)

of public databases. To this end: (i) the proposed method has been evaluated using appropriate datasets from open access mammogram databases (*MIAS*, *DDSM*), (ii) a complete evaluation framework has been introduced, where the datasets are assembled multiple times in a random fashion with the final performance measure defined as the averaged performance on each single dataset, and (iii) further evaluation of the proposed method has been carried out using a large dataset consisted of all *AD* cases available in *DDSM* (163 *AD* vs 375 normal cases), which may provide a proper baseline enabling future comparisons to the current results.

Rigorous evaluation of the proposed pipeline following the aforementioned framework has shown that *LBP* variants and *RL*, when properly combined, can effectively detect *ADs*. Moreover, the combination *PCA* and *t - test* is proven quite capable to successfully transform and identify the most descriptive features especially when adequately large datasets (*F*) are considered. When the smaller datasets (*M1*, *M2*, *SM1*, *SM2*, *H*, *L*) are considered, the performance is deteriorated.

TABLE 5 Performance for dataset *F*

Feature Set	Accuracy (%)	Sensitivity (%)	Specificity (%)	Area Under the Curve (AUC)
<i>DRLE</i>	100.00	100.00	100.00	1.00
<i>DE</i>	100.00	100.00	100.00	1.00
<i>D</i>	99.26	99.40	99.20	0.99

D: only *CLBP* considered.

DE: *CLBP* combined with thresholding and region averaging.

DRLE: *DE* combined with radial lengths.

REFERENCES

- Andreadis, I. I., Spyrou, G. M. and Nikita, K. S. (2011) A comparative study of image features for classification of breast microcalcifications. *Measurement Science and Technology*, **22**, 114005. URL: <http://stacks.iop.org/0957-0233/22/i=11/a=114005>.
- Ayres, F. and Rangayyan, R. (2007) Reduction of false positives in the detection of architectural distortion in mammograms by using a geometrically constrained phase portrait model. *International Journal of Computer Assisted Radiology and Surgery*, **1**, 361–369. URL: <http://dx.doi.org/10.1007/s11548-007-0072-x>.
- Ayres, F. and Rangayyan, R. M. (2005) Characterization of architectural distortion in mammograms. *Engineering in Medicine and Biology Magazine, IEEE*, **24**, 59–67.
- Banik, S., Rangayyan, R. and Desautels, J. (2013) Measures of angular spread and entropy for the detection of architectural distortion in prior mammograms. *International Journal of Computer Assisted Radiology and Surgery*, **8**, 121–134. URL: <http://dx.doi.org/10.1007/s11548-012-0681-x>.
- Beheshti, S., Noubari, H. A., Fatemizadeh, E. and Khalili, M. (2016) Classification of abnormalities in mammograms by new asymmetric fractal features. *Biocybernetics and Biomedical Engineering*, **36**, 56–65. URL: <http://www.sciencedirect.com/science/article/pii/S0208521615000583>.
- Biswas, S. and Mukherjee, D. (2011) Recognizing architectural distortion in mammogram: A multiscale texture modeling approach with GMM. *Biomedical Engineering, IEEE Transactions on*, **58**, 2023–2030.
- Burrell, H. C., Sibbering, D. M., Wilson, A. R., Pinder, S. E., Evans, A. J., Yeoman, L. J., Elston, C. W., Ellis, I. O., Blamey, R. W. and Robertson, J. F. (1996) Screening interval breast cancers: mammographic features and prognosis factors. *Radiology*, **199**, 811–817. URL: <http://dx.doi.org/10.1148/radiology.199.3.8638010>. PMID: 8638010.
- Chatzistergos, S., Andreadis, I. and Nikita, K. (2014) Tensor radial lengths for mammographic image enhancement. In *XIII Mediterranean Conference on Medical and Biological Engineering and Computing 2013* (ed. L. M. Roa Romero), vol. 41 of *IFMBE Proceedings*, 305–308. Springer International Publishing. URL: http://dx.doi.org/10.1007/978-3-319-00846-2_76.
- Cheng, H. D., Shi, X. J., Min, R., Hu, L. M., Cai, X. P. and Du, H. N. (2006) Approaches for automated detection and classification of masses in mammograms. *Pattern Recognition*, **39**, 646–668. URL: <http://www.sciencedirect.com/science/article/pii/S0031320305002955>.
- Guo, Q., Shao, J. and Ruiz, V. (2009) Characterization and classification of tumor lesions using computerized fractal-based texture analysis and support vector machines in digital mammograms. *International Journal of Computer Assisted Radiology and Surgery*, **4**, 11–25. URL: <http://dx.doi.org/10.1007/s11548-008-0276-8>.
- Guo, Z., L., Z. and Zhang, D. (2010) A completed modeling of local binary pattern operator for texture classification. *Image Processing, IEEE Transactions on*, **19**, 1657–1663.
- Heath, M., Bowyer, K., Kopand, D., Heath, R. M. M., Moore, K. B. R. and Kegelmeyer, W. (2001) The digital database for screening mammography. In *5th IWDM*, 212–218. Yaffe M. Medical Physics Publishing.

- Hong, X., Zhao, G., Pietikainen, M. and Chen, X. (2014) Combining lbp difference and feature correlation for texture description. *Image Processing, IEEE Transactions on*, **23**, 2557–2568.
- Jia, X., Yang, X., Zang, Y., Zhang, N., Dai, R., Tian, J. and Zhao, J. (2013) Multi-scale block local ternary patterns for fingerprints vitality detection. In *Biometrics (ICB), 2013 International Conference on*, 1–6.
- Karssemeijer, N. and te Brake, G. (1996) Detection of stellate distortions in mammograms. *Medical Imaging, IEEE Transactions on*, **15**, 611–619.
- Lavoue, V., Fritel, X., Antoine, M., Beltjens, F., Bendifallah, S., Boisserie-Lacroix, M., Boulanger, L., Canlorbe, G., Catteau-Jonard, S., Chabbert-Buffet, N., Chamming's, F., Chereau, E., Chopier, J., Coutant, C., Demetz, J., Guilhen, N., Fauvet, R., Kerdraon, O., Laas, E., Legendre, G., Mathelin, C., Nadeau, C., Naggara, I. T., Ngo, C., Ouldamer, L., Rafii, A., Roedlich, M.-N., Seror, J., Seror, J.-Y., Touboul, C., Uzan, C. and Darai, E. (2016) Clinical practice guidelines from the french college of gynecologists and obstetricians (cngof): benign breast tumors ??? short text. *European Journal of Obstetrics & Gynecology and Reproductive Biology*, **200**, 16 – 23. URL: <http://www.sciencedirect.com/science/article/pii/S0301211516300537>.
- Liu, X. and Tang, J. (2014) Mass classification in mammograms using selected geometry and texture features, and a new svm-based feature selection method. *IEEE Systems Journal*, **8**, 910–920.
- Liu, X. and Zeng, Z. (2015) A new automatic mass detection method for breast cancer with false positive reduction. *Neurocomputing*, **152**, 388 – 402. URL: <http://www.sciencedirect.com/science/article/pii/S0925231214014003>.
- Martinez, A. M. and Kak, A. C. (2001) Pca versus lda. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **23**, 228–233.
- Nemoto, M., Honmura, S., Shimizu, A., Furukawa, D., Kobatake, H. and Nawano, S. (2009) A pilot study of architectural distortion detection in mammograms based on characteristics of line shadows. *International Journal of Computer Assisted Radiology and Surgery*, **4**, 27–36. URL: <http://dx.doi.org/10.1007/s11548-008-0267-9>.
- Ojala, T., Pietikainen, M. and Maenpaa, T. (2002) Multiresolution, gray-scale and rotation invariant texture classification with local binary patterns. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, **24**, 971–987.
- Rangayyan, R., Banik, S., Chakraborty, J., Mukhopadhyay, S. and Desautels, J. (2013) Measures of divergence of oriented patterns for the detection of architectural distortion in prior mammograms. *International Journal of Computer Assisted Radiology and Surgery*, **8**, 527–545. URL: <http://dx.doi.org/10.1007/s11548-012-0793-3>.
- Sickles, E., D'Orsi, C. and Bassett, L. (2013) *ACR BI-RADS? Mammography*. In: *ACR BI-RADS? Atlas, Breast Imaging Reporting and Data System*. Reston, VA, American College of Radiology.
- Suckling, J., Parker, J., Dance, D., Astley, S., Hutt, I. and Boggis, C. (1994) The mammographic images analysis society digital mammogram database. *Exerpta Med.*, **1069**, 375–378.
- Suleiman, W., McEntee, M., Lewis, S., Rawashdeh, M., Georgian-Smith, D., Heard, R., Tapia, K. and Brennan, P. (2016) In the digital era, architectural distortion remains a challenging radiological task. *Clinical Radiology*, **71**, e35 – e40. URL: <http://www.sciencedirect.com/science/article/pii/S0009926015003967>.
- Sundaram, M., Ramar, K., Arumugam, N. and Prabin, G. (2011) Histogram modified local contrast enhancement for mammogram images. *Applied Soft Computing*, **11**, 5809–5816. URL: <http://www.sciencedirect.com/science/article/pii/S1568494611001621>.
- Tan, X. and Triggs, B. (2010) Enhanced local texture feature sets for face recognition under difficult lighting conditions. *Image Processing, IEEE Transactions on*, **19**, 1635–1650.
- Tourassi, G. D., Delong, D. M. and Jr, C. E. F. (2006) A study on the computerized fractal analysis of architectural distortion in screening mammograms. *Physics in Medicine and Biology*, **51**, 1299. URL: <http://stacks.iop.org/0031-9155/51/i=5/a=018>.

- Wajid, S. K. and Hussain, A. (2015) Local energy-based shape histogram feature extraction technique for breast cancer diagnosis. *Expert Systems with Applications*, **42**, 6990 – 6999. URL: <http://www.sciencedirect.com/science/article/pii/S0957417415002997>.
- Wei, L., Yang, Y., Nishikawa, R. M. and Jiang, Y. (2005) A study on several machine-learning methods for classification of malignant and benign clustered microcalcifications. *IEEE Transactions on Medical Imaging*, **24**, 371–380.
- Zyout, I. and Togneri, R. (2015) Empirical mode decomposition of digital mammograms for the statistical based characterization of architectural distortion. In *2015 37th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, 109–112.

SEVASTIANOS CHATZISTERGOS received the diploma in Electrical Engineering (2004) from the Aristotle University of Thessaloniki, Greece and the M.Sc. degree in Biomedical Engineering (2008) from the School of Medicine of University of Patras, Greece and the National Technical University of Athens (NTUA). He is currently pursuing the Ph.D. degree in medical imaging from the NTUA. His main research interests include image processing, pattern recognition and machine learning.

IOANNIS ANDREADIS received the degree in electrical and computer engineering from the NTUA in 2006 and his PhD degree in Biomedical Engineering from the same institution in 2014. Since 2007, he has been with the Biomedical Simulation and Imaging Laboratory, Department of Electrical and Computer Engineering, NTUA where he has worked as senior or post-doc researcher in research projects. He has contributed to a number of papers in journals, book chapters, international and national conferences and he has given two invited talks. He serves as reviewer in 11 international scientific journals and 6 international conferences and participated in the organizing and scientific committees of 6 international conferences. His main research interests include biomedical image processing and analysis, medical decision support systems, machine learning and computer-aided clinical diagnosis.

KONSTANTINA NIKITA received the diploma in Electrical Engineering (1986) and the PhD degree (1990) from the NTUA, Greece. She then received the MD degree (1993) from the Medical School, University of Athens, Greece. Since 1990, she has been working as a researcher at the Institute of Communication and Computer Systems. In 1996, she joined the School of Electrical and Computer Engineering, NTUA, as an assistant professor, and since 2005, she serves as a professor at the same school. Her current research interests include biomedical signal and image processing and analysis, biomedical informatics, simulation of physiological systems, medical imaging, biological effects, and medical applications of radiofrequency electromagnetic fields. Dr. Nikita has authored or coauthored 154 papers in refereed international journals, 38 chapters in books, and over 300 papers in international conference proceedings. She has authored or edited two books (*Simulation of Physiological Systems and Medical Imaging Systems*) in Greek and five books in English published by Springer and Wiley. She holds two patents. She has been the technical manager of several European and National R&D projects. She is an associate editor of the

IEEE Transactions on Biomedical Engineering, the *IEEE* Journal of Biomedical and Health Informatics, Wiley Bioelectromagnetics and the Journal of Medical and Biological Engineering and Computing and a guest editor of several international journals. Dr. Nikita has received various honors or awards, among which, the Bodossakis Foundation Academic Prize for exceptional achievements in Theory and Applications of Information Technology in Medicine (2003).

2 dataset Composition

Uncorrected Proof