IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received February 26, 2020, accepted March 6, 2020, date of publication March 13, 2020, date of current version March 25, 2020.

Digital Object Identifier 10.1109/ACCESS.2020.2980616

An Automated Breast Micro-Calcification Detection and Classification Technique Using Temporal Subtraction of Mammograms

KOSMIA LOIZIDOU[1], (Student Member, IEEE), GALATEIA SKOUROUMOUNI[2], CHRISTOS NIKOLAOU[3], AND COSTAS PITRIS[1], (Member, IEEE)

[1]Department of Electrical and Computer Engineering, KIOS Research and Innovation Center of Excellence, University of Cyprus, 1678 Nicosia, Cyprus
[2]Radiology Department, Nicosia General Hospital, 2029 Nicosia, Cyprus
[3]Radiology Department, Limassol General Hospital, 4131 Limassol, Cyprus

Corresponding author: Kosmia Loizidou (cloizi01@ucy.ac.cy)

This work was funded by the European Union's Horizon 2020 research and innovation program under grant agreement No. 739551 (KIOS CoE) and from the Republic of Cyprus through the Directorate General for European Programs, Coordination and Development.

ABSTRACT Radiologists worldwide use mammography as a reliable tool for breast cancer screening. However, mammography assessment is challenging even for well-trained radiologists, leading to a pressing need for Computer Aided Diagnosis (CAD) systems. In this work, a novel technique for the detection and classification of breast Micro-Calcifications (MCs), which are diagnostically significant but difficult to detect findings, is presented. The proposed method is based on the subtraction of temporally sequential mammogram pairs, after pre-processing and image registration, followed by machine-learning. The classification was performed using several features extracted from the subtracted mammograms and selected during training to optimize the accuracy of the results. Six classifiers were tested in a leave-one-patient-out, 4, 5 and 10 fold cross-validation process. This technique was evaluated on a unique dataset, consisting of temporal sequences of mammograms from 80 patients taken between 1 to 6 years apart. The resulting 320 mammograms were reviewed by 2 radiologists who precisely marked each MC location. The accuracy of classifying MCs as benign or suspicious improved from 91.42% without temporal subtraction and an Ensemble of Decision Trees (EDT), to 99.55% with the use of sequential mammograms and Support Vector Machines (SVMs) with leave-one-patient-out validation. The improvement was statistically significant (p-value < 0.005). These results verify the accuracy and the effectiveness of the proposed technique should to be further evaluated on a larger dataset.

INDEX TERMS Breast cancer, computer-aided diagnosis, digital mammography, micro-calcifications, temporal subtraction.

I. INTRODUCTION

Breast cancer is the most common cancer and the second leading cause of death in women in the United States [1]. Carcinogenesis in the breast leads to an uncontrolled growth of cells, usually forming a tumor. However, the process, by which the cancer initially appears and later develops, can vary between patients. For the detection of breast cancer, radiologists use mammography as the key screening tool [2]. After the mammograms are acquired, expert radiologists review the images to determine whether the patient has any signs of malignancy, followed by appropriate disease management in the case of positive findings. Current protocols require reading of the mammogram by two radiologists (and a third, if consensus is not reached), which is an indication of the challenges faced when attempting to identify probable abnormalities in a mammogram. Appearance of Micro-Calcifications (MCs), which are microscopic deposits of calcium that commonly appear in the breast, is one of the suspicious signs that require further investigation. Most are benign and do not require any intervention. Nonetheless, Micro-Calcification Clusters (MCCs) might indicate an increased chance to develop breast cancer. The morphology of the calcifications is the most crucial parameter in their

The associate editor coordinating the review of this manuscript and approving it for publication was Yongming Li.

TABLE 1. Age of participants included in the dataset.

	Normal Mammograms *(n=40)*	Suspicious Mammograms *(n=40)*	Total
Mean ± STD	62.12 ± 5	63.48 ± 5.2	62.8 ± 5.2
Median	61.5	64.5	62
Range	54-72	55-75	54-75

classification as benign or suspicious [3]. One of the challenges of most Computer Aided Diagnosis (CAD) systems, is the accurate differentiation between benign and suspicious MCs.

MCs appear as bright spots in the mammogram due to the higher X-ray attenuation coefficient of calcium compared to normal tissue. The contrast between MCs and the surrounding tissue varies from low, in dense, to high, in fatty mammograms [4]. Benign MCs are usually larger in size, rounder in shape, smaller in number and have homogeneous sizes and shapes. MCs suspicious for early cancer are clustered, smaller with irregular shapes and sizes and branching in orientation. The appearance of such MCCs, is an important sign for early breast cancer [5], [6].

There are various studies in the literature proposing different approaches for the detection of MCs [7]–[12]. A major disadvantage of these approaches is that they do not compare the recent with the older mammograms, something radiologists routinely do, and are thus unable to detect temporal changes.

The comparison of prior and recent mammograms can convey significant information regarding the development and diagnostic significance of abnormalities. When a possible abnormality in the most recent image matches the location and appearance of a corresponding region in the prior mammogram, that area, which has remained unchanged, is most likely not suspicious. Conversely, if the anomaly is new and, thus, changing rapidly, then that finding warrants further investigation [13]. Prior studies have assessed the use of prior mammograms to identify mass lesions [14]–[19]. In addition, some studies have compared the performance of a CAD system for mass lesion detection, with and without using prior mammograms [20]–[22]. These datasets contained mammograms from two consecutive screening rounds. Expert radiologists assessed the mammograms, identified the abnormalities (mass lesions), in both the recent and prior view, and then single-mammogram and temporal features were extracted for the classification. The overall results confirmed that the temporal information enhanced the algorithm's accuracy and identified subtle signs of malignancy that otherwise might have been overlooked if the prior mammogram was not available. Studies using temporal subtraction, rather than comparison, are not available in the literature.

In this work, an algorithm for the automated detection and classification of MCs, using sequential mammograms, is presented. The introduction of temporal subtraction of images from two screening rounds increases the contrast, removes old MCs that did not change since the previous examination,

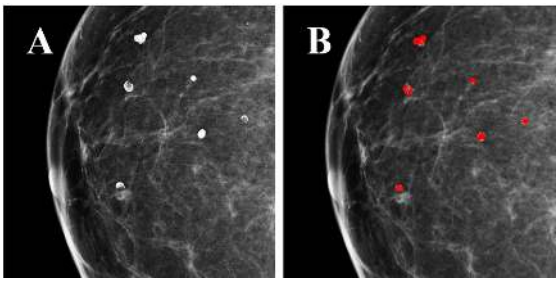

FIGURE 1. Dataset example of a fatty mammogram. (A) Zoomed region of a mammogram with MCs. (B) The same image with precise marking (red dots) of annotated MC locations.

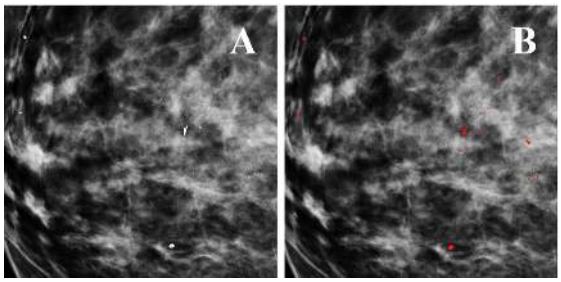

FIGURE 2. Dataset example of a dense mammogram. (A) Zoomed region of a mammogram with MCs. (B) The same image with precise marking (red dots) of annotated MC locations.

eliminates many False Positives (FPs) and improves the classification of the newly developed MCs [23]. The performance of this method was evaluated on a dataset containing temporal sequences of mammograms from 80 patients, for a total of 320 images (two time points and two views of each breast). Of those, 40 patients had at least one suspicious MC or MCC only in the recent mammogram. The remaining 40 had only benign MCs in both recent and prior mammograms.

The rest of the paper is organized as follows: Section II describes the dataset (II-A) and the proposed procedure with (II-B) and without (II-C) temporal subtraction (only on the most recent mammograms). Section III describes the experiments and results with Sections III-A and III-B containing the results with and without temporal subtraction, respectively. Section IV includes a discussion of the findings and the main conclusions are given in Section V.

II. MATERIALS AND METHODS

This section describes, in detail, the creation of the dataset and the proposed methodology.

A. DATASET

For the purposes of this study, a new dataset had to be created since in publicly available databases only one round of mammograms is available for a single patient. For that reason, 80 pairs of full-field digital sequential mammograms were collected from 80 women following their routine screening mammography examinations. For every subject, two mammogaphic views were collected: the CranioCaudal (CC, view from above) and MedioLateral Oblique (MLO, angled view),

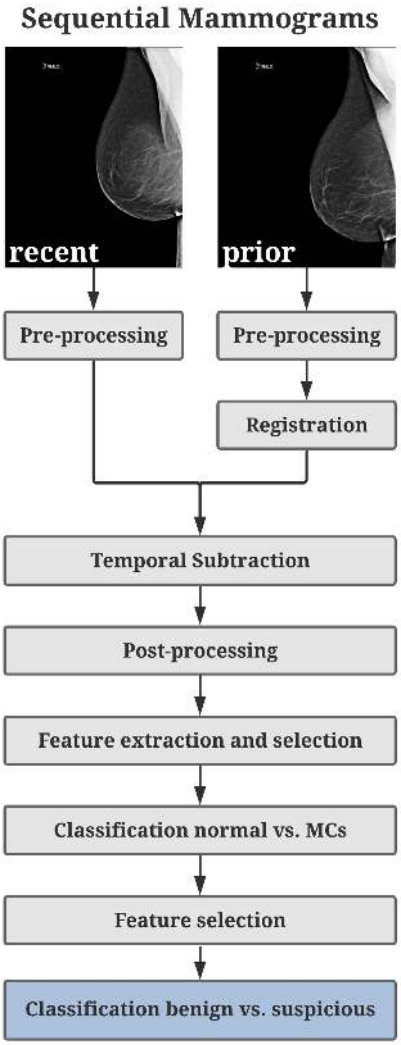

FIGURE 3. Diagram of the proposed methodology for breast MC detection and classification using temporal sequences of digital mammograms.

for two sequential screening rounds (a total of 320 mammograms). The images were collected from hospitals across Cyprus. A radiologist, with nine years of experience, selected and annotated the mammograms and a second radiologist (two years of experience), validated the images. Consecutive mammograms were collected from 2012 to 2018, with an average of 1 to 6 years interval. The age of the participants varied from 54 to 75, with a mean age of 62.8±5.2 years and median age of 62. Half the cases belonged to healthy subjects who did not present suspicious MCCs in either the prior or the recent mammogram. In the remaining 40 pairs, at least one suspicious MC or MCC was present in the most recent mammogram (Table 1). This dataset not only contained temporally sequential mammograms, but also included precise annotation of each individual MC (suspicious as well as benign) to be used as a reference, something not found in publicly available datasets (Fig. 1 & 2). The dimensions of the mammograms were 4096 × 3328 pixels, in an 12-bit DICOM format.

B. MC DETECTION AND CLASSIFICATION USING TEMPORAL SUBTRACTION OF MAMMOGRAMS

For temporal subtraction, both recent and prior mammograms were processed in parallel. Figure 3, shows the proposed methodology for the detection and classification of MCs.

1) PRE-PROCESSING, REGISTRATION, TEMPORAL SUBTRACTION AND POST-PROCESSING

The pre-processing began with the normalization of all mammograms that adjusted the range of pixel intensity values. Then, border removal was applied, to suppress the light structures connected to the border of the image using the following function:

$$F(x,y) = \begin{cases} I(x,y), & \text{if } (x,y) \text{ is on the border of } I. \\ 0, & \text{otherwise.} \end{cases} \quad (1)$$

where, I is the original image and F the border areas. The image was first converted to binary using Otsu's method to get the threshold. Given the dominating intensities of the pectoral muscle and its adjacency to the image border, that region could be isolated by removing F from the original image [24], [25]. This is not be confused with removing bright areas from the periphery of the breast which will be explained later. Contrast adjustment with gamma correction was also applied to account for the non-linear mapping of intensities in the images as described by:

$$T(l) = l_{max}\left(\frac{l}{l_{max}}\right)^{\gamma} \quad (2)$$

where, l_{max} is the maximum intensity value of the input image and γ is the Gamma parameter that maps the pixels toward the darker output values [26].

Subsequently, the most recent and prior mammograms were registered. For an effective subtraction between sequential mammograms, alignment is crucially important. Registration compensates for variations in breast compression, changes in the shape, differences in the amount of the pectoral muscle present in the MLO view and human error [27]. Several registration methods have been proposed [28]. For this study, Demons [29] registration was used. Affine registration was also considered, since it is commonly used in the literature [30], but Demons performed considerably better. To evaluate the performance of the registration, the residual was measured as the sum of the remaining pixels after the subtraction. In all cases, the prior mammogram was registered to the recent one.

Demons registration, is a technique that transforms image pixels locally having an unlike transformation reliant on their regional similarity and location. In contrast to the global methods, local techniques can handle more complicated deformations. Furthermore, this algorithm depends on seeing the registration as a diffusion process influenced by optical flow formulation and sometimes includes regularization to assure smoothness and continuity [29]. Demons registration can be represented as the energy function with respect to the

update field u, using fixed image F, moving image M and a transformation field s, as follows:

$$E^s_{corr}(u) = ||F - M\,o\,(s+u)||^2 + \left(\frac{\sigma_i^2}{\sigma_x^2}\right)||u||^2 \quad (3)$$

where, σ_i^2 is the noise on the image intensity and σ_x^2 the spatial uncertainty. Using Taylor expansion, (3) can be linearized and the energy function will reach its minimum when gradient descent is zero. The registration must be solved iteratively, as the update field is based on local information [31].

Once the images were registered, temporal subtraction was performed. The registered mammogram (prior) was subtracted from the recent one thus removing regions that have not changed since the prior screening. To determine the effectiveness of the subtraction, the Contrast Ratio (CR) of the subtracted image was measured and compared to the CR of the most recent mammogram without subtraction. The ability of subtraction to increase the image contrast was evaluated. In addition, the effectiveness of the subtraction to remove MCs that have not changed since the prior screening, an important step in reducing FPs, was also assessed. To eliminate high intensity regions that might result in areas falsely identified as MCs, high intensity areas on the periphery of the breast were removed. This area corresponds to the skin of the breast, which can not contain MCs, but is not completely removed by registration and subtraction due to misalignment. The skin regions are not removed by border removal since they are not adjacent to the image border. However, these areas can be automatically removed by detecting the breast periphery from a binary image of the entire breast and subsequently removing all the high intensity regions that fall on the periphery.

2) FEATURE EXTRACTION AND SELECTION

A range filter, a local pixel-based filter where every output pixel consists the range value between the maximum and the minimum of a 3-by-3 neighbourhood for every equivalent pixel, was applied [32]. The sub-range was computed by ordering the N pixels inside the window based on the intensity,

$$(f_1, f_2, \ldots, f_N) \quad (4)$$

where,

$$(f_1 \leq f_2 \leq \ldots \leq f_N) \quad (5)$$

and then subtracting the intensity values for two selected positions (i and j) within this list such as that;

$$range(j, i) = f_j - f_i \quad 1 \leq i < j \leq N \quad (6)$$

Next, the image was converted to binary by thresholding to remove low intensity areas. The threshold value was set relatively low to remove only the areas that definitely did not belong to MCs. The value was selected by optimizing the global classification rate. Lastly, the binary image was morphologically processed. The first operation was opening, with a radius of 1 pixel, which is smaller than the radius of a typical MC. Opening removed isolated pixels that did not correspond to MCs in the binary image. Then closing was applied, with a radius of 5 pixels, which connected all grouped pixels together in order to connect the constituents of the MCCs. The proposed algorithm considered the remaining regions as possible MCs.

TABLE 2. Details of extracted features for the detection and classification of MCs.

Name	Description
Shape features	
Area	number of pixels inside a region
Convex area	number of pixels in the convex image
Eccentricity	ratio of distance between the foci of the ellipse and its major axis length
Equivalent diameter	diameter of the circle with the same area as the region
Euler number	subtraction of the number of objects in the region and the number of holes in those objects
Extent	ratio of pixels inside the region to pixels in the bounding box
Filled area	number of pixels inside the filled image
Major axis length	length of the major axis of an ellipse that has same normalized central moment with the region
Minor axis length	length of the minor axis of an ellipse that has same normalized central moment with the region
Orientation	angle between the axis of the ellipse that has same second moments with the region
Perimeter	distance around the boundary of the region
Solidity	relative amount of pixels that appeared in both the convex hull and the region
Intensity-based features	
Minimum intensity	minimum of all the intensity values of the area
Maximum intensity	maximum of all the intensity values of the area
FOS features	
Average intensity	mean of all the intensity values of the area
Entropy	texture measurement
Kurtosis	correlated with probability distribution
Skewness	asymmetry measurement
Standard deviation	variation from the average value
Variance	how far are the pixels values from the average
GLCM features	
Contrast	amount of local variations present in the region
Correlation	measurement of gray tone linear-dependencies of the area
Energy	information measurement
Homogeneity	information about the distribution elements

As with most CAD algorithms, there was a large number of FPs at the detection stage [33]. To eliminate the FPs and identify the true MCs, machine-learning, with 24 features extracted from the subtracted image, was performed. Those features were selected for their capability to identify MCs based on their various characteristics.

Shape features: One important factor for the radiological diagnosis of MCs, is their morphology. Additionally, shape characteristics can differentiate benign and suspicious MCs [34]. In all, 12 shape-based features were extracted for each region (Table 2).

Intensity-based features: MCs appear as bright regions on the mammogram with higher intensity values than their

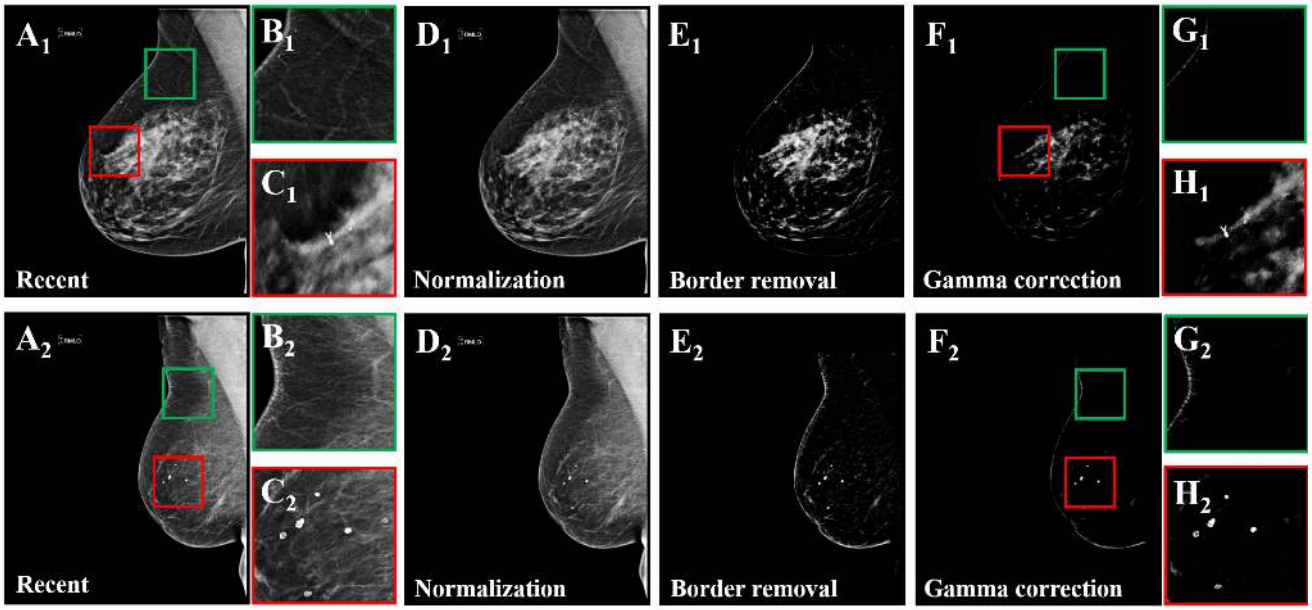

FIGURE 4. Effect of pre-processing on dense (1) and fatty (2) mammograms. (A) Most recent mammogram without any processing. (B) Zoomed region marked by the green square in A, showing an area without MCs. (C) Zoomed region marked by the red square in A, showing an area with benign MCs. (D) Recent image after normalization. (E) Recent image after border removal. (F) Final pre-processed image after gamma correction. (G) Zoomed region marked by the green square in F, showing the same area as (B), after the pre-processing. (C) Zoomed region marked by the red square in F, showing the same area as (C), after the pre-processing.

TABLE 3. Elimination of old MCs that appear in both screening rounds, in dense and fatty mammograms.

Mammograms	New MCs	Old MCs	Overlapping MCs	Overlap [%]	Removed MCs	Reduction [%]
Dense	204	182	44	24.18	13	21.57
Fatty	260	232	65	13	28.02	25
Total	**464**	**414**	**109**	**26.33**	**26**	**23.85**

surroundings. Therefore, intensity and pixel-based features can be useful for their characterization. Two intensity features were extracted (Table 2).

First-Order Statistics features (FOS): The texture of a region can provide important information regarding the distribution of the gray levels of the pixels [35]. In total, 6 FOS features were calculated for every region (Table 2).

Gray Level Co-occurrence Matrix (GLCM) features: GLCM features provide significant information concerning texture distribution. The GLCM was estimated from the image by measuring the pairwise statistics of pixel intensity [36]. Four features were evaluated for the classification procedure (Table 2).

Feature selection was necessary to remove insignificant and unnecessary features and increase the performance of the classifier. For an initial selection of statistically significant features, two algorithms were utilized: hypothesis test (t-test) [37] and multivariate analysis of variance (MANOVA) [38], to identify the features with the bigger contribution based on the training data. Using these approaches, the most statistically significant features ($p < 0.01$) were selected. The optimal combination of features was determined during the classification process by optimizing the accuracy of the results during training.

3) CLASSIFICATION

Different classifiers for the elimination of FPs and the classification of MCs as benign or suspicious, are proposed in the literature [7]. In this study, six classifiers were evaluated: Linear Discriminant Analysis (LDA) [39], k-Nearest Neighbor (k-NN) [2], Naive Bayes (NB) [40], Support Vector Machines (SVMs) [41], Decision Trees (DTs) [42] and Ensemble of Decision Trees (EDTs) [43].

For the training of the classification procedure, mammograms were used in a Leave-One-Patient-Out process (L-O-P-O). During each round, the mammograms from one patient (CC and MLO views of the most recent and prior mammograms) were used as the test sample and the remaining patients were used for the training and the selection of the best features by analyzing those that resulted from the previous feature selection method, until all the cases were classified. This approach is of critical importance to avoid any bias, from including information from the "unknown"

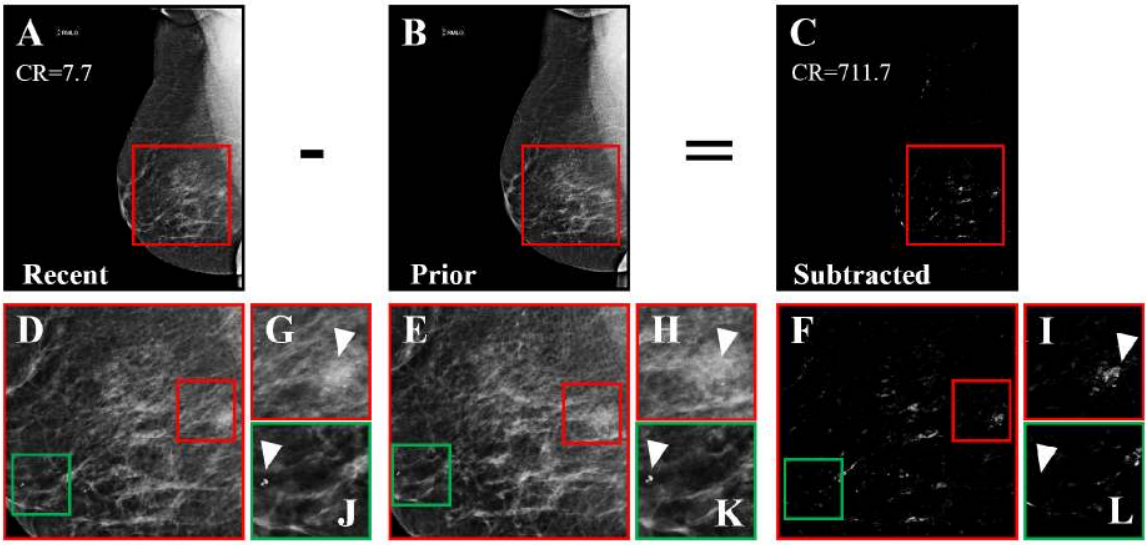

FIGURE 5. Improvement of the CR and subtraction of pre-existing MCs. (A) Most recent mammogram. (B) Prior mammogram. (C) The result of subtracting the registered version of B from A. (D)-(F) Zoomed regions marked by the red squares in A-C. In this example, the CR has increased 92 times after subtraction. The red squares in D-F correspond to the zoomed images (G)-(I) and show a new suspicious MC cluster, which was not subtracted. On the contrary, the green squares, which correspond to (J)-(L), show a pre-existing benign MC, which was completely subtracted.

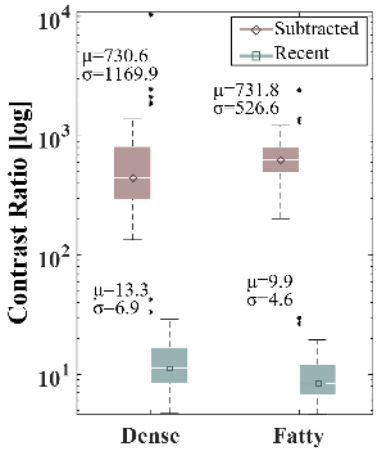

FIGURE 6. Plot comparing the contrast ratio (logarithmic scale) between the unprocessed recent image and the image created from the temporal subtraction, for dense and fatty mammograms.

patient, during the classification process. For comparison, k-fold cross-validation was also considered in order to examine the generalization of the algorithm. All the folds had the same number of patients, since the k-fold was created with the patients and not the regions. Initially, the classifiers were trained and tested for the elimination of FPs resulting from the erroneous identification of normal tissue as MCs in the previous steps. Subsequently, the classifiers were trained and tested to classify the true MCs as benign or suspicious.

The performance of each classification, with and without temporal subtraction, was evaluated by measuring sensitivity, specificity, accuracy and the Area Under the receiver operating characteristics Curve (AUC).

TABLE 4. Selected features for the classification of possible MCs as normal tissue or true MCs (1st round) and of the true MCs as benign or suspicious (2nd round), with and without temporal subtraction.

Features extracted	Selected for Temporal Subtraction		Selected for Recent Mammograms	
	1st round	2nd round	1st round	2nd round
Area		x		x
Convex area				x
Eccentricity		x		x
Equivalent diameter	x	x		x
Euler number	x		x	
Extent	x		x	
Filled area				x
Major axis length		x		x
Minor axis length	x	x	x	x
Orientation				
Perimeter		x		x
Solidity	x	x	x	x
Minimum intensity value				
Maximum intensity value		x		
Average intensity value				
Entropy	x	x		
Kurtosis	x	x		
Skewness	x			
Standard deviation			x	
Variance				
Contrast			x	x
Correlation			x	
Energy			x	
Homogeneity			x	

C. MC DETECTION AND CLASSIFICATION USING ONLY THE MOST RECENT MAMMOGRAMS

For comparison purposes, the detection and classification of MCs was also performed using only the most recent mammograms (without temporal subtraction). The goal was to confirm that despite the elimination of the overlapping MCs, temporal subtraction increases the diagnostic accuracy of the two classification rounds. For the results to be comparable, the same procedures as before were followed. Normalization, pre-processing, post-processing

TABLE 5. Comparison of the accuracy (AC), sensitivity (SE) and specificity (SP) of the algorithm using temporal subtraction and leave-one-patient-out, with and without feature selection. In all cases the selection of features resulted in statistically significant improvement at the performance with $p < 0.01$.

Classifier	normal vs. MCs		benign vs. suspicious	
	without feature selection	with feature selection	without feature selection	with feature selection
LDA	AC: 88.58 SE: 12.19 SP: 97.18	AC: 90.11 SE: 16.55 SP: 99.52	AC: 98.19 SE: 98.82 SP: 1.96	AC: 77.43 SE: 82.35 SP: 76.26
NB	AC: 72.47 SE: 74.72 SP: 72.71	AC: 75.55 SE: 65.46 SP: 76.68	AC: 72.23 SE: 82.35 SP: 69.83	AC: 77.65 SE: 76.47 SP: 77.93
k-NN	AC: 84.50 SE: 66.37 SP: 86.54	AC: 85.78 SE: 79.91 SP: 86.44	AC: 67.04 SE: 61.18 SP: 68.44	AC: 73.81 SE: 84.71 SP: 71.23
SVM	AC: 84.47 SE: 8.13 SP: 93.07	AC: 90.21 SE: 15.42 SP: 99.75	AC: 93.45 SE: 87.06 SP: 94.97	**AC: 99.55** **SE: 98.82** **SP: 99.72**
DT	AC: 91.55 SE: 59.82 SP: 95.12	AC: 94.06 SE: 69.75 SP: 96.80	AC: 81.72 SE: 87.43 SP: 81.72	AC: 90.52 SE: 57.47 SP: 98.60
EDT	AC: 92.97 SE: 75.85 SP: 94.89	**AC: 99.02** **SE: 94.13** **SP: 99.57**	AC: 83.30 SE: 61.18 SP: 88.55	AC: 90.52 SE: 68.24 SP: 95.81

TABLE 6. Comparison of the classification results of the possible MCs as normal tissue or true MCs with temporal subtraction (TS) of mammograms and using only the most recent mammograms (RM).

Classifier	Sensitivity [%]	Specificity [%]	Accuracy [%]	AUC
LDA	TS: 16.55 RM: 11.13	TS: 99.52 RM: 99.95	TS: 90.11 RM: 89.95	TS: 0.53 RM: 0.5
NB	TS: 65.46 RM: 69.53	TS: 76.68 RM: 64.34	TS: 75.55 RM: 64.86	TS: 0.72 RM: 0.67
k-NN	TS: 79.91 RM: 16.8	TS: 86.44 RM: 99.47	TS: 85.78 RM: 89.47	TS: 0.83 RM: 0.50
SVM	TS: 15.42 RM: 12.71	TS: 99.75 RM: 96.06	TS: 90.21 RM: 86.62	TS: 0.52 RM: 0.50
DT	TS: 69.75 RM: 61.17	TS: 96.80 RM: 93.60	TS: 94.06 RM: 90.32	TS: 0.83 RM: 0.80
EDT	**TS: 94.13** **RM: 70.43**	**TS: 99.57** **RM: 88.98**	**TS: 99.02** **RM: 87.10**	**TS: 0.94** **RM: 0.77**

TABLE 7. Comparison of the classification results of the true MCs as benign or suspicious using temporal subtraction (TS) of mammograms and using only the most recent mammograms (RM).

Classifier	Sensitivity [%]	Specificity [%]	Accuracy [%]	AUC
LDA	TS: 82.35 RM: 72.94	TS: 76.26 RM: 76.82	TS: 77.43 RM: 76.07	TS: 0.79 RM: 0.67
NB	TS: 76.47 RM: 81.18	TS: 77.93 RM: 76.26	TS: 77.65 RM: 77.20	TS: 0.79 RM: 0.77
k-NN	TS: 84.71 RM: 43.53	TS: 71.23 RM: 90.78	TS: 73.81 RM: 81.72	TS: 0.78 RM: 0.64
SVM	**TS: 98.82** RM: 65.88	**TS: 99.72** RM: 90.78	**TS: 99.55** RM: 86.00	**TS: 0.99** RM: 0.78
DT	TS: 57.47 RM: 50.59	TS: 98.60 RM: 98.88	TS: 90.52 RM:89.62	TS: 0.77 RM: 0.75
EDT	TS: 68.24 **RM: 58.82**	TS: 95.81 **RM: 99.16**	TS: 90.52 **RM: 91.42**	TS: 0.82 **RM: 0.79**

(filtering, thresholding, morphological operations) and then machine-learning (feature extraction, feature selection and classification) were optimized for FP removal and MC characterization as benign or suspicious, for the single mammogram case.

III. EXPERIMENTS AND RESULTS

A. MC DETECTION AND CLASSIFICATION USING TEMPORAL SUBTRACTION OF MAMMOGRAMS

1) PRE-PROCESSING, REGISTRATION, TEMPORAL SUBTRACTION AND POST-PROCESSING

Pre-processing was performed on all images as described in Section II. Figure 4 illustrates the effects of the pre-processing on dense and fatty mammograms, in regions with and without MCs. Demons registration successfully tracked the temporal MC changes, without introducing any errors. The residuals of the subtraction confirmed that this approach consistently removed unchanged information for both dense and fatty breasts (3.6% vs. 1.82% residual). The performance of the temporal subtraction was evaluated using the CR of the resulting image as compared to the recent mammogram (Fig. 5 & 6). The average CR of the subtracted images was significantly higher (63 times) than that of the corresponding recent ones. The background and other unchanged details were effectively removed and the images were enhanced.

Another advantage of temporal subtraction was the removal of old MCs, which have remained unchanged and do not pose any threat to the patient. Table 3, shows that there was a 23.85% reduction in the number of MCs after temporal subtraction of sequential mammograms, in a total of 26.33% overlapping MCs. This effect is also illustrated in Fig. 5.

2) CLASSIFICATION

For the elimination of FPs, the most valuable features, extracted from the difference image, are shown in Table 4. Applying those in an EDT model with L-O-P-O validation, achieved 99.02% accuracy and 0.94 AUC (Tables 5 & 6) for the identification of the true MCs. As seen in Fig. 7, using 4, 5 and 10 fold cross-validation resulted in lower accuracy due to the smaller dataset. For the classification of MCs as

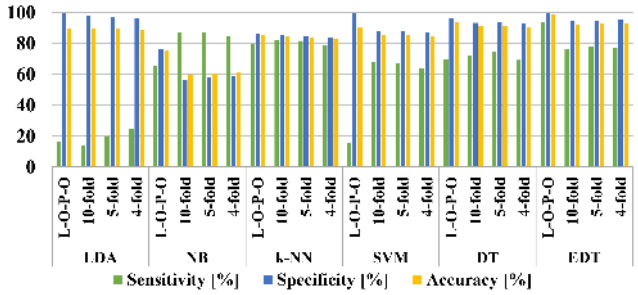

FIGURE 7. Classification results of the possible MCs as normal tissue or true MCs using temporal subtraction of mammograms, for each classifier and different validations.

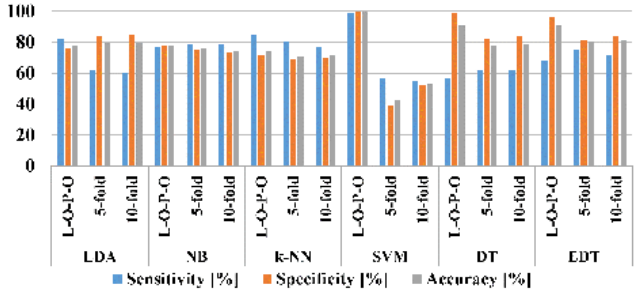

FIGURE 8. Classification results of the true MCs as benign or suspicious using temporal subtraction of mammograms, for each classifier and different validations.

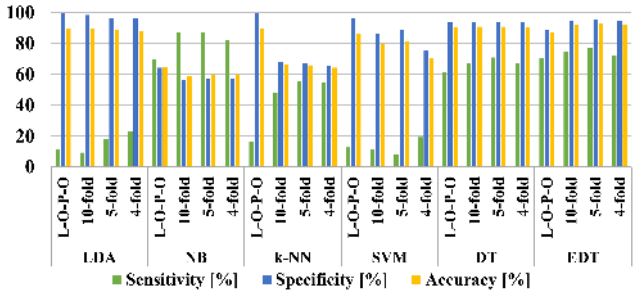

FIGURE 9. Classification results of the possible MCs as normal tissue or true MCs using only the most recent mammograms, for each classifier and different validations.

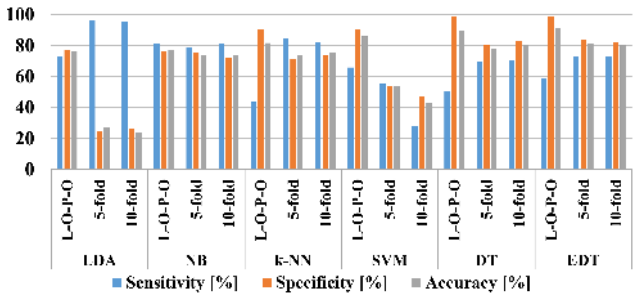

FIGURE 10. Classification results of the true MCs as benign or suspicious using only the most recent mammograms, for each classifier and different validations.

benign or suspicious, a SVM model reached 99.55% accuracy and 0.99 AUC (Tables 5 & 7), using the features indicated in Table 4 and L-O-P-O validation. Applying 5 and 10 fold cross-validation (Fig. 8) dropped the performance due to the lack of suspicious MC data. The selection of features resulted in statistically significant improvement of the performance as shown in Table 5.

B. MC DETECTION AND CLASSIFICATION USING ONLY THE MOST RECENT MAMMOGRAMS

Using the same feature selection methods, the best features for single mammogram MC detection were selected (Table 4). Table 6 shows the classification results for the elimination of FPs using six classifiers in a L-O-P-O validation and the features extracted from the most recent image. From the 443 true MCs, 131 were missed (benign and suspicious) and 434 normal regions were falsely identified as MCs. The accuracy, using an EDT model, was 87.10% and the AUC was 0.77. With the implementation of 4, 5 and 10 fold cross-validation the performance was considerably lower (Fig. 9).

The true MCs were subsequently classified as benign or suspicious and the results are presented in Table 7. EDTs achieved 91.42% accuracy and 0.79 AUC with L-O-P-O validation. However, 3 benign MCs were incorrectly identified as suspicious and 35 suspicious regions were misclassified as benign. The best performance was achieved using the features indicated in Table 4. With 5 and 10 fold cross-validation the performance dropped in terms of sensitivity, specificity and accuracy (Fig. 10).

IV. DISCUSSION

In this paper, the advantages of temporal subtraction on pairs of sequential mammograms from 80 patients were investigated, after registration between the most recent and prior mammograms. Temporal subtraction alone can improve contrast and provide better visualization. After pre-processing, registration and temporal subtraction, the contrast of the subtracted images increased 63 times, compared to the recent mammogram without processing. This contrast improvement can help radiologists identify possible abnormalities easier and faster, since the background and other unchanged regions are removed.

Temporal subtraction also eliminated old MCs and fed the classifier with only the newly developed ones. This step is crucial since the proposed algorithm identifies and then discards all overlapping MCs, to assure that only new MCs have to be evaluated. If the evaluation is to be performed manually, this saves the radiologist effort and time by focusing only on new information without having to refer back to the old mammograms.

The highest classification accuracy (99.02%) was achieved with EDTs for the elimination of FPs (Fig. 11), using an optimized set of features and L-O-P-O validation. The majority of the FPs were effectively removed and only 26 MCs were misclassified as normal regions. Furthermore, SVMs provided 99.55% accuracy in the classification of true MCs as benign or suspicious using L-O-P-O cross-validation (Fig. 12). Only 1 malignancy was misclassified as benign out of a total of 85. Additionally, 1 benign MC was incorrectly

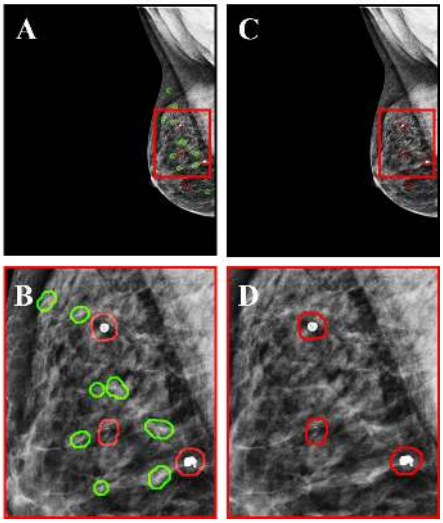

FIGURE 11. Classification of possible MCs as normal tissue or true MCs. (A) Most recent mammogram, with green circles around the false positives (FPs) and red circles in the true MCs. (B) Zoomed view of A, in selected areas. (C) Most recent mammogram after the elimination of FPs. (D) Zoomed view of C, in selected areas.

categorized as suspicious. Such high accuracy exemplifies the effectiveness of the proposed algorithm which is far superior compared to using only the most recent mammogram. Specifically, the average accuracy of classifying true MCs as benign or suspicious, improved by 8%. This improvement was statistically significant with p-value < 0.005.

In all cases, implementation of k-fold cross-validation, using $k = 4$, 5 and 10, resulted in a drop in the performance. This can be explained by the limited number of data. The dataset included 443 MCs, of which only 85 were suspicious thus limiting the effectiveness of the training. Hence, the performance is expected to improve as the size of the dataset increases.

The comparison of the proposed method with other algorithms is not straightforward since this is the first time that temporal subtraction of mammograms and elimination of overlapping MCs was applied. A new and unique dataset was created for the purposes of this study since all publicly available databases related to mammography include only recent mammograms for each patient. Some studies have compared the performance of their algorithms with and without using prior mammograms [20]–[22] for the characterization of mass lesions. Their results (AUC 0.88 [20], AUC 0.77 [21], AUC 0.90 [22]) indicate that the use of temporal features is an effective approach to improve the detection of masses.

A weakness identified in many studies, is in the procedure of the verification of the algorithms' performance. Using cross-validation by randomly dividing the MCs into a training and a test set, or using parts of the same areas as a test set and as a training set, is inappropriate since it can introduce bias. In this study, a more appropriate approach was adopted dividing the patients into training and test groups for the L-O-P-O and k-fold cross-validation. A drawback of this study is the relatively small dataset. In addition, although

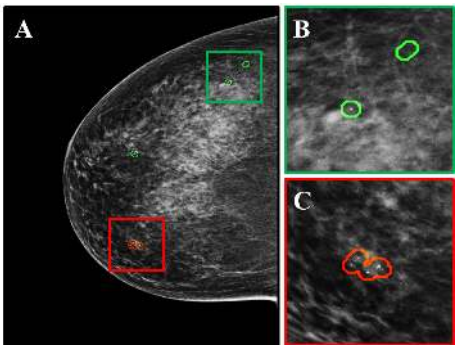

FIGURE 12. Classification of the true MCs as benign or suspicious. (A) Most recent mammogram, with green circles around the benign MCs and a red circle around the suspicious MC. (B) Zoomed view of A, in an area with benign MCs. (C) Zoomed view of A, in an area with suspicious MCC.

the suspicious MCs were identified by two expert radiologists, differences might appear if more experts perform the same task.

V. CONCLUSION

A new method for the detection and classification of MCs on pairs of temporally sequential mammograms was developed. The goal of this work was to combine temporal subtraction with machine-learning in order to increase the CR, remove the old MCs and improve the accuracy of the MC classification. Pre-processing, registration and post-processing were combined to efficiently and effectively subtract the mammograms and improve the MC detection. Machine-learning techniques were used to eliminate the large number of FPs and classify the MCs as benign or suspicious. The proposed method outperforms algorithms without the use of temporal subtraction, as it achieves 99.02% accuracy for the detection of true MCs and 99.55% accuracy for the classification of MCs as benign or suspicious.

Encouraged by this initial success, the technique will be further tested on a larger dataset with more representative age samples. Furthermore, the algorithm can be generalized to other kind of abnormalities (i.e. mass lesions) in mammograms and can provide predictions for the development of new findings in the following screening rounds. Such advancements have the potential to significantly contribute to the development of automated CAD systems, offer invaluable assistance to radiologist and impact patient prognosis.

ACKNOWLEDGMENT

All data included in this study were collected in accordance with the ethical standards of the 1964 Helsinki Declaration and its later amendments or comparable ethical standards.

REFERENCES

[1] R. L. Siegel, K. D. Miller, and A. Jemal, "Cancer statistics, 2015," *CA, A Cancer J. Clinicians*, vol. 65, no. 1, pp. 5–29, Jan. 2015. [Online]. Available: https://onlinelibrary.wiley.com/doi/abs/10.3322/caac.21254

[2] K. Ganesan, U. R. Acharya, C. K. Chua, L. C. Min, K. T. Abraham, and K.-H. Ng, "Computer-aided breast cancer detection using mammograms: A review," *IEEE Rev. Biomed. Eng.*, vol. 6, pp. 77–98, 2013.

[3] P. L. A. Hernández, T. T. Estrada, A. L. Pizarro, M. L. D. Cisternas, and C. S. Tapia, "Breast calcifications: Description and classification according to bi-rads 5th edition," *Rev. Chil. Radiol.*, vol. 22, pp. 80–91, 2016.
[4] R. M. Rangayyan, F. J. Ayres, and J. E. L. Desautels, "A review of computer-aided diagnosis of breast cancer: Toward the detection of subtle signs," *J. Franklin Inst.*, vol. 344, nos. 3–4, pp. 312–348, May 2007.
[5] S. Feig, B. Galkin, and H. Muir, "Evaluation of breast microcalcifications by means of optically magnified tissue specimen radiographs," in *Breast Cancer*, vol. 105. Berlin, Germany: Springer, 1987, pp. 111–123.
[6] E. A. Sickles, "Breast calcifications: Mammographic evaluation," *Radiology*, vol. 160, no. 2, pp. 289–293, Aug. 1986.
[7] H. D. Cheng, X. Cai, X. Chen, L. Hu, and X. Lou, "Computer-aided detection and classification of microcalcifications in mammograms: A survey," *Pattern Recognit.*, vol. 36, no. 12, pp. 2967–2991, Dec. 2003.
[8] A. J. Bekker, M. Shalhon, H. Greenspan, and J. Goldberger, "Multi-view probabilistic classification of breast microcalcifications," *IEEE Trans. Med. Imag.*, vol. 35, no. 2, pp. 645–653, Feb. 2016.
[9] J. Wang, R. M. Nishikawa, and Y. Yang, "Global detection approach for clustered microcalcifications in mammograms using a deep learning network," *J. Med. Imag.*, vol. 4, no. 2, Apr. 2017, Art. no. 024501.
[10] J. Wang and Y. Yang, "A context-sensitive deep learning approach for microcalcification detection in mammograms," *Pattern Recognit.*, vol. 78, pp. 12–22, Jun. 2018.
[11] A. Bria, C. Marrocco, L. R. Borges, M. Molinara, A. Marchesi, J.-J. Mordang, N. Karssemeijer, and F. Tortorella, "Improving the automated detection of calcifications using adaptive variance stabilization," *IEEE Trans. Med. Imag.*, vol. 37, no. 8, pp. 1857–1864, Aug. 2018.
[12] T. M. A. Basile, A. Fanizzi, L. Losurdo, R. Bellotti, U. Bottigli, R. Dentamaro, V. Didonna, A. Fausto, R. Massafra, M. Moschetta, P. Tamborra, S. Tangaro, and D. La Forgia, "Microcalcification detection in full-field digital mammograms: A fully automated computer-aided system," *Phys. Medica*, vol. 64, pp. 1–9, Aug. 2019.
[13] F. Ma, M. Bajger, S. Williams, and M. J. Bottema, "Improved detection of cancer in screening mammograms by temporal comparison," in *Proc. Int. Workshop Digit. Mammography*, vol. 6136. Berlin, Germany: Springer, 2010, pp. 752–759.
[14] S. Timp and N. Karssemeijer, "Interval change analysis to improve computer aided detection in mammography," *Med. Image Anal.*, vol. 10, no. 1, pp. 82–95, Feb. 2006.
[15] M. P. Callaway, C. R. M. Boggis, S. A. Astley, and I. Hutt, "The influence of previous films on screening mammographic interpretation and detection of breast carcinoma," *Clin. Radiol.*, vol. 52, no. 7, pp. 527–529, Jul. 1997.
[16] L. Hadjiiski, H.-P. Chan, B. Sahiner, M. A. Helvie, M. A. Roubidoux, C. Blane, C. Paramagul, N. Petrick, J. Bailey, K. Klein, M. Foster, S. Patterson, D. Adler, A. Nees, and J. Shen, "Improvement in radiologists' characterization of malignant and benign breast masses on serial mammograms with computer-aided diagnosis: An ROC study," *Radiology*, vol. 233, no. 1, pp. 255–265, Oct. 2004.
[17] C. Varela, N. Karssemeijer, J. H. C. L. Hendriks, and R. Holland, "Use of prior mammograms in the classification of benign and malignant masses," *Eur. J. Radiol.*, vol. 56, no. 2, pp. 248–255, Nov. 2005.
[18] J. Bozek, M. Kallenberg, M. Grgic, and N. Karssemeijer, "Use of volumetric features for temporal comparison of mass lesions in full field digital mammograms," *Med. Phys.*, vol. 41, no. 2, Jan. 2014, Art. no. 021902.
[19] T. Kooi and N. Karssemeijer, "Classifying symmetrical differences and temporal change for the detection of malignant masses in mammography using deep neural networks," *J. Med. Imag.*, vol. 4, no. 4, Oct. 2017, Art. no. 044501.
[20] L. Hadjiiski, B. Sahiner, H.-P. Chan, N. Petrick, M. A. Helvie, and M. Gurcan, "Analysis of temporal changes of mammographic features: Computer-aided classification of malignant and benign breast masses," *Med. Phys.*, vol. 28, no. 11, pp. 2309–2317, Nov. 2001.
[21] S. Timp, C. Varela, and N. Karssemeijer, "Temporal change analysis for characterization of mass lesions in mammography," *IEEE Trans. Med. Imag.*, vol. 26, no. 7, pp. 945–953, Jul. 2007.
[22] F. Ma, L. Yu, G. Liu, and Q. Niu, "Computer aided mass detection in mammography with temporal change analysis," *Comput. Sci. Inf. Syst.*, vol. 12, no. 4, pp. 1255–1272, 2015.
[23] K. Loizidou, G. Skouroumouni, C. Nikolaou, and C. Pitris, "A new method for breast micro-calcification detection and characterization using digital temporal subtraction of mammogram pairs," in *Proc. IEEE EMBS Int. Conf. Biomed. Health Informat. (BHI)*, May 2019, pp. 1–4.
[24] P. Soille, *Morphological Image Analysis: Principles and Applications*, 2nd ed. Berlin, Germany: Springer, 2004.
[25] R. Gonzalez, R. Woods, and S. Eddins, *Digital Image Processing Using MATLAB*, 2nd ed. Gatesmark Publishing, 2010.
[26] S.-C. Huang, F.-C. Cheng, and Y.-S. Chiu, "Efficient contrast enhancement using adaptive gamma correction with weighting distribution," *IEEE Trans. Image Process.*, vol. 22, no. 3, pp. 1032–1041, Mar. 2013.
[27] K. Marias, C. Behrenbruch, S. Parbhoo, A. Seifalian, and M. Brady, "A registration framework for the comparison of mammogram sequences," *IEEE Trans. Med. Imag.*, vol. 24, no. 6, pp. 782–790, Jun. 2005.
[28] Y. Guo, R. Sivaramakrishna, C.-C. Lu, J. S. Suri, and S. Laxminarayan, "Breast image registration techniques: A survey," *Med. Biol. Eng. Comput.*, vol. 44, nos. 1–2, pp. 15–26, Mar. 2006.
[29] X. Pennec, P. Cachier, and N. Ayache, "Understanding the 'demon's algorithm': 3D non-rigid registration by gradient descent," in *Proc. Int. Conf. Med. Image Comput. Comput.-Assist. Intervent.* Springer, 1999, pp. 597–605.
[30] Y. Díez, A. Oliver, X. Llado, J. Freixenet, J. Marti, J. C. Vilanova, and R. Marti, "Revisiting intensity-based image registration applied to mammography," *IEEE Trans. Inf. Technol. Biomed.*, vol. 15, no. 5, pp. 716–725, Sep. 2011.
[31] H. Lu, M. Reyes, A. Serifovic, A. Šerifović, S. Weber, Y. Sakurai, H. Yamagata, and P. C. Cattin, "Multi-modal diffeomorphic demons registration based on point-wise mutual information," in *Proc. IEEE Int. Symp. Biomed. Imag., Nano Macro*, Apr. 2010, pp. 372–375.
[32] D. Bailey and R. Hodgson, "Range filters: Localintensity subrange filters and their properties," *Image Vis. Comput.*, vol. 3, no. 3, pp. 99–110, Aug. 1985.
[33] R. M. Nishikawa, M. Kallergi, and C. G. Orton, "Computer-aided detection, in its present form, is not an effective aid for screening mammography," *Med. Phys.*, vol. 33, no. 4, pp. 811–814, Mar. 2006.
[34] B. Singh and M. Kaur, "An approach for classification of malignant and benign microcalcification clusters," *Sādhanā*, vol. 43, no. 3, p. 39, Mar. 2018.
[35] V. Kumar and P. Gupta, "Importance of statistical measures in digital image processing," *Int. J. Emerg. Technol. Adv. Eng.*, vol. 2, no. 8, pp. 56–62, 2012.
[36] R. M. Haralick, K. Shanmugam, and I. Dinstein, "Textural features for image classification," *IEEE Trans. Syst., Man, Cybern.*, vol. SMC-3, no. 6, pp. 610–621, Nov. 1973.
[37] P. Diehr, D. C. Martin, T. Koepsell, and A. Cheadle, "Breaking the matches in a pairedt-test for community interventions when the number of pairs is small," *Statist. Med.*, vol. 14, no. 13, pp. 1491–1504, Jul. 1995.
[38] L. Stahle and S. Wold, "Multivariate analysis of variance (manova)," *Chemometrics Intell. Lab. Syst.*, vol. 9, no. 2, pp. 127–141, 1990.
[39] S. Mika, G. Ratsch, J. Weston, B. Scholkopf, and K.-R. Mullers, "Fisher discriminant analysis with kernels," in *Proc. IEEE Signal Process. Soc. Workshop Neural Netw. Signal Process.*, Aug. 1999, pp. 41–48.
[40] M. Karabatak, "A new classifier for breast cancer detection based on Naïve Bayesian," *Measurement*, vol. 72, pp. 32–36, Aug. 2015.
[41] C. Cortes and V. Vapnik, "Support-vector networks," *Mach. Learn.*, vol. 20, no. 3, pp. 273–297, 1995.
[42] J. R. Quinlan, "Induction of decision trees," *Mach. Learn.*, vol. 1, no. 1, pp. 81–106, 1986.
[43] D. Opitz and R. Maclin, "Popular ensemble methods: An empirical study," *J. Artif. Intell. Res.*, vol. 11, pp. 169–198, Aug. 1999.

KOSMIA LOIZIDOU (Student Member, IEEE) received a B.Sc. in Mechanical and Manufacturing Engineering (minor degree in Biomedical Engineering) from the University of Cyprus, in 2016. In 2018, she obtained an M.Sc. in Electrical Engineering from the University of Cyprus with specialization in Biomedical Engineering. She is currently pursuing a Ph.D. degree at the University of Cyprus. She is also holding the position of a Special Scientist at the KIOS Research and Innovation Centre of Excellence. Her research interests include biomedical engineering and clinical health with a specific focus in medical image processing, machine-learning, and breast cancer diagnosis.

GALATEIA SKOUROUMOUNI received a Medical degree from the Aristotle University of Thessaloniki, Greece. She is currently pursuing a Master's degree in Public Administration–Healthcare Management with Neapolis University Paphos, Cyprus. She is also a Radiology specialist with the Nicosia General Hospital, Nicosia, Cyprus, participating also in the Cyprus Population Breast Screening Program. She has completed her specialty in Radiology at the Carl-Thiem Klinikum in Cottbus, Germany, (16 months), and at the Papageorgiou General Hospital of Thessaloniki, Greece, (50 months). She has participated in multiple oral and poster presentations in Greek and European radiology conferences, published medical case reports in Greek and European journals, and translated scientific books from English to Greek (Making the Diagnosis: A practical guide to breast imaging; ACR BI-RADS Atlas, Fifth Edition). Her current main research interests include machine-learning for the detection and prediction of breast cancer in sequential mammograms.

CHRISTOS NIKOLAOU studied Medicine at the Athens Medical School and has been trained in Radiology at Laikon General and the University Hospital of Athens. He is currently a Consultant Radiologist, the Head of Radiology Department, and the Deputy Scientific Director of Limassol General Hospital. He has been an (Hons.) Assistant Professor with the Nicosia Medical School, since November 2014. He has also been a Senior Consultant Radiologist with the National Mammography Screening Programme, since February 2016. He is also the State Health Services organization (SHSO) coordinator in the eHealth4you project under the State Electronic Health Authority that is setting the specifications for the patient digital records. He is also a member of the Executive board of Cyprus Radiological Society and a member of European Radiological Society. He has attended 19 postgraduate courses and 38 Conferences in which he has 29 presentations and six scientific publications.

COSTAS PITRIS (Member, IEEE) received a B.S. degree (Hons.) in Electrical Engineering and an M.S. degree in Electrical Engineering from the University of Texas at Austin, in 1993 and 1995, respectively, a Ph.D. degree in Electrical and Medical Engineering from the Massachusetts Institute of Technology, in 2000, and an M.D. degree *(magna cum laude)* in Medicine from the Harvard Medical School, in 2002. He is currently an Associate Professor at the KIOS Research and Innovation Centre of Excellence, Department of Electrical and Computer Engineering, University of Cyprus. He is heading the Biomedical Imaging and Applied Optics Laboratory, which he established, in 2004. He has published 39 peer-reviewed journal publications, 123 conference proceedings, four book chapters, and one book. He also holds eight U.S. and European patents and is the co-founder of two start-up companies aiming to commercialize important research findings. The citations to his work have reached more than 12000 (with an h-index of 37) according to Google Scholar. He is also a Grant Reviewer of the European Commission (FP7 - Nanomedicine), National Institutes of Health, USA (Biomedical Optics), and the Cyprus Ministry of Commerce, Industry and Tourism (Start-up Incubators). His main research interests include the areas of optical diagnostics, biomedical imaging and spectroscopy, signal/image analysis and CI. Prof. Pitris has served as a PI or a Co-PI in competitive research grants totaling over €2.5 mil. He is also one of the co-founders and a member of the executive committee of the KIOS Research and Innovation Centre of Excellence, which was a recipient of a H2020 TEAMING grant of over €40 mil.