

Breast Mass Detection and Classification based on Digital Temporal Subtraction of Mammogram Pairs

Kosmia Loizidou¹, Galatea Skouroumouni², Christos Nikolaou³ and Costas Pitris¹

Abstract—Breast cancer is one of the deadliest malignancies worldwide. In mammography, the most reliable screening tool for its diagnosis, expert radiologists review the mammograms to determine whether the patient has any signs of disease. Unfortunately, the evaluation of breast abnormalities is challenging, even for experienced radiologists. Computer-Aided Detection (CAD) systems can assist in the detection of breast cancer. In this work, an algorithm for the automatic detection and classification of masses, based on subtraction of sequential digital mammograms, image registration and machine learning, is presented. Previous studies assessed the use of sequential mammograms to perform temporal analysis by creating a new temporal feature vector. Temporal subtraction registers and subtracts the prior mammogram from the current one, prior to performing mass detection and classification. A new dataset, which includes sequential pairs from 40 patients (160 mammograms) with precisely annotated mass locations (benign and suspicious), was created to assess the performance of the algorithm. For the classification, various features were extracted and six classifiers were used in a leave-one-patient-out cross-validation. The accuracy of the classification of masses as benign or suspicious increased from 90.83% (with the previously described temporal analysis) to 96.51% (with temporal subtraction). The improvement was statistically significant with $p < 0.05$. These results demonstrate the effectiveness of the proposed technique of temporal subtraction of mammograms for the detection of masses.

Index Terms—Breast cancer, computer-aided diagnosis, digital mammography, breast mass, temporal subtraction

I. INTRODUCTION

Breast cancer is the most frequently diagnosed cancer and the leading cause of mortality in women worldwide. Mammography is extensively used as a screening and diagnostic tool [1]. Radiologists evaluate the mammograms manually to determine whether the patient has signs of malignancy. In the case of positive findings, the appropriate disease management is followed. According to the Breast Imaging Reporting and Data System (BI-RADS), an important sign of breast cancer that radiologists look for, is a mass [2]. A breast mass is associated with a localized swelling or lump inside the breast and typically appears as a relatively dense region. Breast masses are defined by their size, shape, density and texture and their characteristics (i.e. contrast, distortion) are important in their assessment.

Radiologists examine the mass properties in order to classify them as benign or suspicious. This categorization is one of the critical tasks in mammography since it determines subsequent management. It is also challenging, not only due to the large variation in size and shape of masses, but also due to their low image contrast [3]. Computer Aided Diagnosis (CAD) systems are being developed to address the issue of the diagnosis of breast masses. Although various algorithms have been proposed [1], [3], [4], there is still room for improvement, especially considering that few existing methods compare the recent with the older mammograms, something that radiologists routinely exploit for more accurate identification of abnormalities.

Only a few studies in the literature assessed the use of sequential mammograms for the detection of breast masses using temporal analysis [5]–[7]. Regions of Interest (ROIs) in the recent image were associated with corresponding ROIs in the prior one. A new temporal feature vector was created by subtracting the numerical values of the features, extracted from each prior image ROI, from the corresponding features of the recent image ROI. This vector was then used for the classification of the ROIs. Overall, the findings confirmed that subtle signs of malignancy could be recognized only with the addition of information from prior mammograms. Moreover, the classification performance increased, compared to using only the features from the recent mammograms. However, only ROIs were included and temporal information from other regions was not available.

In this work, a new approach for the automatic detection and classification of masses, based on subtraction of sequential digital mammograms, is presented. Temporal subtraction, previously demonstrated for micro-calcification detection, unlike temporal analysis, subtracts the entire prior image [8]. With the application of temporal subtraction, unchanged overlapping regions were effectively removed resulting in an increase in the Contrast Ratio (CR). Subsequently, falsely identified regions were eliminated and the remaining masses were classified as benign or suspicious. The results were compared to temporal analysis and proved the proposed technique to be more effective, with a superior classification performance.

II. MATERIALS AND METHODS

A. Dataset

In this study, a new dataset of mammographic data was created since the available open access databases of breast cancer do not include sequential sets of mammograms, but rather only the recent ones. In addition, the images

¹K. Loizidou and C. Pitris, are with the Department of Electrical and Computer Engineering and KIOS Research and Innovation Center of Excellence, University of Cyprus, Nicosia, Cyprus, cloizi01@ucy.ac.cy, cpitris@ucy.ac.cy

²G. Skouroumouni is with the Radiology Department, Nicosia General Hospital, Nicosia, Cyprus, galskour@hotmail.com

³C. Nikolaou is with the Radiology Department, Limassol General Hospital, Limassol, Cyprus, cnikolaou@doctors.org.uk

TABLE I
AGE OF PARTICIPANTS INCLUDED IN THE STUDY.

	Normal Mammograms (n=20)	Suspicious Mammograms (n=20)	Total
Mean \pm STD	52.90 \pm 8.26	60.05 \pm 8.16	56.48 \pm 8.96
Median	49.50	61	56.50
Range	42-74	39-73	39-74

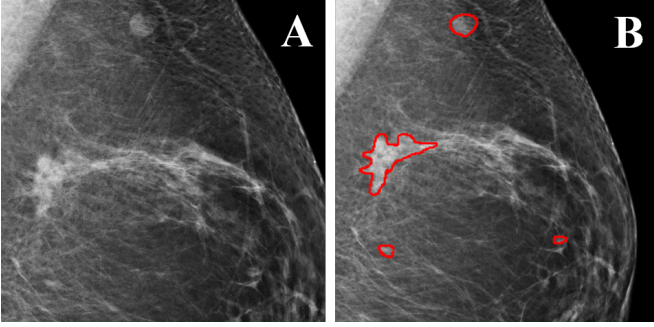


Fig. 1. Dataset example of a fatty mammogram. (A) Zoomed region of a mammogram with breast masses. (B) The same image with precise marking of annotated mass locations.

are not precisely annotated to show the exact margins of a mass, limiting their value as ground truth. The new clinical mammogram dataset developed consists of 40 pairs of full-field digital sequential mammograms from CranioCaudal (CC, above) and MedioLateral Oblique (MLO, side-angle) views. A total of 160 images were taken from various hospitals across the country. The selection and annotation was made by a radiologist with ten years of experience and a second radiologist, with two years of experience, validated the images. Consecutive mammograms were collected from 2012 to 2019, with an average of 1 to 7 years interval. The age of the participants varied from 39 to 74, with a mean age of 56.48 ± 8.96 years and median age of 56.50. More details about the age of participants included in the study can be found in Table I.

Half the cases corresponded to healthy individuals without any signs of malignancy in both screening rounds. The remaining cases belonged to patients with at least one suspicious mass present in the most recent mammogram. This unique dataset, includes detailed annotations of the boundary of each mass (both benign and suspicious) to be used as the ground truth (Fig. 1). The dimensions of the mammograms were 4096×3328 pixels, in a 12-bit DICOM format.

B. Mass Detection and Classification using Temporal Subtraction

In temporal subtraction the two corresponding images of each pair were processed in parallel. Fig. 2 demonstrates the methodology followed. The pre-processing began with the normalization of the images, to adjust the range of pixel intensity values and then Contrast Limited Adaptive Histogram Equalization (CLAHE), Gamma correction and border removal

Sequential Mammograms

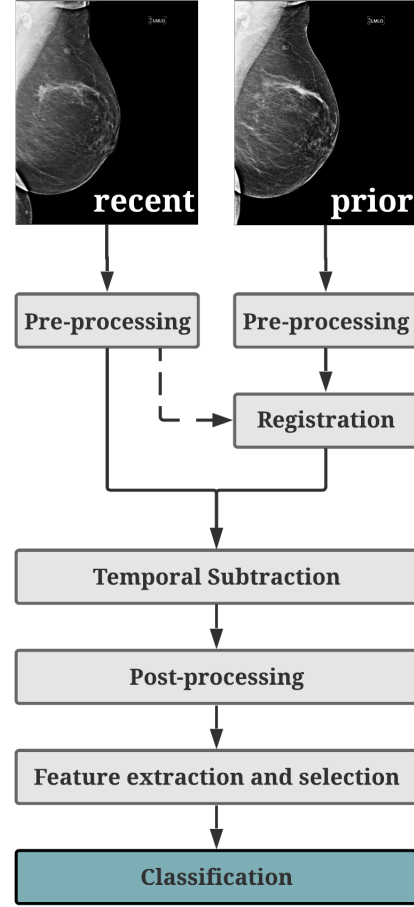


Fig. 2. Diagram of the proposed methodology for breast mass detection and classification using temporal sequential mammograms.

took place. CLAHE is based on the re-allocation of the gray levels inside an image using the probability distribution of the input gray levels. With this technique, the maximum contrast enhancement factor is affected by the clip level that is defined by the user. Hence, CLAHE effectively eliminated the noise and diminished the edge-shadowing effect [9]. Contrast adjustment using Gamma correction was used to account for the non-linear mapping of intensities in the images such as:

$$T(l) = l_{max}(l/l_{max})^\gamma \quad (1)$$

where, l_{max} is the maximum intensity value of the image and γ the Gamma parameter, which was set to 2 [10]. Border removal removed the high intensity areas connected to the border (including the pectoral muscle) using:

$$F(x, y) = \begin{cases} I(x, y), & \text{if } (x, y) \text{ is on the border of } I. \\ 0, & \text{otherwise.} \end{cases} \quad (2)$$

where, I is the original image and F the border areas [11].

For effective image subtraction, precise image registration is required. Registration compensates for patient movement,

shape alterations, differences in the amount of the pectoral muscle in the MLO view, variations in breast compression and human error. Various registration methods are available for the alignment of prior and recent mammograms [12]. In this study, two registration algorithms were assessed: Affine [12] and Demons [13]. The two techniques were compared based on the residuals (sum of the remaining pixels after subtraction) and in all times, the prior image was registered to the recent one.

Affine is an intensity-based linear-mapping registration technique that consists scaling, translation and rotation. Affine secures planes, straight lines and points and it can be explained by:

$$\begin{aligned} x' &= t_x + a_1x + a_2y \\ y' &= t_y + b_1x + b_2y \end{aligned} \quad (3)$$

where, t_x and t_y are the translation of vectors, x and y are the original points, x' and y' are the transformed points and a_1, a_2, b_1, b_2 are the affine parameters [12]. On the contrary, Demons is a local registration technique that aligns the moving image (prior) to the fixed one (recent), using regional similarity and location. In Demons, the registration is seen as a diffusion process and it 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:

$$E_{corr}^s(u) = ||F - Mo(s + u)||^2 + \left(\frac{\sigma_i^2}{\sigma_x^2} \right) ||u||^2 \quad (4)$$

where, σ_i^2 is the noise on the image intensity and σ_x^2 the spatial uncertainty. Using Taylor expansion, (4) can be linearized and the energy function will reach its minimum when gradient descent is zero. The registration must be solved iteratively, since the update field is based on local information [13].

Registration was followed by temporal subtraction, subtracting the prior registered image from the recent one, to eliminate all the areas that remained unchanged. At this point, the CR of the subtracted image was compared to the corresponding CR of the most recent image, to evaluate the image enhancement of the method. High intensity areas on the periphery of the breast were automatically removed by detecting the breast periphery from a binary image of the entire breast, and subsequently by removing all the high intensity regions that fall on the periphery. This approach was critical since the regions that appeared in the periphery corresponded to skin areas that cannot contain masses, and emerged from misalignment.

Afterwards, the subtracted image was processed using unsharp-mask filtering, a spatial filter that enhances a specific range of frequencies inside an image. The form of unsharp masking can be expressed as:

$$f_{sharp}(x, y) = f(x, y) + k * g(x, y) \quad (5)$$

where, k is the scaling constant (can be set from 0.2 to 0.7) and $g(x, y)$ is calculated as follows:

$$g(x, y) = f(x, y) - f_{smooth}(x, y) \quad (6)$$

where, $f(x, y)$ is the input image and $f_{smooth}(x, y)$ is the smoothed version of the input [14].

The first step in the identification of the margin of the breast masses was the conversion to binary using thresholding. To select the most suitable threshold value, Otsu's method was applied using the discriminant criterion [15]. Morphological operations were then applied, such as opening and closing. At this point, the detection was over and the remaining ROIs were considered as possible masses.

Machine learning was performed to reduce the large number of False Positive (FP) detections. In total, 28 features were extracted from four different categories: shape, intensity-based, First-Order Statistics (FOS) and Gray Level Co-occurrence Matrix features (GLCM). Details about the extracted features can be found in Table II. Those features were selected for their capability to identify the masses based on their unique characteristics. Feature selection was then applied for the elimination of the least important features, in order to increase the classification performance. The most statistically significant features ($p < 0.01$) were selected and the optimal combination was determined during the classification process, by optimizing the accuracy of the results during training.

Six classifiers were assessed: Linear Discriminant Analysis (LDA) [16], k-Nearest Neighbors (k-NN) [17], Naive Bayes (NB) [17], Support Vector Machines (SVMs) [18], Decision Trees (DTs) [19] and Ensemble of Decision Trees (EDT) [20], with leave-one-patient-out cross-validation. Each time the four images of a single patient (CC and MLO views of the most recent and prior mammograms) were used as the test sample, while the remaining patients were used for the training of the model, until all the cases were classified. With this approach, bias is avoided since all the ROIs of a single patient were used either in the training or in the testing, but not both. In the first classification round, the detected ROIs were classified as masses or normal tissue and in the second round, the true masses were characterized as benign or suspicious. However, since the dataset was limited, 3-class classification was also performed and the ROIs were automatically classified into normal tissue, benign masses and suspicious masses, in one round. The performance of classification was evaluated by measuring sensitivity, specificity, accuracy and the Area Under the Receiver Operating Characteristics (ROC) Curve (AUC).

C. Mass Detection and Classification using Temporal Analysis

Temporal analysis was proposed in the literature as an effective technique for the detection of breast masses using sequential mammograms. Using this technique, the prior mammogram is used as a reference for the creation of a new feature vector, by subtracting the numerical values of the features extracted at the prior image from the corresponding ones at the recent image. This approach was applied to compare its efficiency to the proposed temporal subtraction. Hence, the same features were extracted from both the recent and prior registered image, and then subtracted, to create a new temporal feature vector. The previously described classification

TABLE II
INFORMATION ABOUT THE EXTRACTED AND SELECTED FEATURES FOR
THE CLASSIFICATION OF BREAST MASSES USING TEMPORAL
SUBTRACTION.

Features Extracted	Selected for Temporal Subtraction	
	1st round	2nd round
Shape features		
Area	X	
Circularity	X	
Compactness		
Convex area	X	X
Eccentricity	X	X
Equivalent diameter		
Euler number	X	X
Extent		X
Filled area	X	
Major axis length	X	X
Minor axis length	X	
Orientation	X	X
Perimeter	X	
Solidity	X	X
Shape ratio	X	X
Minimum intensity	X	
Intensity features		
Maximum intensity	X	
Average intensity		
FOS features		
Entropy		
Kurtosis	X	
Skewness		X
Smoothness		
Standard deviation		
Variance		
GLCM features		
Contrast	X	
Correlation	X	
Energy		
Homogeneity	X	X

methodology was optimized for the temporal analysis method and the same procedure was followed for the validation and evaluation so that the results would be comparable.

III. RESULTS

A. Mass Detection and Classification using Temporal Subtraction

Of the two registration methods evaluated, Demons registration had superior performance as compared to Affine, successfully tracking the image changes that occurred between the two screenings. As seen in Fig. 3, the residuals of the subtraction confirmed that the Demons approach effectively eliminated the unchanged regions in dense and fatty breasts (42% vs. 28.5% overall residual). The performance of

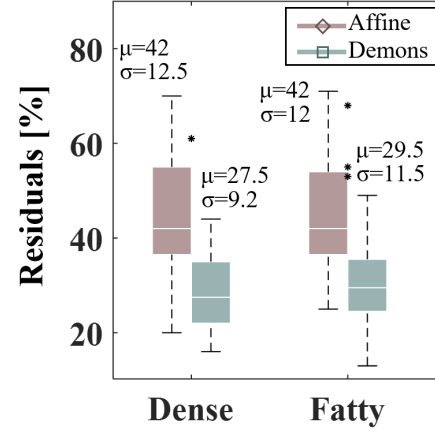


Fig. 3. Plot comparing the residuals remaining after subtraction, between dense and fatty mammograms, using Demons or Affine registration. The box represents the 25th to the 75th percentile of the data. The whiskers indicate the maximum and minimum values. The central mark defines the median of the data and the asterisks the outliers.

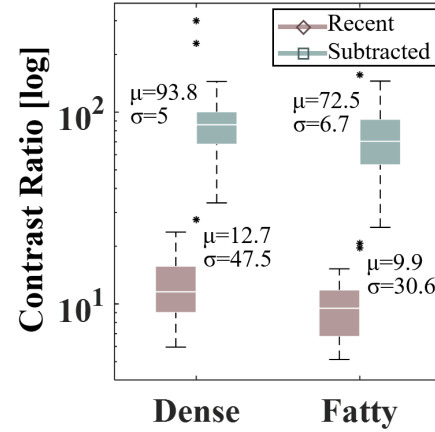


Fig. 4. Plot comparing the contrast ratio (logarithmic scale) between the unprocessed recent image and the image created from temporal subtraction, for dense and fatty mammograms. The box represents the 25th to the 75th percentile of the data. The whiskers indicate the maximum and minimum values. The central mark defines the median of the data and the asterisks the outliers.

temporal subtraction was also evaluated by measuring the CR of the subtracted image and comparing it to the corresponding CR of the most recent mammogram without any processing (Fig. 4 & 5). Overall, the CR increased ~ 7 times (11.29 vs. 83.1), for dense and fatty mammograms, resulting in a new image accentuating the changes between the screenings.

Features were extracted from the images and, using the feature selection algorithms, the best features for the elimination of FPs were determined (Table II). Using those, EDT achieved 97.44% accuracy and 0.92 AUC (Fig. 6). The true masses were then further categorized as benign or suspicious and EDT reached 83.93% accuracy along with 0.74 AUC (Fig. 6). In the 3-class classification (normal - benign - suspicious), the performance, using EDT and the same features

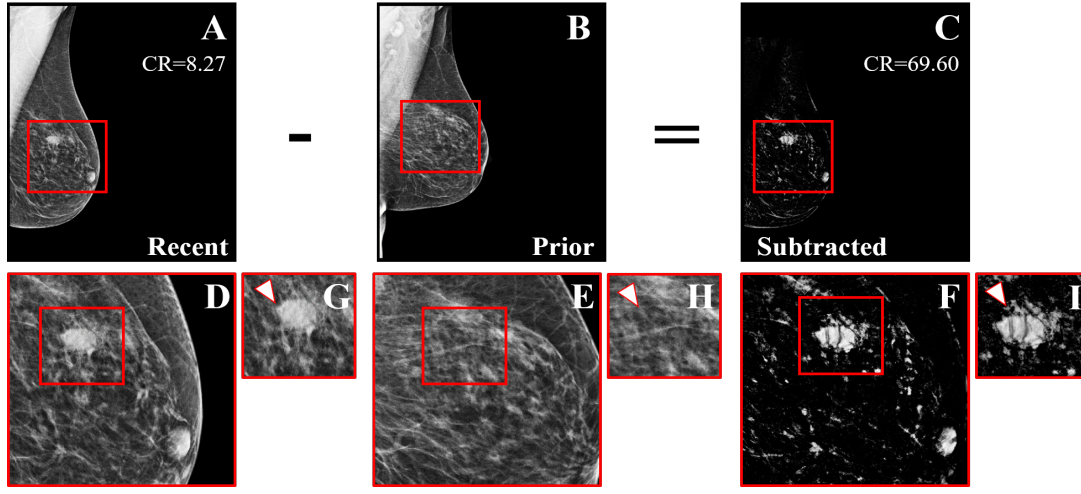


Fig. 5. Improvement of the Contrast Ratio (CR). (A) Most recent mammogram. (B) Prior mammogram. (C) The result of subtracting the registered version of B from A. (D)-(F) are zoomed regions marked by the red squares in A-C. The contrast ratio has increased 8 times, after subtraction. The red squares in D-F correspond to the zoomed images (G)-(I) and show a new suspicious mass, which was not subtracted.

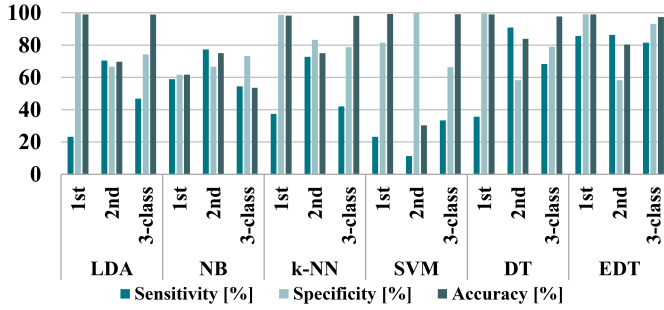


Fig. 6. Classification results for the detection of breast masses using temporal subtraction for the 1st round, 2nd round and 3-class classification.

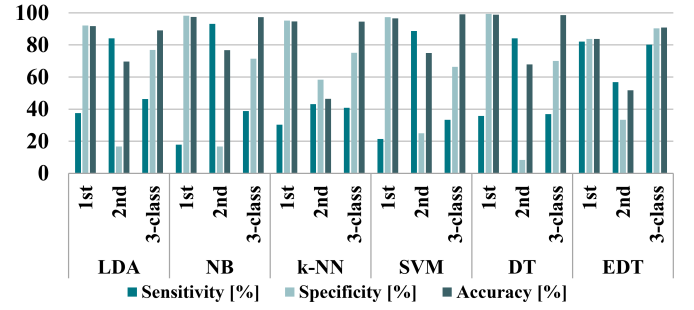


Fig. 7. Classification results for the detection of breast masses using temporal analysis for the 1st round, 2nd round and 3-class classification.

as before, reached a high of 96.51% accuracy and 0.91 AUC (Fig. 6).

B. Mass Detection and Classification using Temporal Analysis

Temporal analysis was also performed for comparison purposes. For the elimination of FPs, the same features as temporal subtraction were used. EDT achieved the highest performance with 87.04% accuracy and 0.85 AUC (Fig. 7). For the classification of masses as benign or suspicious, an EDT model reached 50% accuracy and 0.44 AUC (Fig. 7), using again the same features as temporal subtraction. In the 3-class classification, 90.83% accuracy and 0.87 AUC (Fig. 7) were achieved with the implementation of EDT and, again, the same selected features.

IV. DISCUSSION

As expected, the implementation of temporal subtraction increases the CR, effectively removes the background and renders the new changes increasingly visible to the radiologists. The highest classification performance was achieved using EDT in a 3-class model (96.51%) and there were 2.1 FP detections per image.

These results are far superior compared to those using temporal analysis (Fig. 8). Even though EDT provided 90.83% accuracy in the 3-class classification, 5 benign masses (out of 12) and 5 suspicious masses (out of 44) were misclassified as benign. In total, there were 5.6 FP detections per image. Using temporal subtraction, the average accuracy of mass classification increased by 6%, a statistically significant improvement (p -value < 0.05).

Temporal subtraction has recently been shown to be effective for the detection and classification of micro-calcifications [8]. Studies related to temporal subtraction for masses were not found in the literature, hence, direct comparison with other studies was not possible. Some groups assessed the use of temporal analysis for the detection of masses with promising results (0.88 AUC [5], 0.77 AUC [6], 0.90 AUC [7]), as also shown in this study (0.91 AUC). The results presented in this paper, prove that temporal subtraction can achieve higher performance and can improve the diagnosis of breast masses.

It is important to note that, in some studies, the ROIs were randomly divided for the validation and training. Parts of the same image, thus from the same patient, were used in both

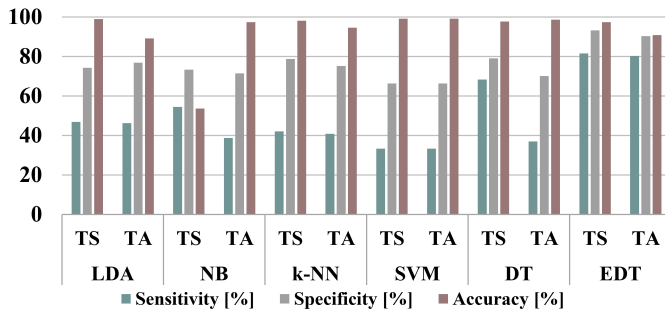


Fig. 8. Comparison between Temporal Subtraction (TS) and Temporal Analysis (TA) for the 3-class classification of breast masses.

training and testing of the classification model. In this work, leave-one-patient-out cross-validation was used, dividing the patients and not the ROIs into testing and training, to avoid any bias.

V. CONCLUSION

In this work, a new dataset consisting of 40 pairs of digital sequential mammograms was created, to evaluate the breast mass classification performance of temporal subtraction, compared to temporal analysis, in combination with image registration and machine learning. With temporal subtraction, the unchanging details in the mammogram, e.g. the background, were eliminated and the CR improved by $\sim 7\times$. The classification performance increased by 6% compared to temporal analysis, achieving 96.51% vs. 90.83% accuracy. Moreover, with the implementation of temporal subtraction, the average number of FP detections per mammogram decreased from 5.6 to 2.1. However, these results are preliminary, given the small dataset considered. As a result, further studies must be conducted to include more cases and different validation methods. The further expansion and improvement of the findings of this study have the potential to substantially contribute to the development of automated CAD systems with significant impact on patient prognosis.

ACKNOWLEDGMENT

This research is 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.

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] 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 Reviews in biomedical engineering*, vol. 6, pp. 77–98, 2012.
- [2] P. 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.

- [3] A. Oliver, J. Freixenet, J. Martí, E. Perez, J. Pont, E. R. Denton, and R. Zwigglar, "A review of automatic mass detection and segmentation in mammographic images," *Medical image analysis*, vol. 14, no. 2, pp. 87–110, 2010.
- [4] S. J. S. Gardezi, A. Elazab, B. Lei, and T. Wang, "Breast cancer detection and diagnosis using mammographic data: Systematic review," *Journal of medical Internet research*, vol. 21, no. 7, p. e14464, 2019.
- [5] 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," *Medical Physics*, vol. 28, no. 11, pp. 2309–2317, 2001.
- [6] S. Timp, C. Varela, and N. Karssemeijer, "Temporal change analysis for characterization of mass lesions in mammography," *IEEE Transactions on Medical imaging*, vol. 26, no. 7, pp. 945–953, 2007.
- [7] F. Ma, L. Yu, G. Liu, and Q. Niu, "Computer aided mass detection in mammography with temporal change analysis," *Computer Science and Information Systems*, vol. 12, no. 4, pp. 1255–1272, 2015.
- [8] K. Loizidou, G. Skouroumouni, C. Nikolaou, and C. Pitris, "An automated breast micro-calcification detection and classification technique using temporal subtraction of mammograms," *IEEE Access*, vol. 8, pp. 52785–52795, 2020.
- [9] S. Agrawal, R. Rangnekar, D. Gala, S. Paul, and D. Kalbande, "Detection of breast cancer from mammograms using a hybrid approach of deep learning and linear classification," in *2018 International Conference on Smart City and Emerging Technology (ICSCET)*, pp. 1–6, IEEE, 2018.
- [10] S.-C. Huang, F.-C. Cheng, and Y.-S. Chiu, "Efficient contrast enhancement using adaptive gamma correction with weighting distribution," *IEEE Transactions on Image Processing*, vol. 22, no. 3, pp. 1032–1041, 2013.
- [11] W. Gonzalez and Eddins, *Digital image processing using MATLAB*. Gatesmark Publishing, 2nd ed., 2010.
- [12] Y. Diez, A. Oliver, X. Llado, J. Freixenet, J. Martí, J. C. Vilanova, and R. Martí, "Revisiting intensity-based image registration applied to mammography," *IEEE Transactions on Information Technology in Biomedicine*, vol. 15, no. 5, pp. 716–725, 2011.
- [13] H. Lu, M. Reyes, A. Šerifović, S. Weber, Y. Sakurai, H. Yamagata, and P. C. Cattin, "Multi-modal diffeomorphic demons registration based on point-wise mutual information," in *2010 IEEE International Symposium on Biomedical Imaging: From Nano to Macro*, pp. 372–375, IEEE, 2010.
- [14] H. P. Chan, C. J. Vyborny, H. MacMahon, C. E. Metz, K. Doi, and E. A. Sickles, "Digital mammography. roc studies of the effects of pixel size and unsharp-mask filtering on the detection of subtle microcalcifications," *Investigative radiology*, vol. 22, no. 7, pp. 581–589, 1987.
- [15] N. Otsu, "A threshold selection method from gray-level histograms," *IEEE transactions on systems, man, and cybernetics*, vol. 9, no. 1, pp. 62–66, 1979.
- [16] S. Mika, G. Ratsch, J. Weston, B. Scholkopf, and K.-R. Mullers, "Fisher discriminant analysis with kernels," in *Neural networks for signal processing IX: Proceedings of the 1999 IEEE signal processing society workshop (cat. no. 98th8468)*, pp. 41–48, Ieee, 1999.
- [17] N. I. Yassin, S. Omran, E. M. El Houbay, and H. Allam, "Machine learning techniques for breast cancer computer aided diagnosis using different image modalities: A systematic review," *Computer methods and programs in biomedicine*, vol. 156, pp. 25–45, 2018.
- [18] C. Cortes and V. Vapnik, "Support-vector networks," *Machine learning*, vol. 20, no. 3, pp. 273–297, 1995.
- [19] J. R. Quinlan, "Induction of decision trees," *Machine learning*, vol. 1, no. 1, pp. 81–106, 1986.
- [20] D. Opitz and R. Maclin, "Popular ensemble methods: An empirical study," *Journal of artificial intelligence research*, vol. 11, pp. 169–198, 1999.