BREAST MASS DETECTION AND CLASSIFICATION ALGORITHM BASED ON TEMPORAL SUBTRACTION OF SEQUENTIAL MAMMOGRAMS

Kosmia Loizidou* Galateia Skouroumouni† Christos Nikolaou† Costas Pitris*

*Department of Electrical and Computer Engineering, KIOS Research and Innovation Center of Excellence, University of Cyprus, Cyprus.
†Radiology Department, Nicosia General Hospital, Cyprus
‡Radiology Department, Limassol General Hospital, Cyprus

ABSTRACT
Breast cancer screening with mammography is the most efficient way to reduce breast cancer mortality. However, the large population and the use of double reading creates a high workload that heavily burdens the efficiency of the radiologists. Hence, Computer-Aided Detection (CAD) systems are being developed to assist the radiologists. In this study, an algorithm for the automatic detection and classification of breast masses, based on temporal subtraction of sequential mammograms, image registration and machine learning, is presented. While, some previous studies in the literature proposed temporal analysis by creating a new feature vector, temporal subtraction takes into consideration the entire prior image. A new dataset, consisting of 40 cases (two time points and two views of each breast per patient, a total of 160 images), with precisely annotated mass locations was created. The accuracy of the classification of masses as benign or suspicious increased from 85.7% (using temporal analysis) to 92.9% (using temporal subtraction). The improvement was statistically significant with p < 0.05. These results demonstrate the effectiveness of temporal subtraction for the diagnosis of breast masses.

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

1. INTRODUCTION
Breast cancer remains a significant medical concern in women’s health despite recent improvements in the survival rates. Radiologists worldwide use mammography as a reliable tool for breast cancer screening [1]. In the case of positive findings, the most appropriate disease management route is followed. According to the Breast Imaging Reporting and Data System (BI-RADS), numerous types of abnormalities can be found in a mammogram, including masses of various sizes and shapes [2]. Masses are characterized by their size, shape, intensity and texture, and radiologists examine those properties in order to classify them as benign or suspicious [3]. This classification is challenging, not only because of the large variation in size and shape of masses, but also because of their low image contrast [4]. Thus, Computer Aided Diagnosis (CAD) Systems are being developed to help radiologists with the detection and classification of breast masses [3].

While most studies in the literature propose different algorithms for the diagnosis of breast masses using the most recent mammogram [1, 3, 5], few methods compare recent and prior mammograms. Radiologists routinely exploit such comparisons to identify newly developing abnormalities. Some studies have assessed the use of temporal sequential images for the creation of a new technique, called temporal analysis [6, 7, 8]. In temporal analysis the detected Regions of Interest (ROIs) from the recent image are associated with the corresponding ROIs in the prior one, using image registration algorithms. Then, a new temporal feature vector is created and used for the classification. Despite the fact that the findings are promising, only parts of the prior image are used in the analysis.

In this study, an automatic algorithm for the detection and classification of breast masses using subtraction of temporally sequential mammograms, is presented. Temporal subtraction was previously developed by this group and applied successfully to the diagnosis of micro-calcifications [9]. Compared to temporal analysis, temporal subtraction uses the entire prior image which, after registration, is subtracted from the recent one. The unchanged regions are effectively removed and the Contrast Ratio (CR) is successfully increased. Subsequently, False Positive (FP) detections are eliminated and the true masses are classified as benign or suspicious. Temporal analysis was also performed in this study for comparison purposes.

2. MATERIALS AND METHODS
2.1. Dataset
For the purposes of this study, a new dataset was created since in the open access databases, only the recent images are available for each patient and those images are not precisely an-
Fig. 1. Diagram of the proposed methodology for breast mass detection and classification using sequential mammograms.

Fig. 2. Dataset example of a 67-year-old woman. (A) Original mammogram. (B) Zoomed region from A showing masses. (C) The region in B with precise marking of mass locations, as annotated by two expert radiologists.

noted to show the exact margins of a mass, limiting their value as ground truth. The dataset consists of 40 pairs of full-field digital sequential mammograms from Craniocaudal (CC, above) and Mediolateral oblique (MLO, side-angle) views, collected from 40 women undergoing routine screening mammography examinations (a total of 160 images). Consecutive mammograms were collected from 2012 to 2019 and the age of the participants varied from 40 to 75, with a mean age of 57.48 ± 8.96 years and median age of 57.50. Half the cases belonged to patients with at least one suspicious mass in the most recent mammogram. The remaining cases were obtained from healthy individuals without any signs of malignancy in either screening. These cases were selected to form a matched group compared to those with suspicious findings. This dataset also includes detailed annotations of the boundary of each mass (both benign and suspicious) to be used as the ground truth (Fig. 2). The selection and annotation was performed by two expert radiologists (ten and two years of experience).

2.2. Mass Detection and Classification using Temporal Subtraction

In temporal subtraction the prior and recent images were processed in parallel. Figure 1 shows the proposed methodology. First, normalization was applied, to adjust the range of pixel intensity values and then Contrast Limited Adaptive Histogram Equalization (CLAHE), Gamma correction and border removal 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 [10]. Contrast adjustment using Gamma correction was used to account for the non-linear mapping of intensities [11]. Border removal eliminated the high intensity areas connected to the border (including the pectoral muscle) [12].

For an effective subtraction of the sequential images, registration is required. Registration compensates for changes that occur in the breast over time and differences in the mammography procedure. Several registration techniques have been proposed in the literature [13]. In this study, Demons registration was used. Demons is a local technique that aligns the moving (prior) image to the fixed (recent) one, using regional similarity and location. In Demons the registration is viewed as a diffusion process and it can be represented as the energy function with respect to the update field, using fixed image, moving image and a transformation field [14].

After the registration, the prior registered image was subtracted from the recent one, to eliminate the regions that remained unchanged between the screenings, and the CR of the subtracted image was compared to that of the most recent image. Afterwards, the subtracted image was processed using unsharp-mask filtering, a spatial filter that subtracts a blurred version of the original image for the creation of a new enhanced image [15]. For the segmentation of breast masses, the difference image was converted to binary using Otsu’s thresholding and then morphological operations (erosion with a radius of 2 pixels and closing with a radius of 10 pixels) were applied. The remaining ROIs were considered as possible masses.

Machine learning was performed to eliminate the large number of FP detections. In total, 28 features were extracted, divided in four major categories: shape, intensity, First-Order Statistics (FOS) and Gray Level Co-occurrence Matrix features (GLCM). Those features were selected for their capability to identify breast masses based on their unique characteristics. They include: area, circularity, compactness, convex area, eccentricity, equivalent diameter, Euler number, ex-
Fig. 3. Improvement of the Contrast Ratio (CR) in a 67-year-old woman. (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. The CR has increased 7 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. 4. Plot comparing the contrast ratio (logarithmic scale) of the unprocessed recent image and the image created using temporal subtraction, for dense and fatty mammograms.

The performance of temporal subtraction was 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. 3 & 4). Overall, the CR increased 11.19 times after subtraction.
resulting in a new image accentuating the changes between the renders the new changes more visible to the radiologists.

purposes. For the elimination of FPs, the selected features pactness, convex area, Euler number, major and minor axis For the first round the selected features included area, com­
tensity, entropy, kurtosis, smoothness, standard deviation, variance, contrast, correlation, energy and homogeneity. Using those, the best combination of features was average and max in­
length, orientation, perimeter, shape ratio and variance. Us­


classification performance was increased by 7% compared to temporal analysis. The pro­
posed methodology was evaluated on a dataset containing 40 patients and in total, 160 images were used (two time points and two views of each breast). With temporal subtraction, the unchanging details in the images were eliminated and the CR improved by ~7 times (11.3 vs. 83.2), for dense and fatty mammograms, resulting in a new image accentuating the changes between the screenings.

As expected, the implementation of temporal subtraction increases the CR, effectively removes the background and renders the new changes more visible to the radiologists. For the first round the selected features included area, compactness, convex area, Euler number, major and minor axis length, orientation, perimeter, shape ratio and variance. Us­
ing those, voting achieved 98.2% accuracy and 0.91 AUC (Table 1). Likewise, for the second round, area, compactness, convex area, equivalent diameter, extent, major and minor axis length, orientation, perimeter and solidity were selected. ANN reached 92.9% accuracy along with 0.92 AUC (Table 2).

Temporal analysis was also performed for comparison purposes. For the elimination of FPs, the selected features included: min and average intensity, entropy, skewness, smoothness, standard deviation, variance, contrast, correlation and homogeneity. Voting achieved 83.8% accuracy and 0.79 AUC (Table 1). Moving on, for the second round the best combination of features was: average and max in­
tensity, entropy, kurtosis, smoothness, standard deviation, contrast, correlation, energy and homogeneity. Using those, ANN reached 85.7% accuracy along with 0.82 AUC (Table 2).

The highest classification performance was achieved using ANN with 92.9% accuracy. One benign mass (out of 12) was misclassified as suspicious and 3 suspicious masses (out of 44) were misclassified as benign. These results are far superior compared to those using temporal analysis. Even though ANN provided 85.7% accuracy, 4 benign and 4 suspi­
cious masses were misclassified. Using temporal subtraction, the average accuracy of mass classification increased by 7%, a statistically significant improvement (p < 0.05).

Temporal subtraction has recently been shown to be effective for the detection and classification of micro-calcifications, only by this group [9]. Studies related to temporal subtraction for diagnosis of breast 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 [6], 0.77 AUC [7], 0.90 AUC [8]), as also shown in this study (0.82 AUC). The results presented here, prove that temporal subtraction can achieve higher classification performance. It is important to note that in some studies, the ROIs were randomly divided for the validation and training. In this work, leave-one-patient-out cross-validation was used, dividing the patients, and not the ROIs, to avoid any bias.

### 4. CONCLUSION

This study illustrates that the use of temporal subtraction for the classification of breast masses, results in significantly higher performance, compared to temporal analysis. The proposed methodology was evaluated on a dataset containing 40 patients and in total, 160 images were used (two time points and two views of each breast). With temporal subtraction, the unchanging details in the images were eliminated and the CR improved by ~7x. Additionally, the classification performance was increased by 7% compared to temporal analysis (92.9% vs. 85.7% accuracy) proving the effectiveness of the proposed technique. Given the small dataset, further studies must be conducted with more image pairs and different validation methods. With further expansion and improvement, the methodology introduced in this study has the potential to substantially contribute to the development of automated CAD systems with significant impact on patient prognosis.

| Table 1. Comparison of the first classification round results using Temporal Subtraction (TS) and Temporal Analysis (TA) of sequential mammograms |
|-----------------|-----------------|-----------------|-----------------|-----------------|
| Classifier      | Accuracy [%]    | Sensitivity [%] | Specificity [%] | AUC             |
| LDA             | 89.30           | 94.64           | 89.25           | 0.92            |
| TA              | 55.39           | 55.36           | 55.39           | 0.55            |
| 5-NN            | 91.73           | 58.93           | 92.01           | 0.75            |
| TA              | 88.71           | 51.79           | 89.03           | 0.70            |
| SVM             | 86.89           | 92.86           | 86.84           | 0.90            |
| TA              | 54.06           | 46.43           | 54.12           | 0.50            |
| DT              | 98.24           | 62.50           | 98.55           | 0.81            |
| TA              | 41.43           | 92.86           | 40.98           | 0.70            |
| ET              | 99.57           | 76.79           | 99.77           | 0.88            |
| Voting          | 85.56           | 71.43           | 85.68           | 0.79            |
| ANN             | 91.73           | 57.14           | 92.03           | 0.75            |

| Table 2. Comparison of the second classification round results using Temporal Subtraction (TS) and Temporal Analysis (TA) of sequential mammograms |
|-----------------|-----------------|-----------------|-----------------|-----------------|
| Classifier      | Accuracy [%]    | Sensitivity [%] | Specificity [%] | AUC             |
| LDA             | 60.71           | 54.55           | 83.33           | 0.69            |
| TA              | 64.29           | 70.45           | 41.67           | 0.56            |
| 5-NN            | 78.57           | 83.33           | 77.27           | 0.80            |
| TA              | 78.57           | 83.33           | 77.27           | 0.80            |
| SVM             | 67.86           | 77.27           | 33.33           | 0.55            |
| TA              | 67.86           | 70.45           | 58.33           | 0.64            |
| DT              | 73.21           | 84.09           | 33.33           | 0.59            |
| ET              | 80.36           | 86.36           | 58.33           | 0.72            |
| Voting          | 73.21           | 86.36           | 25.00           | 0.56            |
| Bagging         | 71.43           | 77.27           | 50.00           | 0.64            |
| Voting          | 71.43           | 79.55           | 41.67           | 0.61            |
| ANN             | 92.86           | 93.18           | 91.67           | 0.92            |
| TA              | 85.71           | 90.91           | 66.67           | 0.82            |
5. COMPLIANCE WITH ETHICAL STANDARDS

The corresponding Institutional Review Board - National Bioethics Committee, approved the collection of the data and their use in this study. 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.

6. ACKNOWLEDGMENTS

This research 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.

7. REFERENCES

[1] K. Ganesan et al., “Computer-aided breast cancer detection using mammograms: a review,” IEEE Reviews in biomedical engineering, vol. 6, pp. 77-98, 2012.

[2] P. A. Hernández et al., “Breast calcifications: description and classification according to bi-rads 5th edition,” Rev. Chil. Radiol., vol. 22, pp. 80-91, 2016.

[3] A. Oliver et al., “A review of automatic mass detection and segmentation in mammographic images,” Medical image analysis, vol. 14, no. 2, pp. 87-110, 2010.

[4] W. Xie, Y. Li, and Y. Ma, “Breast mass classification in digital mammography based on extreme learning machine,” Neurocomputing, vol. 173, pp. 930-941, 2016.

[5] S. J. S. Gardezi et al., “Breast cancer detection and diagnosis using mammographic data: Systematic review,” Journal of medical Internet research, vol. 21, no. 7, p. e14464, 2019.

[6] L. Hadjiiski et al., “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.

[7] 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.

[8] F. Ma et al., “Computer aided mass detection in mammography with temporal change analysis,” Computer Science and Information Systems, vol. 12, no. 4, pp. 1255-1272, 2015.

[9] K. Loizidou et al., “An automated breast microcalcification detection and classification technique using temporal subtraction of mammograms,” IEEE Access, vol. 8, pp. 52785-52795, 2020.

[10] S. Agrawal et al., “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.

[11] 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.

[12] R. C. Gonzalez, R. E. Woods, and S. L. Eddins, Digital image processing using MATLAB. Gatesmark Publishing, 2nd ed., 2010.

[13] Y. Diez et al., “Revisiting intensity-based image registration applied to mammography,” IEEE Transactions on Information Technology in Biomedicine, vol. 15, no. 5, pp. 716-725, 2011.

[14] H. Lu et al., “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.

[15] H. P. Chan et al., “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.