

ORIGINAL ARTICLE

Open Access



# Digital subtraction of temporally sequential mammograms for improved detection and classification of microcalcifications

Kosmia Loizidou<sup>1\*</sup> , Galatea Skouroumouni<sup>2</sup>, Costas Pitris<sup>1</sup> and Christos Nikolaou<sup>3</sup>

## Abstract

**Background:** Our aim was to demonstrate that automated detection and classification of breast microcalcifications, according to Breast Imaging Reporting and Data System (BI-RADS) categorisation, can be improved with the subtraction of sequential mammograms as opposed to using the most recent image only.

**Methods:** One hundred pairs of mammograms were retrospectively collected from two temporally sequential rounds. Fifty percent of the images included no (BI-RADS 1) or benign (BI-RADS 2) microcalcifications. The remaining exhibited suspicious findings (BI-RADS 4-5) in the recent image. Mammograms cannot be directly subtracted, due to tissue changes over time and breast deformation during mammography. To overcome this challenge, optimised preprocessing, image registration, and postprocessing procedures were developed. Machine learning techniques were employed to eliminate false positives (normal tissue misclassified as microcalcifications) and to classify the true microcalcifications as BI-RADS benign or suspicious. Ninety-six features were extracted and nine classifiers were evaluated with and without temporal subtraction. The performance was assessed by measuring sensitivity, specificity, accuracy, and area under the curve (AUC) at receiver operator characteristics analysis.

**Results:** Using temporal subtraction, the contrast ratio improved ~ 57 times compared to the most recent mammograms, enhancing the detection of the radiologic changes. Classifying as BI-RADS benign *versus* suspicious microcalcifications, resulted in 90.3% accuracy and 0.87 AUC, compared to 82.7% and 0.81 using just the most recent mammogram ( $p = 0.003$ ).

**Conclusion:** Compared to using the most recent mammogram alone, temporal subtraction is more effective in the microcalcifications detection and classification and may play a role in automated diagnosis systems.

**Keywords:** Breast cancer, Mammography, Radiographic image interpretation (computer-assisted), Retrospective studies, Machine learning

\* Correspondence: [cloizi01@ucy.ac.cy](mailto:cloizi01@ucy.ac.cy)

<sup>1</sup>KIOS Research and Innovation Center of Excellence, Department of Electrical and Computer Engineering, University of Cyprus, 1 Panepistimiou Avenue, Aglantzia, 2109 Nicosia, Cyprus  
Full list of author information is available at the end of the article



© The Author(s). 2021 **Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

## Key points

- The contrast ratio of the subtracted images was improved  $\sim 57$  times, compared to that of the recent images without preprocessing.
- Eighteen per cent of pre-existing Breast Imaging Reporting and Data System (BI-RADS) category 2 findings were effectively removed and the remaining were detected with an accuracy of 94.1%.
- Accuracy and area under the curve of the classification of microcalcifications as BI-RADS 2 versus BI-RADS 4 or 5 were significantly higher with the use of temporal subtraction, compared to using only the most recent mammogram ( $p = 0.003$ ).

## Background

Breast cancer screening with mammography is an effective approach to reduce breast cancer mortality. However, the large population involved and the use of double reading increases the workload and can limit the efficiency of the screening process [1]. To further exacerbate the challenge, various types of abnormalities are associated with breast cancer, including microcalcifications [2]. They can be benign or form microcalcification clusters possibly suspicious, to be managed appropriately. For the BI-RADS classification of microcalcifications as benign or suspicious, morphology, distribution, and change over time are key parameters [3]. Computer-aided diagnosis (CAD) systems are being explored as a means to improve the specificity of the classification of mammographic anomalies without compromising the sensitivity [4]. Several groups [5–7] have assessed the use of CAD systems for the detection of microcalcifications present in the most recent mammographic views, with sensitivity and specificity in the range of 82–89% and 87–88%, respectively, when discriminating between benign and suspicious microcalcifications. The main drawback of those systems is the considerable number of false positives (FPs) per image, that can range up to 1–3, which reduces their clinical applicability [8]. In addition, they provide no information regarding the presence of those abnormalities in previous mammographic sessions.

In temporal analysis, mammograms from multiple prior examinations are utilised. When prior information is available for direct comparison, abnormalities can be detected at an earlier stage and the radiologists feel more confident of their assessment [9]. Some studies have attempted to combine information from prior and recent mammographic views to reduce the FP and recall rates [9–14]. Prior and recent images were coarsely registered based on anatomical features (e.g., nipple, skin, centre of mass which is the mean value across each dimension), and the locations of recently identified microcalcifications in the

prior image were identified by regional registration, *i.e.*, searching around the location of the recent finding. Combining the features from both images resulted in an improved specificity, specificity, and reduced FPs rates. However, temporal analysis offers no benefit, over using just the recent mammographic view, when the findings are new with no traces of abnormality in the prior screening [11].

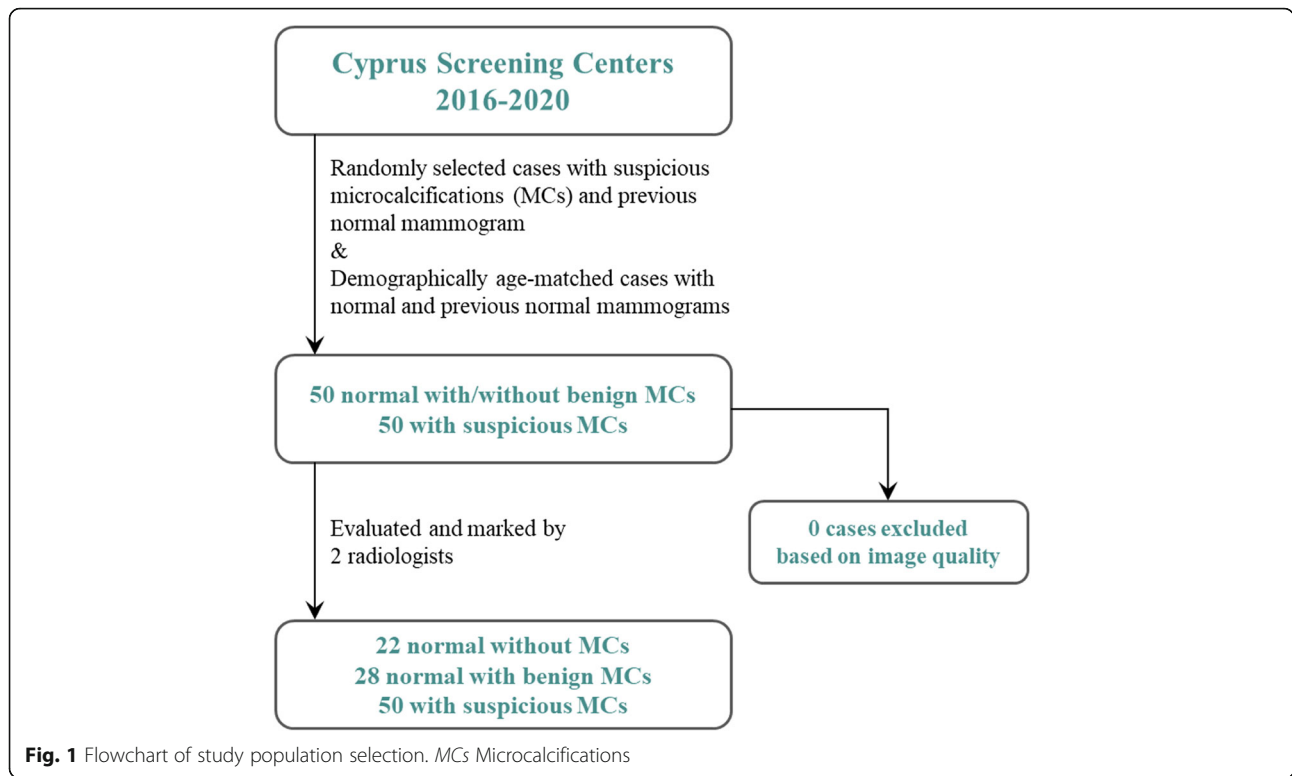
Going a step further, temporal subtraction begins with using both global and local features to register the entire breast areas of the recent and prior mammographic views. This allows direct subtraction of the images, digitally removing unchanged regions from the recent mammographic view, further delineating subtle recent changes, including microcalcifications [15, 16]. The objective of this study was to evaluate the effect of the subtraction of temporally sequential mammograms to eliminate unchanged features and improve the detection and classification of microcalcifications into benign and suspicious, based on their BI-RADS category. For comparison, the detection and classification methodologies were also applied to the most recent images without temporal subtraction.

## Methods

### Study population

The current study expands on prior published work [16]. From the 100 participants that were eventually included, 80 were previously used for the purposes of describing the technical details of the algorithms employed. For this retrospective study, 100 pairs of full-field digital mammograms were collected, between 2012 and 2020, from various local hospitals (Nicosia General Hospital, Limassol General Hospital, Cyprus population screening program Aglantzia and Linopetra), performed by women (38 to 83 years of age,  $60.07 \pm 7.09$ , mean  $\pm$  standard deviation) with either no microcalcifications or BI-RADS benign (normal population) or BI-RADS suspicious microcalcifications (suspicious population) in their current mammograms. A BI-RADS normal or benign prior mammogram (average interval of 2.2 years) was required for inclusion in the study (Fig. 1). The normal population was selected to form an age-matched group compared to the patients with BI-RADS suspicious microcalcifications. The study was approved by the appropriate Institutional Review Board (Cyprus National Bioethics Committee #EEBK EII 2020.01.144) and informed consent was retrospectively collected.

For every participant, two mammographic views (cranio-caudal and medio-lateral oblique) of the breast showing the presence of microcalcifications, from two sequential screening rounds, were included for a total of 400 images. A radiologist (C.N. with 10 years of experience) identified the patients to be included, according to



**Table 1** Characteristics of the population and digital mammography examinations selected for the study

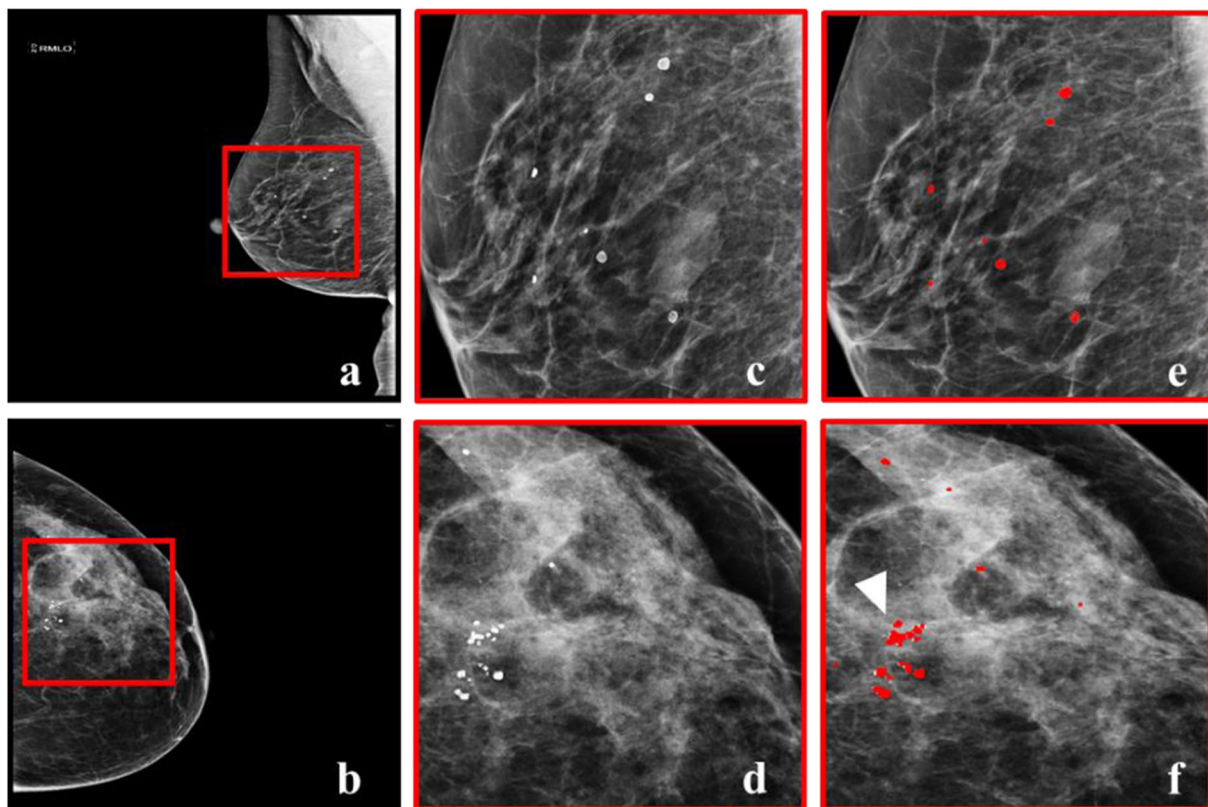
Variable	Population		
	BI-RADS normal (n = 50)	BI-RADS suspicious (n = 50)	Total (n = 100)
<b>Patient age</b>			
Mean ± standard deviation	59.42 ± 5.97	60.72 ± 8.00	60.07 ± 7.09
Median	59	61.5	59.5
Range	47–75	38–83	38–83
Interquartile range	55–64	55.75–66.25	55–65
<b>BI-RADS breast density</b>			
<i>a</i>	5	5	10
<i>b</i>	33	24	57
<i>c</i>	10	18	28
<i>d</i>	2	3	5
<b>BI-RADS classification</b>			
1	22	0	22
2	28	0	28
3	0	0	0
4a	0	27	27
4b	0	15	15
4c	0	6	6
5	0	2	2

the criteria specified above, and along with a second radiologist (G.S. with 2 years of experience), assessed the mammograms for assigning the BI-RADS category and marked the location of the microcalcifications. Interobserver agreement, *i.e.*, microcalcifications marked by both observers, was estimated at 97.1% for benign (BI-RADS 2) and 98.2% for suspicious (BI-RADS 4-5) microcalcifications; differences were resolved by consensus. A summary of the study population is shown in Table 1. Fifty percent of the mammograms came from healthy participants (28 with only BI-RADS benign microcalcifications in the prior and recent mammographic views and 22 with no visible microcalcifications). In the remaining 50%, at least one new BI-RADS suspicious microcalcification was present in the most recent mammographic view. The ground truth was based on the BI-RADS category, as evaluated by the radiologists, without any confirmation by follow-up (for “benign” lesions), or biopsy (for “suspicious” lesions). This data set not only included temporally sequential mammograms, but also precise annotation of each individual microcalcification to be used as a reference (Fig. 2). The data set included a total of 629 microcalcifications, 515 BI-RADS 2 and

114 BI-RADS 4 or 5. The size of the mammographic views was  $4096 \times 3328$  pixels, in an 8-bit DICOM format. This data is publicly available (<https://doi.org/10.5281/zenodo.5036062>).

#### Image registration, subtraction, and segmentation

Supplemental Fig. S1 shows the diagram of the proposed methodology for detection and BI-RADS classification of microcalcifications. The procedure began with image preprocessing for normalisation, border removal [17], and gamma correction [18]. Next, each prior mammographic view was co-registered to the most recent one. Image registration is a critical step in temporal subtraction, since it corrects for the changes that occur in the breast over time and due to deformation between mammograms. Demons registration [19], a non-rigid method based on local flows, was employed due to superior performance compared to other common approaches [20]. The prior registered mammographic view was subsequently subtracted from the recent, effectively removing the regions that have remained unchanged since the previous exam. The contrast ratio of the subtracted image, *i.e.*, the ratio of the maximum divided by the average



**Fig. 2** Dataset examples. **a** Most recent mammographic view of a woman (BI-RADS breast density class *b*) with BI-RADS benign microcalcifications. **b** Most recent mammographic view of woman (BI-RADS breast density class *c*) with BI-RADS benign and suspicious microcalcifications. **c, d** Zoomed regions from **a** and **b** showing microcalcifications. **e, f** The regions in **c** and **d** with precise marking of the location of microcalcifications, as annotated by two expert radiologists. The arrowhead in **f** points to a BI-RADS suspicious microcalcifications

intensity, was compared to that of the recent mammographic view to evaluate the effectiveness of the removal of the background. Furthermore, the capability of the subtraction to remove unchanged microcalcifications, thus reducing the FP rate, was also assessed. After range filtering [21], the intensity values of each image were converted to binary, *i.e.*, 0 or 1, using an intensity threshold, obtained by optimising the global BI-RADS classification rate. The binary image was further processed morphologically. The operations of closing (removing small unconnected regions) and opening (filling small gaps) were applied and the remaining regions were considered as possible microcalcifications.

#### Feature extraction and selection for classification

To differentiate the true microcalcifications from other tissue and, subsequently, the BI-RADS benign from the BI-RADS suspicious microcalcifications, using machine learning algorithms, several features were estimated from every possible region containing microcalcifications. Ninety-six, shape, intensity first order statistic, and grey-level co-occurrence matrix (GLCM) features were extracted [22–24]. The GLCM was calculated at 0°, 45°, 90°, and 135° and 5-, 15-, and 25-pixel offset. The mean and standard deviation of each GLCM texture property were obtained, resulting in 72 features. Hypothesis test (*t*-test) [25] and feature importance [26] were employed to identify the most significant features and further evaluate their contribution to the classification.

#### Training and comparison of classifier designs

For the BI-RADS classification, 9 classifiers were evaluated: 9-nearest neighbors [1], decision trees [27], random forest [27], multi-layer perceptron [28], adaptive boosting [29], bagging [30], gradient boosting [31], and ensemble voting [32]. In addition, different neural network configurations were evaluated using Python (Python Software Foundation, Wilmington, USA, v. 3.7.7) and Keras (Keras Special Interest Group, François Chollet, Mountain View, USA, v. 2.3.1) [33]. The resulting, most suitable, network consisted of 7 fully connected layers, with 986,738 trainable parameters. Rectified linear unit was used as an activation function and adaptive dropout regularisation was included every two hidden layers. Gaussian noise was added after dropout, as a regularisation term, in order to increase the robustness of the network. Batch size was set to 128 and the network was trained for 100 epochs. Traditional classifiers were selected based on their prior application to mammography.

The complete dataset was used during the training stage with leave-one-patient-out (LOPO) cross-validation. This cross-validation approach was critical in order to avoid bias from including images of the same patient in both test and training sets. K-fold cross-validation was also considered, again by dividing the

patients into folds. Initially, the possible microcalcifications were classified as normal tissue or true microcalcifications and, subsequently, the true microcalcifications were classified as BI-RADS benign or suspicious.

#### Detection and classification using the most recent mammogram alone

For comparison purposes, the same classification approach was optimised and applied to the most recent mammograms, without temporal subtraction, to verify the benefit of temporal subtraction.

#### Statistical analysis

The classification performance was evaluated by computing the sensitivity, specificity, accuracy, and the area under the curve (AUC) at receiver operator characteristics analysis. The cut-off values for calculating sensitivity and specificity were selected optimally so that BI-RADS false positive and false negative numbers are minimised, *i.e.*, the cross point of the positive and negative distributions. In the comparison of the results from using just the most recent mammogram, the Fisher test was used with a level of statistical significance set to  $p = 0.05$ .

## Results

#### Image registration, subtraction, and segmentation

Image registration and subtraction yielded an average 72% reduction of image intensity, a result of removing structures that have remained unchanged between screenings. The average contrast ratio of the subtracted images was ~ 57 times higher compared to the recent mammographic view (Fig. 3). Eighteen percent of old BI-RADS benign microcalcifications were effectively removed (Table 2). It is also important to note that none of the BI-RADS suspicious microcalcifications were removed by this process (Fig. 4). The processing time for these operations was an average of ~ 15 min per image pair (Intel® Core™ i7 2 GHz; Intel Corp., Santa Clara, CA, USA).

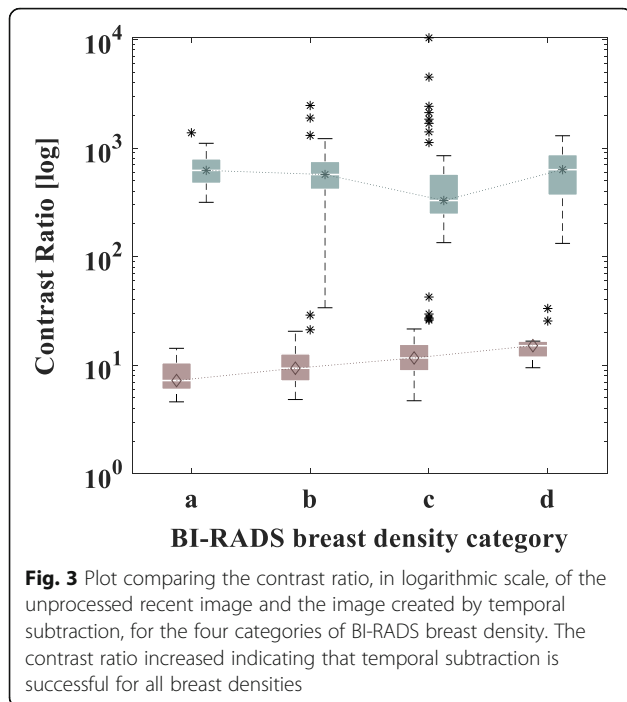
#### Feature extraction and selection for classification

Features of microcalcifications were extracted from the images as described in the previous section. Based on the results of *t*-test and feature importance, the features with the most significant contribution to classification were identified for each classification round. The details of the features that have been selected in each case can be found in Supplemental Table S1.

#### Detection and classification using temporal subtraction

The results of the detection of the microcalcifications are summarised in Table 3. The sensitivity, specificity, accuracy, and AUC of the different methods were in the range of 72.8–82.3%, 83.7–97%, 83.5–95%, and 0.82





−0.88%, respectively. Based on the AUC, the most successful classification scheme was the ensemble voting with 81.4% sensitivity, 95.5% specificity, 94.1% accuracy, and 0.88 AUC. The application of k-fold cross-validation (Fig. 5a), using  $k = 4, 5,$  and  $10,$  verified that the performance remained approximately at the same level.

The optimisation of the various classifiers for the classification of microcalcifications as benign or suspicious according to their BI-RADS category, using LOPO cross-validation, resulted in the outcomes summarised in Table 4. The sensitivity, specificity, accuracy and AUC were in the range of 57.9–84.2%, 77.8–92.2%, 79.0–90.3% and, 0.70–0.87%, respectively. Based on the AUC, the most successful classification scheme was again ensemble voting with 81.6% sensitivity, 92.2% specificity, 90.3% accuracy, and 0.87 AUC. For this classification round, only 13-fold cross-validation was applied (Fig. 6a) due to the smaller number of patients with microcalcifications (*i.e.*, 78 patients with true

microcalcifications, 50 BI-RADS suspicious, and 28 BI-RADS normal with benign microcalcifications). Example images visually demonstrating the classification outcome, using temporal subtraction, where BI-RADS benign and suspicious microcalcifications were correctly identified, are shown in Supplemental Fig. S2.

**Detection and classification using the most recent mammogram alone**

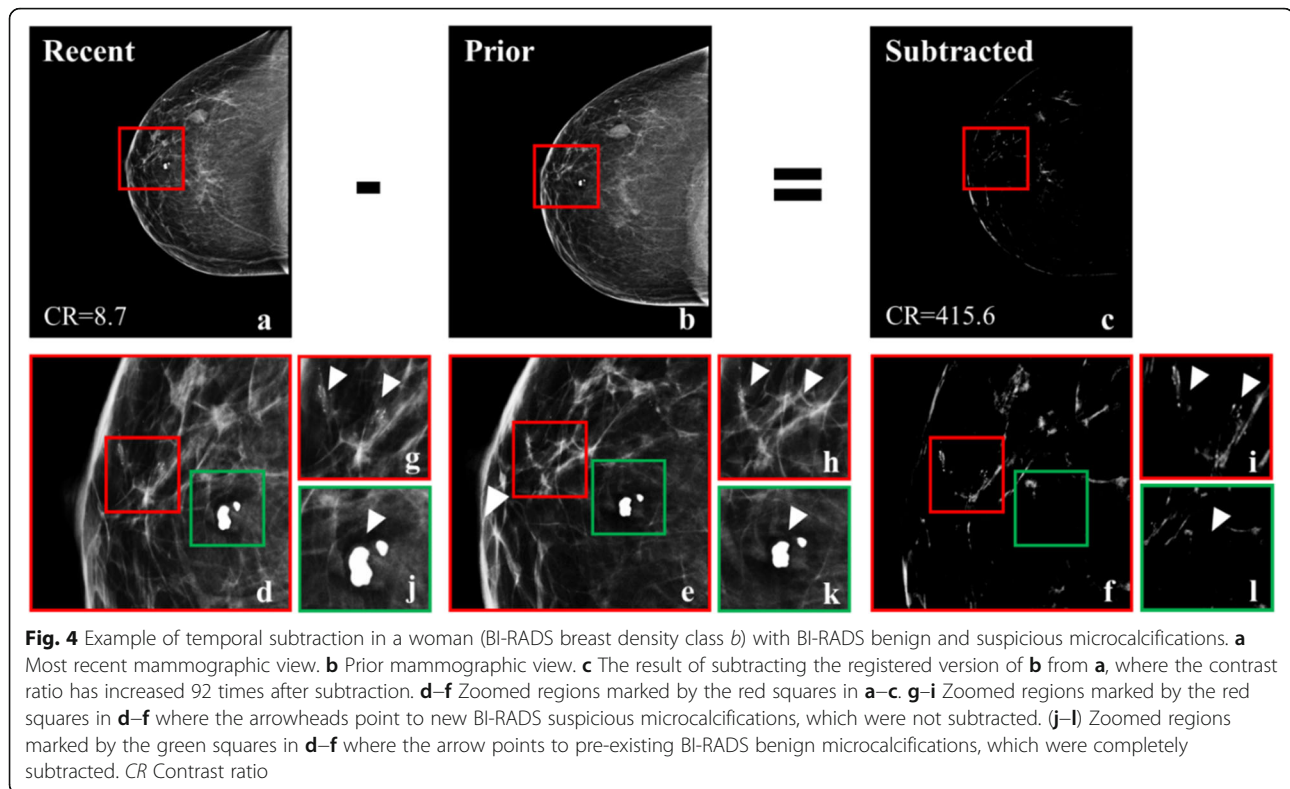
Table 3 shows the classification results for the identification of true microcalcifications using features selected only from the most recent mammographic view (Supplemental Table S1) and the same classifiers as before optimised for these features. The best classification performance was achieved using ensemble voting with 72.7% sensitivity, 96.5% specificity, 94.1% accuracy, and 0.83 AUC. Subsequently, the true microcalcifications were classified as BI-RADS benign or suspicious and the results are presented in Table 4. Again, the ensemble voting was the most successful method, providing 78.9% sensitivity, 83.5% specificity, 82.7% accuracy, and 0.81 AUC. As before, k-fold cross-validation was also performed, using the same values of  $k$  (Figs. 5b and 6b) showing that the algorithm was also stable and robust.

**Discussion**

A method for the detection and classification of microcalcifications, according to their BI-RADS category, from the subtraction of temporally sequential mammographic views was developed. The aim of this work was to combine temporal subtraction with machine learning in order to enhance the contrast ratio, eliminate the radiologically unchanged microcalcifications and, most importantly, improve the classification accuracy of microcalcifications as benign or suspicious, based on their BI-RADS categories. For effective and efficient subtraction of the prior from the most recent mammographic view, preprocessing, registration, and postprocessing procedures were applied. Machine learning techniques were then used to eliminate the FPs, *i.e.*, normal tissue misclassified as microcalcifications, and, furthermore, to classify the true microcalcifications as BI-

**Table 2** Elimination of old microcalcifications that appear in both screening rounds, in BI-RADS normal and suspicious mammograms

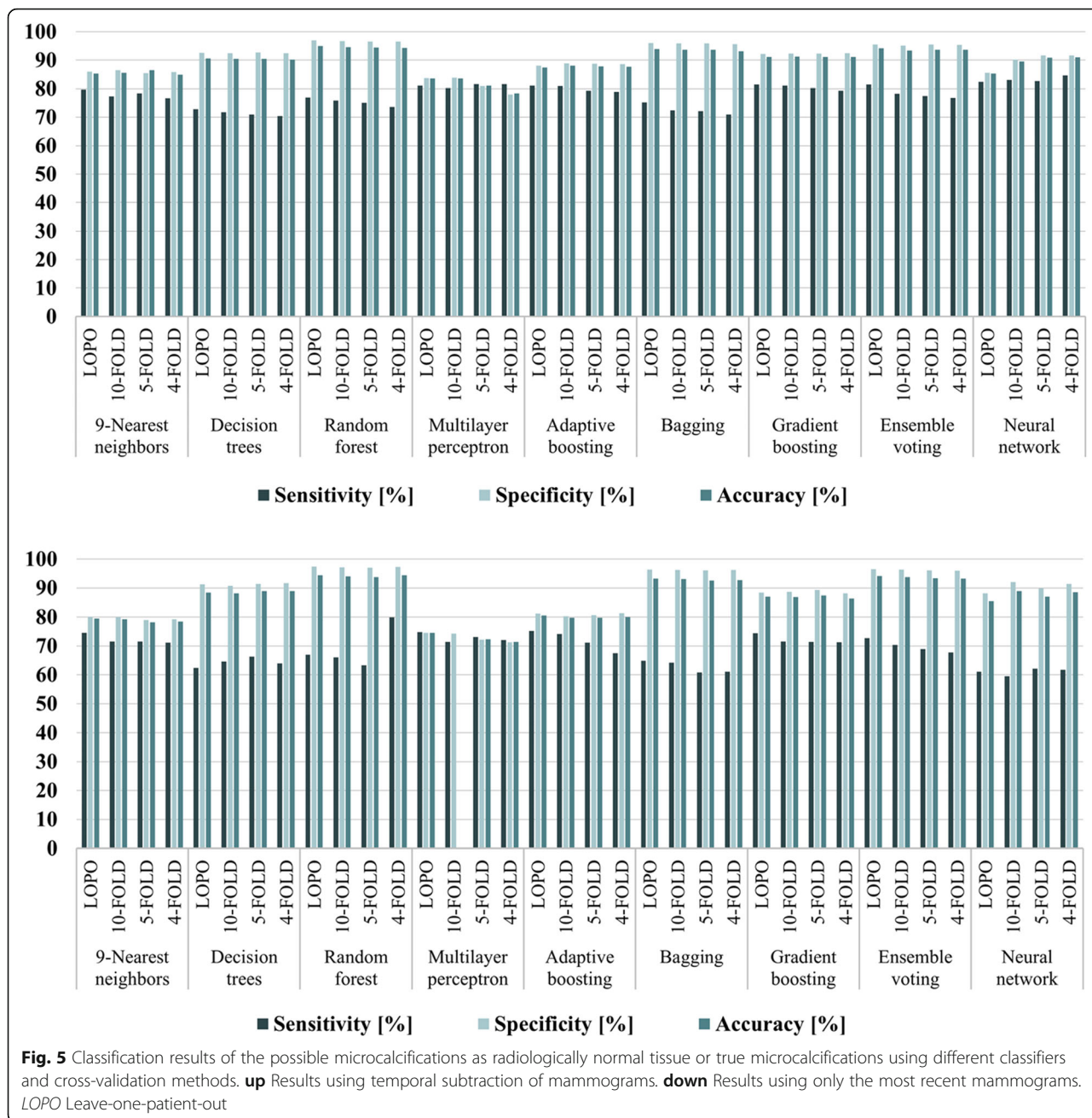
Mammograms	New microcalcifications	Old microcalcifications	Overlapping microcalcifications	Overlap (%)	Reduction (%)
BI-RADS normal (n = 200)	248	224	54	54/224 (24.1)	54/248 (21.8)
BI-RADS suspicious (n = 200)	398	319	64	64/319 (20.1)	64/398 (16.1)
<b>Total (n = 400)</b>	<b>646</b>	<b>543</b>	<b>118</b>	<b>118/543 (21.7)</b>	<b>118/646 (18.3)</b>



**Table 3** Comparison of the classification results of the possible microcalcifications as normal tissue or radiologically true microcalcifications with temporal subtraction (TS) of mammograms or using only the most recent mammograms (RM), in a leave-one-patient-out cross-validation scheme

Classifier	Sensitivity (%)			Specificity (%)			Accuracy (%)			AUC	
<b>9-Nearest neighbors</b>	TS	501/629	(79.7)	TS	4928/5739	(85.9)	TS	5429/6368	(85.3)	TS	0.83
	RM	469/629	(74.6)	RM	4593/5739	(80.0)	RM	5062/6368	(79.5)	RM	0.76
<b>Decision trees</b>	TS	458/629	(72.8)	TS	5311/5739	(92.5)	TS	5769/6368	(90.6)	TS	0.83
	RM	392/629	(62.3)	RM	5234/5739	(91.2)	RM	5626/6368	(88.4)	RM	0.78
<b>Random forest</b>	TS	484/629	(77.0)	TS	5565/5739	(97.0)	TS	6049/6368	(95.0)	TS	0.87
	RM	421/629	(66.9)	RM	5591/5739	(97.4)	RM	6012/6368	(94.4)	RM	0.82
<b>Multilayer perceptron</b>	TS	510/629	(81.1)	TS	4806/5739	(83.7)	TS	5316/6368	(83.5)	TS	0.82
	RM	470/629	(74.7)	RM	4275/5739	(74.5)	RM	4745/6368	(74.5)	RM	0.73
<b>Adaptive boosting</b>	TS	510/629	(81.1)	TS	5561/5739	(88.0)	TS	6071/6368	(87.3)	TS	0.85
	RM	473/629	(75.2)	RM	4659/5739	(81.2)	RM	5132/6368	(80.6)	RM	0.77
<b>Bagging</b>	TS	473/629	(75.2)	TS	5510/5739	(96.0)	TS	5983/6368	(94.0)	TS	0.86
	RM	408/629	(64.9)	RM	5528/5739	(96.3)	RM	5936/6368	(93.2)	RM	0.80
<b>Gradient boosting</b>	TS	512/629	(81.4)	TS	5291/5739	(92.2)	TS	5803/6368	(91.1)	TS	0.87
	RM	468/629	(74.4)	RM	5073/5739	(88.4)	RM	5541/6368	(87.0)	RM	0.80
<b>Ensemble voting</b>	<b>TS</b>	<b>512/629</b>	<b>(81.4)</b>	<b>TS</b>	<b>5480/5739</b>	<b>(95.5)</b>	<b>TS</b>	<b>5992/6368</b>	<b>(94.1)</b>	<b>TS</b>	<b>0.88</b>
	<b>RM</b>	<b>457/629</b>	<b>(72.8)</b>	<b>RM</b>	<b>5536/5739</b>	<b>(96.5)</b>	<b>RM</b>	<b>5993/6368</b>	<b>(94.1)</b>	<b>RM</b>	<b>0.83</b>
<b>Neural network</b>	TS	518/629	(82.4)	TS	4911/5739	(85.6)	TS	5429/6368	(85.3)	TS	0.84
	RM	384/629	(61.1)	RM	5058/5739	(88.1)	RM	5448/6368	(85.5)	RM	0.76

*AUC* Area under the curve



RADS benign or suspicious. Bagging, gradient boosting, ensemble voting, and neural networks are more recent additions to machine learning and were chosen for their potential to provide improved classification. Ideally, the proposed method should have also been verified on an independent external dataset. However, publicly available datasets do not provide sequential mammograms or images annotated at the level of individual microcalcifications.

Demons registration [19] was very effective in matching the prior to the recent mammographic views, since it

accounted for the complex transformations and distortions that appear between screenings. Using the proposed technique, the contrast ratio improved ~ 57 times enhancing the contrast of the recent changes in the images. The elimination of most of the background and unchanged BI-RADS benign microcalcifications can make the radiologic evaluation of mammograms and the detection of even subtle abnormalities, easier and faster. This reduces the effort and time expended by the radiologist by enhancing the new and, most probably, more diagnostically useful information.



**Table 4** Comparison of the classification results of the true microcalcifications as BI-RADS benign or suspicious using temporal subtraction (TS) of mammograms and using only the most recent mammograms (RM), in a leave-one-patient-out cross-validation scheme

Classifier	Sensitivity (%)			Specificity (%)			Accuracy (%)			AUC	
<b>9-Nearest neighbors</b>	TS	96/114	(84.2)	TS	393/515	(76.3)	TS	489/629	(77.7)	TS	0.80
	RM	73/114	(64.0)	RM	357/515	(69.3)	RM	430/629	(68.4)	RM	0.67
<b>Decision trees</b>	TS	66/114	(57.9)	TS	446/515	(86.6)	TS	512/629	(81.4)	TS	0.72
	RM	65/114	(57.0)	RM	445/515	(86.4)	RM	510/629	(81.1)	RM	0.72
<b>Random forest</b>	TS	73/114	(64.0)	TS	461/515	(89.5)	TS	534/629	(84.9)	TS	0.77
	RM	70/114	(61.4)	RM	452/515	(87.8)	RM	522/629	(83.0)	RM	0.75
<b>Multilayer perceptron</b>	TS	93/114	(81.6)	TS	411/515	(79.8)	TS	504/629	(80.1)	TS	0.81
	RM	69/114	(60.5)	RM	302/515	(58.6)	RM	371/629	(59.0)	RM	0.6
<b>Adaptive boosting</b>	TS	86/114	(75.4)	TS	433/515	(84.1)	TS	519/629	(82.5)	TS	0.8
	RM	80/114	(70.2)	RM	430/515	(83.5)	RM	510/629	(81.1)	RM	0.77
<b>Bagging</b>	TS	69/114	(60.5)	TS	458/515	(88.9)	TS	527/629	(83.8)	TS	0.75
	RM	65/114	(57.0)	RM	450/515	(87.4)	RM	515/629	(81.9)	RM	0.72
<b>Gradient boosting</b>	TS	82/114	(71.9)	TS	447/515	(86.8)	TS	529/629	(84.1)	TS	0.79
	RM	77/114	(67.5)	RM	438/515	(85.1)	RM	515/629	(81.9)	RM	0.76
<b>Ensemble voting</b>	<b>TS</b>	<b>93/114</b>	<b>(81.6)</b>	<b>TS</b>	<b>475/515</b>	<b>(92.2)</b>	<b>TS</b>	<b>568/629</b>	<b>(90.3)</b>	<b>TS</b>	<b>0.87</b>
	<b>RM</b>	<b>90/114</b>	<b>(79.0)</b>	<b>RM</b>	<b>430/515</b>	<b>(83.5)</b>	<b>RM</b>	<b>520/629</b>	<b>(82.7)</b>	<b>RM</b>	<b>0.81</b>
<b>Neural network</b>	TS	89/114	(78.1)	TS	450/515	(87.4)	TS	539/629	(85.7)	TS	0.83
	RM	83/114	(72.8)	RM	485/515	(94.2)	RM	568/629	(90.3)	RM	0.83

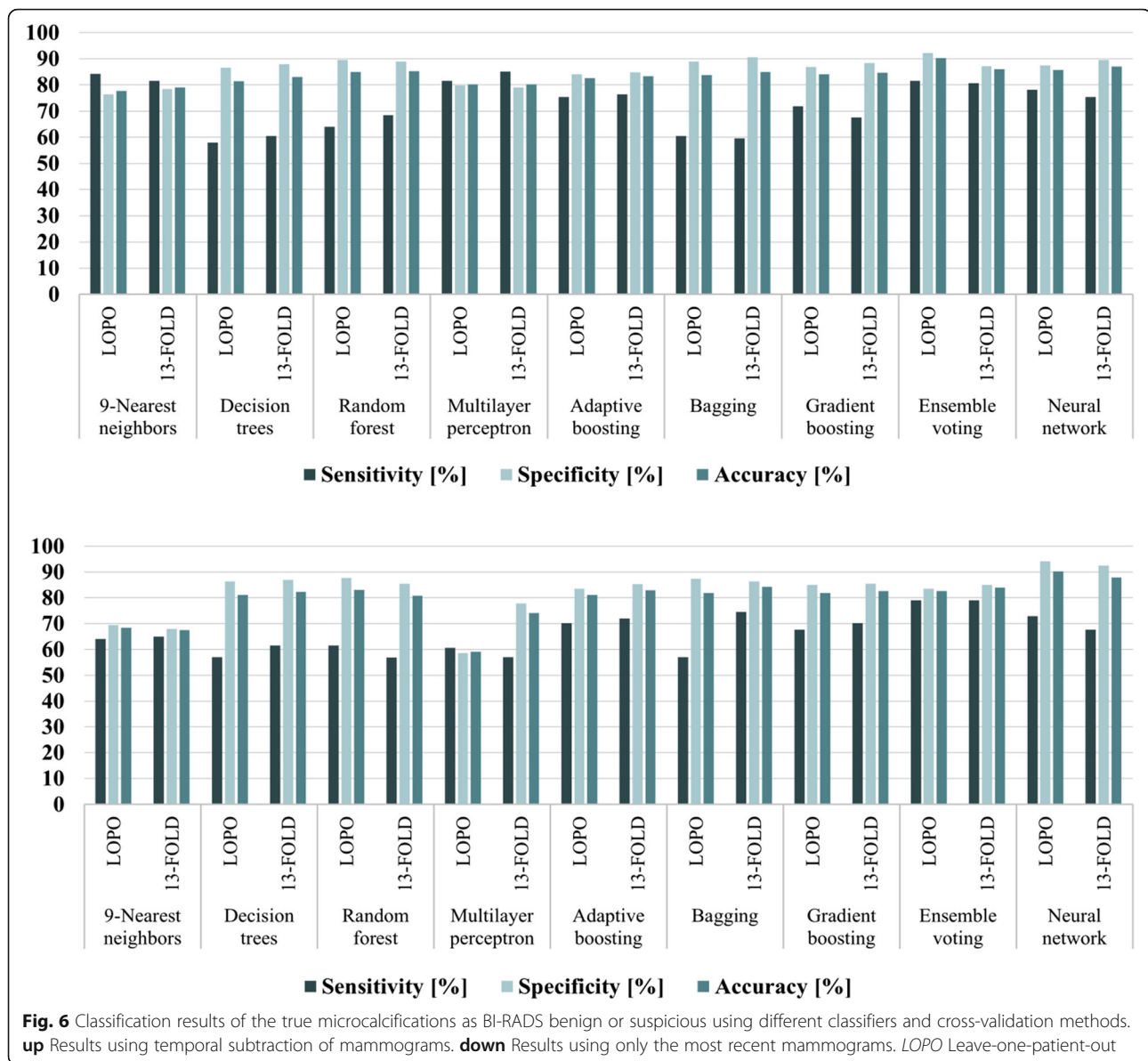
AUC Area under the curve

With automated BI-RADS classification, the proposed method achieved 81.4% sensitivity, 95.5% specificity and 94.1% accuracy for the detection of true microcalcifications and 81.6% sensitivity, 92.2% specificity, and 90.3% accuracy for the classification of microcalcifications as BI-RADS benign or suspicious, both using an optimised feature set and an ensemble voting model. This high accuracy demonstrates the effectiveness of the algorithm, which provided statistically superior performance over using only the most recent mammograms. The average classification accuracy for the characterisation of true microcalcifications as BI-RADS benign or suspicious improved by 7% with the introduction of temporal subtraction ( $p = 0.003$ ). To evaluate the robustness of the method, k-fold cross-validation was implemented, in addition to LOPO cross-validation. The algorithm performed at approximately the same level in all the cross-validation scenarios, indicating that the proposed method is robust and should be able to function equally well as new data becomes available.

The main limitation of this study is the relatively limited dataset acquired from local hospitals with a single protocol. Even though the results presented here are promising, more sequential pairs are required to definitively prove the generalizability of the proposed algorithm. Unfortunately, publicly available databases cannot be exploited for the purposes of this project, since they neither contain sequential mammograms nor they include images annotated at the level of individual

microcalcifications as in this study. Another limitation is the fact that although the suspicious microcalcifications were identified by two expert radiologists, the BI-RADS classification of clusters of microcalcifications not only varies from one radiologist to another, but might also be disproved by follow-up or pathology. In addition, the adoption of the BI-RADS classification as the ground truth, without any confirmation by follow-up or pathology, limits the generalizability and the ultimate utility of the tool. This is a consequence of mimicking the human reader rather than offering an unbiased opinion based on the true and confirmed state of the microcalcifications.

The results of this study cannot be readily and directly compared to other state-of-the-art techniques described in the literature for various reasons. The existing and freely available image databases contain only one mammogram per patient (*i.e.*, no prior information). In addition, in several cases, entire images are classified rather than individual microcalcifications [34]. Furthermore, in most state-of-the-art algorithms, cross-validation is implemented by randomly dividing the microcalcifications into training and test sets, or by using part of the same image in the test and another part in the training set [35]. Such approaches can introduce bias, which results in artificially improved classification results. In general, most studies in the literature report accuracy and AUC, which in the case of benign *versus* malignant classification of microcalcifications range



between 80–89% and 0.86–0.92% (Supplemental Table S2). In this study, a more appropriate approach was selected, which assigned the entire data set corresponding to a patient either to the training or the test set, and performed LOPO cross-validation. In order to prove the effectiveness of temporal subtraction in a fair manner, the results were compared to an optimised algorithm using only the most recent mammograms of the same dataset.

There are no examples of temporal analysis for microcalcification identification and classification in the literature. Furthermore, there are only a few studies on the use of temporal analysis for the characterisation of mass lesions [9–11, 13]. Rather than image subtraction, their approach was to extract several features from the recent and prior mammograms separately and, then, combine

them to improve the classification accuracy. Their results confirmed that the use of prior information could improve the detection and classification of mass lesions. However, temporal analysis offers no benefit when there is no abnormality in the prior screening.

In conclusion, the proposed technique demonstrates that temporal subtraction achieves superior performance in the detection and classification of microcalcifications, based on their BI-RADS category, compared to using only the most recent mammograms. The inclusion of more patients as well as the extension of the approach to detect and classify other abnormalities in mammograms (*e.g.*, breast masses or distortions) can further enhance the diagnostic potential of temporal subtraction. In the future, the proposed methodology has the

potential to substantially contribute to the development of automated CAD systems to assist in the radiologic classification of breast abnormalities and serve as a “second reader” or a “tie breaker” especially in low resource settings.

#### Abbreviations

AUC: Area under the curve; BI-RADS: Breast Imaging Reporting and Data System; CAD: Computer-aided diagnosis; FP: False positive; GLCM: Grey-level co-occurrence matrix; LOPO: Leave-one-patient-out

## Supplementary Information

The online version contains supplementary material available at <https://doi.org/10.1186/s41747-021-00238-w>.

**Additional file 1** Supplemental Fig. S1 Diagram of the proposed methodology for the detection and BI-RADS classification of breast microcalcifications using temporal subtraction of sequential mammograms. Supplemental Table S1 Features selected for the 1st and 2nd round of BI-RADS classification using temporal subtraction and only the most recent mammograms. Supplemental Fig. S2 Results of the classification of the radiologically true microcalcifications as BI-RADS benign or suspicious in the most recent mammographic view of a woman (BI-RADS breast density class b). (a) Most recent mammographic view, with green circles around the BI-RADS benign microcalcifications, and a red circle around the BI-RADS suspicious microcalcifications. (b) Zoomed view of the red square in a with BI-RADS suspicious microcalcifications. (c) Zoomed view of the green square in a with BI-RADS benign microcalcifications. Supplemental Table S2 Comparison of accuracy and AUC of different state-of-the-art techniques for the classification of benign versus malignant microcalcifications [36–40].

#### Acknowledgements

This study 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.

#### Authors' contributions

CP was the guarantor of integrity of the entire study. KL, CP, and CN defined the basic concepts of the study. KL analysed the literature. GS and CN collected and selected the dataset. KL and CP developed the algorithm. KL performed the statistical analysis and analysed the results. KL prepared the first draft of the manuscript with all the figures and tables. KL, GS, CP, and CN edited the manuscript. All authors read and approved the final manuscript.

#### Funding

This study has received funding 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.

#### Availability of data and materials

The dataset generated during the current study is publicly available through Zenodo. The dataset can be accessed using the following link: <https://doi.org/10.5281/zenodo.5036062>.

#### Declarations

##### Ethics approval and consent to participate

The study was approved by the local Institutional Review Board. Written informed consent was obtained from all patients.

##### Consent for publication

Not applicable.

#### Competing interests

The authors declare that they have no competing interests.

#### Author details

<sup>1</sup>KIOS Research and Innovation Center of Excellence, Department of Electrical and Computer Engineering, University of Cyprus, 1 Panepistimiou Avenue, Aglantzia, 2109 Nicosia, Cyprus. <sup>2</sup>Nicosia General Hospital, 215 Nicosia-Limassol Old Road, Strovolos, 2029 Nicosia, Cyprus. <sup>3</sup>Limassol General Hospital, 1 Nikaias, 4131 Limassol, Cyprus.

Received: 30 March 2021 Accepted: 4 August 2021

Published online: 14 September 2021

#### References

- Ganesan K, Acharya UR, Chua CK, Min LC, Abraham KT, Ng KH (2013) Computer-aided breast cancer detection using mammograms: a review. *IEEE Rev Biomed Eng* 6:77–98. <https://doi.org/10.1109/RBME.2012.2232289>
- Arancibia Hernández PL, Taub Estrada T, López Pizarro A, Díaz Cisternas ML, Sáez Tapia C (2016) Breast calcifications: description and classification according to BI-RADS 5th edition. *Rev Chil Radiol* 22:80–91. <https://doi.org/10.1016/j.rchira.2016.06.004>
- Burnside ES, Ochsner JE, Fowler KJ, et al (2007) Use of microcalcification descriptors in BI-RADS 4th edition to stratify risk of malignancy. *Radiology* 242:388–395. <https://doi.org/10.1148/radiol.2422052130>
- Singh S, Maxwell J, Baker JA, Nicholas JL, Lo JY (2011) Computer-aided classification of breast masses: Performance and interobserver variability of expert radiologists versus residents. *Radiology* 258:73–80. <https://doi.org/10.1148/radiol.10081308>
- Cheng HD, Cai X, Chen X, Hu L, Lou X (2003) Computer-aided detection and classification of microcalcifications in mammograms: a survey. *Pattern Recogn* 36:2967–2991. [https://doi.org/10.1016/S0031-3203\(03\)00192-4](https://doi.org/10.1016/S0031-3203(03)00192-4)
- Rodríguez-Ruiz A, Krupinski E, Mordang JJ, et al (2019) Detection of breast cancer with mammography: Effect of an artificial intelligence support system. *Radiology* 290:305–314. <https://doi.org/10.1148/radiol.2018181371>
- Kumar MNA, Kumar MNA, Sheshadri HS (2019) Computer aided detection of clustered microcalcification: a survey. *Curr Med Imaging Rev* 15:132–149. <https://doi.org/10.2174/1573405614666181012103750>
- Nishikawa RM, Kallergi M, Orton CG (2006) Computer-aided detection, in its present form, is not an effective aid for screening mammography. *Med Phys* 33:812–814. <https://doi.org/10.1118/1.2168063>
- Hadjjiiski L, Sahiner B, Chan HP, Petrick N, Helvie MA, Gurcan M (2001) Analysis of temporal changes of mammographic features: computer-aided classification of malignant and benign breast masses. *Med Phys* 28:2309–2317. <https://doi.org/10.1118/1.1412242>
- Varela C, Karssemeijer N, Hendriks JHCL, Holland R (2005) Use of prior mammograms in the classification of benign and malignant masses. *Eur J Radiol* 56:248–255. <https://doi.org/10.1016/j.ejrad.2005.04.007>
- Timp S, Varela C, Karssemeijer N (2007) Temporal change analysis for characterization of mass lesions in mammography. *IEEE Trans Med Imaging* 26:945–953. <https://doi.org/10.1109/TMI.2007.897392>
- Roelofs AAJ, Karssemeijer N, Wedekind N, et al (2007) Importance of comparison of current and prior mammograms in breast cancer screening. *Radiology* 242:70–77. <https://doi.org/10.1148/radiol.2421050684>
- Timp S, Varela C, Karssemeijer N (2010) Computer-aided diagnosis with temporal analysis to improve radiologists interpretation of mammographic mass lesions. *IEEE Trans Inf Technol Biomed* 14:803–808. <https://doi.org/10.1109/TITB.2010.2043296>
- Hakim CM, Catullo VJ, Chough DM, et al (2015) Effect of the availability of prior full-field digital mammography and digital breast tomosynthesis images on the interpretation of mammograms. *Radiology* 276:65–72. <https://doi.org/10.1148/radiol.15142009>
- Loizidou K, Skouroumouni G, Nikolaou C, Pitris C (2019) A new method for breast micro-calcification detection and characterization using digital temporal subtraction of mammogram pairs. In: 2019 IEEE EMBS International Conference on Biomedical and Health Informatics, BHI 2019 - Proceedings, pp 1–4
- Loizidou K, Skouroumouni G, Nikolaou C, Pitris C (2020) An automated breast micro-calcification detection and classification technique using temporal subtraction of mammograms. *IEEE Access* 8:52785–52795. <https://doi.org/10.1109/ACCESS.2020.2980616>

17. Gonzalez RC, Woods RE, Eddins SL (2010) Digital image processing using Matlab, 2nd edn. Tata McGraw-Hill
18. Huang SC, Cheng FC, Chiu YS (2012) Efficient contrast enhancement using adaptive gamma correction with weighting distribution. *IEEE Trans Image Process* 22:1032–1041. <https://doi.org/10.1109/TIP.2012.2226047>
19. Pennec X, Cachier P, Ayache N (1999) Understanding the “demon’s algorithm”: 3D non-rigid registration by gradient descent. *Lect Notes Comput Sci* 1679:597–606. [https://doi.org/10.1007/10704282\\_64](https://doi.org/10.1007/10704282_64)
20. Diez Y, Oliver A, Lladó X et al (2011) Revisiting intensity-based image registration applied to mammography. *IEEE Trans Inf Technol Biomed* 15: 716–725. <https://doi.org/10.1109/TITB.2011.2151199>
21. Bailey D, Hodgson R (1985) Range filters: Localintensity subrange filters and their properties. *Image Vis Comput* 3:99–110. [https://doi.org/10.1016/0262-8856\(85\)90058-7](https://doi.org/10.1016/0262-8856(85)90058-7)
22. Singh B, Kaur M (2018) An approach for classification of malignant and benign microcalcification clusters. *Sadhana* 43:1–18. <https://doi.org/10.1007/s12046-018-0805-2>
23. Kumar V, Gupta P (2012) Importance of statistical measures in digital image processing. *Int J Emerg Technol Adv Eng* 2:56–62
24. Haralick RM, Dinstein I, Shanmugam K (1973) Textural features for image classification. *IEEE Trans Syst Man Cybern SMC*-3:610–621. <https://doi.org/10.1109/TSMC.1973.4309314>
25. Diehr P, Martin DC, Koepsell T, Cheadle A (1995) Breaking the matches in a paired t-test for community interventions when the number of pairs is small. *Stat Med* 14:1491–1504. <https://doi.org/10.1002/sim.4780141309>
26. Pedregosa F, Varoquaux G, Gramfort A et al (2011) Scikit-learn: machine learning in Python. *J Mach Learn Res* 12:2825–2830
27. Alam N, Denton ERE, Zwiggelaar R (2019) Classification of microcalcification clusters in digital mammograms using a stack generalization based classifier. *J Imaging* 5. <https://doi.org/10.3390/jimaging5090076>
28. Mohebian MR, Marateb HR, Mansourian M, Mañanas MA, Mokarian F (2017) A hybrid computer-aided-diagnosis system for prediction of breast cancer recurrence (HPBCR) using optimized ensemble learning. *Comput Struct Biotechnol J* 15:75–85. <https://doi.org/10.1016/j.csbj.2016.11.004>
29. Pak F, Kanan HR, Alikhassi A (2015) Breast cancer detection and classification in digital mammography based on Non-Subsampled Contourlet Transform (NSCT) and Super Resolution. *Comput Methods Prog Biomed* 122:89–107. <https://doi.org/10.1016/j.cmpb.2015.06.009>
30. Breiman L (1996) Bagging predictors. *Mach Learn* 24:123–140. <https://doi.org/10.1007/bf00058655>
31. Friedman JH (2002) Stochastic gradient boosting. *Comput Stat Data Anal* 38:367–378. [https://doi.org/10.1016/S0167-9473\(01\)00065-2](https://doi.org/10.1016/S0167-9473(01)00065-2)
32. Raschka S (2015) Python machine learning: unlock deeper insights into machine learning with this vital guide to cutting-edge predictive analytics, 1st edn. Packt Publishing
33. Yassin NIR, Omran S, El Houby EMF, Allam H (2018) Machine learning techniques for breast cancer computer aided diagnosis using different image modalities: a systematic review. *Comput Methods Prog Biomed* 156: 25–45. <https://doi.org/10.1016/j.cmpb.2017.12.012>
34. Li H, Zhuang S, Dao L, Zhao J, Ma Y (2019) Benign and malignant classification of mammogram images based on deep learning. *Biomed Signal Process Control* 51:347–354. <https://doi.org/10.1016/j.bspc.2019.02.017>
35. Yu S, Guan L (2000) A CAD system for the automatic detection of clustered microcalcifications in digitized mammogram films. *IEEE Trans Med Imaging* 19:115–126. <https://doi.org/10.1109/42.836371>
36. Ren J (2012) ANN vs. SVM: Which one performs better in classification of MCCs in mammogram imaging. *Knowledge-Based Syst* 26. <https://doi.org/10.1016/j.knsys.2011.07.016>
37. Khehra BS, Partap A, Pharwaha S (2013) Least-Squares Support Vector Machine for Characterization of Clusters of Microcalcifications. 7:1612–1621
38. Strange H, Chen Z, Denton ERE, Zwiggelaar R (2014) Modelling mammographic microcalcification clusters using persistent mereotopology. *Pattern Recogn Lett* 47:157–163. <https://doi.org/10.1016/j.patrec.2014.04.008>
39. Chen Z, Strange H, Oliver A, Denton ERE, Boggis C, Zwiggelaar R (2015) Topological modeling and classification of mammographic microcalcification clusters. *IEEE Trans Biomed Eng* 62:–1214. <https://doi.org/10.1109/TBME.2014.2385102>
40. Fanizzi A, Basile TMA, Losurdo L, et al (2020) A machine learning approach on multiscale texture analysis for breast microcalcification diagnosis. *BMC Bioinformatics* 21:91. <https://doi.org/10.1186/s12859-020-3358-4>

## Publisher’s Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

**Submit your manuscript to a SpringerOpen® journal and benefit from:**

- Convenient online submission
- Rigorous peer review
- Open access: articles freely available online
- High visibility within the field
- Retaining the copyright to your article

---

Submit your next manuscript at ► [springeropen.com](https://www.springeropen.com)

---