Available online at www.sciencedirect.com

ScienceDirect

journal homepage: www.elsevier.com/locate/bbe

Original Research Article

Breast cancer diagnosis using abnormalities on ipsilateral views of digital mammograms

Suhas Sapate [a,c,*], Sanjay Talbar [a], Abhishek Mahajan [b], Nilesh Sable [b], Subhash Desai [b], Meenakshi Thakur [b]

[a] SGGS Institute of Engineering & Technology, Nanded, Maharashtra, India
[b] Department of Radiodiagnosis, Tata Memorial Centre, Parel, Mumbai, Maharashtra, India
[c] Dept. of CSE, Ashokrao Mane Group of Institutions, Vathar, Kolhapur, Maharashtra, India

ARTICLE INFO

Article history:
Received 26 October 2018
Received in revised form
23 April 2019
Accepted 29 April 2019
Available online 18 May 2019

Keywords:
Breast cancer
Digital mammograms
Ipsilateral views
Radiomic features
Image-based sensitivity
Case-based sensitivity

ABSTRACT

Ipsilateral views of digital mammograms help radiologists to localize and confirm abnormal lesions during diagnosis of breast cancers. This study aims at developing algorithms which improve accuracy of computer-aided diagnosis (CADx) for analyzing breast abnormalities on ipsilateral views. The proposed system is a fusion of single and two view systems. Single view approach detects and characterizes suspicious lesions on craniocaudal (CC) and mediolateral oblique (MLO) view separately using geometric and textural features. Lesions detected on each view are paired with potential lesions on another view. The proposed algorithm computes the correspondence score of each lesion pair. Single view information is fused with two views correspondence score to discriminate malignant tumours from benign masses using the SVM classifier. Performance of SVM classifier is assessed using five-fold cross validation (CV), Kappa metric and ROC analysis. Algorithms are applied to 110 pairs of mammograms from local dataset and 74 pairs from open dataset. Single view scheme yielded image-based sensitivity of 91.63% and 88.17% at 1.35 and 1.51 false positives per image (FPs/I) on local and open dataset respectively. Single view classification yielded FPs/I of 1.03 and 1.20 with sensitivity of 70%. Fusion based two views scheme using SVM classifier produced average case-based sensitivity of 75.91% at 0.69 FPs/I and 73.65% at 0.72 FPs/I on local and open dataset respectively. Fusion of single view features with two view correspondence score leads to improved case-based detection sensitivity. Proposed fusion based approach results into accurate and reliable diagnosis of breast abnormalities than single view approach.

* *Corresponding author at*: SGGS Institute of Engineering & Technology, Nanded, Maharashtra 431606, India.
E-mail address: sapatesuhas@sggs.ac.in (S. Sapate).

https://doi.org/10.1016/j.bbe.2019.04.008
0208-5216/

1. Introduction

Early detection is viewed as the best hope to reduce the breast cancer mortality by allowing timely treatment at an earlier cancer progression stage. The imaging techniques such as X-ray, MRI, ultrasound, etc. are widely used for detection, diagnosis, and treatment of abnormalities that can be cancerous [1]. Mammography is the most reliable cost-effective imaging modality used by choice for early detection of breast cancer [2]. Full-field digital mammography (FFDM) provides a high contrast high-resolution images which are found useful in improving an efficacy and performance of mammography [3]. Particularly, the FFDM images are more significant in the mammographic screening of dense breasts of younger women [4]. The majority of breast cancers detected in symptomatic patients during screening mammography appear as masses within a range of 5–20 mm in diameter. Visibility of masses smaller than 5 mm is rare and sometimes the nonpalpable masses larger than 20 mm are not visible on mammogram [5]. During mammography, two projection views namely CC and MLO views, known as ipsilateral views, are routinely obtained for each breast [6]. As ipsilateral views encompass the entire breast parenchyma from two different sides, the suspicious lesions, if are there, are visible and can be detected on either or both the views [7]. Medical practitioners usually analyze ipsilateral views to describe the breast composition, morphological descriptors such as bilateral asymmetry, abnormal lesions, architectural distortion, microcalcifications, etc. Then they make their clinical decisions regarding grading the abnormality or performing further diagnostic tests. The mammogram interpretation may go wrong or inaccurate due to fatigue or more workload [8]. The CAD system using ipsilateral views can be a second pair of eyes to diagnose the breast anomalies and improve the positive predictions while reducing false predictions during diagnosis [9].

However, currently radiologists use single view CAD systems as a second reader for breast cancer diagnosis using FFDM images [10]. These CAD systems have high false positive (FP) and false negative (FN) rates because of varying subtle nature of abnormalities and limitations of imaging system [5,11]. The single view CAD systems for mass detection have good diagnostic accuracy but less positive predictive value; and hence are not as accurate and reliable as like the experienced radiologist [12].

Some findings reveal that the sensitivity of digital mammography to detect breast lesions obscured by surrounding and overlapping dense parenchymal tissues is as low as 30–48% [13]. Many of the patients are found with nonpalpable breast lesions during screening using mammography [14]. The diagnosis of these nonpalpable lesions is confirmed after the histopathological reports (HPRs) of the sample tissues obtained using surgical or needle biopsy. However, the patients as well as radiologists wish to avoid biopsy of benign mass but without missing an opportunity to detect the malignant tumour. Thus, accurate detection and characterization of the suspicious lesions can help to minimize the degree of cosmetic disfigurement of the breast during treatment. In short, improvement in the detection performance of CAD with proper balance of sensitivity and specificity is crucial for accurate diagnosis and treatment [15]. This becomes easy with CC and MLO views together. The recent studies have shown that the two view systems can improve the detection and diagnosis of breast abnormalities and can become the effective alternative as a second reader to radiologists [16].

The CAD systems combining the advantages of single view and two view systems can be useful as a simultaneous mammograms analysis tool for the radiologists [17]. Accordingly, we propose a different fusion based CAD system that leverages advantages of existing approaches with improved case-based lesion detection and diagnosis at a comparable false positive (FP) rate. The dual view analysis framework presented in this article is different than other conventional ipsilateral mammograms analysis techniques as follows.

- It selects the width of annular region dynamically based on size of the lesion and the breast to match the corresponding pairs of suspicious lesions from CC and MLO view.
- Idea of template matching using selected set of similarity features towards improving lesion pairing on ipsilateral views.
- It utilizes radiomics features and location-based features to improve film-based single view performance of cancer detection.
- It combines single view features and two view correspondence score to improve the case-based two views performance of cancer diagnosis.

The remaining part of this article is structured as follows. Section 2 gives an overview of existing CAD approaches using ipsilateral views. Section 3 covers the clinical data used and the methodology is explained in Section 4. The experimental results are analyzed in Section 5. The discussion on the various key issues is presented in Section 6. Section 7 concludes the article.

2. Literature survey

To overcome the limitations of prompting single view CAD systems, many researchers have attempted to develop multiview-based CAD systems. The clinical studies on breast cancer have indicated that two view mammograms help in achieving more diagnostic accuracy than with one view [18]. In their pioneer research, Good et al. [19] demonstrated that the multiview features of a single physical lesion on ipsilateral views distinguishes T and F pairs of lesions with $A_z = 0.82 \pm 0.03$ on ROC curves. This approach has limitations on number of features and number of lesion pairs. Chang et al. [20] investigated arc method with $A_z = 0.79$ and Cartesian straight-line method with $A_z = 0.78$ to localize and match breast lesions on 571 CC views with lesions on 571 MLO views. In this scheme, the measurements related to nipple location and chest wall orientation were not automatic. Paquerault et al. [21] developed two-view detection technique and reduced FP rate from 2.1 FPs/I to 1.2 FPs/I with sensitivity of 80%. Sahiner et al. [22] devised a new method

combining results of single view and two views analysis to reduce the FPs/I from 1.5 at 80% sensitivity to 0.7 FPs/I at 70% sensitivity for the overall detection of breast cancer. The only drawback of these approaches is that a true mass may be missed when the detected lesion is on either view and not on both views. Paquerault et al. [9] incorporated the geometric locations, textural and morphological features to correlate the suspicious lesions detected on ipsilateral views for improving CAD performance. They improved the case-based detection sensitivity from 77% to 91% using fusion of two-view information. However, the annular search region with 80 pixel radial width is a problem for smaller breasts. The nipple detection is manual and not automatic as required for a CAD. Engeland et al. [23] presented a method to find corresponding suspicious lesions on CC and MLO view with 82% accuracy. The technique faces the problem of one-to-one correspondence link of TP lesion with FP lesion instead of actual TP lesion. A new multiview-based CAD system presented by Zheng et al. [24] detected cue suspected lesions on both views at reduced FPs/I of 0.91 with 74.4% sensitivity. The matching areas of the suspicious lesions are identified approximately and not exactly; this leaves a gap for increasing the robustness of the scheme. Yuan et al. [25] presented a two stage computerized framework to match suspicious lesion on CC view to its correlative counterpart on MLO view with area under ROC curve (AUROC) of 0.87 ± 0.03. However, the metrics such as sensitivity, specificity, accuracy, false positives are not assessed to show diagnostic abilities of their CAD system. Wei et al. [6] analyzed two views of mammograms to improve case-based sensitivities up to 73.4% and 85.7% at 0.5 and 1.0 FPs/I respectively for mass detection. Nipple detection is manual and not fully automatic. Padayachee [26] explored GLCM textural features, distance similarity, and mutual information for matching suspicious lesion on ipsilateral views for a mammographic CAD system with AUROC of 0.77 ± 0.25. The scheme reported two shortcomings, one related to ground truth and the other regarding smaller dataset. Zheng et al. [27] investigated three methods for deciding width of search area used to match the suspicious lesion from ipsilateral views and showed that straight strip search method works better with case-based sensitivity of 86% at 0.29 FPs/I and hence is suitable for CAD systems. Samulski and Karssemeijer [28] proposed a multiview CAD system for mass detection with improved case-based sensitivity by 4.7% at 0.5 FPs/I using information collected from ipsilateral views of mammograms. If the chest wall is not parallel to the border of CC view or nipple on larger breast is not detected, the performance of the system may degrade. Tanner et al. [29] exploited region-based 71 features combined with modified search area techniques for correlating mammographic masses on ipsilateral views with improved performance in terms of AUROC from 0.895 to 0.915. In this approach, distance to manually delineated location of nipple seems to degrade the performance.

The literature surveyed has helped us to identify the research gap and we have attempted to bridge this gap by devising computer algorithms which can improve performance of CAD systems using two views mammogram analysis.

3. Clinical dataset

The clinical data for this study is taken from Tata Memorial Centre (TMC), Mumbai, India. Institutional Research Ethics Committee of TMC, Mumbai has approved the usage of all the biopsy-proven mammograms of breast cancer patients. The performance of the proposed two views CAD scheme is tested on 110 pairs of CC and MLO views of FFDM obtained from 110 patients at TMC Mumbai. Enough care is taken to hide the identity of all the patients prior to start the work. All the images selected for this study have proper visible border and the nipple is also in the breast profile. FFDM images cover patients with different densities of breasts belonging to all classes with respect to Breast Imaging Reporting and data system (BIRADS) lexicons. From 110 CC and 110 MLO views, 88 (176 lesions) are malignant tumours and 22 (44 lesions) are benign masses. The abnormalities on the images cover 2–6 categories of BIRADS mammographic assessment. At least one abnormal lesion is present on either CC or MLO views considered in this study. The size of abnormal lesions on the selected images ranges from 5 mm to 35 mm. All the images in DICOM (Digital Imaging and Communications in Medicine) format captured on 'Hologic Selenia System' and on 'GE Medical Senograph System' contain 4096×3328 pixels each of size $50\ \mu m \times 50\ \mu m$. The ground truths (GT) on all the images are marked under the supervision of expert radiologists participating in this study. Another set consists of 148 mammograms selected from publicly available DDSM dataset [30]. This set comprises 38 pairs benign masses and 36 pairs malignant tumours visible on both CC and MLO views. The lesions on all these images are delineated by the radiologists in the DDSM dataset.

4. Methods

Methods for detection of breast abnormalities require that suspicious anatomical lesions in mammograms be segmented and characterized. The locations specific information of lesions are then used to restrict processing to relevant areas, to perform location-dependent processing, or to correlate findings in different views. This section describes the proposed method covering four stages such as pre-processing, single view diagnosis, ipsilateral view diagnosis, fusion of single with two view diagnosis. Pre-processing of medical images increases the efficiency and accuracy of the CAD algorithms in identifying the abnormalities easily [31,32]. Pre-processing stage includes detection of air–skin interface, nipple in the profile and pectoral muscle on MLO views, etc. It is covered in brief in the subsection as below.

4.1. *Pre-processing*

On an average, the actual size of the FFDM image is almost double the size of the breast parenchyma as shown in Fig. 1(a). The proposed work starts with minimizing the size of the input mammogram to reduce the computational burden during segmentation phase. The air–skin interface and nipple of the breast are detected using fuzzy *c*-means clustering technique.

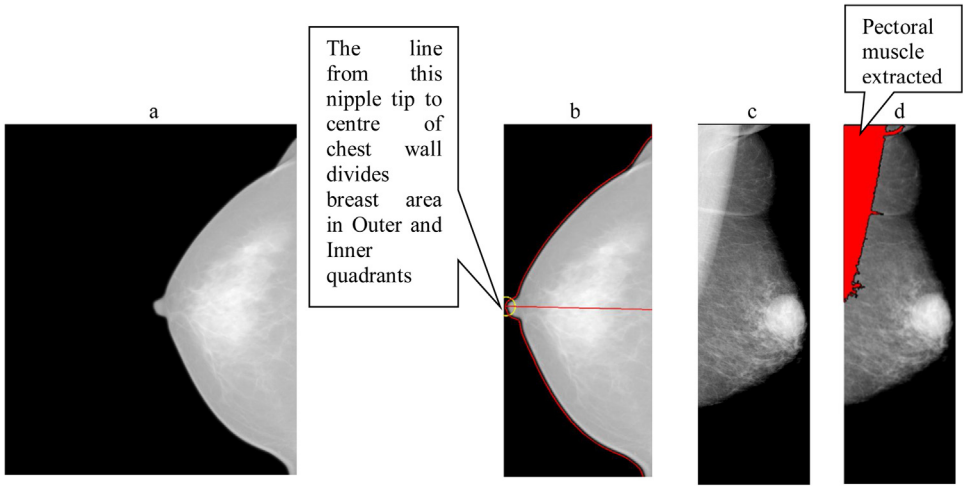

Fig. 1 – Results of pre-processing: (a) original left CC view of size 4096 × 3328; (b) cropped image of size 4096 × 3328 with border, nipple, and midline detected; (c) cropped left MLO view image; (d) pectoral muscle extracted out on right MLO view.

The most popular fuzzy *c*-means algorithm makes use of the reciprocal of distances to decide the cluster centres of all points, weighted by their degree of belonging to the cluster. This results in a class membership to more than one cluster at the same time with different degrees of association [33]. This multiple association represents an important feature for increasing sensitivity in medical image segmentation. Furthermore, the fuzzy logic allows to model non-linear, imprecise, complex systems by applying human knowledge and experience as a set of fuzzy rules that use linguistic variables for better segmentation [15,34]. The modified fuzzy clustering function for segmenting image with n pixels into *c* clusters minimizing the within-class sum-of-square errors and is defined as Eq. (1):

$$I_f(U,V) = \sum_{i=1}^{n}\sum_{j=1}^{c}\mu_{ij}^{m} d^2(X_i, V_j) \tag{1}$$

where μ_{ij}^{m} represents the degree of membership of the ith pixel X_i to jth class V_j which is the centre of jth class and $d^2(X_i, V_j)$ is the square of the Euclidian distance between X_i and V_j. The function is minimized under the following conditions:

$$V_j = \frac{\sum_{i=1}^{n}(\mu_{ij})^m X_i}{\sum_{i=1}^{n}(\mu_{ij})^m} \quad 1 \le j \le c \tag{2}$$

where

$$\mu_{ij} = \left(\sum_{k=1}^{c}\left\{\left(\frac{d_{ij}}{d_{ik}}\right)^{2/(m-1)}\right\}\right)^{-1} \quad 1 \le i \le n; \quad 1 \le k \le c \tag{3}$$

The experiment is performed by selecting three clusters of the given mammographic image. The three cluster centres separate the mammogram into three partitions background, breast profile and a dense portion in the profile.

In most of the cases, the leftmost pixel on right views or the rightmost pixel on left views is enough to determine the required width of the mammogram and to crop out the essential breast parenchyma region. Exceptionally, on a few images, the farthest pixel from breast wall lies on the border instead of the nipple of the breast. The part of the breast parenchyma with a margin of 30 pixels wide beyond this parenchyma is included in the cropped image. The resulting image of cropped parenchyma region is depicted in Fig. 1(b). Thus, the width of the image is reduced and height remains intact. The nipple position is required in locating the lesion position in the breast and the breast quadrant the lesion belongs to. The horizontal line passing through the nipple divides the breast views in two sections. The section above the line is called 'outer quadrant' and section below the line is 'inner quadrant' on CC view. Similarly, the section above the line is 'upper quadrant' and section below the line is 'lower quadrant' on MLO view. The pectoral muscle on each MLO view obstructs the segmentation of breast for identifying suspicious lesions [35]. Hence, it is extracted using the RANdom SAmpling Concensus (RANSAC) algorithm [36], and the results are shown in Fig. 1(c) and (d).

4.2. *Single view diagnosis*

Accurate segmentation of the suspicious lesions with fuzzy, ill-defined and diffused boundaries especially on the low contrast CC or MLO view images is challenging task [37]. The literature survey reveals that the fuzzy-based region growing methods are effective in segmentation of such cases [38]. Therefore, the hybrid segmentation method in the proposed work leverages the advantage of fuzzy-based systems for segmenting FFDM images. The segmentation algorithm consists of four major steps.

- Identifying the suspicious abnormal lesions with dynamically selected seed pixel

- Expanding the suspicious lesions using adaptive fuzzy region growing algorithm
- Merging the similar small and close neighbouring regions and
- Delineating the border of abnormal lesions.

It is assumed that the malignant tumours are the regions with dense core surrounded by the slowly extended transitions till their boundaries. The regions on mammogram with textural homogeneity, maximum entropy, maximum average intensity and minimum variance are selected as centroids of candidate lesions. The scheme assumes the fuzzy connectedness of neighbouring pixels with homogeneity using intensity values constrained by mean and standard deviation. The procedure starts with dynamically chosen centroid of the suspicious lesion. The region being defined starts growing with neighbouring pixels surrounding the designated centroid. Subsequently, the pixel which has intensity near to the mean μ of the region is included in the region. The traditional algorithms with 8-connected or 6-connected region growing patterns incorporate a recursive process to determine the neighbouring pixels. However, the recursive algorithms are inefficient as far as time and computational power are concerned. To overcome this pitfall, a novel scheme called ''Sapate's neighbouring pixels selection scheme (SNPSS)'' for selecting unique pixels along eight radial directions is proposed in our earlier work [39].

The literature surveyed reveals that the tumour starts with a tiny core and slowly starts growing in a radial pattern. This type of biological tumour growth pattern has motivated us to come out with SNPSS for modifying region growing method. SNPSS allows us to get rid of the curse of recursive expansion of region growing algorithm. The selection of neighbouring pixels minimizes the repetition of the pixels which results in computational effectiveness and requires no recursion at all. Thus, segmentation using SNPSS plays a very important role in delineating the exact boundary of the mass or tumour which is crucial in further characterization and treatment planning [39]. The results of mammogram segmentation are depicted in Fig. 2(a) and (c).

The segmented images with multiple candidate lesions undergo one more stage to reduce the false positive lesions. The proposed system filters out some of the candidate lesions using few empirical rules. Empirical rules are usually used for predicting some data which have no numeric expression. This happens when it is very difficult to obtain right data which follows normal distribution. Empirical rules applied to any random variable can give rough estimate of what your data collection may look like if you were able to survey the large population [40]. In our study, these parameters include area less than 2500 pixels, extent smaller than 0.15, perimeter twice that of image width and eccentricity greater than 0.98. All parameters and their respective values are selected under the guidance of the expert radiologists involved in this study. While applying the empirical rule to any of above parameters, the condition that 'approximately 68% of the data falls within one standard deviation of the mean' is confirmed to be true.

The experimental results obtained are shown in Fig. 2(b) and (d). Another case with multiple lesions is shown in Fig. 3. The lesions in Fig. 3(b) are paired with lesions on MLO view in Fig. 3(d) using a procedure described in Section 4.3.2. Seven selected geometric features concerning shape, size, margin, contrast and six textural features along 3 distances, 4 directions based on GLCM are calculated to characterize and classify the detected lesions. The features selected have shown highest classification accuracy. A pool of actively discriminative geometric and textural features is fed to the simple SVM classifier to achieve the classification results. The pixels in the background of the suspicious lesion do not play a vital role in characterizing the mass or tumour; hence

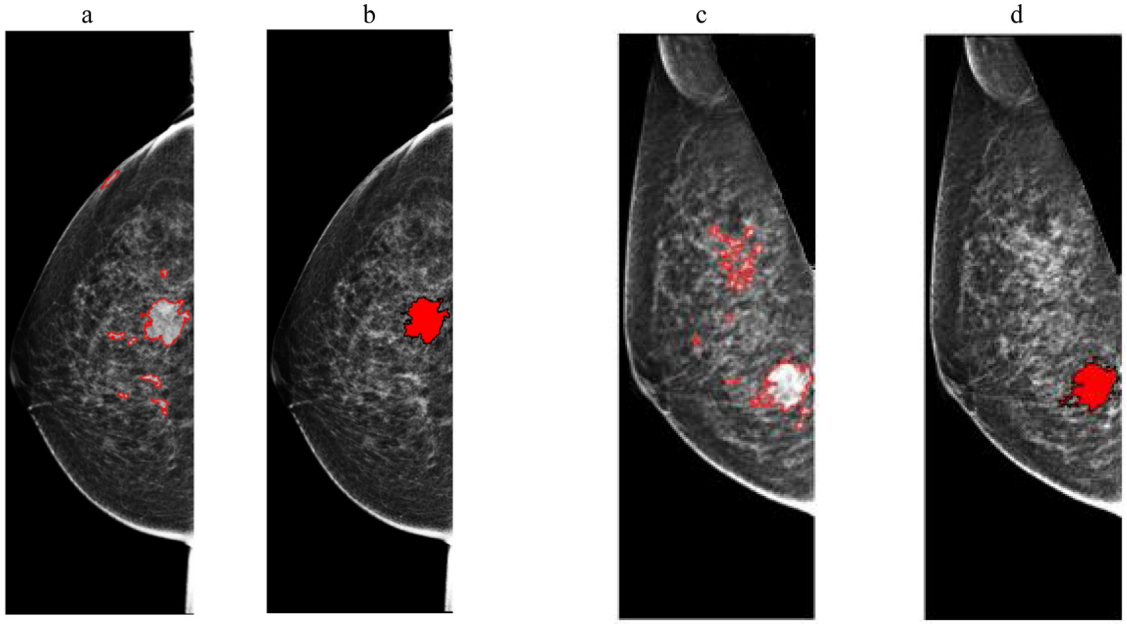

Fig. 2 – FP reduction using empirical rules during the initial detection stage: (a) CC view with suspicious lesion detected; (b) selected lesion after applying empirical rules for FP reduction; (c) MLO view of the same breast; (d) lesion after FP reduction.

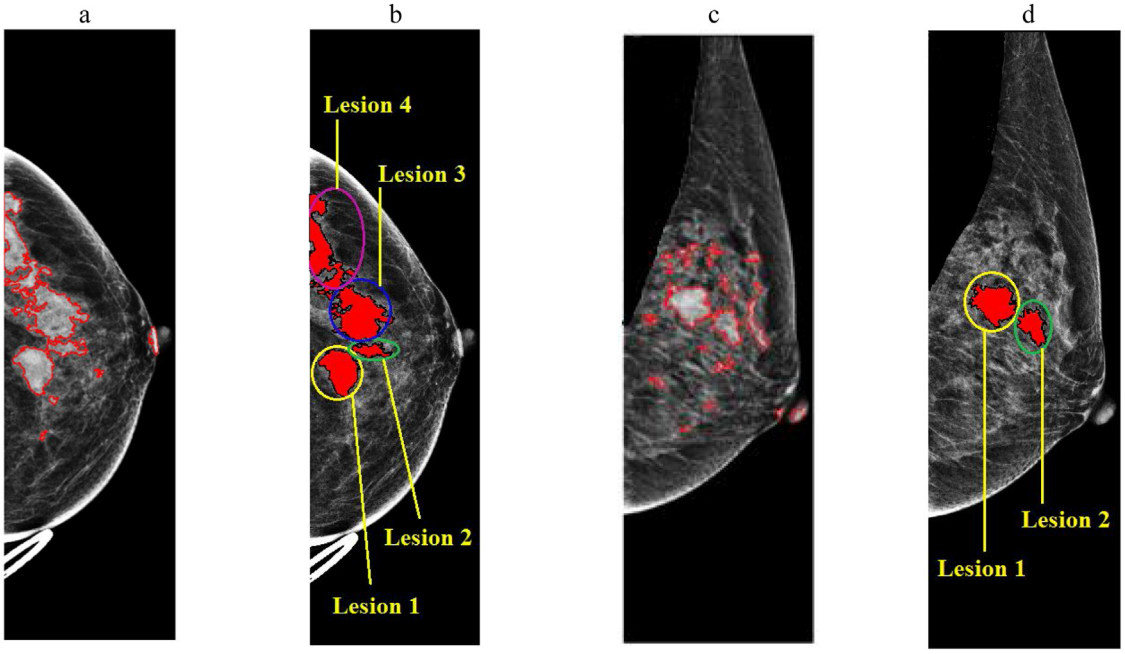

Fig. 3 – FP reduction using empirical rules during the initial detection stage: (a) CC view with suspicious lesion detected; (b) multiple selected lesions after applying empirical rules for FP reduction; (c) MLO view of the same breast; (d) two lesions selected after FP reduction.

the ROI is taken excluding the background. The textural and geometric features extracted are enlisted and defined in Appendix A and Appendix B respectively.

Support Vector Machine (SVM) is the simple, most popular yet strong classifier for a small set of images [41]. It focuses more on confidence of classification to discriminate the benign mass and malignant tumours using the respective feature vectors as input. The decision function is defined as Eq. (4):

$$f(x) = \sum_{i=1}^{S} y_i \alpha_i K(x_i, x) + b \tag{4}$$

where $x_i \in \{-1,+1\}$ represents training vectors, S is the size of training set, α_i is the parameter for optimizing the margin, $K(x_i, x)$ is the kernel function and b is the regularization parameter for minimizing training error [42]. In this study, radial basis function (RBF) with a variety of combinations of (α_i, b) is obtained once at fixed interval values of α_i and then with fixed interval values of b. The problem of SVM related to trade-off between the large margin and few misclassifications is solved here by optimized combination of (α_i, b) parameters that achieves highest accuracy [43,44]. The feature vectors of the segmented lesions are split into two classes namely benign mass and malignant tumour. Both these classes are then divided into five groups or folds randomly and evenly. Every time the classifier trained with 4 folds and the remaining fold is applied to test the classifier. Thus every fold is used as a test set only once. This cross-validation step is repeated five times. The average of the accuracy achieved for each fold is taken as a final result of the classifier.

4.3. *Ipsilateral-view diagnosis*

The overall design of proposed CAD system using ipsilateral views is depicted in Fig. 4. The scheme generates a correspondence score for each suspicious lesion from two views. The lesion is localized and confirmed as benign mass or malignant tumour as a final result which is achieved with four steps: (1) The suspicious lesions on both views are located using arc-based method and are paired using width of arc pair, (2) similarity features of the selected pairs are calculated, (3) correspondence score of every lesion is generated and (4) single view features and two view correspondence score are combined to confirm the lesion either benign mass or malignant tumour using SVM classifier. The detailed description of each of these steps is as follows.

4.3.1. *Localizing suspicious lesions*

Due to the compression of deformable breast during imaging and variable landmarks, it is difficult to pinpoint exact location of the physical lesion on ipsilateral views detected in Section 4.2. If two or more nonpalpable suspicious lesions are detected on mammogram, the use of preoperative localization may become painful for the patient. In this section, a technique to determine precise radiologic localization on both the views is proposed to help the radiologists in locating the lesion being biopsied. The suspicious lesion on either view can be matched to the same physical lesion, if visible, on the other view. The proposed technique deals with locating the suspicious lesions on CC and MLO views of the same breast and forming the candidate pairs of lesions representing the single physical abnormality.

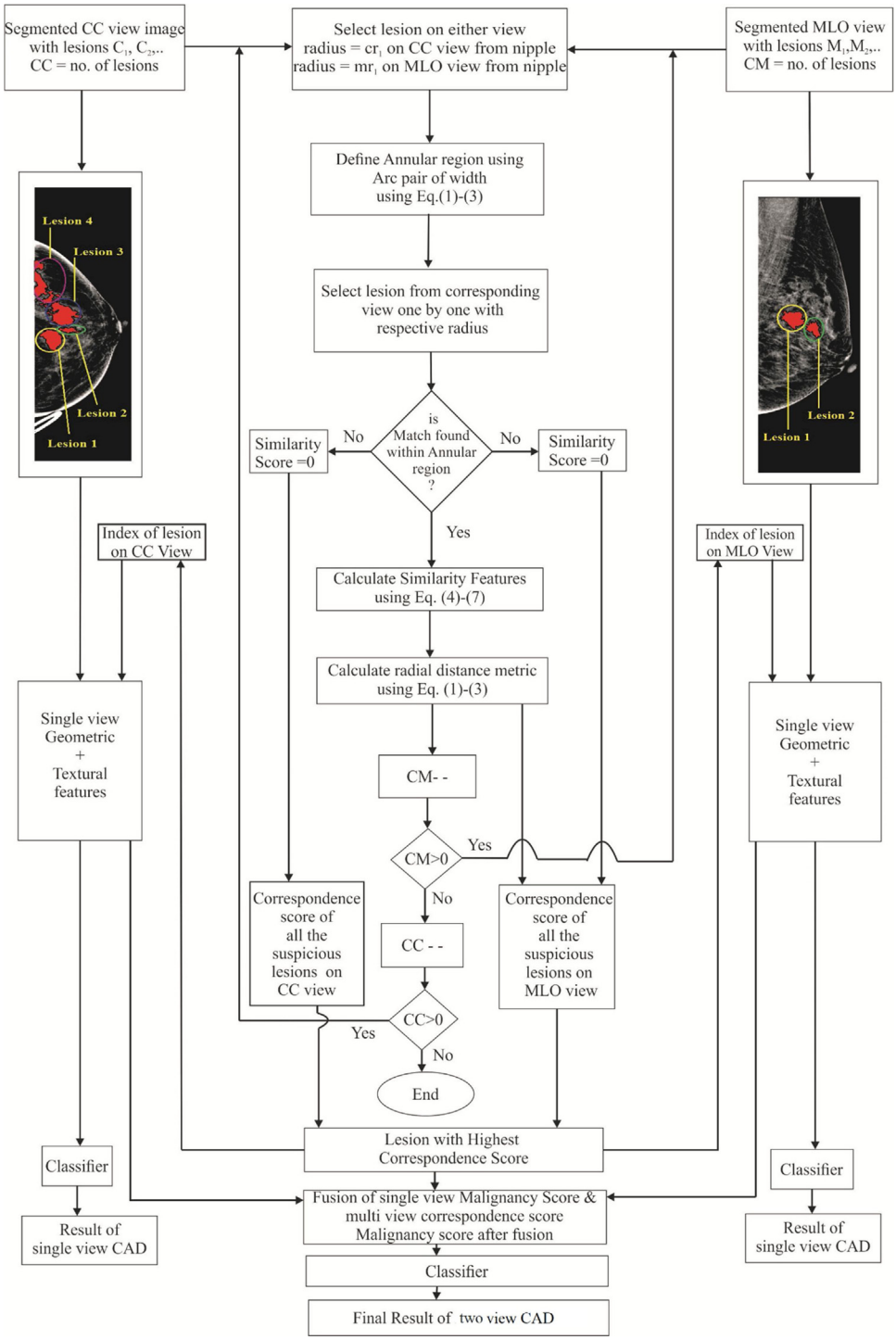

Fig. 4 – Schematic flow of proposed fusion-based two views CAD system.

There are two popular methods in the literature: straight line-based method and arc-based method [9,27] for locating the lesions and confirming their geometric relations. The straight line-based method requires two landmarks on the mammograms: chest wall on CC or pectoral muscle on MLO view and the nipple centre. The chest wall in CC view may not be parallel and pectoral muscle may not be perfectly a straight line which may lead to chaos in determining straight-line distances. On the other hand, arc-based method uses only nipple and centroid of the suspicious lesion and is comparatively easy; hence is incorporated in this work to locate the suspicious lesions and to eliminate false positives using reduced annular search region. Secondly, in order to expose the maximum portion of the breast for imaging, the nipple during compression is pulled away from the chest wall equally in CC and MLO views [28]. This results in an approximately constant distance between centre of nipple and the centroid of suspicious lesion on both the views. This distance is useful

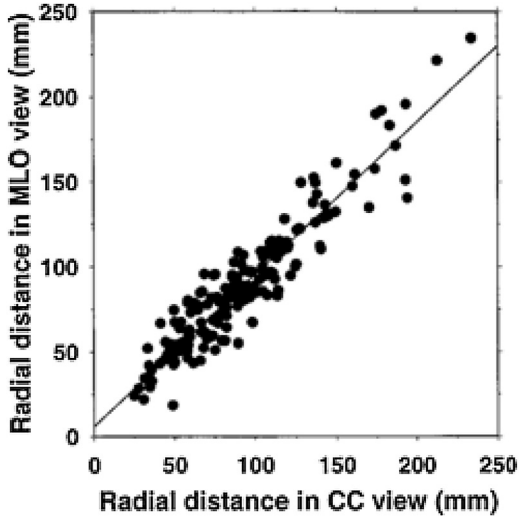

Fig. 5 – The radial distances of the detected lesions on CC view versus that of the MLO view from the nipple centre.

during mapping the corresponding lesions on ipsilateral views. The centre of nipple and centroid are marked under the supervision of the radiologists involved in this study. A scatter plot of these radial distances on the radiologist-identified objects on the ipsilateral views is shown in Fig. 5. It is revealed that there is a high correlation of the radial distances of the correlated lesions on the ipsilateral views. However, the angular coordinates are found with a significant correlation. Consequently, the radial distance is selected as the metric for determining the correlation of a lesion on one view with the same physical lesion on the other view.

The suspicious lesions are found in a wide variety of shapes and size. Moreover, the radial distance of a physical lesion on CC view is different than that of on MLO view. Thus, both these facts are considered for deciding the width of annular region for optimal search. The first coefficient is calculated by minimizing mean square error between the radial distance of ith true lesion on CC, R^i_{cc}, with its jth true distance on a MLO view, R^j_{mlo} as given in Eq. (5):

$$R_y = \sqrt{\left(R^i_{cc} - R^j_{mlo}\right)^2} \tag{5}$$

where i, j = 1, 2, . . . The second coefficient is the length of minor axis of the lesion on CC view, L_c. The width of annular region, ΔW, is defined in Eqs. (6) and (7).

$$\Delta W = 2 \times max(L_c, R_y), \quad \text{if } L_c < R_y \tag{6}$$

$$\Delta W = 2 \times min(L_c, R_y), \quad \text{if } L_c < R_y \tag{7}$$

The annular width is dynamically selected and is limited within the breast parenchyma. The limited width reduces the search region and eliminates all the false positive candidates from set of suspicious lesions on the ipsilateral views. The matching process is depicted in Fig. 6. It starts with a breast coordinate system which has its origin on the central outer tip of the nipple. The distance between the centroid of the suspicious lesions C_1 and nipple tip N_c, is defined as cr_1, and ΔC_1 is the length of minor axis of the lesion, C_1, measured in terms of pixels as shown in Fig. 6(a). The arc (green line) with radius, cr_1, is drawn with nipple tip, N_m, on MLO view as centre as shown in Fig. 6(b). An arc pair (yellow band) around this central arc is drawn with width ΔW which is calculated using Eqs. (6) and (7). The arc intersects the breast border at two points, the centre of which is used to define breast midline (maroon line) passing from nipple tip to the centre of chest wall as shown in Fig. 6(b). This midline divides the MLO view in two sections known as Upper and Lower quadrants, as shown in Fig. 6(b). The location of the lesion in a breast quadrant is used for preoperative localization before biopsy. Similarly, the distance between the centroid of the suspicious lesions on MLO view, M_1 and nipple tip N_m is defined as mr_1. ΔM_1 is the length of minor axis of the lesion, M_1, measured in terms of pixels/Euclidean distance as shown in Fig. 6(b). The arc (green line) with radius, mr_1, is drawn with nipple tip, N_c, on CC view as centre is shown in Fig. 6(b).

An arc pair (yellow band) around this central arc is drawn with width ΔW which is calculated using Eqs. (6) and (7). The arc intersects the breast border at two points, the centre of which is used to define breast midline (maroon line) from nipple tip to centre of chest wall as shown in Fig. 6(c). This midline divides the CC view in two sections known as Outer and Inner quadrants, as shown in Fig. 6(c). The location of the lesion in a breast quadrant is used for preoperative localization before biopsy.

Thus the annular regions for all the suspicious lesions are drawn to limit the number of lesions to be paired. The square templates on each of the paired lesions are selected for extracting the features shown in Fig. 7(a) and (b).

The template is the square region centred on the centroid of the selected suspicious lesion belonging to either view. To calculate the similarity features, the templates of the suspicious lesions on both views must be of equal size. The size of the templates is equal to minimum of the minor axis lengths the respective lesions. The square template within minor and major axis ensures the maximum of the common portion of the lesions is selected for extracting similarity features. We conducted a small experiment on few templates and calculated score by applying similarity features and traditional pixel by pixel template matching [45]. The similarity features of the template are more effective instead of pixel by pixel score as like in traditional template matching.

4.3.2. *Similarity features of paired lesions*

An important task in two view diagnosis is to find the same abnormal lesion on both CC and MLO views. This is achieved using similarity features of all possible pairs of lesions, one from CC and other from MLO view. Every lesion on CC view is paired with every other lesion lying within the annular region of MLO of the same breast. Similarly, every lesion on MLO view is paired with every other lesion lying within the annular region of CC of the same breast. To correlate the lesions on CC and MLO views, a few similarity features with highest rank as explained in Ref. [46] are extracted from every possible pair of lesions. As experienced by the radiologists involved in this

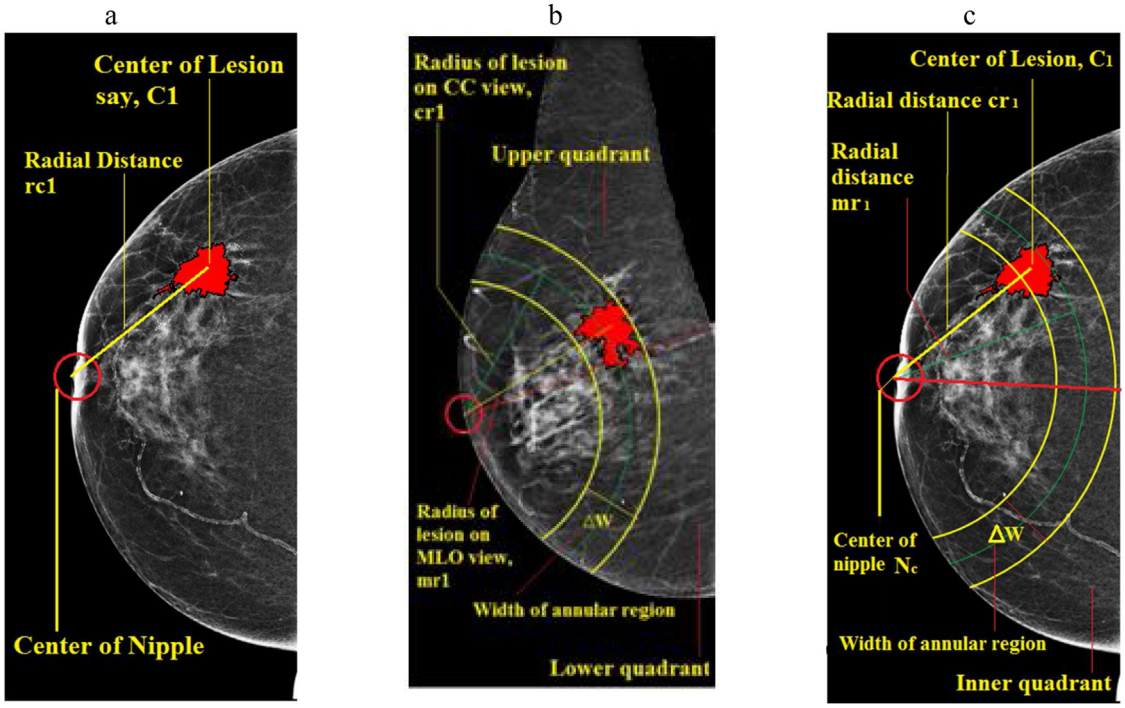

Fig. 6 – Breast coordinate system: (a) CC view measuring R^1_{cc}, R^2_{cc}; (b) MLO view measuring R^1_{mlo}, R^2_{mlo} annular region within yellow arcs, midline (red) showing upper/lower quadrants; (c) CC view annular region within yellow arcs, midline (green) showing outer/inner quadrants.

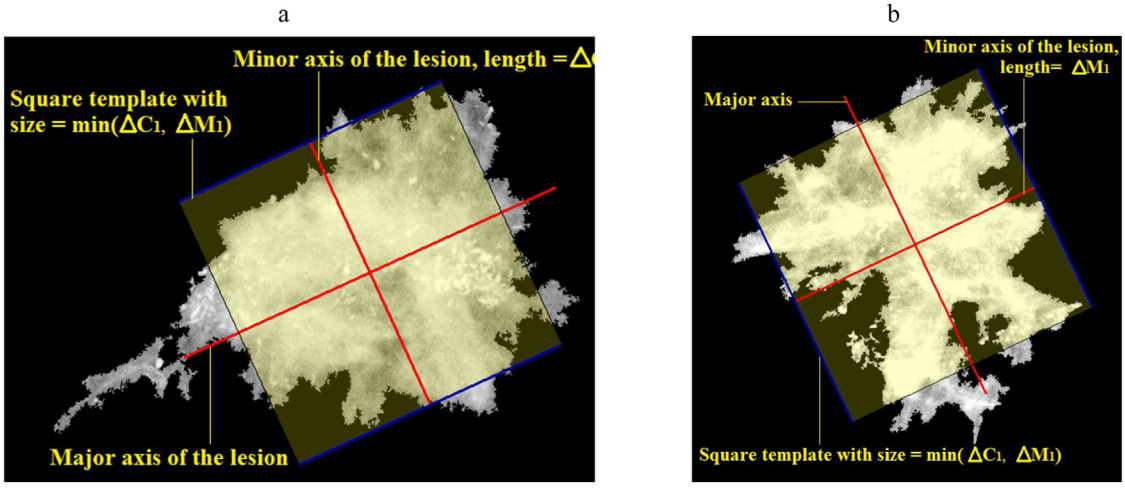

Fig. 7 – The lengths of minor axis of the sample lesions: (a) on CC view; (b) on MLO view.

study, the TP–TP pairs demonstrate dominant geometrical and textural features than that of TP–FP or FP–FP pairs and hence are easy to identify. In this regard, a set of features which are invariant to the positioning and the compression are chosen. The features chosen based on an intuitive probability that they would extract useful information are explained briefly as below.

- *Pixelwise correlation (SF1)*: This similarity measure is based on Pearson's correlation coefficient. Pixelwise correlation of a suspicious lesion L_{cc} on cc view with any suspicious lesion L_{mlo} on MLO view mammogram is defined using Eq. (8).

$$SF1 = r = \frac{\sum_{i,j}(L_{cc}(i,j) - L_{cc})(L_{mlo}(i,j) - L_{mlo})}{\sqrt{\sum_{i,j}(L_{cc}(i,j) - L_{cc})^2}\sqrt{\sum_{i,j}(L_{mlo}(i,j) - L_{mlo})^2}} \quad (8)$$

where $L_{cc}(i,j)$ and $L_{mlo}(i,j)$ are the pixel value at ith row jth column of lesion template whereas L_{cc} and L_{mlo} are the mean

pixel values of the lesion template of the suspicious lesions being evaluated, on CC and MLO view respectively.

- *Correlation standardized by median (SF2)*: It is a variation of Pearson's correlation coefficient and is defined using Eq. (9).

$$SF2 = r_{med} = \frac{\sum_{i,j}(L_{cc}(i,j) - \hat{L_{cc}})(L_{mlo}(i,j) - \hat{L_{mlo}})}{\sqrt{\sum_{i,j}(L_{cc}(i,j) - \hat{L_{cc}})^2}\sqrt{\sum_{i,j}(L_{mlo}(i,j) - \hat{L_{mlo}})^2}} \quad (9)$$

where $\hat{L_{cc}}$ and $\hat{L_{mlo}}$ are the median pixel values of the lesion template of the suspicious lesions being evaluated, on CC and MLO view respectively.

- *Pixelwise mutual information (SF3)*: It is assumed that the similar lesions contain a lot of intensity pixels which are redundant. The purpose of this information measure is to maximize the redundant information between the participating lesions $L_{cc}(i,j)$ and $L_{mlo}(i,j)$. The definition for the mutual information is given in Eq. (10).

$$SF3 = MI = \sum_{L_{cc}(i,j),L_{mlo}(i,j)} p(L_{cc}(i,j), L_{mlo}(i,j)) \times \log \frac{p(L_{cc}(i,j), L_{mlo}(i,j))}{p(L_{cc}(i,j))\, p(L_{mlo}(i,j))} \quad (10)$$

where p denotes probability. The probabilities in Eq. (10) are calculated based on a joint histogram of intensities of pixels levels of $L_{cc}(i,j)$ taken along the x axis of the histogram and the pixel levels of $L_{mlo}(i,j)$ are taken along the y axis.

- *Difference in entropy (SF4)*: Shannon's entropy is a measure of the average information carried in a pattern, which is widely used to quantify the smoothness of image texture. Entropy usually is relatively higher in heterogeneous regions and is low in homogeneous patterns. The absolute difference between the entropy of the regions on ipsilateral views is considered as a measure of dissimilarity. It is presented in Eq. (11).

$$SF4 = H(L_{cc}(i,j)|L_{mlo}(i,j)) = \sum_{i=1}^{n} L_{cc}(i,j) \log_2 \frac{L_{cc}(i,j)}{L_{mlo}(i,j)} \quad (11)$$

4.3.3. Computing correspondence score

The location-based information and the similarity features of the corresponding pairs of suspicious lesions from ipsilateral views are used to compute the correspondence score for all lesions. The features are assigned with the weight factors based on their rank as explained in Ref. [47]. The procedure followed to compute the correspondence score of a lesion pair is depicted in Fig. 4 and the same is explained with the help of two scenarios as below.

(1) A lesion on CC view is paired with a lesion on MLO view and for this pair –
   (i) Similarity features SF1, SF2, SF3, and SF4 are calculated;

$$SC(i,j) = \frac{(4 \times SF1 + 3 \times SF2 + 2 \times SF3 + SF4)}{10} \quad (12)$$

   where 4, 3, 2, 1 are the weight factors based on ranking of the features.

   (ii) Radial distance-based feature RF is calculated as

$$RF(i,j) = 2 \times (1 - R_y) \quad (13)$$

   where R_y is calculated using Eq. (1).

   (iii) Correspondence score of this pair is calculated as

$$CS(i,j) = \frac{(RF(i,j) + SC(i,j))}{2} \quad (14)$$

   Similarly, correspondence score $CS(1,2), \ldots, CS(1,n), \ldots, CS(m,n)$ are calculated.

   The pair with highest score is assumed to be a TP–TP pair, rest all are false pairs.

(2) A lesion on CC view paired with no lesion on MLO view: the correspondence score is set to zero.

Though the lesions from ipsilateral views are taken, the aim here is to assess the suspiciousness of the individual lesion on one (CC) view using its correlation with lesion on the other (MLO) view as additional information. Firstly, the highest correspondence score is assumed as belonging to a true positive lesion on the CC view. Secondly, the correspondence score for true positive lesion on MLO view is calculated. The correspondence score can act as an extended feature of that individual lesion. Hence, it is used in fusion of single and two view detection scheme covered in the next section.

4.4. Fusion of single and two view information

As illustrated in Fig. 4, the radiomic features of suspicious lesions are fused with the two views correspondence score to generate the final feature vector. For the single view CAD system, a total of 72 textural features and 7 geometric features are extracted from each of suspicious lesions. The maximum 4 lesions per image are considered for the sake of simplicity. This limit allows us to consider lesions on both views for pairing while keeping the FP rate as low as possible. The selected suspicious lesions are allowed to undergo two views analysis as illustrated in Fig. 4. As described in Section 4.3.2, a correspondence score is calculated using radial distance-based features and similarity features of lesions on CC and MLO views. This correspondence score is extended with single view-based feature vector of 79 features. In order to improve the overall detection performance, malignancy score of a single view CAD system is combined with correspondence score of two view CAD system. The detected lesions are classified as either malignant tumour or benign mass using SVM classifier. To avoid the biasing in the SVM classifier and to test its performance, 5 random partitions in the training to testing ratio of 4:1 are employed [48]. For each partition, a score of classifier is recorded separately of all partitions and average score is calculated. This approach of classification is similar to single view diagnosis described in Section 4.2.

5. Experimental results

5.1. Image-based detection

As a first part of our experiment, the individual images of ipsilateral views are pre-processed and then segmented to

Table 1 – Results after segmentation on TMC dataset.

True disease status w.r.t. HPR	Results by proposed segmentation method			Sensitivity	FPs/I
	Abnormal	Normal	Total		
Abnormal	241 (TP)	22 (FN)	263	91.63%	1.35
Normal	297 (FP)	–	297		

Table 2 – Results after segmentation on DDSM dataset.

True disease status w.r.t. HPR	Results by proposed segmentation method			Sensitivity	FPs/I
	Abnormal	Normal	Total		
Abnormal	164 (TP)	22 (FN)	186	88.17%	1.51
Normal	246 (FP)	–	224		

identify the suspicious lesions. Some of these identified lesions which are actually not true are known as false positives (FP). The total number of FPs from all the images divided by number of images give rise to the term false positives per image (FPs/I). One of the goals in this article is to reduce the FPs/I to the extent possible. A set of empirical rules is employed to reduce FPs/I to some extent as demonstrated in Figs. 2 and 3. The results of subjective evaluation of the automatic segmentation for detecting suspicious lesions on images of TMC dataset by the expert radiologists concerning the biopsy-proven HPRs are enlisted in Table 1.

As shown in Table 1, the segmentation stage yields total (241 + 297) 538 suspicious lesions from 220 abnormal mammograms. There are 262 lesions on CC view and 276 lesions are on MLO view. Thus, the single view CAD algorithm has the image-based detection sensitivity of (241/263) 91.63% at the cost of (297/220) 1.35 FPs/I of TMC dataset of 220 images. Similarly, the results of subjective evaluation of the automatic segmentation for detecting suspicious lesions on images of DDSM dataset by the expert radiologists are enlisted in Table 2.

The single view CAD system detected (164 + 246) 410 lesions (198 on CC and 212 on MLO view) with image-based detection sensitivity of (164/186) 88.17% at the cost of (246/148) 1.66 FPs/I on DDSM dataset of 148 images.

5.2. *Image-based diagnosis*

Image-based diagnosis includes classifying the 241 TP lesions as either benign mass or malignant tumour. For the single view CADx system, a total of 72 textural features and 7 geometric features are extracted from all the lesions detected as suspicious. The performance of the classifier is tested using combined features only. Feature vectors of all 241 TP suspicious lesions are divided into 4 sets of 48 lesions for training and remaining set of 49 for testing in a five-fold cross validation using SVM classifier. Five folds allow more variety of samples in a each single fold than that of ten folds. The testing set is selected in such a way that the proportion of actual benign masses to malignant tumours (1:4) is maintained approximately. The experiment is performed five times. The results are depicted in Table 3.

The image-based FPs/I in the single view diagnosis is 1.08 with 74.09% sensitivity. A similar experiment is performed on DDSM dataset and results are stored.

5.3. *Localizing and pairing*

The radiologists during their clinical practice observe carefully a single view, either CC or MLO, to find out the abnormal lesions. If they find the abnormal lesions on one view, they also carefully observe the second view. If the abnormality present on both the views is likely to represent the same physical lesion, the suspect is confirmed. Then it is characterized to decide its abnormality. This analysis using two views of same breast together is referred as case-based analysis and we have tried to mimic the same as explained below.

The annular region defined by the arc pair with radial distance as arc radius is used to locate the suspicious lesion on the MLO view which corresponds to the lesion on CC view. The prediction is based on location-based properties as well as similarity features extracted from the suspicious lesion pairs on ipsilateral views. The total of 262 lesions on CC views and 276 lesions on MLO view are the candidate lesions allowed to undergo two views analysis. The annular or search region is

Table 3 – Performance of SVM classifier in single view scheme.

Trial	True malignant	False malignant	True benign	False benign
1	27	14	7	1
2	24	15	7	3
3	25	15	8	1
4	26	13	8	2
5	24	17	7	1
Total	126	74	37	8

Table 4 – Assessment of pairing between CC and MLO view lesions on TMC dataset.

Correspondence from	Lesions on CC view	Lesions on MLO view	Possible pairs	Pairs within annular region	TP–TP pairs with highest score	TP–FP pairs	FPs/I
CC to MLO	262	276	860	520	155	365	1.66
MLO to CC	262	276	924	560	163	397	1.80

Table 5 – Assessment of pairing between CC and MLO view lesions on DDSM dataset.

Correspondence from	Lesions on CC view	Lesions on MLO view	Possible pairs	Pairs within annular region	TP–TP pairs with highest score	TP–FP pairs	FPs/I
CC to MLO	198	212	638	376	113	263	1.78
MLO to CC	198	212	684	401	102	299	2.08

Table 6 – Performance of SVM classifier using fusion scheme on TMC dataset.

	True malignant	False malignant	True benign	False benign	Total pairs	Sensitivity	FPs/I	Kappa
CC to MLO	69	58	15	13	155	76.36	0.65	0.0047
MLO to CC	70	63	13	17	163	75.45	0.73	0.0565

dynamic and is based on the length of minor axis of the lesions participating in the pairing process. Hence, the missing of any lesion due to static parameter of region width is avoided which results in appropriate and limited pairing. There are two experiments performed in this clinical study. First, all suspicious lesions on CC view are selected one by one and are paired with every other lesion from MLO view and subsequently the correspondence score of the pair is calculated. The lesions on the MLO view which are lying outside the annular regions defined by the radial distances of all the lesions on CC view are discarded. Thus, there are limited number of pairs from which one with highest correspondence score is selected. This highest score actually belongs to the most likely abnormal lesion from suspicious lesions on CC view. In the second experiment, all suspicious lesions on MLO view are selected one by one and are paired with every other lesion from annular region on CC view. Subsequently, the correspondence score of the pair is calculated. All the possible pairs of lesions from CC to MLO view are divided into 4 sets of 104 pairs for training and remaining set of 104 pairs for five-fold cross validation using SVM classifier. SVM classified 155 of all the pairs as TP–TP and 365 as TP–FP. All the possible pairs of lesions from MLO to CC view are divided into 4 sets of 112 pairs for training and remaining set of 112 pairs for five-fold cross validation using SVM classifier. SVM classified 163 of all the pairs as TP–TP and 397 as TP–FP. The results are depicted in Table 4.

Due to limiting the number of lesions per mammogram, two view analysis of only 860 pairs is performed which yielded with 6.81 FPs/I before classification. After geometric elimination using arc-based method, pairing from CC to MLO yields 365 false pairs leading to (365/220) 1.66 FPs/I while pairing from MLO to CC yields (397/220) 1.80 FPs/I. Among 220 images, the average FPs/I is reduced to 1.73. The pair is declared as likely to be malignant or benign based on classification using fusion of single and two view scores as described in the next section. Similarly, the results on DDSM dataset are depicted in Table 5.

5.4. *Case-based analysis*

Once the TP–TP and TP–FP pairs are finalized, the task is to classify the pair of benign mass from the pair of malignant tumours. This is nothing but identifying correct TP–TP pairs. The 155 TP–TP pairs from CC to MLO are now divided into 4 sets of 31 pairs for training and set of 31 for testing using five-fold cross validation using SVM classifier. Similarly, 163 TP–TP pairs from MLO to CC are divided into 4 sets of 32 pairs for training and set of 35 for testing using five-fold cross validation using SVM classifier. The correspondence score of the lesion on CC view extended by combined features set forms a new feature vector. The same experiment for lesions on MLO view is performed. The performances of a SVM classifier to distinguish the suspicious lesions on CC and MLO view are depicted in Table 6.

The average case-based lesions diagnosis sensitivity is 75.91% (from 110 cases) at the cost of 0.69 FPs/I on an average. The sensitivity here is calculated based on the number of TP–TP pairs classified correctly either as benign mass or malignant tumour from actual 110 pairs. The specificity in this case is not meaningful as we have not selected true negatives (TN) for classification and hence not considered. Similarly, the results on DDSM dataset are depicted in Table 7.

The average case-based lesions diagnosis sensitivity is 73.65% (from 110 cases) at the cost of 0.72 FPs/I on an average. Sometimes the classifier performance metrics such as sensitivity, specificity and accuracy are not enough. Especially, the classifiers built and evaluated on multiple imbalanced data sets can be compared more reliably through the Kappa statistics. It gives a better indicator of how the classifier performed across all instances of different class distributions. In our experiment as shown in Tables 6 and 7, the values can be interpreted as indicators of correlation between observed and expected accuracy.

Fig. 8(a) depicts the performance of the SVM classifier on single, two view and fusion image-based CAD scheme using

Table 7 – Performance of SVM classifier using fusion scheme on DDSM dataset.

	True malignant	False malignant	True benign	False benign	Total pairs	Sensitivity	FPs/I	Kappa
CC to MLO	28	46	27	12	113	74.32	0.78	0.2815
MLO to CC	26	38	28	10	102	72.97	0.65	0.3138

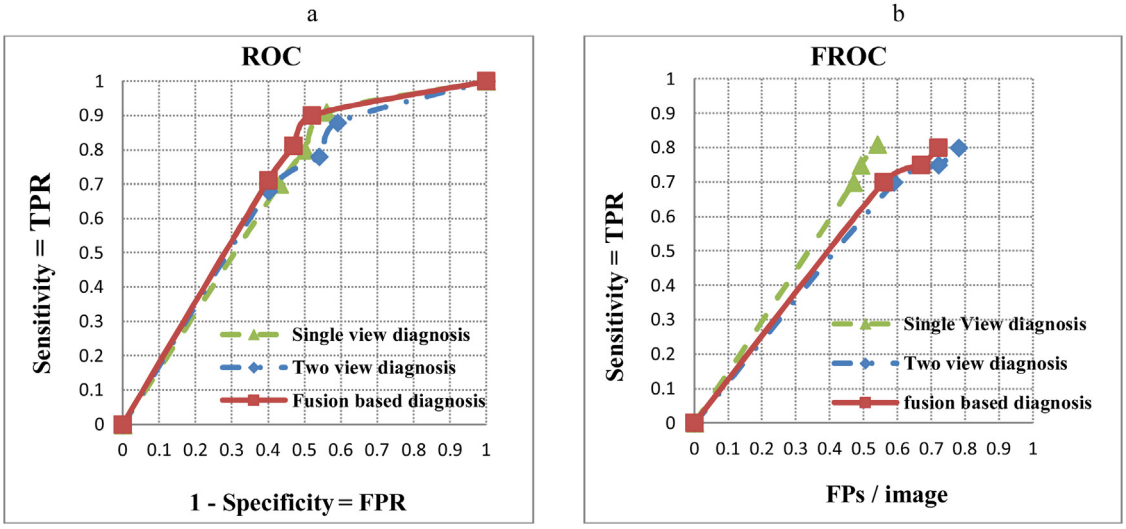

Fig. 8 – Performance of SVM classifier using: (a) ROC plot; (b) FROC plot.

ROC curves for the classifying malignant tumours from benign masses. Fig. 8(b) shows the case-based performance of SVM classifier on single, two views and fusion-based scheme assessed using average of the free-response ROC (FROC) curves of the two subsets. The comparison as such shows slight improvement in the case-based sensitivity; but case-based performance is more significant as far as reliable aid to the radiologists in clinical diagnosis is concerned. Area under ROC curve (AUROC) showing performance of classifier on TMC dataset is $A_z = 0.74 \pm 0.23$ for single view, $A_z = 0.79 \pm 0.12$ for two view and $A_z = 0.82 \pm 0.27$ for fusion-based scheme. Two sample *t*-test is conducted to understand whether the difference between single view and fusion-based scheme is statistically significant or not. Usually the difference denoted by p value less than 0.05 indicates that results from clinical trials are statistically significant. In our experiment, the p value is found less than 0.02. Hence the proposed fusion-based scheme is superior to the single view approach.

Table 8 compares the results of classifier in terms of FPs/I for single view, two view and fusion-based scheme at three different adjusted sensitivity levels.

5.5. *Computational complexity*

The preprocessing using fuzzy *c*-means algorithms has the complexity of $O = (i\,d\,n\,c^2)$ where i is number of iterations, d is number of dimensions ($d = 2$ in our case), n is number of pixels in the given image and c is number of clusters. The computational complexity of segmentation algorithm is $O = (n)$. The computational complexity of feature extraction with GLCM of a image matrix size $M \times M$ is $O = (M^2)$. The complexity of SVM classifier with RBF kernel is $O = (d\,n_s)$ where d is number of input dimensions and n_{sv} is the number of support vectors [49].

6. Discussion

FFDM images used in our study cover larger breast area and the nipple is in the profile most of the time. This fact makes the proposed task using arc-based method relatively easy to implement compared to that of in Ref. [20]. One of the limitations of the proposed methods lies with the detection accuracy of nipple centre and lesion centre. As mentioned in the dataset, the border and nipple of the selected images is clearly visible in the profile and hence detected successfully. Actually, in some cases the nipple may not be in the profile or border sometimes is not detectable successfully. In all such cases the radial distance of the lesions from nipple centre may be wrong. The width of annular region can also be incorrect leading to the failure of the proposed method. The accuracy may come down with increasing number of such cases. However, we are trying to modify our earlier algorithm to correctly delineate the border and identify the nipple position correctly especially with difficult cases in our future work. Consistent and accurate positioning of the breast during mammography is essential for our method to acquire accurate distance metrics for correspondence score. Incorrect and inconsistent positioning may lead to lower accuracy that was obtained on the dataset in this study.

Table 8 – Comparison of single view, two view and fusion-based scheme.

	Sensitivity	Single view FPs/I	Two-view FPs/I	Fusion FPs/I
TMC dataset	80%	1.17	1.08	0.74
	75%	1.10	1.02	0.69
	70%	1.03	0.95	0.64
DDSM dataset	80%	1.37	0.97	0.77
	75%	1.28	0.95	0.72
	70%	1.20	0.92	0.67

The authors in Refs. [9,46] showed that there is a high correlation between distances from the nipple to the centroid of lesions on CC and MLO views. The proposed work has leveraged this idea to devise a mechanism to choose lesion-based parameters as specified through Eqs. (5)–(7) in Section 4.3.1 to determine the width of annular region which is different for every individual lesion on the either views. The results shown in Tables 4 and 5 prove that our selection of annular region is appropriate and works well on mammograms of larger as well as smaller breast area. It is observed that in few cases the correspondence score of TP–FP pair is better than that of TP–TP pair. We are investigating the reasons behind such exceptional cases.

The purpose of minimizing the false pairs of lesions using arc-based method is served well by selecting an appropriate width of the annular region. The prediction of the radial distances of the lesion from nipple on CC and MLO views and the width of annular search region are determined statically in Refs. [9,28] which are applicable only to the images in their respective studies. In this regard, we have selected the width of annular region dynamically based on the length of minor axis of both lesions involved in the process. The success of this task is dependent on the accurate positioning of the lesion and its centre during the segmentation phase.

During matching the lesions on ipsilateral views in two view analysis, one lesion may be paired with two or more lesions at approximately the same distance in the annular search region. Obviously, the only distance metric is not sufficient to select the real TP–TP pair. We have selected four simple similarity measures based on our intuitive expectation that they would provide useful information for matching the corresponding lesions. The radial distance-based score combined with a similarity score, as described in Section 4.3.3 to generate a correspondence score, represent suspiciousness of the candidate lesion on either of the ipsilateral views. Consequently, this correspondence score of a lesion helps in matching it with the most appropriate candidate on the other view. Although there is no substantial improvement observed in the results (sensitivity) in Tables 6 and 7 as compared to that of in Tables 4 and 5, the FPs/I is reduced to a satisfactory extent. More number of possible pairs in the proposed method are the bottleneck in the overall performance. Even though the lesion on one view is not paired with any other lesion on other view, the single view characterization can yield the classification results as shown in Fig. 4.

The final stage in the proposed work for improving lesion detection is the fusion of two view correspondence scores with the one-view abnormality score. The fusion process is simple and computationally simpler than approaches described in the literature. The metric shown in terms of Kappa indicates the slight correlation between observed and expected accuracy which can be improved further. However, better features can be selected to improve the correspondence score and subsequently the accuracy in matching the lesion pairs.

7. Conclusion

The significantly vital concurrent analysis for estimating likelihood of cancer of the abnormal lesions on the ipsilateral views of the mammogram is described in this article. The proposed approach is more realistic as it closely mimics the radiologist's clinical practice of two views-based analysis of breast anomalies. The preliminary results after fusion of single view features with two views information have demonstrated a slight improvement in the case-based diagnostic accuracy and FPs/I close to the film-based scheme. The findings of proposed two view scheme are significantly useful compared to the traditional single view CAD systems. The proposed algorithms are helpful to improve the outcomes and efficiency of two views CAD system for assisting the radiologists during their diagnostic decisions in clinical practice. However, the results obtained can be further optimized in terms of case-based sensitivity and specificity. The perfection and accuracy in automatic nipple detection can be improved further and we are attempting to do the same to enhance the performance of our proposed approach. The overall performance improvements in our proposed algorithms are encouraging to increase the potential usage of our two views CAD system as a ''second reader'' in breast cancer diagnosis.

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at doi:10.1016/j.bbe.2019.04.008.

REFERENCES

[1] Mahajan A, Goh V, Basu S, Vaish R, Weeks A, Thakur M, et al. Bench to bedside molecular functional imaging in translational cancer medicine: to image or to imagine? Clin Radiol 2015;70(10):1060–82.

[2] Pawar MM, Talbar SN, Dudhane A. Local binary patterns descriptor based on sparse curvelet coefficients for false-positive reduction in mammograms. J Healthcare Eng 2018;1–16.
[3] Tzias D, O'Flynn EAM, Allen SD, Wilson ARM. Current status and new developments in breast cancer diagnosis and detection. Eur Oncol Haematol 2013;9(1):21–6.
[4] He W, Juette A, Denton ERE, Oliver A, Martí R, Zwiggelaar R. A review on automatic mammographic density and parenchymal segmentation. Int J Breast Cancer 2015;1–31 [article ID 276217].
[5] Giger ML, Karssemeijer N, Schnabel JA. Breast image analysis for risk assessment, detection, diagnosis, and treatment of cancer. Annu Rev Biomed Eng 2013;15:327–57.
[6] Wei J, Chan HP, Sahiner B, Zhou C, Hadjiiski LM, Roubidoux MA, et al. Computer-aided detection of breast masses on mammograms: dual system approach with two-view analysis. Med Phys 2009;36(10):4451–60.
[7] Sun X, Wei Q, Song D. Ipsilateral-mammogram computer-aided detection of breast cancer. Comput Med Imaging Graph 2004;28:151–8.
[8] Goudarzi M, Maghooli K. Extraction of fuzzy rules at different concept levels related to image features of mammography for diagnosis of breast cancer. Biocybern Biomed Eng 2018;38(4):1004–14.
[9] Paquerault S, Petrick N, Chan HP, Sahiner B, Helvie MA. Improvement of computerized mass detection on mammograms: fusion of two-view information. Med Phys 2002;29:238–47.
[10] Celaya-Padilla Jose M, Guzmán-Valdivia Cesar H, Galván-Tejada Carlos E, Galván-Tejada Jorge I, Gamboa-Rosales Hamurabi, Garza-Veloz Idalia. Contralateral asymmetry for breast cancer detection: a CADx approach. Biocybern Biomed Eng 2018;38:115–25.
[11] Qian W, Song D, Lei M, Sankar R, Eikman E. Computer-aided mass detection based on ipsilateral multiview mammograms. Acad Radiol 2007;14(5):530–8.
[12] Blanks RG, Wallis MG, Given-Wilson RM. Observer variability in cancer detection during routine repeat (incident) mammographic screening in a study of two versus one view mammography. J Med Screen 1999;6:152–8.
[13] Kim SA, Chang JM, Cho N, Yi A, Moon WK. Characterization of breast lesions: comparison of digital breast tomosynthesis and ultrasonography. Korean J Radiol 2015;16(2):229–38.
[14] Houssami N, Abraham LA, Miglioretti DL, Sickles EA, Kerlikowske K, Buist DSM, et al. Accuracy and outcomes of screening mammography in women with a personal history of early-stage breast cancer. J Am Med Assoc 2011;305(8):790–9.
[15] Qian W, Song D, Lei M, Sankar R, Eikman E. Computer-aided mass detection based on ipsilateral multiview mammograms. Acad Radiol 2007;14:530–8.
[16] Sun W, Tseng T, Qian W, Saltzstein EC, Zheng B, Yu H, et al. A new near-term breast cancer risk prediction scheme based on the quantitative analysis of ipsilateral view mammograms. Comput Methods Programs Biomed 2018;155:29–38.
[17] Vignesh AA, Bonnie NJ, Natalya ML, Jessica WTL, James Brenner R, Flowers CI, et al. Benefit of semiannual ipsilateral mammographic surveillance following breast conservation therapy. Radiology 2012;264(2):371–7.
[18] Doi K. Computer-aided diagnosis in medical imaging: historical review, current status and future potential. Comput Med Imaging Graph 2007;31:198–211.
[19] Good WF, Zheng B, Chang YH, Wang ZH, Maitz GS, Gur D. Multi-image CAD employing features derived from ipsilateral mammographic views. Proc. 1999 SPIE 3661 Conference; 1999.
[20] Chang YH, Good WF, Sumkin JH, Zheng B, Gur D. Computerized localization of breast lesions from two views: an experimental comparison of two methods. Invest Radiol 1999;34(9):585–8.
[21] Paquerault S, Petrick N, Chan HP, Sahiner B, Dolney AY. Improvement of mammographic lesion detection by fusion of information from different views. Proc. 2001 SPIE 4322; 2001.
[22] Sahiner B, Petrick N, Chan HP, Paquerault S, Helvie MA, Hadjiiski LM. Recognition of lesion correspondence on two mammographic views – a new method of false-positive reduction for computerized mass detection. Proc. 2001 SPIE 4322; 2001.
[23] Engeland S, Timp S, Karssemeijer N. Finding corresponding regions of interest in mediolateral oblique and craniocaudal mammographic views. Med Phys 2006;33:3203.
[24] Zheng B, Leader JK, Abrams GS, Lu AH, Wallace LP, Maitz GS, et al. Multiview-based computer-aided detection scheme for breast masses. Med Phys 2006;33(9):3135–43.
[25] Yuan Y, Giger ML, Li H, Sennett C. Correlative feature analysis on FFDM. Med Phys 2008;35(12):5490–500.
[26] Padayachee J, Alport MJ, Rae WD. Mammographic CAD: correlation of regions in ipsilateral views – a pilot study. South Africa J Radiol 2009;48–54.
[27] Zheng B, Tan B, Ganott MA, Chough DM, Gur D. Matching breast masses depicted on different views: a comparison of three methods. Acad Radiol 2009;16(11):1338–47.
[28] Samulski M, Karssemeijer N. Optimizing case-based detection performance in a multiview CAD system for mammography. IEEE Trans Med Imaging 2011;30 (4):1001–9.
[29] Tanner C, Schie G, Lesniak J, Karssemeijer N, Székely G. Improved location features for linkage of regions across ipsilateral mammograms. Med Image Anal 2013;17:1265–72.
[30] Heath M, Bowyer K, Kopans D. Current status of the digital database for screening mammography. Proc. 1998 4th International Workshop on Digital Mammography; 1998. pp. 457–60.
[31] Pawar MM, Talbar SN. Local entropy maximization based image fusion for contrast enhancement of mammogram. J King Saud Univ - Comput Inform Sci 2018.
[32] Gandhamal AP, Talbar SN, Gajre SS, Hani AFM, Kumar D. Local gray level S-curve transformation – a generalized contrast enhancement technique for medical images. Comput Biol Med 2017;83:120–33.
[33] Baid U, Talbar SS, Talbar SN. Comparative study of K-means, GMM and fuzzy clustering for brain tumour segmentation. Proc. 2016 ICCASP BATU, Lonere International Conference; 2016.
[34] Bezdek J, Ehrlich R, Full W. Cluster validity with fuzzy sets. Comput Geosci 1984;10(3):191–203.
[35] Oliver A, Tortajada M, Lladó X, et al. Breast density analysis using an automatic density segmentation algorithm. J Digit Imaging 2015;28(5):604–12.
[36] Sapate SG, Talbar SN. An overview of pectoral muscle extraction algorithms applied to digital mammograms. In: Dey N, et al., editors. Medical imaging in clinical applications. Singapore: Springer; 2016 [chapter 2].
[37] Baid U, Talbar SS, Talbar SN. Brain tumour segmentation based on non-negative matrix factorization and fuzzy clustering. Proc. 2017 Int. Conf. on BioImaging; 2017.
[38] Shirley SS, Shenbaga D. Automatic seed point selection in ultrasound echography images of breast using texture features. Biocybern Biomed Eng 2015;35(3):157–68.
[39] Sapate SG, Mahajan A, Talbar SN, Sable N, Desai S, Thakur M. Radiomics based detection and characterization of suspicious lesions on full field digital mammograms. Comput Meth Program Biomed 2018;163:1–20.

[40] Wierzchowska MD, Borys D, Billewicz BB, Jarzab M, Swierniak A. Simplification of breast deformation modelling to support breast cancer treatment planning. Biocybern Biomed Eng 2016;36:531–6.
[41] Lahmiri S, Boukadoum M. DWT and RT-based approach for feature extraction and classification of mammograms with SVM. Proc. 2011 IEEE Biomedical Circuits and Systems Conference (BioCAS); 2011.
[42] Lahmiri S, Boukadoum M. Hybrid discrete wavelet transform and Gabor filter banks processing for mammogram features extraction. Proc. 2011 IEEE 9th International New Circuits and Systems Conference; 2011.
[43] Vert JP, Tsuda K, Bernhard S. A primer on kernel methods. Kernel Meth Comput Biol 2004;47:35–70.
[44] Lahmiri S, Boukadoum M. Hybrid discrete wavelet transform and Gabor filter banks processing for features extraction from biomedical images. J Med Eng 2013;13 [article ID 104684].
[45] Tourassi GD, Vargas-Voracek R, Catarious DM, Floyd CE. Computer-assisted detection of mammographic masses: a template matching scheme based on mutual information. Med Phys 2003;30(8):2123–30.
[46] Filev P, Hadjiiski L, Sahiner B, Chan H, Helvie M. Comparison of similarity measures for the task of template matching of masses on serial mammograms. Med Phys 2005;32(2):515–29.
[47] Samulski M, Karssemeijer N. Matching mammographic regions in mediolateral oblique and cranio caudal views: a probabilistic approach. In: Giger ML, Karssemeijer N, editors. Proc. 2008 SPIE Medical Imaging Conference. 2008.
[48] Lahmiri S, Boukadoum M. Hybrid cosine and Radon transform-based processing for digital mammogram feature extraction and classification with SVM. Proc. 2011 Annual International Conference of the IEEE Engineering in Medicine and Biology Society; 2011.
[49] Chapelle O. Training a support vector machine in the primal. Neural Comput 2007;19(5):1155–78.