IET Image Processing

Research Article

Context-based ensemble classification for the detection of architectural distortion in a digitised mammogram

ISSN 1751-9659
Received on 23rd May 2019
Revised 17th September 2019
Accepted on 13th November 2019
E-First on 3rd February 2020
doi: 10.1049/iet-ipr.2019.0639
www.ietdl.org

Yusuf Akhtar[1] ✉, *Dipti Prasad Mukherjee*[1]

[1]*Electronics and Communication Sciences Unit, Indian Statistical Institute, Kolkata, India*

✉ *E-mail: yusuf.eciit@gmail.com*

Abstract: The problem of computer-aided detection of architectural distortion (AD) in a digitised mammogram has been attempted in this manuscript. In examining a mammogram, the decision regarding a particular region of interest (RoI) is dependent on the appearance of the surrounding regions. However, in existing methods to detect AD the inference about an RoI is dependent on the appearance of this RoI alone. In addition, multiple radiologists infer the same mammogram in coming to a final decision about the mammogram. Contrary to popular ensemble classifiers like Adaboost and Random Forest, the authors propose an ensemble based method (imitating multiple radiologists by classifiers) for detecting AD such that the decision on a test RoI is dependent on the decisions of the surrounding RoIs in the proposed ensemble classifier. The proposed context-based ensemble classifier has been validated on two mammographic databases. The proposal shows promising results in both the databases.

1 Introduction

Architectural distortion (AD) is an abnormality of the breast that is detected in a mammogram (an X-ray imaging modality of the breast). AD [1] is defined as a focal retraction of the breast tissue or a radiating pattern of ridges referred by spicules in the medical literature (refer Fig. 1). The detection of AD is challenging because of its subtle appearance [1]. The early detection of AD is crucial because of its relevance to the survival chances of a cancer patient [3]. To provide an inexpensive large-scale screening of mammograms, several computer-aided detection (CAD) algorithms [4–7] have been designed for the detection of AD.

There are two issues with the preceding proposals to detect AD. First is that in the existing works for detecting AD, the inference about a region of interest (RoI) in the test set is dependent on the feature vector of this RoI alone (refer Fig. 1*e*). In practice, however, the decision regarding an RoI is influenced by the appearance of surrounding RoIs. None of the existing approaches to detect AD incorporate this vital observation of the preceding statement.

Second weighing the decisions regarding a particular RoI from atleast two radiologists is commonplace in coming to a final decision regarding this particular RoI. The number of radiologists could be increased theoretically; however practical issues like cost incurred in examining the mammograms limit such increase in the number of radiologists. The correct label of a particular RoI is likely to occur the most number of times in the interpretations of an RoI from multiple radiologists. An attempt of the detection of AD that looks similar to the example on multiple radiologists is weighing the decisions of multiple classifiers (in place of radiologists) in coming to a final decision regarding a particular RoI. Results on different pattern recognition problems show that an ensemble performs better [8] than a single classifier in certain problems.

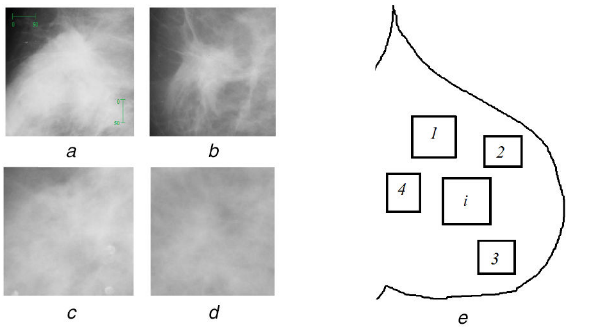


Fig. 1 *Illustration of the examination of a mammogram*
(a), (b) Examples of prominent AD sites (mdb115.pgm and mdb117.pgm [2]), ***(c), (d)*** Examples of ambiguous AD sites (mdb121.pgm and mdb130.pgm [2]), ***(e)*** Example showing multiple RoIs: decision of RoI *i* depends on decisions on RoIs 1, 2, 3 and 4

The reader may note that in the machine learning literature there are two popular ensemble classifiers like Adaboost [8] and Random Forest [9]. In this manuscript, a method has been proposed that incorporates the observation of Fig. 1*e*. The proposed ensemble classifier assumes that every classifier in an ensemble would behave consistently on a large set of mammograms. Till date, to the best of the authors' knowledge, there has been no proposal to detect AD with the aid of ensemble classifiers.

The objective of an ensemble classifier is to assign a class label (AD/normal) to a RoI in a test set of RoIs. Let the class label assigned to the *i*th RoI in the test set be denoted by c_i. Let us denote the class label assigned to the *j*th RoI in the training set by a radiologist by the same notation c_j. The class label c_i (c_j) can take value -1 (representing normal) and 1 (representing AD). Let us assume that there are K number of classifiers. Each classifier assigns a class label to an RoI in a set of RoIs.

Let the class label assigned to the *i*th RoI by the *k*th classifier be denoted by c_{ki}. Therefore, each RoI has K number of class labels assigned to it which may not be consistent with one another. It is apparent that c_j of an RoI in the training set is available with us from the ground truth. However if the *i*th RoI belongs to the test set, c_i is not available; the problem is to predict the class label c_i.

How is the proposal different: The proposed ensemble approach differs from Adaboost and Random Forest as the decision regarding multiple RoIs influence the decision of an individual RoI in the proposal. The above formulation is assumed because the labelling of an RoI depends not only on the pixel values within it but also on the context of the problem that may be elicited from the appearance of the surrounding regions. Hence, the proposed ensemble approach is referred as context based ensemble approach to classification. Specifically, the proposed ensemble formulation can be viewed as follows. It is apparent that multiple RoIs are extracted from the mammograms in the test set. Let $\boldsymbol{F}_i$ denote the

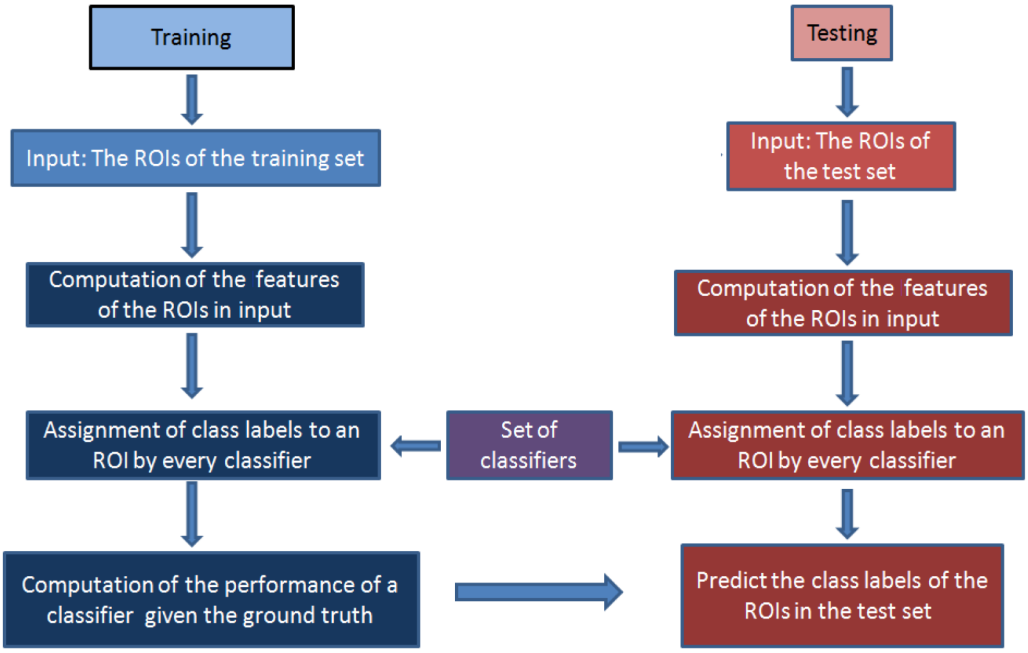

Training
Input: The ROIs of the training set
Computation of the features of the ROIs in input
Assignment of class labels to an ROI by every classifier
Computation of the performance of a classifier given the ground truth
Set of classifiers
Testing
Input: The ROIs of the test set
Computation of the features of the ROIs in input
Assignment of class labels to an ROI by every classifier
Predict the class labels of the ROIs in the test set

Fig. 2 *Overview of our ensemble based approach. The proposed ensemble classification assumes decision on multiple RoIs (RoIs with indices 1, 2, 3 and 4 in Fig. 1e) affect the decision on a particular ith RoI*

feature vector of the *i*th RoI. In AdaBoost and Random Forest, the class for the *i*th RoI, c_i is given by some function (designed during training) of the feature vector $\boldsymbol{F}_i$. Let this function be denoted by $G()$. We have

$$c_i = G(\boldsymbol{F}_i)\,. \tag{1}$$

In Adaboost, $G()$ is a weighted function of c_{ki} for all classifiers; each classifier referred by an index say k in an ensemble. The class labels c_{ki} are dependent on $\boldsymbol{F}_i$. Whereas in Random Forest, $G()$ returns the class label to which the trees (the classifiers in an ensemble; the forest) put the maximum number of votes. Note that the weights in Adaboost and the votes in Random Forest are computed during training and is used as it is during testing. In contrast to the above formulation, the class label that is assigned to an RoI in the proposal, is a function of $\boldsymbol{F}_i$, the feature vectors $\boldsymbol{F}_j$ ($j \neq i$) and also the class labels that we assign to the other RoIs of the test set. If the preceding function is denoted $H()$, for the proposal, we have

$$c_i = H(\{\boldsymbol{F}_i\} \cup \{\boldsymbol{F}_j, c_j; j \neq i\})\,. \tag{2}$$

For example in Fig. 1*e*, the class label of the *i*th RoI depends on $\boldsymbol{F}_i$ and features $\boldsymbol{F}_1$, $\boldsymbol{F}_2$, $\boldsymbol{F}_3$ and $\boldsymbol{F}_4$ and the class labels of RoIs with the indices 1, 2, 3 and 4.

In practice, decision regarding multiple RoIs influence the decision of an individual RoI. In (1), there is no provision to adopt the preceding observation evident from (2).

How does the proposal solve the ensemble classification: To solve (2), the proposal assumes that a classifier behaves consistently over a large set of mammograms. The first step in solving the preceding equation lies in computing the performance (sensitivity and specificity) of a classifier on the RoIs in the training set as we have c_j if the *j*th RoI belongs to the training set. In the proposal it is assumed that a classifier would produce the same sensitivity and specificity in the test set. This implies that the class label c_i of the *i*th RoI in the test set should be such that a classifier should produce the same sensitivity and specificity as in the training set. The preceding conclusion should also be true for every classifier in an ensemble. The AD detection problem then reduces to making an educated guess of c_i. Details of this educated guess are presented in Section 3. An illustration of the methodology is shown in Fig. 2.

In Adaboost and in Random Forest, a training model is built. The training model is used as it is in the test set. The proposal of this manuscript differs in this regard since the only information the proposal incorporates from training is the sensitivity and the specificity of a classifier and their consistency. In the proposal, a classifier recomputes its model in the test set. In brief, this recomputation involves the computation of a threshold based on the features of all RoIs in the test set. Comparing the features of an RoI with this threshold, a classifier assigns a class label to the RoI in the test set.

The methodology of the proposal is discussed in Section 3. The methodology is validated on two mammographic databases. The results are discussed in Section 4. In the next section, we present existing works that are also related to the detection of AD.

2 Literature survey

Most of the existing CAD approaches detect AD by defining features on the spicules (faint ridges). These features assist in the distinction between an AD site and a normal site. For example in [4–7], ridges are extracted with the aid of Gabor filters. These Gabor filters return the orientation of the ridge (if any) at a pixel of an RoI. The phase portrait model is used to encapsulate these orientations (at all the pixels) of an RoI. The parameters of the phase portrait model are used along with a classifier such as support vector machine (SVM) to distinguish between an AD site and a normal site. In [10, 11], ridges are extracted with the aid of linear directional filters. The concentration (evaluating the radiating pattern of linear structures) and isotropic indices (evaluating the distribution of the orientations of the linear structures) are used to evaluate the pattern of the ridges. Using discriminant analysis on the concentration index, the isotropic index and seven other features obtained from the pixel values in an RoI, the RoI is labelled as an AD site or a normal site. In [12], ridges are detected with the aid of Radon transform and a filter that resembles a striped disc. Features obtained by applying the Radon transform and the disc like filter are taken as input to SVM classifier (that is trained on the feature vectors of the training set) to return the AD sites in a test mammogram.

In [13], linear structures are detected with the aid of spiculated lesion filters. The probability of the occurrence of a pattern of the linear structures is characterised with the aid of a Gaussian mixture model (GMM). The parameters of the GMM model that represent the chance of occurrence of such a pattern (of the linear structures in an RoI) are used along with the SVM classifier to label an RoI as an AD site or a normal site. In [14], a modified hierarchical clustering has been adopted to extract spicules. Segmentation of these extracted spicules with the aid of active contours is implemented to isolate the spiculated regions from other normal regions.

Other approaches to detect AD opine that an AD site is associated with a bright patch. For example in [15], blobs are detected with the aid of Otsu [16] method of clustering. This is followed by area based thresholding of the blobs to report the AD sites. The drawback of [15] lies in the definition of an AD site that does not relate the appearance of an AD site to a bright patch.

Some approaches do not rely on the extraction of linear structures to detect AD. These approaches use an abstract representation of the pixel values of an RoI. For example in [17], the texture of an RoI is evaluated with the aid of difference of Gaussian (DoG) based filters. This adoption of the DoG model decomposes an image into sub-images. The probabilities of the occurrence of the sub-images are characterised with the aid of a GMM. The parameters of the GMM model are used along with the SVM classifier to label an RoI as an AD site or a normal site. In [18], an RoI is decomposed into sub-images by using bi-dimensional empirical mode decomposition. Three features are defined on these sub-images. Using SVM on these features an RoI is labelled as an AD RoI or a normal RoI. In [19], an RoI is transformed to polar co-ordinates from the Cartesian co-ordinates. The texture in this newly transformed RoI is evaluated with the aid of monogenic binary coding (MBC). The RoI is labelled as an AD or a normal site with the features from the MBC and a SVM classifier. The drawbacks of the preceding approaches discussed in this subsection is that the decision regarding a particular RoI in the test set is independent of the appearance of the surrounding regions; which is however not true when a mammogram is examined.

As mentioned in Section 1, we use an ensemble of classifiers to label an RoI as an AD or a normal site. There are two popular ensemble classifiers, AdaBoost [8, 20, 21] and Random Forest [9], both of which build a classifier with multiple weak learners. The reader may note that weak learners in AdaBoost are trained on feature vectors that are assigned unequal weights based on their difficulty level in classification. In contrast in Random Forest, equal weights are assigned to the feature vectors. In addition, the weak classifiers (the trees) are trained on different subsets of the training samples.

The proposal differs from AdaBoost and Random Forest since contrary to the preceding ensemble classifiers, the decision regarding an RoI in the test set is influenced by the decision of the other RoIs in the test set. This observation is incorporated in the proposal's ensemble classifier by assuming that all classifiers in an ensemble behave consistently over a large dataset of mammograms. We discuss the methodology of the proposal in the next section.

3 Methodology

The design of a classifier: It is apparent that the strong presence of certain features asserts whether an RoI belongs to a particular class. For example, a radiating pattern of ridges [1] indicates an AD site. The presence of a certain feature may be quantified by a measure. In the context of the detection of AD there are two classes we are interested in. One is the AD class and the other is the normal class. The aforesaid discussion indicates that we should be using two measures on the pixel values of an RoI; one measure for each class (AD/normal). These two measures are some functions, say $f1()$ and $f2()$, of the pixel values of an RoI.

Consider a set of RoI in the training set. Let us assume that we have some method to extract features from an RoI say R. Let the vector consisting of the features of R be referred by $\boldsymbol{F_R}$. Let us use two functions $f1()$ and $f2()$ to map $\boldsymbol{F_R}$ to a scalar. The function $f1()$ is used to map $\boldsymbol{F_R}$ to a scalar which specifies the class label of R as AD. This scalar is denoted by A_1. The AD class label is denoted by C_1. The function $f2()$ is used to map $\boldsymbol{F_R}$ to a scalar when the class label of R is set as normal. This scalar is denoted by A_2. The normal class label is denoted by C_2.

Let i assume the value 1 or 2. In the proposal it is assumed that R belongs to the class C_i if A_i is less than a threshold say T. If this condition is not satisfied R does not belong to C_i. The preceding statements are mathematically stated as

If $A_i < T$, R belongs to class C_i.

If $A_i \geq T$, R does not belong to class C_i.

There are three cases that arise when the threshold is varied. If the threshold is less than both A_1 and A_2 then $A_1 > T$ and $A_2 > T$. Therefore, R does not belong to either of C_1 or C_2. If the threshold is greater than both A_1 and A_2, then $A_1 < T$ and $A_2 < T$. Therefore, R belongs to both C_1 and C_2. The preceding two cases make our decision regarding the class to which this RoI belongs ambiguous.

However if the threshold is chosen between A_1 and A_2, either $A_1 < T < A_2$ or $A_1 > T > A_2$. Consider the first relation $A_1 < T < A_2$. Since $A_1 < T$, R is interpreted as belonging to the AD class. Since $A_2 > T$, R is interpreted as not belonging to the normal class.

Now consider the inequality $A_1 > T > A_2$. Since $A_1 > T$, R is interpreted as not belonging to the AD class. Since $A_2 < T$, R is interpreted as belonging to the normal class. In the former relation $A_1 < T < A_2$, R belongs to the AD class (C_1); in the latter inequality $A_1 > T > A_2$, R belongs to the normal class (C_2). This motivates the choice of the threshold as $(A_1 + A_2)/2$ since this expression always lies between A_1 and A_2. Let the aforesaid average of A_1 and A_2, referring to RoI R, be denoted by τ_R. Since there are multiple RoIs in the training set, we will have many such τ_R with us. We are constrained to choose only one value as the threshold T for multiple RoIs.

From the preceding discussion it is reasonable to conclude that selecting the τ_R (as the threshold T) that occurs most frequently will reduce the number of ambiguous decisions the most (where an RoI belongs to or does not belong to both the AD and normal classes) regarding the class labels of the RoIs. This motivates the selection of the mode of the expressions of the form $(A_1 + A_2)/2$ as the threshold. This mode(threshold) is denoted τ_X. The process is shown in Fig. 3.

The labelling of an RoI by a classifier: On selecting τ_X as the threshold, if the AD class label of the RoI R is correct and the normal class label of the RoI R is incorrect, this RoI is assigned the label 1. If the normal class label of the RoI R is correct and the AD class label of the RoI R is incorrect, this RoI is assigned the label -1. For the other cases when both the class labels are correct or incorrect, the RoI R is assigned the ambiguous class label, 0. This class label (1, -1 or 0) of RoI R is referred by lbn_R. A block diagram of this discussion is shown in Fig. 3.

Changing the functions $f1()$ and $f2()$ may generate a different lbn_R. Let there be N_A (N_N) number of AD (normal) sites in the training set. Let $N_R = N_A + N_N$ denote the total number of RoIs in the training set.

It is reiterated that the intention of the proposal in this manuscript is to incorporate the observation of Fig. 1*e* with the aid of an ensemble classifier. A classifier can be characterised with the pairs $f1()$ and $f2()$. Let us adopt K number of different pairs of $f1()$ and $f2()$. A pair of $f1()$ and $f2()$ is also referred by function pair.

Notations for an ensemble of classifiers: In order to unify the observation of Fig. 1*e* with the proposed ensemble classifier, a data structure (two- dimensional matrix) referred by $\boldsymbol{M}$ is used. Along a row of $\boldsymbol{M}$ we have class labels assigned to multiple RoIs by a single classifier. Along a column of $\boldsymbol{M}$ we have the class labels assigned to a particular RoI by all the classifiers in the ensemble.

It is now apparent that the dimensions of the matrix $\boldsymbol{M}$ is $K \times N_R$. Let the class label assigned to RoI R by the kth function pair be denoted by $lbn_R(k)$. If this RoI is the jth RoI in the training set, $\boldsymbol{M}(k, j) = lbn_R(k)$. It may be noted that a radiologist interprets an RoI as either belonging to the AD class (label=1) or the normal class ($label = -1$). Considering the ground truth that is given in the publicly available mammography databases [2, 22], the ambiguous class (label=0) is not present in the ground truth. The assessment by the radiologist needs to be represented in the

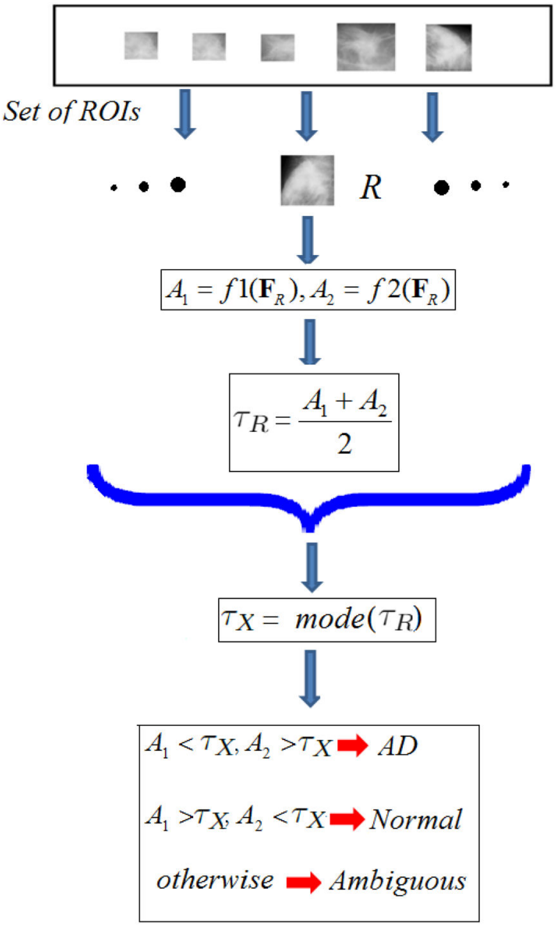

Set of ROIs
R
$A_1 = f1(\mathbf{F}_R), A_2 = f2(\mathbf{F}_R)$
$\tau_R = \frac{A_1 + A_2}{2}$
$\tau_X = mode(\tau_R)$
$A_1 < \tau_X, A_2 > \tau_X$ → AD
$A_1 > \tau_X, A_2 < \tau_X$ → Normal
otherwise → Ambiguous

Fig. 3 *Assignment of a class label to an RoI R by a classifier. The classifier in the image is characterised by the pair of functions $f1()$ and $f2()$. For details refer Section 3*

classification model. Given this, let $\boldsymbol{P}$ denote a $N_R \times 1$ column vector whose value at the jth index is equal to the class label (1 for AD/-1 for normal) of the jth RoI as per radiologist's assessment.

Analogous to the notations $\boldsymbol{M}$, N_R, N_A, N_N and $\boldsymbol{P}$ for the training set, let us define notations for the test set. Let $\boldsymbol{M}_e$ contain the class labels that are assigned to the test RoIs by all the classifiers in the ensemble. Specifically, $\boldsymbol{M}_e(k, j)$ contains the class label assigned to the jth RoI of the test set by the kth classifier. Let N_{Re} denote the total number of RoIs in the test set. Let N_{Ae} denote the number of AD RoIs in the test set. Let N_{Ne} denote the number of normal RoIs in the test set. Let $\boldsymbol{P}_e$ contain the class labels of the test RoIs as per ground truth.

It is apparent that the goal of the AD problem is that of discerning $\boldsymbol{P}_e$ (since this vector contains the true class labels of the RoIs in the test set as per radiologist's assessment). Let the estimate of $\boldsymbol{P}_e$ obtained from the proposal (discussed below) be denoted $\tilde{\boldsymbol{P}}_e$.

In the next subsection, the method of finding the estimate $\tilde{\boldsymbol{P}}_e$ is discussed.

3.1 Estimation of $\tilde{\boldsymbol{P}}_e$

Motivation: It is reiterated that the design of the proposal's ensemble classifier is based on the principle that a classifier in an ensemble would behave consistently on a large number of RoIs. Following derivation relates to estimation of performance.

The class label assigned to an RoI by the kth classifier may not be consistent with the radiologist's assessment. From this we imply that the kth classifier will assign correct labels to a certain fraction of the number of AD (as per ground truth) RoIs. Let this fraction (sensitivity) be denoted sen_k. Consider the following sum:

$$\sum_{j \in J1} \boldsymbol{M}(k, j), \quad J1 = \{j \text{ such that } \boldsymbol{P}(j) = 1\}. \tag{3}$$

This sum contains all the class labels that are assigned to the AD (as per ground truth) RoIs by the kth classifier. Expression (3) can be decomposed into

$$\sum_{j \in J2} 1 - \sum_{j \in J3} 1, \tag{4}$$

where $J2 = \{j$ such that $\boldsymbol{P}(j) = 1$ and $\boldsymbol{M}(k, j) = 1\}$; $J3 = \{j$ such that $\boldsymbol{P}(j) = 1$ and $\boldsymbol{M}(k, j) = -1\}$. Expression (4) denotes the number (the first summation or the LHS of (4)) of AD RoIs in the training set that have been correctly classified by the kth function pair minus the number (the second summation or the RHS of (4)) of AD RoIs that have been misclassified by the kth function pair. We may assume that by our judicious selection (median of the average of A_1 and A_2) of τ_X, the number of AD RoIs (as per ground truth) that are assigned the ambiguous class label (0) is very small compared to the number of AD RoIs that are classified as normal RoIs by the kth classifier. Therefore, the LHS term of expression (4) is equal to $N_A sen_k$ and the RHS term of the aforesaid expression is approximated as $N_A(1 - \text{sen}_k)$; the entire term equates to $N_A(\text{sen}_k - (1 - \text{sen}_k))$. If $(\text{sen}_k - (1 - \text{sen}_k))$ is denoted by sn_k (or $\text{sen}_k = (sn_k + 1)/2$), expression (4) becomes $N_A sn_k$.

Similarly, the kth classifier will assign correct labels to a certain fraction of the number of normal (as per ground truth) RoIs. Let this fraction (specificity) be denoted by spec_k. Consider the following sum:

$$-\sum_{j \in J4} \boldsymbol{M}(k, j), \quad J4 = \{j \text{ such that } \boldsymbol{P}(j) = -1\}. \tag{5}$$

This sum contains all the class labels that are assigned to the normal (as per ground truth) RoIs by the kth classifier. Expression (5) can be decomposed into

$$\sum_{j \in J5} 1 - \sum_{j \in J6} 1, \tag{6}$$

where $J5 = \{j$ such that $\boldsymbol{P}(j) = -1$ and $\boldsymbol{M}(k, j) = -1\}$; $J6 = \{j$ such that $\boldsymbol{P}(j) = -1$ and $\boldsymbol{M}(k, j) = 1\}$.

Expression (6) denotes the number (the first summation or the LHS of (6)) of normal RoIs that have been correctly classified by the kth function pair minus the number (the second summation or the RHS of (6)) of normal RoIs that have been misclassified by the kth function pair. We may assume that by our judicious selection of τ_X, the number of normal RoIs (as per ground truth) that are assigned the ambiguous class label (0) is very small compared to the number of normal RoIs that are classified as AD RoIs by the kth classifier. Therefore, the LHS of this expression (6) is equal to $N_N spec_k$ and the RHS of the aforesaid expression is approximated as $N_N(1 - \text{spec}_k)$; the entire expression equates to $N_N(\text{spec}_k - (1 - \text{spec}_k))$. If $(\text{spec}_k - (1 - \text{spec}_k))$ is denoted by sp_k (or $\text{spec}_k = (sp_k + 1)/2$), this expression (6) equates to $N_N sp_k$.

Mathemetical formulations for measuring the performance of classifiers: Now we shall explore the concise representation of the discussion related to (3) and (5). This concise representation will aid us in solving the estimate $\tilde{\boldsymbol{P}}_e$. To do this concise representation we need the aid of four additional variables. One is the vector of length K containing the sn_k of all the classfiers in the ensemble. This is represented by $\boldsymbol{S}_A$. Second is the vector cotaining the sp_k of all the classifiers in the ensemble. This is represented by $\boldsymbol{S}_N$. The third (fourth) is a vector of dimensions $N_R \times 1$ ($N_{Re} \times 1$) containing ones. This vector is represented by $\boldsymbol{O}$ ($\boldsymbol{O}_e$). The roles that the notations $\boldsymbol{S}_A$, $\boldsymbol{S}_N$, $\boldsymbol{M}$, $\boldsymbol{M}_e$, $\boldsymbol{P}$ and $\tilde{\boldsymbol{P}}_e$ play in the proposal has been illustrated in Fig. 4.

It follows that the kth element in the vector $(1/2)\boldsymbol{M}(\boldsymbol{P} + \boldsymbol{O})$ is equal to $N_A sn_k$ which is shown in (3). It also follows that the kth

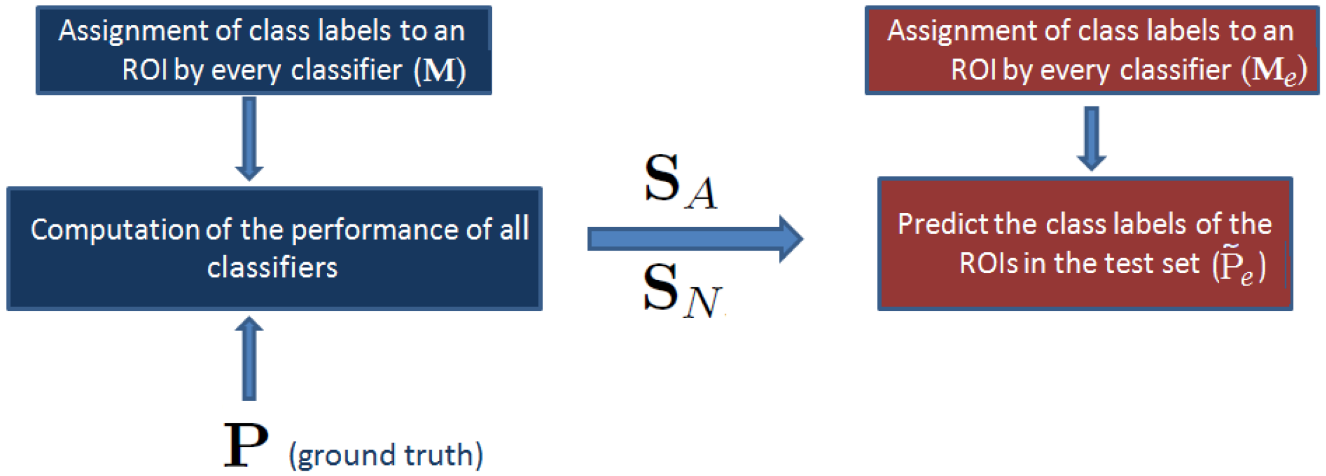

Assignment of class labels to an ROI by every classifier (M)
Computation of the performance of all classifiers
P (ground truth)
S_A
S_N
Assignment of class labels to an ROI by every classifier (M_e)
Predict the class labels of the ROIs in the test set (P̃_e)

Fig. 4 *Explanation of the bottom four blocks of Fig. 2. The symbol* ***P*** *represents the class label of the RoIs of the training set as per ground truth. The symbols* $\boldsymbol{S}_A$ *and* $\boldsymbol{S}_N$ *are vectors whose elements are* sn_k *and* sp_k*, respectively, of all the classifiers in the ensemble. For details refer Section 3.1*

element in the vector $(1/2)\boldsymbol{M}(\boldsymbol{P}-\boldsymbol{O})$ is equal to $N_N sp_k$ which is shown in (5).

The following relations then hold:

$$\frac{1}{2}\boldsymbol{M}(\boldsymbol{P}+\boldsymbol{O}) = N_A\boldsymbol{S}_A, \\ \frac{1}{2}\boldsymbol{M}(\boldsymbol{P}-\boldsymbol{O}) = N_N\boldsymbol{S}_N\,. \tag{7}$$

Analogous to the notations $\boldsymbol{S}_A$ and $\boldsymbol{S}_N$, let $\boldsymbol{S}_{Ae}$ and $\boldsymbol{S}_{Ne}$ denote the vectors containing sn_k and sp_k, respectively, that are evaluated on the test set. Similar to the above equations, the following relations hold for test cases:

$$\frac{1}{2}\boldsymbol{M}_e(\boldsymbol{P}_e+\boldsymbol{O}_e) = N_{Ae}\boldsymbol{S}_{Ae}, \\ \frac{1}{2}\boldsymbol{M}_e(\boldsymbol{P}_e-\boldsymbol{O}_e) = N_{Ne}\boldsymbol{S}_{Ne}\,. \tag{8}$$

How does the proposal solve the estimate of $\tilde{\boldsymbol{P}}_e$: Since $\boldsymbol{S}_{Ae}$, $\boldsymbol{S}_{Ne}$, N_{Ae} and N_{Ne} are unknown beforehand, we may assume that the sensitivities (fraction of the number of AD RoIs classified as AD) and the specificities (fraction of the number of normal RoIs classified as normal) of an approach (corresponding to a function pair of $f1()$ and $f2()$) are constant irrespective of the set of mammograms when the number of mammograms in the set is large. The consequence of these assumptions are $\boldsymbol{S}_{Ae} = \boldsymbol{S}_A$ and $\boldsymbol{S}_{Ne} = \boldsymbol{S}_N$ because $(sn_k+1)/2$ represents sensitivity and $(sp_k+1)/2$ represents specificity. The terms N_{Ae} and N_{Ne} can be equated to the number of class labels equal to $1(-1)$ in $\tilde{\boldsymbol{P}}_e$ (an estimate of $\boldsymbol{P}_e$). The equations in (8) are rewritten below:

$$\frac{1}{2}\boldsymbol{M}_e(\tilde{\boldsymbol{P}}_e+\boldsymbol{O}_e) = N_{Ae}\boldsymbol{S}_A, \\ \frac{1}{2}\boldsymbol{M}_e(\tilde{\boldsymbol{P}}_e-\boldsymbol{O}_e) = N_{Ne}\boldsymbol{S}_N\,. \tag{9}$$

The estimate $\tilde{\boldsymbol{P}}_e$ would be ideal if the equalities of (9) are satisfied. In the non-ideal case there would be a deviation of LHS from the RHS of (9). It is assumed that lesser the deviation of LHS from the RHS of (9), the better is the estimate. In the proposal, a greedy approach is used to minimise the sum of the LHS–RHS of the above equations (refer (9)) when the goal is to label not all RoIs in the test set but a prespecified fixed number of RoIs as AD sites. The sum of difference between the LHS and the RHS of (9) is represented in the following expression that is referred by *Loss*:

$$\text{Loss} = \frac{1}{N_{Ae}+1}\left\|\frac{1}{2}\boldsymbol{M}_e(\tilde{\boldsymbol{P}}_e+\boldsymbol{O}_e) - N_{Ae}\boldsymbol{S}_A\right\| \\ +\frac{1}{N_{Ne}+1}\left\|\frac{1}{2}\boldsymbol{M}_e(\tilde{\boldsymbol{P}}_e-\boldsymbol{O}_e) - N_{Ne}\boldsymbol{S}_N\right\|\,. \tag{10}$$

In the proposal, $\tilde{\boldsymbol{P}}_e$ is initialised to $-\boldsymbol{O}_e$. By exhaustive search of the N_{Re} positions (where by constraint $N_{Re} = N_{Ae} + N_{Ne}$), the position in $\tilde{\boldsymbol{P}}_e$ where if a 1 is inserted *Loss* will be the minimum is found. Let us call this optimum position $o1$. After finding $o1$, $\tilde{\boldsymbol{P}}_e(o1)$ is set to 1 (the AD class label). Let the now updated $\tilde{\boldsymbol{P}}_e$ be referred by $\tilde{\boldsymbol{P}}_{e(1)}$. The next position in $\tilde{\boldsymbol{P}}_{e(1)}$ is then established where if a 1 is inserted *Loss* is further minimised. Let us call this optimum position $o2$. Next $\tilde{\boldsymbol{P}}_{e(1)}(o2)$ is set to 1. Let the updated estimate $\tilde{\boldsymbol{P}}_{e(1)}$ be referred by $\tilde{\boldsymbol{P}}_{e(2)}$. This is continued until N_{Ae} is equal to some prespecified fraction of the total number of RoIs in the test set *viz.* N_{Re}. In this manuscript, this prespecified fraction is represented by θ.

At iteration a, let the updated estimate of $\boldsymbol{P}_e$ be referred by $\tilde{\boldsymbol{P}}_{e(a)}$. Let the value of the position where if a 1 is inserted in $\tilde{\boldsymbol{P}}_{e(a)}$, Loss is minimised be referred by oa. It is apparent that the value of oa is modulated by the vector $\tilde{\boldsymbol{P}}_{e(a)}$ which contains the class label 1 at positions $o1, o2, \ldots, o(a-1)$ and -1 at the remaining positions. Note that the preceding statement shows that the optimum position oa is found by taking all the class labels of the test RoIs into consideration. This is unlike Adaboost and Random Forest where the class label assigned to a test RoI is dependent on the RoI's feature vector and a classifier model that is built during training.

As θ increases we will necessarily obtain some AD RoIs which are AD as per radiologist's assessment. The remaining AD RoIs are labelled as normal RoIs as per ground truth. The fraction of number of AD RoIs labelled as AD sites by the proposal to the total number of AD RoIs as per ground truth represents the sensitivity (sen) of our proposal (note that sen_k represents the sensitivity of the kth classifier while sen represents the sensitivity of the entire proposal). Similarly, we will necessarily obtain some normal (as per ground truth) RoIs that are labelled as AD RoIs by the proposal. This number of normal RoIs per mammogram represents the number of false positives per mammogram referred by FP. The pair of $\{\text{sen}, FP\}$ will constitute one point in the free receiver operating characteristic (FROC) curve of our proposal. The proposal is validated with this FROC curve in Section 4.

It is apparent that the FROC is obtained on mammograms whose ground truth is available. The multiple co-ordinates of the FROC are obtained on choosing multiple values of θ; one co-ordinate for one value of θ. The question arises that which value of θ should the user use if the proposal needs to be deployed in real practice. In this regard, refer Fig. 5. Once the FROC is obtained, the user would choose an operating point on this FROC that best

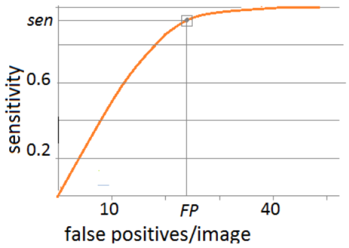


Fig. 5 *Illustration of choosing the operating point on the FROC for deployment of the proposal in the medical practice*

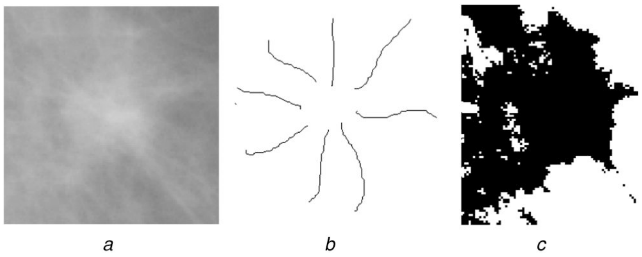


Fig. 6 *Illustration of the adoption of features from an RoI*
(a) Spiculated AD RoI as per ground truth (mdb158.pgm, miniMIAS [2]), ***(b)*** Synthetic radiating pattern of ridges that resembles an AD site, ***(c)*** Segmented image: Pixels in (a) having grey scale value greater than 165 lie within the black region. The choice of threshold 165 is explained in Section 3.2. Note the roughness of the boundary of the black region. The eccentricity of the boundary (contour) of this black region is used as one of the features in the proposal

reflects the optimum performance of the proposal. This operating point would refer to a particular value of sen and *FP*. For example, choosing an operating point with a higher value of sen would yield a greater value of *FP*. Thus, the user has to make a compromise in choosing the operating point. The operating point so chosen would have been obtained for a specific value of θ. The operator would use this value of θ to implement the proposal in medical practice.

In the next subsection, the implementation of the function pairs $f1()$ and $f2()$ is discussed.

3.2 Implementation of $f1()$ and $f2()$

AD is related to the presence of bright lines in the mammogram (refer Fig. 6).These lines are referred by spicules in the medical literature. Technically a bright line is referred by a ridge in the image processing literature. Since AD is defined as a radiating pattern of spicules, in the proposal the concentration index [23] has been adopted as the function $f1()$. The concentration index is aimed at evaluating the pattern of lines in an RoI; lower values indicate a radiating pattern, higher values indicate a parallel set of lines. Referring [23], let the magnitude and the orientation of the ridge at pixel i in RoI R be m_i and α_i, respectively. These magnitudes and the orientations at every pixel in R constitute the feature vector of R. Let the co-ordinates of i be referred by x_i and y_i. The concentration index [23] is defined as follows:

$$\mathscr{C}\mathscr{I}_R = \frac{\sum_i m_i (x_i \sin\alpha_i - y_i \cos\alpha_i)^2}{\sum m_i}. \qquad (11)$$

The term $\mathscr{C}\mathscr{I}_R$ is used to refer to A_1. It is observed that normal tissues tend to have smooth circular shaped contours in contrast to the corrugated contours in an AD site (refer Fig. 6*c*). It would then make sense that the contours in a normal region would be more circular in shape compared to the AD region. This motivates the author to measure the eccentricity of the contours of an RoI and use these eccentricity measures as inputs to $f2()$.

For simplicity, the eccentricity of one contour has been measured. It makes sense that there are two extreme cases that might arise in choosing the preceding contour. One when all the pixels within an RoI are included in the contour. The other when almost no pixels are included within the contour. In either case both an AD site and a normal site would return the same eccentricity values; thus differentiating between a normal site and an AD site would be difficult.

This motivates the author to consider that contour which encloses half of the number of pixels within an RoI. It is imperative that there might be cases where such a realisation of a contour might not be possible. For example, consider the case when all the pixel values within an RoI are identical. Either all the pixels must be considered or no pixels should be considered. Such cases are however rare. It is possible to realise the contour that encloses approximately half of the number of pixels within an RoI as follows.

Let us sort the pixel values of an RoI in a descending order. Let the number of pixels in an RoI be denoted as n. The pixel value that is located at $n/2$ position in the aforementioned order is used as the threshold to binarise the RoI. It is apparent that the number of white pixels in this binary image will be approximately $n/2$. Since there may be many connected components, the connected component with the largest area is chosen. The eccentricity of the largest ellipse that fits within the preceding mentioned connected component is measured. This eccentricity is denoted as e_R. This is how the threshold 165 has been chosen in Fig. 6*c*.

Until now the measures $\mathscr{C}\mathscr{I}_R$ and e_R are characteristics of the pixel values in R. The function $f1()$ was made to assume the value of $\mathscr{C}\mathscr{I}_R$. If $f2()$ assumes the value of e_R alone, we have created only one classifier till now.

The question arises that how do we design different classifiers with $\mathscr{C}\mathscr{I}_R$ and e_R. For simplicity $f2()$ is chosen to implement a linear transformation on e_R. Different classifiers can be designed to implement different linear transformations in $f2()$ keeping the function $f1()$ fixed. A linear transformation is characterised by a slope and a y-intercept. The slope is referred by w_1 and the y-intercept is referred by w_2. In the next subsection, the choice of the parameters w_1 and w_2 that characterise the function $f2()$ of a classifier in the proposal is discussed.

3.2.1 Choice of w_1 and w_2: Ideally, we would like to construct a classifier that assigns class labels to RoIs of the training set in perfect accordance with the ground truth. Had we chosen e_R of an RoI R as A_2 we would have been able to construct only one classifier, say $\mathscr{C}$; that computes τ_X based on A_1 and A_2 of all RoIs and then assigns the class label to an RoI based on the comparison of the RoI's A_1 and A_2 with τ_X. This computation has been illustrated in Fig. 3. There is no guarantee that the class labels so assigned by the preceding classifier matches with the ground truth. We would try to create a classifier whose assignment of class labels to RoIs of the training set is uncorrelated with the class labels assigned by $\mathscr{C}$.

This is because if the classifiers assign class labels to RoIs in a correlated way, these class labels would create similar rows in $\boldsymbol{M}$. This would generate the same sn_k and sp_k (when the index k refers to the similar classifiers) in $\boldsymbol{S}_A$ and $\boldsymbol{S}_N$, respectively.

Let $(1/2N_{Ae})\boldsymbol{M}_e(\tilde{\boldsymbol{P}}_e + \boldsymbol{O}_e)$ be referred by $\tilde{\boldsymbol{S}}_{Ae}$. Similarly, let $(1/2N_{Ae})\boldsymbol{M}_e(\tilde{\boldsymbol{P}}_e - \boldsymbol{O}_e)$ be referred by $\tilde{\boldsymbol{S}}_{Ne}$. Similar classifiers would produce similar rows in $\boldsymbol{M}_e$. Therefore, $\tilde{\boldsymbol{S}}_{Ae}$ and $\tilde{\boldsymbol{S}}_{Ne}$ would contain similar values at the indices corresponding to the similar classifiers. The term *Loss* in (10) can also be rewritten as $(1/N_{Ae} + 1)\,\|\,\tilde{\boldsymbol{S}}_{Ae} - \boldsymbol{S}_A\,\| \;+ (1/(N_{Ne} + 1))\,\|\,\tilde{\boldsymbol{S}}_{Ne} - \boldsymbol{S}_N\,\|$. Keeping similar information in most indices in $\tilde{\boldsymbol{S}}_{Ae}$ and $\tilde{\boldsymbol{S}}_{Ne}$ would imply recomputing similar information for the same indices within the norms in *Loss*. This would result in increasing space and computational complexity without gaining in performance.

The purpose of using w_1 and w_2 is to construct classifiers such that the classifiers assign class labels to RoIs of the training set in an uncorrelated way amongst one another. The question arises that how do we do it?

Choice of $\{w_1, w_2\}$ *if there are two RoIs in the training set:* It is reiterated here that the function $f1()$ that is applied on the pixels values in an RoI is identical to the concentration index of the RoI. The concentration index is a scalar. The output of $f1()$ is also referred by A_1. The function $f2()$ is a linear transformation of another feature of an RoI, say R referred by e_R. The value of this linear transformation is a scalar. This scalar is equal to $w_1 e_R + w_2$.

Now let there be two RoIs in the training set; $R1$ and $R2$. In this context let e_R, A_1 and A_2 of $R1$ be denoted by e_{R1}, A_{1R1} and A_{2R1} respectively. Let the parameters w_1 and w_2 of the kth classifier be denoted by w_{1k} and w_{2k}, respectively. Let the difference between A_{1R1} and A_{2R1} be denoted by λ_{kR1}. Therefore

$$\begin{aligned} A_{1R1} &= A_{2R1} + \lambda_{kR1}, \\ A_{1R1} &= w_{1k} e_{R1} + w_{2k} + \lambda_{kR1} \, . \end{aligned} \tag{12}$$

If $A_{1R1} > A_{2R1}$, three cases might arise. First $A_{1R1} > \tau_X > A_{2R1}$. Second $A_{1R1} > A_{2R1} > \tau_X$. Third $\tau_X > A_{1R1} > A_{2R1}$. This means $R1$ can be assigned either the normal class label (-1) or the ambiguous class label (0). Similarly if $A_{1R1} < A_{2R1}$, $R1$ can be assigned either the AD class label (1) or the ambiguous class label (0). If an ensemble contains classifiers that assign class labels to an RoI in the training set in an uncorrelated way, the AD class label and the normal class label would have a 50–50 chance in getting assigned to a particular RoI in the training set. We infer that for the classifiers to be uncorrelated amongst one another we need to adjust w_1 and w_2 such that the two conditions $A_{1R1} > A_{2R1}$ and $A_{1R1} < A_{2R1}$ have a 50–50 chance.

Referring (12) let the expression $A_{1R1} - w_{1k} e_{R1} - w_{2k}$ be denoted by L_{R1}. Let the region in the w_{1k} and w_{2k} space where $L_{R1} > 0$ be denoted by Q^1_{R1}. Similarly let the region in the w_{1k} and w_{2k} space where $L_{R1} < 0$ be denoted by Q^0_{R1}. The pair $\{w_{1k}, w_{2k}\}$ should be chosen with equal probability from Q^1_{R1} and Q^0_{R1}. The preceding analysis should be true for $R2$. There are four regions that arise in the context of bringing the new RoI ($R2$). The first region is $Q^1_{R1} \cap Q^1_{R2}$. The second region is $Q^1_{R1} \cap Q^0_{R2}$. The third region is $Q^0_{R1} \cap Q^0_{R2}$. The fourth region is $Q^0_{R1} \cap Q^1_{R2}$. These regions are illustrated in Fig. 7*a*.

Choice of $\{w_1, w_2\}$ *if there are more than two RoIs in the training set:* Analogous to the four regions for two RoIs, for N number of RoIs there would arise 2^N regions. The pair $\{w_{1k}, w_{2k}\}$ should be chosen with equal probability from each of the 2^N regions. Let these regions be referred by $Q^1, Q^2, \ldots, Q^{2^N}$. The method of choosing $\{w_{1k}, w_{2k}\}$ based on the preceding analysis is now discussed.

Let the intersection point of $L_{R1} = 0$ and $L_{R2} = 0$ be denoted by $W_{R1,R2}$. The symbol $W_{R1,R2}$ has two components. One component refers to the value along the w_{1k} axis. The other component refers to the value along the w_{2k} axis. These two components are, respectively

$$\left\{ \frac{A_{1R1} - A_{1R2}}{e_{R1} - e_{R2}}, \frac{A_{1R1} e_{R2} - A_{1R2} e_{R1}}{e_{R2} - e_{R1}} \right\}. \tag{13}$$

For all pairs of RoIs we would find points analogous to $W_{R1,R2}$. These clusters of points due to $W_{R1,R2}$ may be represented by one cluster. Let the convex hull of this cluster be denoted by $\mathscr{V}(W)$.

All the lines L_{R1} corresponding to the RoIs in the training set will create multiple triangles (refer Fig. 7*b*) in the space of $\{w_{1k}, w_{2k}\}$. The vertices of these triangles will be one of $W_{R1,R2}$. We can assume that $\mathscr{V}(W)$ will include some or all of $Q^1, Q^2, \ldots, Q^{2^N}$.

The process of choosing the pairs $\{w_{1k}, w_{2k}\}$ should begin by randomly generating pairs within $\mathscr{V}(W)$. The number of pairs (so generated) within each of $Q^1, \ldots, Q^{2^N}$ should be proportional to the area of the respective region within $\mathscr{V}(W)$. Let the number of pairs randomly generated in the vth region (Q^v) be denoted by n_v. We can choose ν number of pairs from n_v pairs in Q^v; thus representing every region Q^v with equal probability and making the classifiers uncorrelated amongst one another.

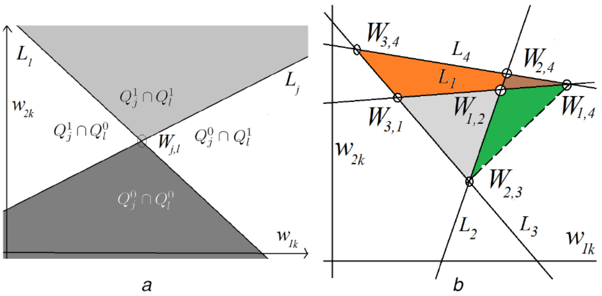


Fig. 7 *Illustration of decomposition of the* $\{w_{1k}, w_{2k}\}$ *space when there are*
(a) Two RoIs (for simplicity of illustration $R1$ and $R2$ has been replaced by j and l, respectively), ***(b)*** Four RoIs in the training set (for simplicity of illustration the RoIs are referred by 1, 2, 3 and 4)

```
1. Input: $\mathcal{R}1$= The set of RoIs. $f1()$ and the scalars $w_1$ and $w_2$ that characterizes $f2()$.
2. Procedure: $L_\mathcal{T} = L_l = \phi$.
  (a) For every $R$ in $\mathcal{R}1$,
    (i) Compute $\mathcal{CI}_R$ from (11). $A_1 = \mathcal{CI}_R$.
    (ii) Compute $e_R$, the eccentricity of the contour that encloses half of the number of pixels of $R$. $A_2 = w_1 e_R + w_2$.
    (iii) $L_\mathcal{T} = [L_\mathcal{T}\ A_1\ A_2]$.
    (iv) $\tau_R = \frac{A_1 + A_2}{2}$. $L_l = [L_l\ \tau_R]$.
  (b) $\tau_X = mode(\tau_R), \tau_R \in L_l$.
3. Output: $\tau_X$ and $L_\mathcal{T}$.
```

Fig. 8 *Algorithm 1: Establishment of* τ_X

```
1. Input: $\mathcal{N}$= The set of function pairs. $\mathcal{R}1$= The set of RoIs.
2. Procedure: For every function pair in $\mathcal{N}$. (Let the index of this function pair in $\mathcal{N}$ be $k$.)
  (a) $[\tau_X, L_\mathcal{T}] = Algorithm1(\mathcal{R}1, \mathcal{N}(k))$.
  (b) For the $j$th RoI in $\mathcal{R}1$,
    (i) $A_1 = L_\mathcal{T}(2j - 1)$.
    (ii) $A_2 = L_\mathcal{T}(2j)$.
    (iii) if $A_1 < \tau_X$ and $A_2 > \tau_X$, $label = 1$ else if $A_1 > \tau_X$ and $A_2 < \tau_X$, $label = -1$ else $label = 0$.
    (iv) $\mathbf{M}(k, j) = label$.
3. Output: $\mathbf{M}$.
```

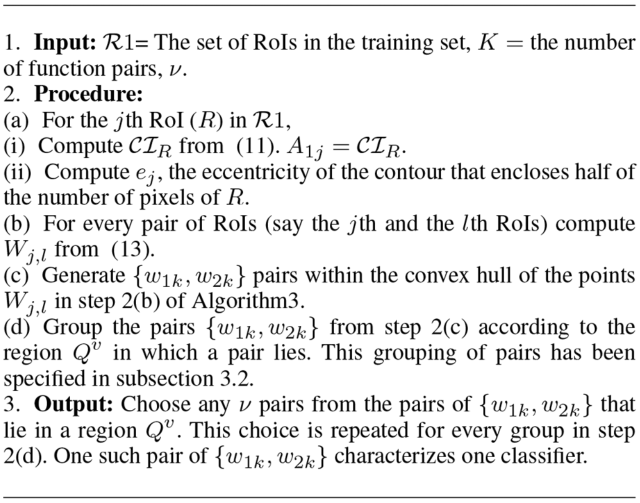
Fig. 9 *Algorithm 2: Establishment of* $\boldsymbol{M}$

The preceding discussions of our proposal in the form of algorithms are now outlined in the next subsection.

3.3 Training and testing

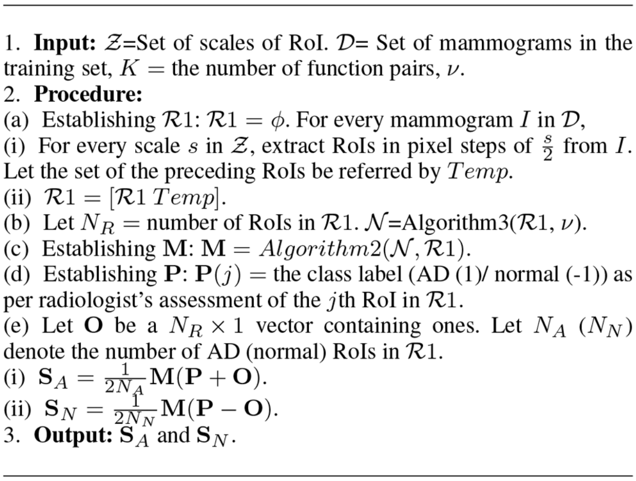
Establishment of τ_X: First an outline of the establishment of τ_X given a set, say $\mathscr{R}1$ of RoIs and a function pair is presented. In this algorithm, two storage variables L_l and L_τ are used. The pairs A_1 and A_2 for an RoI R are inserted in L_l. The average (τ_R) of A_1 and A_2 is inserted in L_τ. Let R denote an RoI in the set $\mathscr{R}1$. In step 2(a)(i) and step 2(a)(ii) of Algorithm 1 (see Fig. 8), A_1 and A_2 are computed. In step 2(a)(iv) the mean of A_1 and A_2 is computed. In step 2(b) the mode of τ_R is assigned to τ_X. The output of Algorithm 1 (Fig. 8) is the value of τ_X and the set L_τ.

Establishment of $\boldsymbol{M}$: It is reiterated here that $\boldsymbol{M}$ contains the class labels assigned to all RoIs of the test set by the classifiers in an ensemble. In Algorithm 2 (see Fig. 9), the establishment of $\boldsymbol{M}$ has been outlined. The input to this algorithm is the set $\mathscr{R}1$ and the set of K number of function pairs. Note that the function pairs differ from one another in the scalars w_1 and w_2. The function pairs are similar to one another since by design all of the pairs use $\mathscr{CI}$ and e_R to compute A_1 and A_2. This set of function pairs is referred by $\mathscr{N}$.



```
1. Input: $\mathcal{R}1$= The set of RoIs in the training set, $K$ = the number of function pairs, $\nu$.
2. Procedure:
  (a) For the $j$th RoI ($R$) in $\mathcal{R}1$,
    (i) Compute $\mathcal{CI}_R$ from (11). $A_{1j} = \mathcal{CI}_R$.
    (ii) Compute $e_j$, the eccentricity of the contour that encloses half of the number of pixels of $R$.
  (b) For every pair of RoIs (say the $j$th and the $l$th RoIs) compute $W_{j,l}$ from (13).
  (c) Generate $\{w_{1k}, w_{2k}\}$ pairs within the convex hull of the points $W_{j,l}$ in step 2(b) of Algorithm3.
  (d) Group the pairs $\{w_{1k}, w_{2k}\}$ from step 2(c) according to the region $Q^v$ in which a pair lies. This grouping of pairs has been specified in subsection 3.2.
3. Output: Choose any $\nu$ pairs from the pairs of $\{w_{1k}, w_{2k}\}$ that lie in a region $Q^v$. This choice is repeated for every group in step 2(d). One such pair of $\{w_{1k}, w_{2k}\}$ characterizes one classifier.
```

Fig. 10 *Algorithm 3: Establishment of* w_1 *and* w_2 *of all classifiers*



```
1. Input: $\mathcal{Z}$=Set of scales of RoI. $\mathcal{D}$= Set of mammograms in the training set, $K$ = the number of function pairs, $\nu$.
2. Procedure:
  (a) Establishing $\mathcal{R}1$: $\mathcal{R}1 = \phi$. For every mammogram $I$ in $\mathcal{D}$,
    (i) For every scale $s$ in $\mathcal{Z}$, extract RoIs in pixel steps of $\frac{s}{2}$ from $I$. Let the set of the preceding RoIs be referred by $Temp$.
    (ii) $\mathcal{R}1 = [\mathcal{R}1\ Temp]$.
  (b) Let $N_R$ = number of RoIs in $\mathcal{R}1$. $\mathcal{N}$=Algorithm3($\mathcal{R}1$, $\nu$).
  (c) Establishing $\mathbf{M}$: $\mathbf{M} = Algorithm2(\mathcal{N}, \mathcal{R}1)$.
  (d) Establishing $\mathbf{P}$: $\mathbf{P}(j)$ = the class label (AD (1)/ normal (-1)) as per radiologist's assessment of the $j$th RoI in $\mathcal{R}1$.
  (e) Let $\mathbf{O}$ be a $N_R \times 1$ vector containing ones. Let $N_A$ ($N_N$) denote the number of AD (normal) RoIs in $\mathcal{R}1$.
    (i) $\mathbf{S}_A = \frac{1}{2N_A}\mathbf{M}(\mathbf{P} + \mathbf{O})$.
    (ii) $\mathbf{S}_N = \frac{1}{2N_N}\mathbf{M}(\mathbf{P} - \mathbf{O})$.
3. Output: $\mathbf{S}_A$ and $\mathbf{S}_N$.
```

Fig. 11 *Algorithm 4: Establishment of* $\boldsymbol{S}_A$ *and* $\boldsymbol{S}_N$

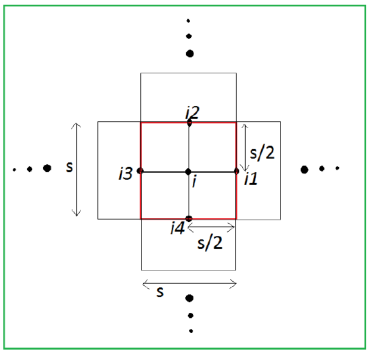


Fig. 12 *Extraction of RoIs of size $s \times s$ pixels from a mammogram. The outline of the mammmogram is indicated by a green rectangle. The notations i, i1, i2, i3 and i4 denote pixel positions*

Based on the value of τ_X, the values of A_1 and A_2 of the RoI R, we can compute the class label $(0/1/-1)$ of R as per the conditions in step 2(b)(iii) of Algorithm 2 (Fig. 9). This label of the jth RoI in $\mathscr{R}1$ by the kth function pair is assigned to the element at the kth row and jth column in $\boldsymbol{M}$ *viz.* $\boldsymbol{M}(k, j)$. This assignment is implemented in step 2(b)(iv) of Algorithm 2 (Fig. 9). The output of this Algorithm is $\boldsymbol{M}$.

Establishment of $\{w_1, w_2\}$*:* In Algorithm 3 (see Fig. 10), the scalars w_1 and w_2 that characterise a classifier following our discussion in Section 3.2.1 is established. In step 2(a) A_{1j} and e_j of the jth RoI are computed. In step 2(b) $W_{j,l}$ is computed from (13) for all possible pairs of RoIs. In step 2(c) pairs of $\{w_{1k}, w_{2k}\}$ are generated at random within the convex hull $\mathscr{V}(W)$. In step 2(d) the pairs of $\{w_{1k}, w_{2k}\}$ in step 2(c) are grouped according to the region Q^v in which a pair lies. The output of this algorithm contains ν pairs of $\{w_{1k}, w_{2k}\}$ from each group of pairs in step 2(d).

Establishment of $\boldsymbol{S}_A$ *and* $\boldsymbol{S}_N$*:* It is reiterated here that the performances of all the classifiers in an ensemble are encapsulated in $\boldsymbol{S}_A$ and $\boldsymbol{S}_N$. In Algorithm 4 (see Fig. 11), the vectors $\boldsymbol{S}_A$ and $\boldsymbol{S}_N$ of the aforesaid classifiers that are generated from Algorithm 3 (Fig. 10) are established. The input to this algorithm is the set of mammograms in the training set. The radiologist's assessment is available for the mammograms in this training set, say D.

From the mammograms in D, RoIs at varying scales (values of the height and width of an RoI) are extracted (refer Fig. 12). These scales are stored in $\mathscr{Z}$. Let the value of a scale be referred by s. The method of extracting RoIs (from a mammogram) is identical to a sliding window approach that slides a window over the mammogram in pixel steps of $s/2$ both horizontally and vertically. For example, if the value of a scale refers to s, and an RoI of size $s \times s$ centred at a pixel, say i in Fig. 12 is extracted then RoIs of size $s \times s$ are also extracted at pixels $i1$, $i2$, $i3$ and $i4$ that are situated at a distance of $s/2$ from the pixel i.

The set of the preceding RoIs is referred by $\mathscr{R}1$. The construction of $\mathscr{R}1$ is outlined in step 2(a) of Algorithm 4 (Fig. 11). In step 2(b) different classifiers are generated; assigning these classifiers to $\mathscr{N}$. In step 2(c) $\boldsymbol{M}$ for the preceding set of RoIs *viz.* $\mathscr{R}1$ and the set of function pairs $\mathscr{N}$ are computed. In step 2(d) the vector $\boldsymbol{P}$ which contains the class labels of the RoIs in $\mathscr{R}1$ as per radiologist's assessment is constructed. In step 2(e) $\boldsymbol{S}_A$ and $\boldsymbol{S}_N$ are computed. The output of this algorithm is $\boldsymbol{S}_A$ and $\boldsymbol{S}_N$.

Establishment of $\tilde{\boldsymbol{P}}_e$*:* In Algorithm 5 (see Fig. 13) $\tilde{\boldsymbol{P}}_e$ is established. The input to this algorithm is the set of mammograms in the test set D_e. The radiologist's assessment is unavailable for the mammograms in D_e. In addition to D_e we also take $\mathscr{N}$ and a threshold θ as inputs to this algorithm. The threshold θ sets the number of RoIs to be labelled as an AD site by the proposal.

Similar to the preceding discussion referring to Fig. 12, from the mammograms in D_e RoIs at varying scales are extracted. These scales are stored in $\mathscr{Z}_e$. The set of the preceding RoIs is referred by $\mathscr{R}1_e$. The construction of $\mathscr{R}1_e$ is outlined in step 2(a) of Algorithm 5 (Fig. 13). Let N_{Re} represent the number of RoIs in $\mathscr{R}1_e$. In step 2(c) $\boldsymbol{M}_e$ for the preceding set of RoIs *viz.* $\mathscr{R}1_e$ and the set of function pairs, $\mathscr{N}$ is computed. In step 2(d) the vector $\tilde{\boldsymbol{P}}_e$ is established. To establish, the following approach has been taken in the proposal.

The vector $\tilde{\boldsymbol{P}}_e$ has been initialised to $-\boldsymbol{O}_e$, a vector of the same dimension as $\tilde{\boldsymbol{P}}_e$ and containing all ones. Procedure 1 (see Fig. 14) has been used to find the optimum position in $\tilde{\boldsymbol{P}}_e$ by exhaustive search such that if a 1 is inserted at this position *Loss* is minimum. This is repeated until the number of RoIs that are labelled as 1 equals $[\theta N_{Re}]$. Let a denote an integer. The value of $\tilde{\boldsymbol{P}}_e$ at iteration a is denoted by $\tilde{\boldsymbol{P}}_{e(a)}$.

The output of this algorithm is the assertion that the RoIs with label 1 in $\tilde{\boldsymbol{P}}_e$ are the AD sites.

Notice that in step 2(b) of Algorithm 1 (Fig. 8), the computation of τ_X is dependent on the feature vectors $\boldsymbol{F}_R = [\mathscr{C}\mathscr{I}_R e_R]$ of each RoI in the training or test set. This justifies why the proposal incorporates $\boldsymbol{F}_j, j \neq i$ in (2). In step 2(d)(ii) of Algorithm 5 (Fig. 13) $\tilde{\boldsymbol{P}}_{e(a+1)}$ is computed based on the value of $\tilde{\boldsymbol{P}}_{e(a)}$. This implies that the index j in step 2(d)(i) of Algorithm 5 (Fig. 13) is dependent on the class labels of all the RoIs in the test set. This justifies why the proposal incorporates $c_j, j \neq i$ in (2).

The experimental results are now discussed in the next section.

4 Experimental results

Necessary details of datasets and assumptions in validating the proposal are provided in the next subsection. The experiments are presented in Section 4.2. Discussions on the results (time complexity) are provided in Section 4.3 (4.4).

```
1. Input: $\mathcal{Z}_e$=Set of scales of RoI. $\mathcal{D}_e$= Set of mammograms in the test set. $\mathcal{N}$= Set of function pairs from step 2(b) of Algorithm4. $\mathbf{S}_A$ and $\mathbf{S}_N$ from Algorithm4. A threshold $\theta \in [0, 1]$.
2. Procedure:
(a) Establishing $\mathcal{R}1_e$: $\mathcal{R}1_e = \phi$. For every mammogram $I$ in $\mathcal{D}_e$,
    (i) For every scale $s$ in $\mathcal{Z}_e$, extract RoIs in pixel steps of $\frac{s}{2}$ from $I$. Let the set of the preceding RoIs be referred by $Temp$.
    (ii) $\mathcal{R}1_e = [\mathcal{R}1_e \, Temp]$.
(b) Let $N_{Re}$ = number of RoIs in $\mathcal{R}1_e$.
(c) Establishing $\mathbf{M}_e$: $\mathbf{M}_e = Algorithm2(\mathcal{N}, \mathcal{R}1_e)$.
(d) Establishing $\tilde{\mathbf{P}}_e$: Let $\mathbf{O}_e$ be a $N_{Re} \times 1$ vector containing ones. Set $\tilde{\mathbf{P}}_{e(0)} = -\mathbf{O}_e$. Set $a = 0$ and $N_{Ae} = 0$. For iteration $a$,
    (i) $j$ =Procedure1$(\tilde{\mathbf{P}}_{e(a)}, \mathbf{M}_e)$.
    (ii) $\tilde{\mathbf{P}}_{e(a+1)} = \tilde{\mathbf{P}}_{e(a)}. \tilde{\mathbf{P}}_{e(a+1)}(j) = 1$. Increment $N_{Ae}$ by 1.
    (iii) $a = a + 1$. Goto step 2(d)(i) if $N_{Ae} < [\theta N_{Re}]$ else goto step 2(d)(iv).
    (iv) $\tilde{\mathbf{P}}_e = \tilde{\mathbf{P}}_{e(a)}$.
3. Output: The $n$th RoI with $\tilde{\mathbf{P}}_e(n) = 1$ is an AD site.
```

Fig. 13 *Algorithm 5: Establishment of* $\tilde{\boldsymbol{P}}_e$

4.1 Databases

There are two popular mammography databases in the medical literature. These two databases are the Mammographic Image Analysis Society Database (MIAS) [2] and the Digital Database for Screening Mammography (DDSM) [22, 24]. Both these databases are publicly available. The proposal has been evaluated on MIAS and DDSM.

The MIAS database comprises of 322 mammograms having the grey-level resolution of 8 bits, image size of 1024 × 1024 pixels and the portable grey map file format of the images. Nineteen mammograms out of 322 mammograms contain 19 AD RoIs, each having one AD RoI. Out of these 19 images, 10 AD RoIs are malignant and the remaining are benign. The MIAS database describes the outline of an AD site by a circle. In addition to these 19 mammograms containing AD, there are also 19 mammograms containing spiculations, 23 mammograms containing circumscribed masses, 25 mammograms containing calcifications and 15 mammograms containing breast asymmetry. The remaining mammograms are normal. In the MIAS database only the MLO mammographic views of the breasts are present.

The DDSM database comprises 2620 cases having the grey-level resolution of 16 bits and lossless jpeg (LJPEG) file format of the images. Each case contains the two mammographic views (CC and MLO) of each breast of a patient. Fifty cases out of 2620 contain malignant AD. Out of 200 mammograms in the aforesaid 50 cases [note that a case contains four mammograms, two views (CC and MLO) from each breast], 90 mammograms contain one AD RoI each and two mammograms contain two AD RoIs each. The rest of these 200 mammograms are normal. The DDSM database describes the outline of an AD site by a chain code. In addition to the 50 cases there are about 700 normal cases, 970 malignant cases, 830 benign cases and 140 benign cases without any callback (i.e. these patients did not have any follow-up mammograms).

The acquisition of the mammograms in the DDSM is made with either of three scanners. These scanners are HOWTEK, DBA and LUMISYS. The spatial pixel resolutions of these scanners are, respectively, 43.5, 42 and 50 μm. The mammograms in the MIAS database are generated by the Joyce-Lobel (JL) microdensitometer. The scanner JL has a pixel resolution of 50 μm. In addition, certain de-identified patient details such as age and imaging characteristics such as intensity resolution and type of scanner are provided in the DDSM database. In the MIAS database only the scanner with which the mammograms were generated is known. In DDSM the breast tissue density is inferred by the radiologist who assigns a number to the inference. In MIAS, the breast tissue is inferred as either fatty or glandular. In addition to this breast density, the BI-RADS score of the evaluation of a mammogram is given in the DDSM.

```
1. Input: $\tilde{\mathbf{P}}_{e(a)}$ and $\mathbf{M}_e$.
2. Procedure:
(a) $N_{Ae}$ = number of elements with 1 in $\tilde{\mathbf{P}}_{t(a)}$.
(b) $N_{Ne}$ = number of elements with -1 in $\tilde{\mathbf{P}}_{t(a)}$.
(c) $N_{Re} = N_{Ae} + N_{Ne}$.
(d) For all $k \in [1, N_{Re}]$ and $\tilde{\mathbf{P}}_{e(a)}(k) = -1$,
    (i) $\tilde{\mathbf{P}}_{e(a+1)} = \tilde{\mathbf{P}}_{e(a)}. \tilde{\mathbf{P}}_{e(a+1)}(k) = 1$.
    (ii) Compute ((10)) replacing $Loss$ by $Loss(k)$ and $\tilde{\mathbf{P}}_e$ with $\tilde{\mathbf{P}}_{e(a+1)}$.
(e) $j =\overset{k}{argmin} \, Loss(k)$.
3. Output: $j$.
```

Fig. 14 *Procedure 1: Establishing index position j in* $\tilde{\boldsymbol{P}}_{e(a)}$

Table 1 False positives per mammogram at 80% sensitivity of our proposal in the MIAS database

x%	1	10	40	100	600
20	23.4	23.4	23.4	21.4	20.4
50	20.4	20.4	18.4	18	17.7
80	20	16.4	15.7	19.4	18.7

Recall that the sets $\mathscr{Z}$ and $\mathscr{Z}_e$ in Algorithms 4 and 5 (Figs. 11 and 13) contain the dimensions of the RoI extracted from a mammogram. Noting that the average size of an AD RoI is 40 mm in the MIAS and the DDSM database, the spatial size of the RoIs extracted from a mammogram have been allowed to take one of 45 and 57 mm. It is apparent that the aforesaid two spatial values can be converted to their equivalent dimensions in pixels by the relation: dimension in pixels = (spatial dimension)/(pixel resolution). These pixel dimensions have been adopted in the sets $\mathscr{Z}$ and $\mathscr{Z}_e$.

In the proposal it has been assumed that an AD site (as per radiologist's assessment) is detected if there is at least one RoI whose centre pixel is contained within the AD site (as per radiologist's assessment). The RoIs that are labelled as AD sites by our proposal but whose centre pixel does not lie within any AD site as per radiologist's assessment are counted as a false positive. But false positives which share a common area of more than 70% (of the area of the smallest RoI) are counted as one false positive.

Recall that the proposal relied on the assumption that the parameters $\boldsymbol{S}_A$ and $\boldsymbol{S}_N$ are constant for a large set of mammograms. The question arises that how large should this training set be? As a part of the validation process, 20, 50 and 80% of the available mammograms are chosen as the mammograms in the training set. Let this percentage be denoted as *x*. The remaining (100 − *x*)% is chosen for testing. For each *x*, three combinations of training and testing are created. This is done by choosing the first, the middle and the last *x*% of the available mammograms as the mammograms in the training set. The validation process is repeated three times (for the first, the middle and the last *x*% of the available mammograms) and the average result (the false positives per mammogram) of three runs is reported. All images of a patient (two MLO views per patient in MIAS, two MLO and two CC views per patient in DDSM) are placed in any one of training and testing set. The experiments are discussed next.

4.2 Experiment

In this experiment, all the 19 mammograms in the MIAS database that contain AD have been adopted. The notation *K* (number of function pairs) has been assigned one amongst the values 1, 10, 40, 100 or 600. The terms w_1 and w_2 have been generated as per Algorithm 3 (Fig. 10). For each combination of *x* and *K*, the number of false positives per mammogram (*FP*) at 80% sensitivity (the fraction of AD sites detected by the proposal) has been included in Table 1. In a different experiment along with the previous 19 mammograms containing AD, 20 normal mammograms have been included from MIAS. Now the discussion

Table 2 False positives per mammogram at 80% sensitivity of our proposal in the MIAS database

x%	1	10	40	100	600
20	16.6	19	18	18	21.3
50	23	24.3	24	23.6	24.3
80	17.6	17.6	17.6	15.3	20.6

Table 3 False positives per mammogram at 80% sensitivity of our proposal in the MIAS database

x%	1	10	40	100	600
20	17	18.3	18	18.3	19.6
50	24	23.5	24	24.5	24.8
80	18	19	19.6	19	20

Table 4 False positives per mammogram at 85% sensitivity of our proposal in the DDSM database

x%	1	10	40	100	600
20	18	18	17	16	16
50	18	18	18.3	17.3	15.3
80	22	22	21	20	17.7

Table 5 False positives per mammogram at 80% sensitivity of our proposal in the DDSM database

x%	1	10	40	100	600
20	16.6	17.3	18.6	16.6	18.6
50	24.6	24	24	27	26.3
80	31	31.3	32.3	29.6	30

Table 6 False positives per mammogram at 80% sensitivity of our proposal in the DDSM database

x%	1	10	40	100	600
20	17	17.3	17.3	18	18.3
50	24	24.6	25	25	25.3
80	31.6	30.6	33	32.3	27.6

Table 7 False positives per mammogram at 80% sensitivity of our proposal

Train	Test	method	1	10	40	100	600
MIAS	DDSM	mode	13	15	14	17	16
DDSM	MIAS	mode	14	15	14	14	18

Table 8 Comparison of the proposed approach with competing approaches

Method	MIAS		DDSM	
	90%	85%	90%	85%
real AdaBoost	20	19	19.3	18.3
gentle AdaBoost	20	19	19	18
modest AdaBoost	21	20	20	19
parametrised AdaBoost	20	19	19	18
margin-pruning AdaBoost	21	20	21	20
penalised AdaBoost	20	19	19	18
boosted random forest [9]	22	21	21	20
consensus ensemble [26]	20	19	19	18
phase portrait [6]	> 30	> 30	> 30	> 30
proposal	19	18.3	17.3	16.8

in the last paragraph of Section 4.1 has been modified by keeping *x* % of the mammograms containing AD and *x*% of the normal mammograms in the training set. The remaining mammograms containing AD and the normal mammograms are kept in the test set. We have evaluated our proposal with these additional normal mammograms and have shown the results in Table 2. The same using the median in step 2(b) of Algorithm 1 (Fig. 8) has been shown in Table 3.

The process in previous paragraph was repeated on 55 mammograms in the DDSM dataset. All these 55 mammograms contain AD. The results have been shown in Table 4. In a separate experiment along with 51 mammograms of DDSM containing AD, 49 normal mammograms have been included from DDSM. We have evaluated our proposal with these additional normal mammograms and have shown the results in Table 5. The same using the median in step 2(b) of Algorithm 1 (Fig. 8) has been shown in Table 6. We have also trained on the 39 chosen (referring Tables 2 and 3) mammograms of MIAS and tested on the 100 chosen (referring Tables 5 and 6) mammograms of DDSM and vice versa. The results are shown in Table 7. In Tables 2– 7, the chosen mammograms from DDSM for the evaluation of the proposal have been down-sampled to a pixel resolution of 200 μm for a fair comparison with the mammograms in MIAS which also have a pixel resolution of 200 μm.

In Tables 12 and 13 of [25], surveys of different ensemble approaches in breast cancer classification have been provided. In Tables 3–5 of [26], the performance of a newly developed ensemble classifier on breast cancer has also been presented. In Tables 1–5 of [8], the performance of different Adaboost variants on breast cancer has been provided. With the aid of these results the performance of different Adaboost variants in [8] and boosted Random Forest [9] has been estimated for our mammograms from MIAS and DDSM. For example, the area under the receiver operating characteristics curve (ROC) is provided in [25]. The plot of the ROC is not provided in [25]. We can approximate the ROC from the given area under the ROC (AUC) as follows.

We first consider two variables say $\{yv, yf\}$ that denote some pair of sensitivity and specificity in the ROC. We draw a line segment from the origin to $\{yv, yf\}$ and from $\{yv, yf\}$ to $\{1, 1\}$. Let the two line segments so obtained represent the approximated ROC. The area under this approximated ROC is made equal to the AUC given in [9, 26, 25] by adjusting *yv* and *yf*. We solve for *yv* and *yf* to satisfy the condition of the preceding statement. Once the ROC is obtained the FROC is obtained by equating (1-specificity) × total number of RoIs per mammogram to *FP* and using sensitivity as it is.

The estimated results along with the results of the phase portrait approach have been given in Table 8 at $K = 100$ and $x = 50\%$ for the MIAS and the DDSM database. The FROC curves are plotted in Fig. 15*a*.

It is reiterated that false positives which share a common area of more than 70% are counted as one false positive. We have evaluated our proposal on the 55 mammograms of DDSM containing AD for false positives sharing a common area of more than 50 and 25%. The FROC results are shown in Fig. 15*b*. Plots of *Loss* for $x = 50\%$ and $K = 100$ have been shown in Fig. 15*c*.

4.3 Discussions

Tables 1 and 4 show the false positives (*FP*) per mammogram in the MIAS and the DDSM database, respectively, for different values of *K* (varying along a row) and *x* (varying along a column). The *FP* values in Table 4 are lesser than the same in Table 1. There can be two reasons for this result. These are as follows.

The pixel resolution in MIAS images is approximately four to five times the pixel resolution in DDSM images. In addition, the grey-level resolution in DDSM images is higher by eight bits than the same in MIAS. Given this, lines representing spicules and vessels within dense breast tissues are much better captured in DDSM than in MIAS. These strong signature of ridges in DDSM would be reflected in the value of the concentration index in our proposal. The consequence is that the proposal would be able to distinguish a radiating pattern of ridges from other normal patterns thus affecting the label lbn_R (the class label assigned to RoI *R* by a classifier in an ensemble). This may be a reason for the better performance of our proposal in the DDSM database. The aforesaid inability is one limitation and calls for modifying the proposal's

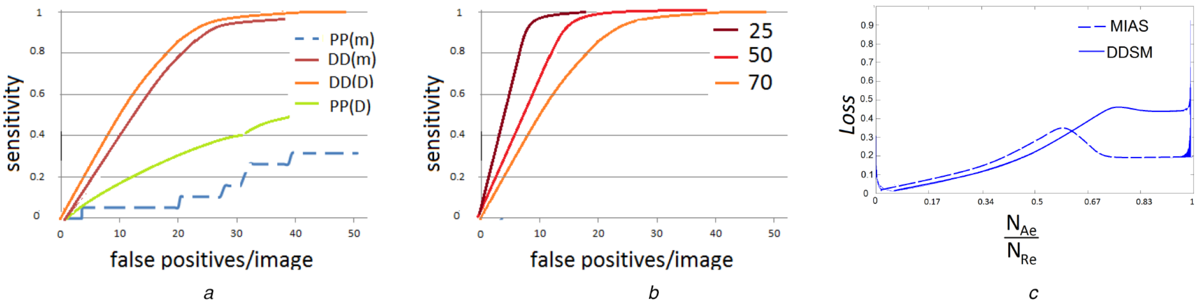


Fig. 15 *Illustration of the different plots of our experiments in Section 4*
(a) FROC curves on the MIAS (m) database and the DDSM (D) database for the proposed method (DD) and the phase portrait approach (PP), ***(b)*** Illustration of the proposal's performance in counting false positives sharing a common area of more than 70% (50%, 25%) of the smallest RoI as one false positive, ***(c)*** Plots of *Loss* in the MIAS and the DDSM for $x = 50\%$ and $K = 100$. Note that because of oscillation of *Loss* towards the end, the lines appear thicker

concentration index which will focus on faint signatures in the breast tissue.

The second reason could be related to the fact that a mammogram is a 2D projection of a 3D object: the breast. It is apparent that a normal site in the breast and an abnormal site containing AD might be located at different regions in the breast. However, the 2D projections of both these regions might overlie on one another in the mammogram. A site (RoI) containing both these projections would be interpreted as an AD site as well as a normal site yielding a label lbn_R of 0. The chances of this happening decrease if two views (projections) are taken for a breast. This is because the chance of both views failing to capture the AD and the normal regions at different sites in the mammogram is lesser than that of a single view capturing the image of the breast. The consequence is that the proposal failing to detect the AD site in both the views is unlikely. In the DDSM we have two views of a breast. This is unlike in MIAS where only the medio-lateral-oblique view of a breast is present. This observation might account for the lesser values of *FP* in Table 4.

In the context of the validation of our proposal, it is apparent that there are three parameters in our proposal. These parameters are K, the number of classifiers in the ensemble, and the number of mammograms in the training and test set as judged by x, the percentage of the total number of mammograms kept in the training set. If the aforesaid parameters are kept constant, the proposal performs better when the mode (of τ_R) is chosen (as τ_X) instead of choosing the median in step 2(b) of Algorithm 1 (Fig. 8). The better performance of the proposal in choosing the mode is evident from the lesser (in majority cases) values (3% on the average) of false positives per mammogram (*FP*) in Tables 2, 5 and 7 compared to the same in Tables 3, 6 and 7. A lesser value of *FP* in choosing the mode in Algorithm 1 (Fig. 8) is expected. This is because as discussed in the beginning of Section 3, the frequently occurring value of τ_R must be chosen as the threshold in Algorithm 1 (Fig. 8) to generate less number of ambiguous labels (0) in matrix $\boldsymbol{M}$ or $\boldsymbol{M}_e$ (a unified representation of the class labels assigned to all ROIs in a training/test set by all the classifiers in an ensemble) of the proposal. The mode by definition represents the frequently occurring value of a variable unlike the median.

It is also observed that the proposal performs better when the number of mammograms in both the training and test set is fairly large. For instance, the *FP* in Table 7 is obtained from 100 mammograms of DDSM and 39 mammograms of MIAS; a total of 139 mammograms. The *FP* in Tables 2–6 is obtained from either 100 mammograms of DDSM or 39 mammograms of MIAS. The values of *FP* in Tables 2–6 are greater by 18% (on the average) than the same in Table 7. It is reiterated that the estimation of the class labels of the RoIs of the test set is made on the assumption that a classifier would perform consistently when the number of mammograms in the training and the test set is fairly large. The argument that our proposal would perform well for a large number of mammograms in the training and test set is promoted by the greater *FP* values in Tables 2–6 compared to the same in Table 7.

In Table 1, we did not incorporate normal mammograms contrary to Tables 2–7. We observe that the incorporation of normal mammograms yields lesser *FP* (by 10% on the average). The reason for this is that the normal mammograms tend to serve better training examples in the form of normal RoIs. These normal RoIs show circularly shaped contours (measured in the form of ellipticity; refer Section 3.2) compared to the contours of the normal RoIs in the mammograms containing AD. The ellipticity of the contours of some normal RoIs in the mammograms containing AD resemble closely the same (in magnitude) of the contours of the AD RoIs. This makes the discrimination of normal RoIs (in the mammograms containing AD) from the AD RoIs difficult (for the experimental setting in Table 1); leading to greater *FP*.

However on the contrary, Table 4 which does not incorporate normal mammograms yields lesser *FP*. The reason could be the features in the mammograms of Table 4 are prominent since the pixel resolution is four times lesser than the mammograms in Tables 5–7. It is reiterated that the mammograms in Tables 5–7 have been down-sampled to $200\,\mu$m for a fair comparison with the mammograms in MIAS which also have a pixel resolution of $200\,\mu$m.

It may be observed that in Tables 1–7 the performance of the proposal is consistent for low values of K such as $K < 100$. A question may arise why keeping one classifier in the ensemble performs as well as keeping multiple classifiers. We argue that in choosing $K = 1$ the class labels assigned to the test RoIs is not the class labels assigned by the single classifier in the ensemble. There is an intermediate process that predicts the class labels of the test RoIs by assuming that the single classifier would perform consistently in both training and test sets. Therefore, the class labels assigned to the test RoIs for $K = 1$ is not the labels in $\boldsymbol{M}_e(1, 1{:}N_{Re})$ but the class labels in $\tilde{\boldsymbol{P}}_e$.

Let us say that $FP < 19.6$ is low, $FP > 19.6$ and $FP <= 25$ is moderate and $FP > 25$ is high. We observe from the tables that $x = 20$ (recall that x is the percentage of total number of mammograms kept in the training set) yields low *FP* values in all the tables except for the moderate *FP* values in Table 1. We notice that $x = 50$ yields moderate *FP* values except for the low *FP* values in Tables 1 and 4. We also observe that $x = 80$ yields low *FP* values except for the moderate *FP* values in Table 4 and high *FP* values in Tables 5 and 6. In the discussion of the preceding paragraphs we have shown that there are at least two factors that have played a role in the values of *FP* in these tables. For example, the spatial pixel resolution of a mammogram and the inclusion of normal mammograms have explained some anomalies as elucidated in the preceding paragraphs of this subsection. Due to the presence of such multiple factors we cannot interrelate the magnitude of *FP* values with a particular value of x. Therefore, there may not be a fixed guideline of how many mammograms are suited in the test (training) set for the proposal. However, given the results in Tables 1– 7 and the imaging quality (a spatial resolution of approximately $50\,\mu$m) of existing scanners we can expect that the proposal fares well for number of mammograms (in the training and test set) > 20.

In Table 8, the proposal extracts the least number of *FP* in the DDSM database whereas in the MIAS database the proposal extracts greater number of *FP*. The reason for the poor performance of the proposal in the MIAS database can be attributed to the fact that the function pair and method of creating the label lbn_R does not differentiate between difficult to classify RoIs and other easier to classify RoIs; which is done in boosting [8]. However, the proposal performs better in case of DDSM because as highlighted earlier the signatures in the breast tissue are prominent in the DDSM; hence all RoIs are of the same difficulty level (if the ridges and the contours are used to describe an RoI) in the context of labelling an RoI as an AD site.

4.4 Time complexity

The time complexity of Algorithm 1 (Fig. 8) (Algorithm 2 (Fig. 9)) is $N_R \log N_R$ ($KN_R \log N_R$). The time complexity of Procedure 1 (Fig. 14) (Algorithm 3 (Fig. 10)) is $N_{Re}K$ ($KN_{Re} \log N_{Re} + N_{Re}K[\theta N_{Re}]$). The execution time of an unoptimised Matlab code of Algorithm 2 (Fig. 9) (Algorithm 3 (Fig. 10)) is 3 (20) s per mammogram with a Xeon(R) CPU, 3 GHz clock frequency, 24GB RAM and Windows 7 operating system. The same for a phase portrait approach is of the order of minutes.

5 Conclusion

We have proposed an ensemble classifier for the detection of AD in a mammogram. Our approach performs better than variants of AdaBoost in one mammographic database. The proposal performs better than Random Forest and phase portrait approach in two mammographic databases. We intend to integrate the proposal with a decision tree based ensemble classifier in order to reinforce the classification result.

6 References

[1] Zonderland, H., Smithuis, R.: Bi Rads for mammography and ultrasound. Availabel at http://radiologyassistant.nl/en/p53b4082c92130/bi-rads-for-mammography-and-ultrasound-2013.html. Accessed as in May 2019

[2] Suckling, J., Parker, J., Dance, D.R., *et al.*: 'The mammographic image analysis society digital mammogram database'. Exerpta Medica, Int. Congress Series, York, UK, 1994

[3] American Cancer Society: Breast cancer early detection and diagnosis. Available at https://www.cancer.org/cancer/breast-cancer/screening-tests-and-early-detection.html. Accessed as in May 2019

[4] Ayres, F.J., Rangayyan, R.M.: 'Characterization of architectural distortion in mammograms', *IEEE Eng. Med. Biol. Mag.*, 2005, **24**, pp. 59–67

[5] Rangayyan, R.M., Banik, S., Desautels, J.E.L.: 'Detection of architectural distortion in prior mammograms using measures of angular dispersion'. IEEE Int. Symp. on Medical Measurements and Applications Proc. (MeMeA), Rome, May 2012, pp. 1–4

[6] Banik, S., Rangayyan, R.M., Desautels, J.E.L.: 'Detection of architectural distortion in prior mammograms', *IEEE Trans. Med. Imaging*, 2011, **30**, (2), pp. 279–294

[7] Shanthi and, S., Muralibhaskaran, V.: 'Automatic detection and classification of microcalcification, mass, architectural distortion and bilateral asymmetry in digital mammogram', *Int. J. Med. Health Biomed. Bioeng. Pharm. Eng.*, 2014, **8**, pp. 818–823

[8] Wu, S., Nagahashi, H.: 'Analysis of generalization ability for different adaboost variants based on classification and regression trees', *J. Electr. Comput. Eng.*, 2015, **2015**, p. 17

[9] Mishina, Y., Tsuchiya, M., Fujiyoshi, H.: 'Boosted random forest'. Int. Conf. on Computer Vision Theory and Applications (VISAPP), Lisbon, Portugal, 2014

[10] Matsubara, T., Ito, A., Tsunomori, A., *et al.*: 'An automated method for detecting architectural distortions on mammograms using direction analysis of linear structures'. , 37th Annual Int. Conf. of the IEEE Engineering in Medicine and Biology Society, Milano, Italy, 2015, pp. 2661–2664

[11] Ichikawa, T., Matsubara, T., Hara, T., *et al.*: 'Automated detection method for architectural distortion areas on mammograms based on morphological processing and surface analysis', *Proc. SPIE Med. Imag.: Image Process.*, 2004, **5370**, pp. 920–925

[12] Sampat, M.P., Whitman, G.J., Markey, M.K., *et al.*: 'Evidence based detection of spiculated masses and architectural distortion', *Proc. SPIE Med. Imag.: Image Process.*, 2005, **5747**, pp. 26–37

[13] Khoubani, S., Nadjar, H.S., Fatemizadeh, E., *et al.*: 'A two layer texture modeling based on curvelet transform and spiculated lesion filters for recognizing architectural distortion in mammograms'. Middle East Conf. on Biomedical Engineering, Doha, Qatar, 2014, pp. 21–24

[14] Boonthong, P., Benchaporn, J., Rasmequan, S., *et al.*: 'Semi-automated detection of breast mass spiculation using active contour'. 2014 Annual Summit and Conf. on SPIE Medical Imaging, Chiang Mai, Thailand, 2014

[15] Lakshmanan, R., Shiji, T.P., Thomas, V., *et al.*: 'A preprocessing method for reducing search area for architectural distortion in mammographic images'. Fourth Int. Conf. on Advances in Computing and Communications, Kochi, Kerala, India, 2014, pp. 101–104

[16] Otsu, N.: A threshold selection method from gray-level histograms. Available at https://engineering.purdue.edu/kak/computervision/ECE661.08/OTSU_paper.pdf. Accessed as in May 2019

[17] Biswas, S.K., Mukherjee, D.P.: 'Recognizing architectural distortion in mammogram: A multiscale texture modeling approach with GMM', *IEEE Trans. Biomed. Eng.*, 2011, **58**, pp. 2023–2030

[18] Zyout, I., Togneri, R.: 'Empirical mode decomposition of digital mammograms for the statistical based characterization of architectural distortion', *Annual Int. Conf. IEEE Eng. Med. Biol. Soc.*, 2015, **2015**, pp. 109–112

[19] Mohammadi, E., Fatemizadeh, E., Sheikhzadeh, H., *et al.*: 'A textural approach for recognizing architectural distortion in mammograms'. 8th Iranian Conf. on Machine Vision and Image Processing, Zanjan, Iran, 2013, pp. 136–140

[20] Freund, Y.: 'An adaptive version of boost by majority algorithm'. Proc. of the Twelfth Annual Conf. on Computational Learning Theory, Santa Cruz, California, US, 1999

[21] Friedmann, J., Hastle, T., Tibshirani, R.: 'Additive logistic regression: a statistical view of boosting', *Ann. Stat.*, 2000, **28**, (2), pp. 337–407

[22] Heath, M., Bowyer, K., Kopans, D., *et al.*: 'The digital database for screening mammography'. Proceedings of the Fifth Int. Workshop on Digital Mammography, Toronto, Canada, 2001, pp. 212–218

[23] Akhtar, Y., Mukherjee, D.P.: 'Detection of architectural distortion from the ridges in a digitized mammogram', *Signal Image Video Process.*, 2018, **12**, (7), pp. 1285–1292

[24] Heath, M., Bowyer, K., Kopans, D., *et al.*: 'Current status of the digital database for screening mammography'. Digital Mammography, Nijmegen, Netherlands, 1998, pp. 457–460

[25] Nahid, A.A., Kong, Y.: 'Involvement of machine learning for breast cancer image classification: a survey', *Comput. Math. Methods Med.*, 2017, **2017**, p. 29

[26] Alzubi, O., Alzubi, J., Tedmori, S., *et al.*: 'Consensus-based combining method for classifier ensembles', *Int. Arab. J. Inf. Technol.*, 2018, **15**, (1), pp. 76–85