



## Aberystwyth University

### *Classification of Micro-calcification in Mammograms using Scalable Linear Fisher Discriminant Analysis*

Suhail, Zobia; Denton, Erika R. E.; Zwiggelaar, Reyer

*Published in:*

Medical and Biological Engineering and Computing

*DOI:*

[10.1007/s11517-017-1774-z](https://doi.org/10.1007/s11517-017-1774-z)

*Publication date:*

2018

*Citation for published version (APA):*

Suhail, Z., Denton, E. R. E., & Zwiggelaar, R. (2018). Classification of Micro-calcification in Mammograms using Scalable Linear Fisher Discriminant Analysis. *Medical and Biological Engineering and Computing*, 56(8), 1475-1485. <https://doi.org/10.1007/s11517-017-1774-z>

#### Document License

CC BY

#### General rights

Copyright and moral rights for the publications made accessible in the Aberystwyth Research Portal (the Institutional Repository) are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the Aberystwyth Research Portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the Aberystwyth Research Portal

#### Take down policy

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

tel: +44 1970 62 2400  
email: [is@aber.ac.uk](mailto:is@aber.ac.uk)



# Classification of micro-calcification in mammograms using scalable linear Fisher discriminant analysis

Zobia Suhail<sup>1</sup> · Erika R. E. Denton<sup>2</sup> · Reyer Zwiggelaar<sup>1</sup>

Received: 8 June 2017 / Accepted: 13 December 2017 / Published online: 25 January 2018  
© The Author(s) 2018. This article is an open access publication

## Abstract

Breast cancer is one of the major causes of death in women. Computer Aided Diagnosis (CAD) systems are being developed to assist radiologists in early diagnosis. Micro-calcifications can be an early symptom of breast cancer. Besides detection, classification of micro-calcification as benign or malignant is essential in a complete CAD system. We have developed a novel method for the classification of benign and malignant micro-calcification using an improved Fisher Linear Discriminant Analysis (LDA) approach for the linear transformation of segmented micro-calcification data in combination with a Support Vector Machine (SVM) variant to classify between the two classes. The results indicate an average accuracy equal to 96% which is comparable to state-of-the art methods in the literature.

**Keywords** Micro-calcification · Classification · Fisher discriminant analysis · Principal component analysis · Computer aided detection · Dimensionality reduction

## 1 Background

Machine Learning is widely being used to solve problems involving high dimensional data. In a large number of cases, the dimension of the data is much larger than the sample size, which is referred to as the undersampling problem [1]. High dimensionality and undersampling occurs in many applications [2, 3]. One of the solutions to deal with this undersampling problem is dimensionality reduction [3, 4].

Principal component analysis (PCA) is a procedure to convert a number of correlated variables into fewer variables called principal components [5], commonly used in fields of pattern recognition and computer vision [6, 7]. The purpose of PCA is to transform data to some low dimensional space and subsequently apply a classification method. Fisher

Linear Discriminant Analysis (LDA) has been around for a long time with applications found in face recognition [3, 8], marketing [9], and biomedical studies [10]. LDA is a classical approach used for feature extraction and dimensionality reduction [11, 12]. The objective function of conventional LDA is to find a linear transformation, where the class separation is maximized while keeping the in-class variance small [5]. One of the major problems associated with LDA in the singularity issue [13]: LDA requires the scatter matrices of the training data to be non singular but the training samples are from a high dimensional space and in most cases the sample size is smaller than this dimension leading to a sparse matrices and potential singularity issues. Many LDA extensions have been proposed to resolve this singularity issue. PCA+LDA [3] and LDA/QR [4] are some of these two-stage extensions. The purpose of the two-stage approach is to convert the data into some intermediate form before applying the actual LDA. While applying these two-stage algorithms some of the important information may be lost in the first dimensionality reduction stage that may be beneficial for the subsequent LDA [14].

Zhang et al. [14] proposed a fast two-stage LDA algorithm as an alternative to the PCA+LDA or QR/LDA solutions. They claimed with the theoretical analysis that their algorithm outperforms the other two-staged

---

✉ Zobia Suhail  
zoa1@aber.ac.uk  
Erika R. E. Denton  
erika.denton@nnuh.nhs.uk  
Reyer Zwiggelaar  
rrz@aber.ac.uk

<sup>1</sup> Aberystwyth University, Aberystwyth, UK

<sup>2</sup> Norfolk and Norwich University Hospital, Norwich, UK

algorithms [3, 4] with the same scalability. They also provided the theoretical bound on the approximation of two-staged LDA. We propose a novel application of the scalable-LDA approach [14] for the classification of breast calcifications. To our knowledge, such method of feature extraction through dimensionality reduction has never been used for the problem of classifying micro-calcifications. The proposed method provides a way of encoding binary calcification data to a single value, which is a one-dimensional representation of the high dimension micro-calcification data. Instead of using a large number of features, the proposed classification approach used only a single feature to distinguish between the benign and malignant micro-calcifications. In terms of classification accuracy, the algorithm is giving good results compared to other state-of-the-art approaches developed for the classification of malignant and benign micro-calcifications (an overview of state-of-the-art approaches developed for the classification of benign and malignant micro-calcification as well as comparison with the current method can be found in Section 6.1).

## 2 Dataset

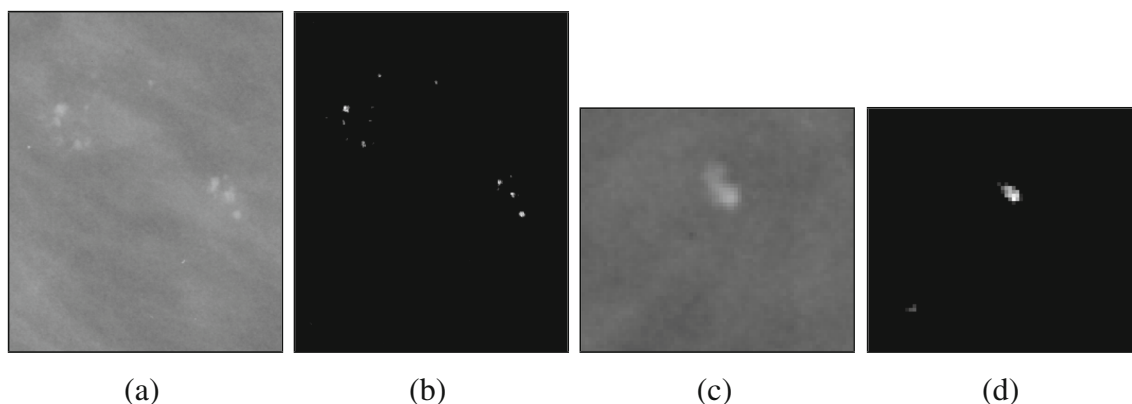
We used data from the Digital Database for Screening Mammography (DDSM) [15], which contains identified micro-calcifications. The mammograms in the DDSM database were digitized by one of four scanners: DBA M2100 ImageClear (42  $\mu\text{m}$  per pixel, 16 bits), Howtek 960 (43.5  $\mu\text{m}$  per pixel, 12 bits), Lumisys 200 Laser (50  $\mu\text{m}$  per pixel, 12 bits), and Howtek MultiRad850 (43.5  $\mu\text{m}$  per pixel, 12 bits). The patches are extracted from the whole mammogram containing micro-calcifications according to the available information regarding position of micro-calcification clusters. In addition, we excluded cases

with overlapping mass regions. Subsequently, the micro-calcification clusters are either automatically detected [16] or manually annotated by expert radiologists. The automatic approach for segmenting micro-calcification clusters [16] involves local feature extraction using filter banks (including delta, Gaussian and Laplacian filters). After performing training to select salient features, a boosting classifier is used to detect individual micro-calcifications. Segmented images are representing the micro-calcification in binary form, where 0's indicate the absence of micro-calcifications and 1's the presence of micro-calcifications. The images in this dataset are of variable sizes. The average image size is (482  $\times$  450), whereas the maximum height of all the images from the dataset is 2754 and the maximum width is 3778. The dataset has in total 288 ROIs (139 malignant and 149 benign). Each of the 288 ROI belongs to a different woman. Some sample variations in the size of the ROIs used in this study can be seen in Figs. 1 and 2 covering benign and malignant classes, respectively.

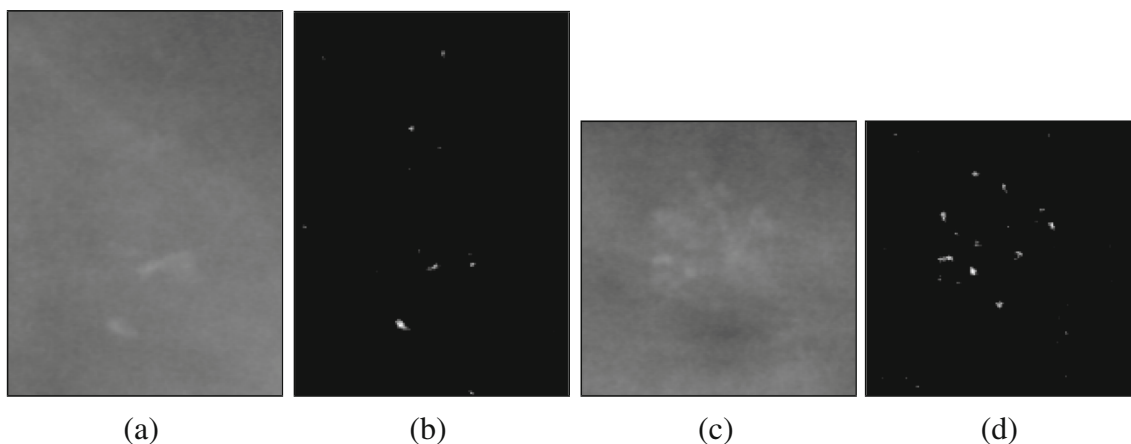
## 3 Fisher discriminant analysis

Given a data matrix  $D = [d_1, d_2, \dots, d_n]^T \in \mathbb{R}^{n \times m}$ , where  $d_i$  is a  $m$ -dimensional vector, let the data matrix  $D$  be partitioned into  $k$  classes as  $D^T = [D_1^T, D_2^T, \dots, D_k^T]$ , where  $D_i \in \mathbb{R}^{n_i \times m}$  and  $\sum_{i=1}^k n_i = n$ . The objective of conventional LDA is to compute the optimal linear transformation  $G \in \mathbb{R}^{m \times l}$  such that the class structure of the original space is preserved in the low-dimensional space. So,  $G$  maps each  $d_i$  of  $D$  in the  $m$ -dimensional space to a vector  $y_j$  in  $l$ -dimensional space.

$$G : d_i \in \mathbb{R}^m \rightarrow y_j = G^T d_i \in \mathbb{R}^l (l < m)$$



**Fig. 1** Sample ROIs from the benign class: **a** and **c** are showing original mammographic ROIs, whereas **b** and **d** are showing the segmented micro-calcifications



**Fig. 2** Sample ROIs from the malignant class: **a** and **c** are showing original mammographic ROIs, whereas **b** and **d** are showing the segmented micro-calcifications

For discriminant analysis [12], two scatter matrices (between class and total scatter matrices) are defined as:

$$S_b = \frac{1}{n} \sum_{i=1}^k \sum_{d \in D_i} (c_i - c)(c_i - c)^T = \frac{1}{n} \sum_{i=1}^k n_i (c_i - c)(c_i - c)^T \quad (1)$$

$$S_t = \frac{1}{n} \sum_{i=1}^n (d_i - c)(d_i - c)^T \quad (2)$$

where  $c_i = \frac{1}{n_i} \sum_{d \in D_i} d$  is the mean of the  $i^{th}$  class and  $c = \frac{1}{n} \sum_{d \in D} d$  is the mean of the whole data set. In the low-dimensional space, obtained as a result of linear transformation  $G$ , the scatter matrices become:

$$S_b^L = G^T S_b G, S_t^L = G^T S_t G$$

The calculation of the scatter matrices can be simplified through using precursors  $H_b$  and  $H_t$  as:

$$H_b = \frac{1}{\sqrt{n}} (\sqrt{n_1}(c_1 - c) \dots \sqrt{n_k}(c_k - c)) \quad (3)$$

$$H_t = \frac{1}{\sqrt{n}} (D^T - ce^T) \quad (4)$$

where  $e = [1, \dots, 1]^T \in \mathbb{R}^n$ , then the scatter matrices  $S_b$  and  $S_t$  can be expressed as:

$$S_b = H_b H_b^T, S_t = H_t H_t^T$$

An optimal transformation  $G$  can be obtained by using the optimization from classical discriminant analysis [12]:

$$\arg \max_G \{trace((S_t^L)^{-1} S_b^L)\}$$

The solution to this optimization can be obtained by applying eigen-decomposition on the matrix  $S_t^{-1} S_b$ , if  $S_t$  is non-singular [12]. However, if  $S_t$  is singular, we can use the eigen-decomposition of  $S_t^\dagger S_b$ , where  $S_t^\dagger$  is the pseudo-inverse of  $S_t$ . The use of pseudo-inverse for LDA has been studied in the literature [17, 18]. Ye [19] proposed a

SVD-based solution for the eigen-decomposition (Algorithm 1), according to which the optimal transformation are the top  $q$  eigenvectors of  $S_t^\dagger S_b$ , where  $q$  is equal to rank ( $H_b$ ), which in most cases is equal to  $k - 1$  ( $k$  is the number of classes). Zhang et al. [14] proposed a two-stage LDA algorithm as a scalable version of the conventional LDA. At the first stage of their algorithm, they introduced a linear transformation  $Z \in \mathbb{R}^{m \times r}$  to reduce the data dimensionality to some intermediate dimension  $r$ , and then applied conventional LDA on the reduced total scatter matrix  $\tilde{S}_t = Z^T S_t Z$  and reduced between-class scatter matrix  $\tilde{S}_b = Z^T S_b Z$  (by representing them as precursors  $\tilde{H}_b$  and  $\tilde{H}_t$ ) in order to get the linear transformation  $\tilde{G}$ . In the final stage they produced the reduced transformation as  $\hat{G} = Z \tilde{G}$ . The pseudocode of this approach is presented in Algorithm 2.

**Algorithm 1** The SVD based LDA algorithm

- 1: Calculate  $H_b$  and  $H_t$  as in Eqs. 1 and 3;
- 2: Compute the reduced SVD of  $H_t$  as  $H_t = U_1 \sum_t V_1^T$ ;
- 3:  $B \leftarrow \sum_t^{-1} U_1^T H_b$ ;
- 4: Compute reduced SVD of  $B$  as  $B = P \sum Q^T$ ;  $q \leftarrow rank(B)$ ;
- 5:  $X \leftarrow U_1 \sum_t^{-1} P$ ;
- 6:  $G \leftarrow X_q$ ;

**Algorithm 2** SVD-QR-LDA algorithm

- Input:** Target dimension  $r$  for the first stage
- 1: Calculate  $H_b$  and  $H_t$  as in Eqs. 1 and 3,  $q \leftarrow rank(H_b)$ ;
  - 2:  $Z_1 \leftarrow$  the top  $r - q$  left singular vectors of  $H_t$ ;
  - 3: Compute QR decomposition with pivoting of  $H_b - Z_1 Z_1^T H_b$  as  $QR[\cdot] = H_b - Z_1 Z_1^T H_b$
  - 4:  $Z_2 \leftarrow$  first  $q$  columns of  $Q$ ,  $Z \leftarrow [Z_1, Z_2]$ ;
  - 5:  $\tilde{H}_b \leftarrow Z^T H_b$ ,  $\tilde{H}_t \leftarrow Z^T H_t$ ;
  - 6:  $\tilde{G} \leftarrow$  result of LDA on  $\tilde{H}_b$  and  $\tilde{H}_t$ ;

**Return:**  $\hat{G} \leftarrow Z \tilde{G}$

## 4 Scalable LDA

In this section, we will explain the overall experimental setup for the Scalable-LDA approach implementation on the given dataset of segmented micro-calcifications with the aim of classifying them as benign or malignant micro-calcifications. The flow of the whole process can be seen in Fig. 3.

### 4.1 Resizing the Rols

As explained in Section 2, we have 288 RoIs from the DDSM database. The basic purpose of this research was to devise a way to distinguish between the two classes of micro-calcification (i.e., benign versus malignant). Algorithm 2 required for all images to be equal size. For this, we resized all the images to max-height  $\times$  max-width so that we are not losing any single piece of information from any image within the dataset. We did this resizing by retaining the original image data  $I(x,y)$  at the center and fill the extra max-height- $I_x$  and max-width- $I_y$  pixels with the value 0. The purpose of using the proposed 0-padding technique instead of any other resizing approach (e.g. interpolation) is to retain the original data. The resizing through this 0-padding will keep the data at the centre of the image frame without adding additional bits to the original data or removing image information. After resizing the images in the dataset, all the images have size equal to  $2754 \times 3778$ .

### 4.2 Getting the intermediate dimensions [two-stage LDA]

Subsequently, we vectorize each image by converting it from  $2754 \times 3778$  to  $10404612 \times 1$  dimensions. In

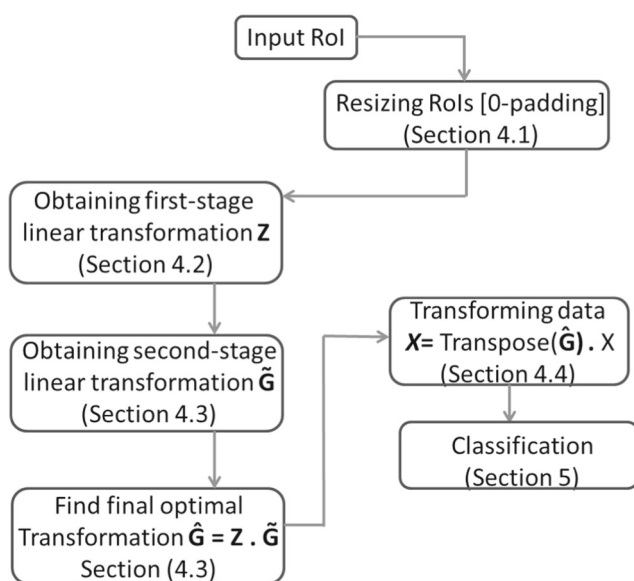


Fig. 3 Process flow of the proposed method

order to apply Fisher LDA for the dataset with this large dimensionality, we need to reduce the dimensionality of the data. As explained in Section 2, we have 149 images from the benign class and 139 from the malignant class, so we have 149 and 139 feature vectors from the benign and malignant classes. Here, we used terms  $D_{benign}$  to represent the data belonging to the benign class and  $D_{malignant}$  represents data from the malignant class, which are:

$$D_{benign} = [d_1, d_2, \dots, d_q]^T \in \mathbb{R}^{q \times m}, q = 149, \text{ and}$$

$$D_{malignant} = [d_{q+1}, d_{q+2}, \dots, d_{q+l}]^T \in \mathbb{R}^{l \times m}, l = 139.$$

The total dataset is then represented as a data matrix  $D^T$  as  $[D_{benign}^T, D_{malignant}^T]$ , where the total size of  $D_{benign}$  is  $149 \times 10404612$  and  $139 \times 10404612$  for  $D_{malignant}$  according to the size of the dataset for each class. The total size of the final data matrix  $D$  is  $288 \times 10404612$  (containing data vectors from both benign and malignant classes). As the sample size is much less than the dimensionality of the feature space, we can not apply the conventional LDA to solve this problem. To execute Algorithm 2, we need to compute the precursors  $H_b$  and  $H_t$  (Eqs. 3 and 4) which requires the mean of the two classes. These precursors i.e.  $H_b$  and  $H_t$ , are required by Algorithms 1 and 2 to obtain the intermediate and final linear transformations.

We compute the mean for dataset  $D_{benign}$ ,  $D_{malignant}$  and  $D$  as:

$$c_{benign} = \frac{1}{q} \sum_{d \in D_{benign}} d$$

$$c_{malignant} = \frac{1}{l} \sum_{d \in D_{malignant}} d \text{ and}$$

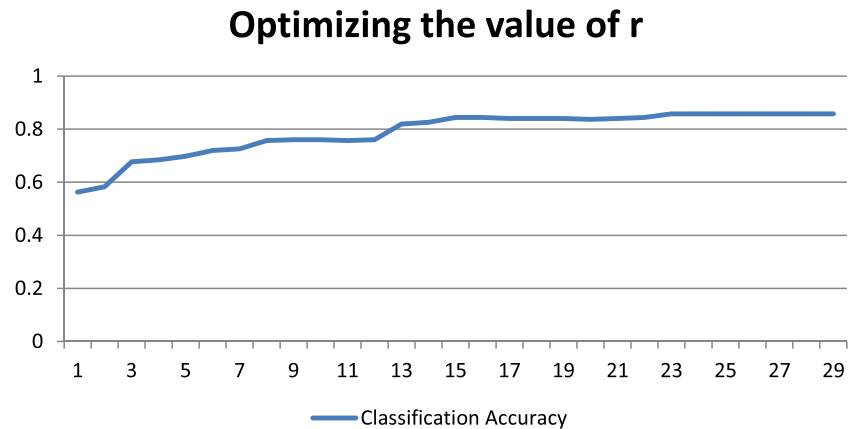
$$c_{total} = \frac{1}{q+l} \sum_{d \in D} d.$$

By using means  $c_{benign}$ ,  $c_{malignant}$  and  $c_{total}$ , we compute the precursors through Eqs. 1 and 2. We run Algorithm 2 by setting the value of  $r = 23$  for the first stage (details of finding optimal value of  $r$  can be found in Section 5). For Algorithm 2 to perform well, the value of  $r \gg q$  [14] and a value to 23 is much larger than  $q$  ( $q = 1$  for our data, as  $q = \text{number of classes} - 1$ ). After executing step 1–4, we obtained the linear transformation  $Z \in \mathbb{R}^{10404612 \times 23}$ , to reduce the dimensionality to 23.

### 4.3 Final linear transformation [two-stage LDA]

After obtaining the intermediate linear transformation  $Z$ , the next step is to find the final optimal transformation. Subsequently, after executing step 5 of Algorithm 2, we obtained  $\tilde{H}_b$  and  $\tilde{H}_t$  (having dimensionality  $(23 \times 2)$  and  $(23 \times 288)$ , respectively) which will be used in step 6 of Algorithm 2, where we used the method proposed in Algorithm 1 for performing LDA (using precursors  $\tilde{H}_b$  and  $\tilde{H}_t$ ) instead of using conventional LDA. The result of

**Fig. 4** Selecting the optimal value of  $r$ . As can be seen the Classification Accuracy gradually increases by strating from value  $r = 1$ , after  $r = 23$  the accuracy becomes stable



this LDA operation at step 6 of Algorithm 2 is a linear transformation  $\tilde{G}$  having dimensionality  $23 \times 1$ . Whereas the dimensionality of transformation matrix  $Z$  at step 4 is  $(10404612 \times 23)$ .

After getting the linear transformations matrix  $Z$  and  $\tilde{G}$ , the next step is to obtain the final optimal transformation  $\hat{G}$ , which is computed by multiplying the transformation matrices  $Z$  ( $10404612 \times 23$ ) and  $\tilde{G}$  ( $23 \times 1$ ). The dimensionality of this final optimal transformation matrix is  $(10404612 \times 1)$ .

#### 4.4 Transforming the data to the reduced dimension

Next we transform our data (being represented as a vector  $[1 \times 10404612]^T$ ) using the linear transformation  $\hat{G}$ . We get the transformation as  $\hat{G}^T d_i$ , where  $d_i \in d_1, d_2 \dots d_q, d_{q+1}, d_{q+2} \dots d_{q+l}$ , where  $\hat{G}$  transformed each vector  $d$  from  $D$  to a one-dimensional space. The result of this projection is a single value for each data vector. The next step is to verify and check whether the data is linearly separable or not, and to use this to classify the data into the benign and malignant classes.

### 5 Performance evaluation

We classified our projected data with five different classifiers, which are Support Vector Machine (SVM), Bayesian Network (BN), K-Nearest Neighbour (K-NN), Decision Table (DT) and ADTree. We used Weka [20] for all these experiments (developer version 3.7.2) and we left all the classifier’s parameters at the default setting.

As the projected data is one-dimensional, one of the ways to classify the data is to use a Support Vector Machine (SVM) [5]. We used Sequential Minimal Optimization (SMO) which is available in Weka [20] as an efficient way to solve the SVM problem [21]. We used a 10 runs

10-fold cross-validation (10-FCV) scheme for the performance evaluation of our results.

**Optimizing the parameter  $r$**  As explained in Section 4.2, we selected the value of  $r$  equal to 23. In order to select this optimal value of  $r$ , we executed the algorithm for a range of values for  $r$  starting from 1. After projecting the data to a single dimension, we used SMO at each value  $r$  from  $1 \dots 29$ . At  $r = 23$ , we found stable as well as reliable classification results (Classification Accuracy equal to 85.57%) using a SMO classifier. The details of the classification results achieved for each value of  $r$  can be seen in Fig. 4.

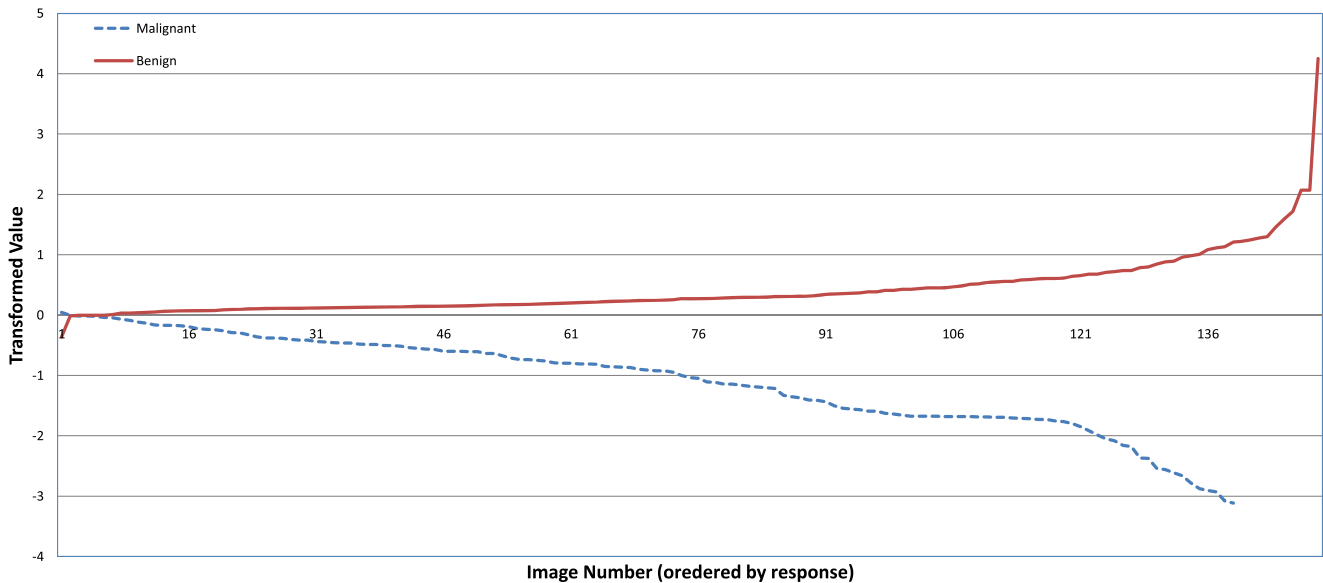
Classification results for the SMO classifier as well as for 4 other classifiers, i.e., BN, KNN, DT, and ADTree (using 10-runs 10-FCV) on the transformed data by using the Scalable-LDA approach can be found in the first column of Table 1.

The overall results were improved when compared to SVM (on average the accuracy was 96%).

In addition to classification accuracy, another commonly used evaluation metric is area under the ROC curve (AUC) (normally denoted as  $A_z$ ), which is used to explain the diagnostic ability of the classifier [22]. The values of  $A_z$  by using 10-FCV classification scheme for 5 classifiers: SMO, BN, KNN, DT, ADTree are 0.853, 0.975, 0.972, 0.975,

**Table 1** Classification accuracy (%) for SMO, BN, KNN, DT and ADTree

	Scalable-LDA	PCA-LDA
SMO	85.57 ± 5.63%	55.85 ± 7.84%
BN	98.61 ± 1.97%	51.27 ± 1.06%
KNN	97.64 ± 2.64%	49.53 ± 8.63%
DT	98.61 ± 1.97%	51.27 ± 1.06%
ADTree	98.30 ± 2.17%	52.36 ± 8.16%



**Fig. 5** Transformed micro-calcification data. Almost all the malignant images have been transformed to the negative axis, whereas the benign images have been transformed to the positive axis. Only a small

amount of data has been misclassified that have been shown at the left hand side of each curve for the benign and malignant class

and 0.985, respectively, which indicates the stability of the classifiers (excluding SMO).

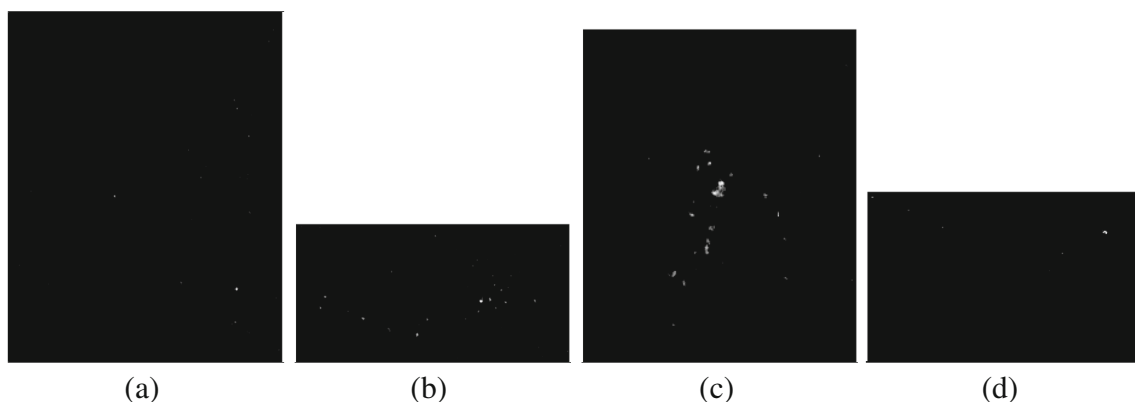
that have been correctly classified by using the developed approach. The projected value for each of these RoIs can also be seen in the figure's captions.

**About projected data** While examining the projected data, we found that the values  $\geq 0$  are classified as benign, whereas the values  $< 0$  are classified as malignant, which is shown in Fig. 5. The result in Fig. 5 cover 139 images from the malignant and 149 images from the benign class. It can also be clearly seen from Fig. 5 that this linear transformation of the segmented micro-calcification data is linearly separable.

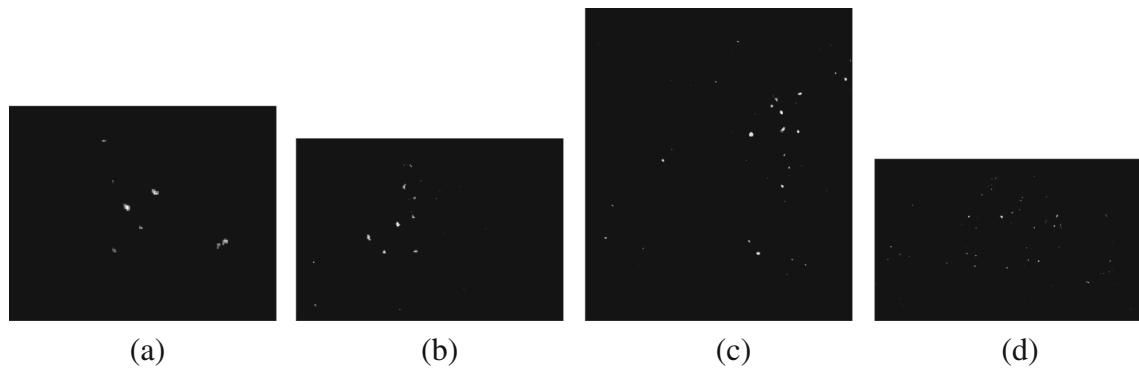
Figures 6 and 7 are showing sample cases (for both the benign and malignant classes) from the referenced dataset

## 6 Discussion on misclassified data

While observing the misclassified data, we observed characteristics of misclassified instances for all the classifiers used in our experiments (SMO, BN, KNN, DT and ADTree). We found 3 instances in all the classifiers (1 from the benign and 2 from the malignant class). Upon investigating the images from these misclassified data, the one benign image is very similar in appearance to the malignant images (having



**Fig. 6** Sample RoIs from the benign class that have been correctly classified: The projected values for the RoIs shown in a–d are 0.125, 0.416, 1.295, and 0.109, respectively



**Fig. 7** Sample ROIs from the malignant class that have been correctly classified: The projected values for the ROIs showed in the figure from **a–d** are  $-0.207$ ,  $-0.607$ ,  $-2.517$ , and  $-1.615$ , respectively

a dense cluster of micro-calcifications) and one malignant image has similarity with benign ROIs, whereas one of the malignant image have no detected micro-calcification. These three misclassified ROIs are shown in Fig. 8.

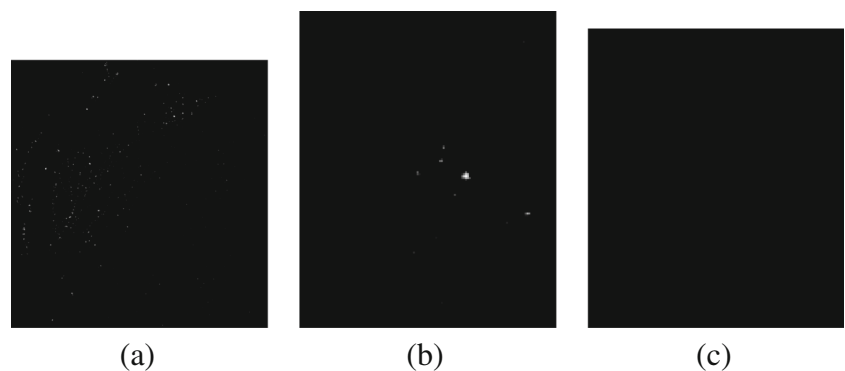
As stated in Section 1, several two stage approaches have also been proposed for LDA implementation [3, 4], which reduce the dimensionality of the data at the first stage before applying the actual LDA algorithm. We applied another version of LDA on our micro-calcification data for the benign and malignant classification that first transformed data to some low-dimensional space by using PCA before applying linear transformation of data (using LDA).

We used the Gram matrix (Appendix A) here for the computation of eigen vectors required for PCA. If  $D$  is representing a data matrix, then the Gram matrix for  $D$  is represented as  $DD^T$ . The Gram matrix has been used in the literature for solving such eigenvalue problems and dimensionality reduction for large data sets [23, 24]. As our data matrix  $D$  is having dimension  $n \times m$  and  $m \gg n$  (for our dataset  $m = 10404612$ ,  $n = 288$ ), it is better to compute the eigen vectors of matrix  $DD^T$  instead of  $D^T D$ . The relationship between the two matrices  $D^T D$  and  $DD^T$  can be found in Appendix B. After computing the eigen vectors  $v$  of  $DD^T$  we left multiply  $v$  by  $D^T$  in order to get the

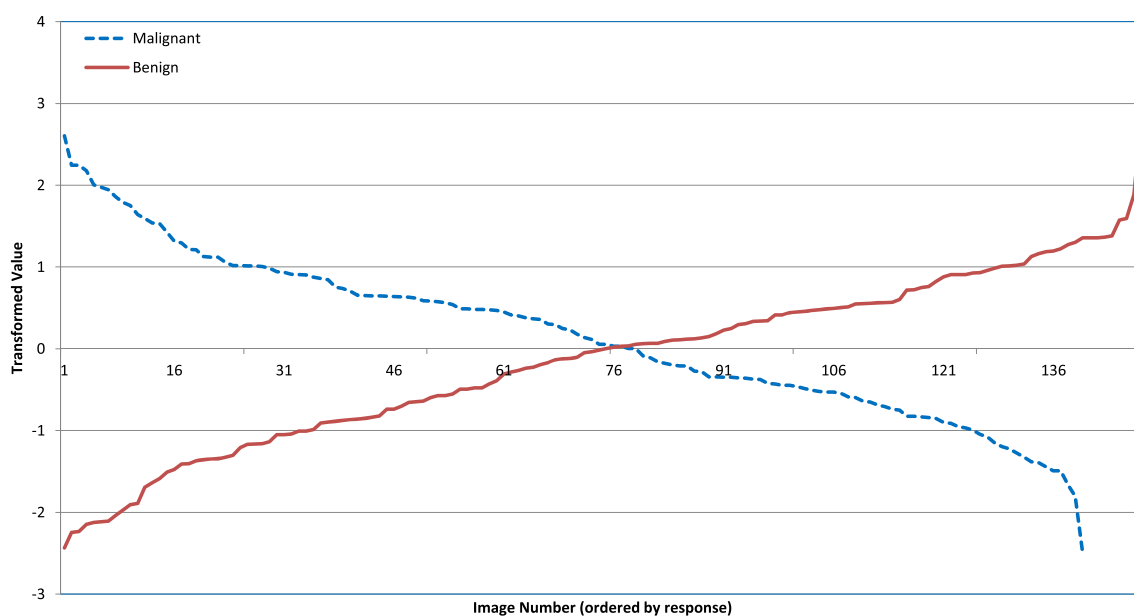
eigen vectors of  $D^T D$ . In order to consider matrix  $DD^T$ , we already normalized the data matrix  $D$  to zero mean in order to consider  $DD^T$  as the covariance matrix of  $D$ .

First we apply PCA on our data by setting the intermediate dimension to 10. We tried to set this dimension to 23, in order to be consistent with the Scalable-LDA approach (Section 4), but due to the limitations posed by the current memory we had to set the intermediate dimensions to 10. It should be noted here that we used the Gram matrix here for computing the eigen vectors, in order to cope with the memory requirements in computing eigen vectors of  $D^T D$ . After that we transformed this reduced dimensional data to another linear space by using LDA. We used the LDA algorithm available with sklearn which is a machine learning library available with Python. The python version that was used in this experiment was version 2.7.0. The results of transforming the data matrix  $D$  to a linear dimension after applying applying LDA can be seen in Fig. 9. As can be clearly seen from the figure, the data is not linearly separable. We used the same classifiers (SMO, BN, KNN, DT, and ADT) and the average classification accuracy was 52% which is far less than the accuracy achieved by the developed approach (i.e. 96%), with details provided

**Fig. 8** Three mis-classified ROIs found in the classification results of all the five classifiers (SMO, BN, KNN, DT, ADTree). **a** Misclassified benign ROI (projected value  $-0.359$ ), **b** Misclassified malignant ROI (projected value:  $0.046$ ), **c** Misclassified malignant ROI (projected value:  $0$ )







**Fig. 9** Transformed micro-calcification data by using the PCA-LDA approach. As can be seen, the data could not be classified correctly based on the transformed data. This is also in accordance with the

results presented in column 2 of Table 1, according to which on average 52% of classification's accuracy could be achieved from the transformed data by using the PCA-LDA approach

in Table 1. The results of the proposed scalable-LDA approach with setting the intermediate dimension  $r$  equal to 10 resulted in classification accuracy of 75% using the SMO classifier, whereas the results for BN, KNN, DT and ADTree classifiers were similar as in Table 1.

## 6.1 Comparative analysis

Multiple methods have been developed in the past for the purpose of classifying benign and malignant micro-calcifications [25, 26]. Some of these approaches rely on the features extracted from individual calcifications [27, 28], whereas some focused on extracting global features from clusters of calcifications [29–32]. Ma et al. [27], suggested the roughness of the individual micro-calcification as a discriminatory property for classifying benign and malignant micro-calcifications. Similarly, other shape features (measures of compactness, moments and Fourier descriptors) have also been studied in the past [28] to measure the roughness of the contours of calcifications and then used as a measure of classifying the benign and malignant micro-calcifications. In a work related to extracting cluster level features of micro-calcifications [29], 23 features were extracted from micro-calcifications. The features were categorized into three general types as: intensity statistics, shape features and linear structure features. They also used balanced learning and optimized decision making for the classification of the micro-calcification clusters. Final results used two classifiers

(Artificial Neural Network (ANN) and Support Vector Machine (SVM)) for classifying benign and malignant micro-calcifications. In addition, topological models have also been studied in the past for modeling and classification of micro-calcifications [30–32]. Both fixed-scale [32] and multi-scale approaches [30, 31] have been proposed in the past for classifying the micro-calcifications with promising results. In addition, the topological models are also useful for radiologists/doctors for visual interpretation of the underlying micro-calcification's structure.

Apart from the methods developed for manual feature extraction (both for individual calcifications or cluster level feature extraction), to our knowledge no method has been developed so far that focussed on the automatic features extraction through dimensionality reduction techniques for classifying the benign and malignant micro-calcifications. The work in this paper presents a novel application of feature extraction through dimensionality reduction techniques for classifying the micro-calcifications as benign or malignant. The results are comparable with other state-of-the-art approaches developed to solve the same problem. The detailed comparison of the proposed technique with the state-of-the-art approaches is shown in Table 2.

The performance measures used for the comparison are Classification Accuracy and area under the ROC curve ( $A_z$ ). The comparison has been made with 6 existing methods. For the current work, average values for the classification accuracy and  $A_z$  are reported for the 5 classifiers used in the experiments.

**Table 2** Results comparison with existing state-of-the-art approaches for the classification of benign and malignant micro-calcification

Author	Database	No. of cases	Accuracy (%)	$A_z$
Shen et al. [28]	Unspecified	18	100	–
Ma et al. [27]	DDSM	183	80	0.76
Ren et al. [29]	DDSM	150	–	avg (SVM & ANN) 0.94
Chen et al. [30]	DDSM	300	86	0.90
Strange et al. [31]	DDSM	300	80	0.82
Suhail et al. [32]	DDSM	129	91	–
Scalable-LDA	DDSM	288	96	0.95

For the Scalable-LDA approach, average results for the 5 classifiers has been reported

## 7 Future work

The basic purpose of this research was to develop a method to classify binary images which contain benign or malignant micro-calcifications. In the future, we will try to use this scalable-LDA approach for the classification of normal and abnormal mammographic images from the DDSM database: to convert the mammographic images into some binary representation before applying this scalable-LDA approach.

Although the results are satisfactory (average accuracy 96%), other methods exist in the literature for making LDA/PCA scalable [3, 4, 33], and we will do some comparative work by applying these approaches to the same dataset.

## 8 Conclusions

We executed a two-stage LDA approach on binary data representing benign and malignant micro-calcifications. The idea was to project the data onto a low dimensional space and then use the resulting information to classify the data into two classes, but undersampling caused problems. By implementing a two-stage LDA approach, we achieved results that are comparable to state-of-the art approaches. The current method also presents a way to encode binary micro-calcification data as a single value. We achieved an accuracy of 96% on average for 5 classifiers, with the best performance at 98.6% by applying a scalable LDA approach for the classifications of benign and malignant micro-calcification. We compared the results with applying PCA-LDA on the same data, indicating a clear difference between the two approaches (Table 1, column 2).

**Data availability** The data that support the findings of this study are available from Prof. Reyer Zwiggelaar, Aberystwyth University but restrictions apply to the availability of these data, which were used under license for the current study, and so are not publicly available. Data are however available from the authors upon reasonable request and with permission of Prof. Reyer Zwiggelaar, Aberystwyth University.

**Acknowledgements** The authors would like to thank Dr. N. Mac Parthaláin for providing assistance to solve the classification problem as well as discussion regarding feature selection and optimization. The authors would also like to acknowledge Sandy Spence and Alun Jones for providing extensive hardware support to complete the experiments.

**Author contributions** ZS carried out the implementation as well as the analysis and results evaluation. ED did the data augmentation and provided the class labels for the dataset. RZ suggested the overall model to solve the classification problem along with the overall coordination. All authors contributed to the written manuscript.

## Compliance with ethical standards

**Competing interests** The authors declare that they have no competing interests.

**Open Access** This article is distributed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made.

## Appendix A: Gram matrix

Consider  $m$  vectors  $\in \mathbb{R}^n [x_1, \dots, x_m]$ . The Gram matrix for this data collection is the  $m \times m$  matrix  $G$  with elements  $G_{ij} = x_i^T x_j$ . The matrix can be expressed as matrix  $X = [x_1, \dots, x_m]$ , as

$$G = X^T X = \begin{bmatrix} x_1^T \\ \vdots \\ x_m^T \end{bmatrix} [x_1 \dots x_m]$$

## Appendix B: Eigenvectors from Gram matrix

Let  $A$  be the matrix having dimension  $m \times n$ . Gram matrix of  $A$  can be computed as  $AA^T$ . If  $\nu$  and  $\lambda$  is representing eigen values and eigen vector of Gram matrix  $AA^T$ , then

$$AA^T \nu = \lambda \nu$$

Left multiply both sides by  $A^T$ , we get

$$A^T(AA^T v) = A^T(\lambda v)$$

Then by re-arranging terms:

$$(A^T A)(A^T v) = \lambda(A^T v)$$

Therefore, if  $v$  is an eigenvector of  $AA^T$ , then  $A^T v$  is an eigenvector of  $A^T A$  with the same eigenvalue  $\lambda$ .

## References

- Krzanowski WJ, Jonathan P, McCarthy WV, Thomas MR (1995) Discriminant analysis with singular covariance matrices: methods and applications to spectroscopic data. *Appl Stat* 44:101–115
- Berry MW, Dumais ST, O'Brien GW (1995) Using linear algebra for intelligent information retrieval. *SIAM Rev* 37(4):573–595
- Belhumeur PN, Kriegman DJ, Hespanha JP (1997) Eigenfaces vs. fisherfaces: recognition using class specific linear projection. *IEEE Trans Pattern Anal Mach Intell* 19(7):711–720
- Ye J, Li Q (2005) A two-stage linear discriminant analysis via qr-decomposition. *IEEE Trans Pattern Anal Mach Intell* 27(6):929–941
- Bishop MC (2006) Pattern recognition. *Mach Learn* 128:179–599
- Li C, Diao Y, Ma H, Li Y (2008) A statistical pca method for face recognition. In: Second international symposium on intelligent information technology application, 2008 IITA'08, vol 3. IEEE, pp 376–380
- Karamizadeh S, Abdullah SM, SM AA, Zamani M, Hooman A (2013) Manaf an overview of principal component analysis. *J Signal Inform Process* 4(3B):173
- Etemad K, Chellappa R (1997) Discriminant analysis for recognition of human face images. *JOSA A* 14(8):1724–1733
- Crask MR, Perreault WDJr (1977) Validation of discriminant analysis in marketing research. *J Mark Res* 14:60–68
- Dudoit S, Fridlyand J, Speed TP (2002) Comparison of discrimination methods for the classification of tumors using gene expression data. *J Am Stat Assoc* 97(457):77–87
- Duda RO, Hart PE, Stork DG (2000) Pattern classification, 2nd edn. Wiley, New York, p 153
- Fukunaga K (1990) Introduction to statistical pattern classification
- Thomaz CE, Kitani EC, Gillies DF (2006) A maximum uncertainty lda-based approach for limited sample size problems-with application to face recognition. *J Braz Comput Soc* 12(2):7–18
- Tu B, Zhang Z, Wang S, Qian H (2014) Making fisher discriminant analysis scalable. In: Proceedings of the 31st international conference on machine learning, pp 964–972
- Heath M, Bowyer K, Kopans D, Moore R, Kegelmeyer WP (2000) The digital database for screening mammography. In: Proceedings of the 5th international workshop on digital mammography, pp 212–218
- Oliver A, Torrent A, Lladó X, Tortajada M, Tortajada L, Sentis M, Freixenet J, Zwiggelaar R (2012) Automatic microcalcification and cluster detection for digital and digitised mammograms. *Knowl-Based Syst* 28:68–75
- Raudys S, Duin R (1998) Expected classification error of the Fisher linear classifier with pseudo-inverse covariance matrix. *Pattern Recogn Lett* 19(5):385–392
- Skurichina M, Duin R (1996) Stabilizing classifiers for very small sample sizes. In: Proceedings of the 13th international conference on pattern recognition, 1996, vol 2. IEEE, pp 891–896
- Ye J (2005) Characterization of a family of algorithms for generalized discriminant analysis on undersampled problems. *J Mach Learn Res* 6:483–502
- Hall M, Frank E, Holmes G, Pfahringer B, Reutemann P, Witten IH (2009) The weka data mining software: an update. *ACM SIGKDD Explorations Newsletter* 11(1):10–18
- Platt JC (1999) 12 fast training of support vector machines using sequential minimal optimization. *Advances in Kernel Methods*, pp 185–208
- Beck JR, Shultz EK (1986) The use of relative operating characteristic (ROC) curves in test performance evaluation. *Arch Pathol Lab Med* 110(1):13–20
- Saul LK, Weinberger KQ, Ham JH, Sha F, Lee DD (2006) Spectral methods for dimensionality reduction. *Semisupervised Learning*:293–308
- Weinberger KQ, Saul L (2006) Unsupervised learning of image manifolds by semidefinite programming. *Int J Comput Vis* 70(1):77–90
- Elter M, Horsch A (2009) Caxd of mammographic masses and clustered microcalcifications: a review. *Med Phys* 36(6):2052–2068
- Cheng HD, Cai X, Chen X, Hu L, Lou X (2003) Computer-aided detection and classification of microcalcifications in mammograms: a survey. *Pattern Recogn* 36(12):2967–2991
- Ma Y, Tay PC, Adams RD, Zhang JZ (2010) A novel shape feature to classify microcalcifications. In: 2010 17th IEEE international conference on image processing (ICIP). IEEE, pp 2265–2268
- Shen L, Rangayyan RM, Desautels JL (1994) Application of shape analysis to mammographic calcifications. *IEEE Trans Med Imaging* 13(2):263–274
- Ren J (2012) ANN vs SVM: which one performs better in classification of mcs in mammogram imaging. *Knowl-Based Syst* 26:144–153
- Chen Z, Strange H, Oliver A, Denton ER, Boggis C, Zwiggelaar R (2015) Topological modeling and classification of mammographic microcalcification clusters. *IEEE Trans Biomed Eng* 62(4):1203–1214
- Strange H, Chen Z, Denton ER, Zwiggelaar R (2014) Modelling mammographic microcalcification clusters using persistent mereotopology. *Pattern Recogn Lett* 47:157–163
- Suhail Z, Denton ER, Zwiggelaar R (2017) Tree-based modelling for the classification of mammographic benign and malignant micro-calcification clusters. *Multimedia Tools and Applications*, pp 1–14
- Zabalza J, Ren J, Yang M, Zhang Y, Wang J, Marshall S, Han J (2014) Novel folded-pca for improved feature extraction and data reduction with hyperspectral imaging and sar in remote sensing. *ISPRS J Photogramm Remote Sens* 93:112–122

**Zobia Suhail** received her M.Phil degree in Computer Science in 2014 from Punjab University College of Information Technology (PUCIT), University of the Punjab, Pakistan. She is a lecturer in the Computer Science department of Punjab University (PUCIT), Pakistan since 2011. Currently she is working towards her PhD in computer science from Aberystwyth University, UK. Her current research is in Computer Vision, Machine Learning and Medical Image Processing (specifically mammographic images). She is first author of 2 journal and 1 conference papers related to mammogram image processing.



**Erika R. E. Denton** received the M.B.B.S. degree from St. Thomas' Hospital Medical School, London, U.K., in 1989. She was trained in radiology at Guys and St. Thomas' Hospital, receiving M.R.C.P. and F.R.C.R. degrees in 1992 and 1994, respectively. From 2005 to 2016, she has been National Clinical Lead, and subsequently the National Clinical Director, for Imaging at the Department of Health. She has been the Chairman of the

Royal College of Radiologist Breast Group and the Vice-President and subsequently the President of the U.K. Radiological Congress. Prior to her national role, she was Divisional Clinical Director for support services in Norwich, the Head of the Norwich Radiology Academy, and the Head of training in Norwich. She is now Clinical Advisor for Imaging to NHS improvement, Honorary Professor of Radiology at UAE and Associate Medical Director at the Norfolk & Norwich University Hospital.



**Reyer Zwigelaar** received the Ir. degree in Applied Physics from the State University Groningen, Groningen, The Netherlands, in 1989, and the Ph.D. degree in Electronic and Electrical Engineering from University College London, London, UK, in 1993. He is currently a Professor in the Department of Computer Science, Aberystwyth University, UK. He is the author or co-author of more than 250 conference and journal papers. His current

research interests include Medical Image Understanding, especially focusing on Mammographic and Prostate Data, Pattern Recognition, Statistical Methods, Texture-Based Segmentation, Biometrics, Manifold Learning and Feature-Detection Techniques. He is an Associate Editor of Pattern Recognition and Journal of Biomedical and Health Informatics (JBHI).