Topological Modelling and Classification of Mammographic Microcalcification Clusters

Chen, Zhili; Strange, Harry; Oliver, Arnau; Denton, Erika R. E.; Boggis, C. R. M.; Zwiggelaar, Reyer

Published in:
IEEE Transactions on Biomedical Engineering

DOI:
10.1109/TBME.2014.2385102

Publication date:
2015

Citation for published version (APA):
Chen, Z., Strange, H., Oliver, A., Denton, E. R. E., Boggis, C. R. M., & Zwiggelaar, R. (2015). Topological Modelling and Classification of Mammographic Microcalcification Clusters. *IEEE Transactions on Biomedical Engineering*, *62*(4), 1203-1214. https://doi.org/10.1109/TBME.2014.2385102

Topological Modelling and Classification of Mammographic Microcalcification Clusters

Zhili Chen, Harry Strange, Arnau Oliver, Erika R. E. Denton, Caroline Boggis, and Reyer Zwiggelaar

Abstract*—*Goal: **The presence of microcalcification clusters is a primary sign of breast cancer; however, it is difficult and time consuming for radiologists to classify microcalcifications as malignant or benign. In this paper, a novel method for the classification of microcalcification clusters in mammograms is proposed.** ***Methods:*** **The topology/connectivity of individual microcalcifications is analysed within a cluster using multiscale morphology. This is distinct from existing approaches that tend to concentrate on the morphology of individual microcalcifications and/or global (statistical) cluster features. A set of microcalcification graphs are generated to represent the topological structure of microcalcification clusters at different scales. Subsequently, graph theoretical features are extracted which constitute the topological feature space for modelling and classifying microcalcification clusters. k-Nearest Neighbours based classifiers are employed for classifying microcalcification clusters.** ***Results:*** **The validity of the proposed method is evaluated using two well-known digitised datasets (MIAS and DDSM) and a full-field digital dataset. High classification accuracies (up to 96%) and good ROC results (area under the ROC curve up to 0.96) are achieved. A full comparison with related publications is provided, which includes a direct comparison.** ***Conclusion:*** **The results indicate that the proposed approach is able to outperform the current state-of-the-art methods.** ***Significance:*** **This work shows that topology modelling is an important tool for microcalcification analysis not only because of the improved classification accuracy but also because the topological measures can be linked to clinical understanding.**

***Index Terms*—mammography, microcalcifications, graphs, topology, classification.**

I. Introduction

BREAST cancer is currently the most common cancer affecting women worldwide [1]. In European women it is the leading cause of cancer death, causing 1 in 6 of all deaths from cancers [2]. In the United States, a woman has a 12.15% (about 1 in 8) risk of developing breast cancer during her lifetime [3]. Mammography is one of the most reliable and effective methods for detecting breast cancer at its early stages [4]. In developed countries, population-based mammography screening programmes have been implemented [1]. Women are encouraged to participate in regular breast examinations through mammography. In the United States, annual mammographic screening is recommended for women at normal risk, beginning at age 40 [5]. In the UK, women aged between 50 and 70 years are invited for breast screening every three years [6].

Microcalcifications are small deposits of calcium salts within breast tissue that appear as small bright spots in mammograms [7]–[10]. The presence of microcalcification clusters is a primary sign of breast cancer. The radiological definition of a microcalcification cluster is an area of 1 cm^2 that contains ,in general, no fewer than 3 microcalcifications [10], [11]. The spatial resolution of mammography is very high (normally in the range of $40-100\mu m$ per pixel) and therefore mammography enables the detection of microcalcifications at an early stage. However, not all microcalcification clusters necessarily indicate the presence of cancer, only certain kinds of microcalcifications are associated with a high probability of malignancy [12], [13]. The first column of Fig. 1 shows two mammographic image patches taken from the Mammographic Image Analysis Society (MIAS) database [14], containing a malignant microcalcification cluster and a benign microcalcification cluster, respectively. In clinical practice, it is difficult and time consuming for radiologists to distinguish malignant from benign microcalcifications. This results in a high rate of unnecessary biopsy examinations [9], [11]. In order to improve the diagnostic accuracy of radiologists interpreting microcalcifications in mammograms, computer-aided diagnosis (CAD) systems have been applied to reduce the false positive rate while maintaining sensitivity [9], [15].

Many methods for computer-aided diagnosis of microcalcifications in mammograms have been proposed [9], [17]. A variety of features have been studied in the literature to characterise microcalcifications and classify these abnormalities into malignant and benign, such as shape, morphological, cluster, intensity-based, and texture features [9], [17]. Early research showed how the morphological characteristics of microcalcifications could be used to differentiate between malignant and benign cases [18]. The shape and morphological features are mainly extracted from individual microcalcifications and describe the morphological characteristics of individual microcalcifications, such as roughness, size, and shape [7], [10], [19]–[21]. Complementary to the shape and morphological features, cluster features concentrate on the global properties of microcalcification clusters [8], [15], [20], [22]–[25]. Some were used to describe the morphology of mi-

Z. Chen is with the Faculty of Information and Control Engineering, Shenyang Jianzhu University, Shenyang, 110168 China e-mail: zzc@sjzu.edu.cn

H. Strange and R. Zwiggelaar are with the Department of Computer Science, Aberystwyth University, Penglais Campus, Aberystwyth, SY23 3DB UK e-mail: hgs08@aber.ac.uk, rrz@aber.ac.uk

A. Oliver is with the Department of Computer Architecture and Technology, University of Girona, Girona, 17071 Spain e-mail: aoliver@eia.udg.edu

E. R. E. Denton is with the Department of Radiology, Norfolk and Norwich University Hospital, Norwich, NR4 7UY UK e-mail: erika.denton@nnuh.nhs.uk

C. Boggis is with the School of Medicine, The University of Manchester, Manchester, M13 9PT UK email: caroline.boggis@manchester.ac.uk

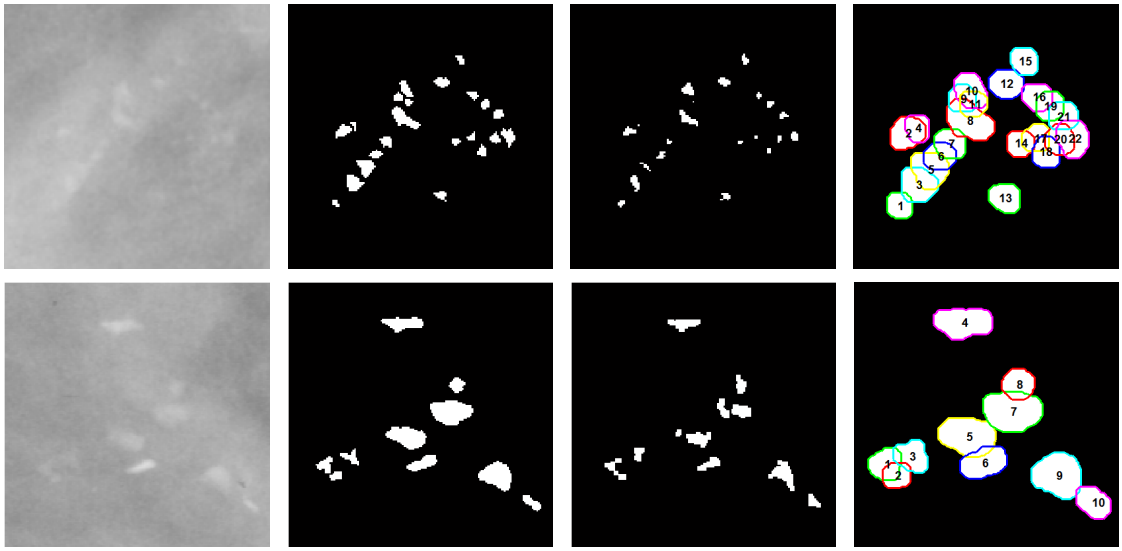


Fig. 1. Example microcalcification clusters: malignant (top row) and benign (bottom row). First column: mammographic image patches; second column: manual annotations; third column: automatic detections [16]; fourth column: dilated microcalcifications using a disk-shaped structuring element (radius equal to 6 pixels).

crocalcification clusters, such as cluster area, cluster perimeter, cluster diameter, cluster circularity, cluster eccentricity, and cluster elongation. Others were intended to capture the spatial distribution of individual microcalcifications within a cluster, such as average and standard deviation of distances between microcalcifications. In addition, a novel model-based method was presented to reconstruct and analyse microcalcification clusters in 3D from two mammographic views [26].

Although a broad variety of techniques for computer-aided diagnosis of breast cancer have been developed in the past two decades, some of which have achieved a high sensitivity and specificity for specific abnormalities, the automatic and accurate classification of microcalcification abnormalities as malignant or benign remains a challenge due to their inherent nature; furthermore, most of the existing approaches have their own specific disadvantages. Firstly, for the approaches based on the shape/morphology of individual microcalcifications [7], [10], informative features cannot be attained when microcalcifications are very small (occupying only a few pixels) so that it seems meaningless to analyse the shape/morphological properties of such small objects. Secondly, microcalcifications may have very low contrast with respect to the surrounding tissue especially when microcalcifications form within dense tissue which has high and homogeneous intensity. As such, the lack of useful texture information within the background region affects the capability of the approaches based on the intensity variations and texture features [8], [27]. In addition, for the approaches describing the spatial distribution of microcalcifications within a cluster, the global cluster features were computed based on a fixed resolution and the distance-based features rely on the original spatial resolution of mammography. This results in a lack of robustness and adaptiveness to different spatial resolutions of mammograms in particular screen-film mammograms acquired by different digitisers.

According to some studies on the evaluation of breast microcalcifications, malignant microcalcifications tend to be small, numerous (> 5 per focus within 1 cm^2) and densely distributed because they lie within the milk ducts and associated structures in the breast and follow the ductal anatomy [12], [13], [28]. However, benign microcalcifications are generally larger, smaller in number ($< 4-5$ per 1 cm^2) and more diffusely distributed as these microcalcifications arise within the breast stroma, benign cysts or benign masses [12], [13], [28]. These differences result in variations in the distribution and closeness of microcalcifications within the clusters and provide radiologists with information which enables decisions regarding the need for further assessment and possible breast biopsy. Hence, we propose a novel method for modelling and classifying microcalcification clusters in mammograms based on their topological properties. The topology of microcalcification clusters is analysed at multiple scales using a graph-based representation of their topological structure. This method is distinct from existing approaches that mainly concentrate on the morphology of individual microcalcifications and only compute the distance-based cluster features at a fixed scale. In this method, a set of topological features are extracted from microcalcification graphs at multiple scales and a multiscale topological feature vector is subsequently generated to discriminate between malignant and benign cases.

A preliminary version of this work has been reported in [29], where the idea of analysing microcalcification clusters using their topological structure is initially investigated based on a small number of cases. In this paper, the evaluation has been extended by including additional data (from several databases). We have also investigated the effect of variation in microcalcification segmentation, the dataset size, the individual significance of eight graph metrics for malignancy diagnosis, and a direct comparison with state-of-the art methods.

II. Data

The data used in the experiments consists of three datasets which are composed of image patches of different cases (taken

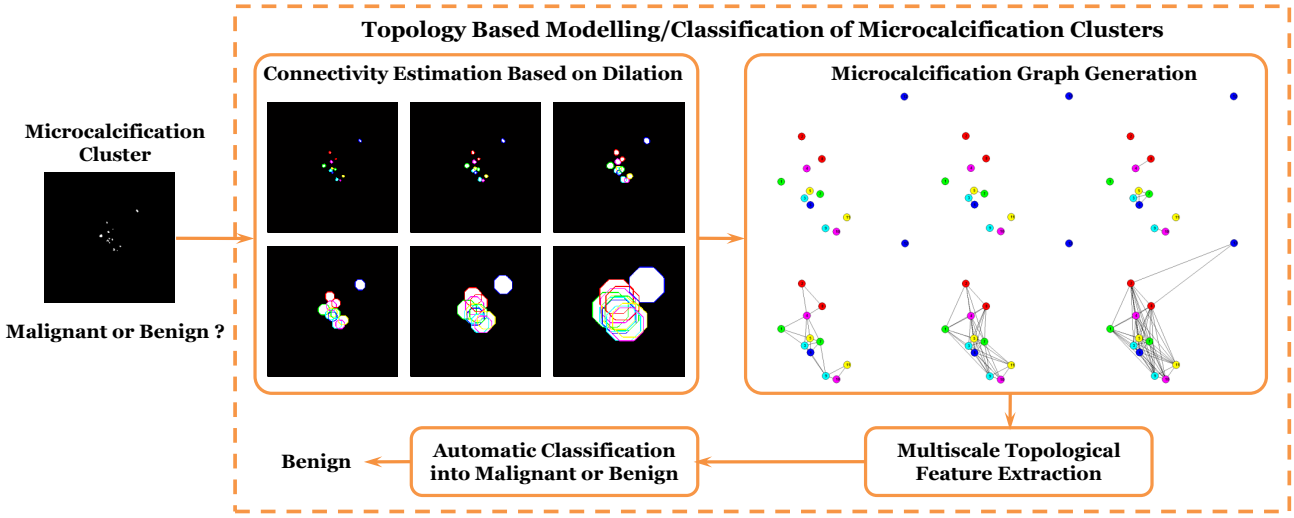


Fig. 2. Methodology framework for topology based modelling and classifying malignant and benign microcalcification clusters in mammograms. The scales shown are $s = 1, 4, 8, 16, 32, 64$ displayed left-to-right and top-to-bottom.

from different mammograms). The first dataset was taken from the MIAS database [14], containing 20 image patches with the same size of 512×512 pixels. The mammograms were digitised to 50 microns per pixel with a linear optical density in the range $0 - 3.2$. The second dataset was extracted from the Digital Database for Screening Mammography (DDSM) database [30], containing 300 image patches with varied sizes (the average size of these image patches is 482×450 pixels). The mammograms in the DDSM database were digitised by one of four different scanners: DBA M2100 ImageClear (42 microns per pixel, 16 bits), Howtek 960 (43.5 microns per pixel, 12 bits), Lumisys 200 Laser (50 microns per pixel, 12 bits), and Howtek MultiRad850 (43.5 microns per pixel, 12 bits). In contrast to the first two datasets, the third dataset contains 25 full-field digital image patches extracted from a non-public mammographic database. These mammograms were acquired using a Hologic Selenia mammography unit, with a resolution of 70 microns per pixel and a depth of 12 bits. The size of these image patches also varies and the average size is 352×301 pixels. In this work, all microcalcifications in each image patch are considered to be part of a single microcalcification cluster. The diagnostic gold standard (benign or malignant) of all microcalcification clusters in this study has been provided by biopsy: there are 9 malignant and 11 benign clusters in the MIAS dataset, 141 malignant and 159 benign clusters in the DDSM dataset, and 14 malignant and 11 benign clusters in the Digital dataset, respectively.

The proposed method works on binary images where 0s stand for "normal" tissue and 1s represent microcalcifications that can be automatically detected by an automatic detection approach or manually annotated by experts. The approach developed by Oliver et al. [16] for automatic detection of microcalcifications is applied to the three datasets (the original work by Oliver et al. [16] showed better results for digital data when compared to digitised data). For the MIAS dataset, in addition to automatic detection, the exact location of individual microcalcifications was manually annotated by an expert (each microcalcification in every image patch was labelled and segmented from the surrounding tissue). The manual annotations and the automatic detection results of the example microcalcification clusters are shown in the second and third columns of Fig. 1, respectively. It appears that the automatic detection approach tends to under-segment individual microcalcifications, such that the pixels close to the boundaries of individual microcalcifications are lost.

III. Methodology

We propose to investigate the potential correlation between the topology of microcalcification clusters and their pathological type. We construct a series of microcalcification graphs to describe the topological structure of microcalcification clusters at different scales. A set of graph theoretical features are extracted from these graphs for modelling and classifying microcalcification clusters. The proposed methodology consists of four main phases: estimating the connectivity between microcalcifications within a cluster using morphological dilation at multiple scales; generating a microcalcification graph at each scale based on the spatial connectivity relationship between microcalcifications; extracting multiscale topological features from these microcalcification graphs; and using the extracted features to build classifier models of malignant and benign microcalcification clusters. The framework of our methodology is summarised in Fig. 2. All image analysis development work was done within MATLAB 7.8.0.

A. Connectivity Estimation Using Morphological Dilation

Morphological dilation [31] is performed on each individual microcalcification using a disk-shaped structuring element at multiple scales. Here, the scale corresponds to the radius of the structuring element measured in pixels. The effect of multiscale morphological dilation on a microcalcification cluster is shown in Fig. 2. It can be seen that the multiscale morphological dilation continuously absorbs neighbouring pixels

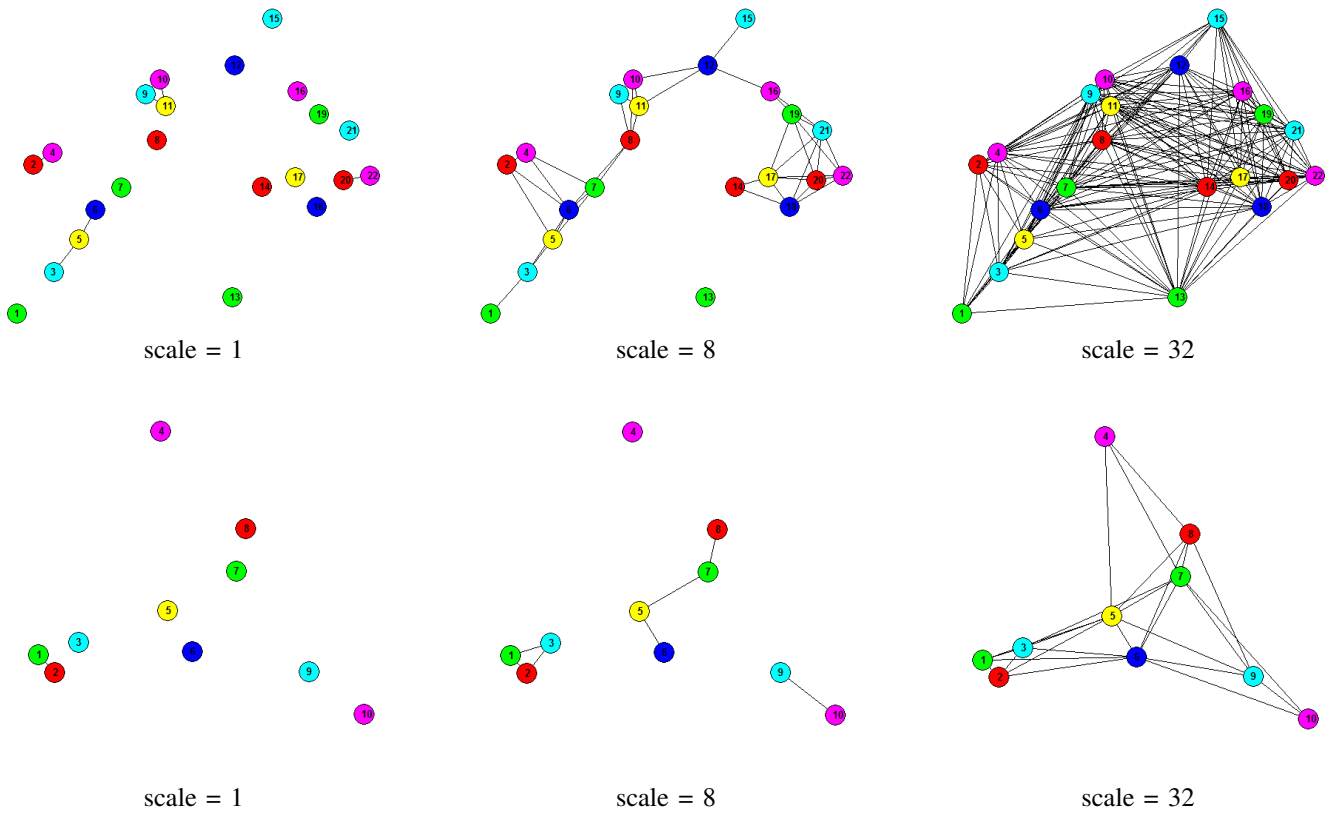


Fig. 3. Microcalcification graphs of the example malignant cluster (top row) and benign cluster (bottom row) in Fig. 1 generated at three scales based on the connectivity of the dilated cluster at these scales. The node colours and sequences are consistent with the corresponding microcalcifications in Fig. 1.

into individual microcalcifications resulting in a change in the connectivity between microcalcifications within the cluster. To illustrate the connectivity of microcalcification clusters with respect to malignant and benign cases, the morphological dilation results of the two example microcalcification clusters are shown in the fourth column of Fig. 1, where the radius of the structuring element is equal to six pixels (i.e. scale = 6). The boundaries of dilated microcalcifications are displayed using different colours and each individual microcalcification is labelled with a sequential number which is ordered according to the spatial location of the corresponding microcalcification in the image patch. As indicated in Section I, the malignant cluster contains a larger number of microcalcifications that are located more closely together within the cluster, while the benign cluster contains fewer microcalcifications that are more diffusely distributed within the cluster.

B. Microcalcification Graph Generation

We propose to represent the topology of microcalcification clusters in graph form. A microcalcification graph is generated based on the spatial connectivity relationship between microcalcifications within a cluster. In a microcalcification graph, each node represents an individual microcalcification, and an edge between two nodes is created if the two corresponding microcalcifications are connected or overlap in the 2D image plane. The resulting microcalcification graphs corresponding to the two example microcalcification clusters in Fig. 1 are shown in Fig. 3. The node locations in these two graphs are in accordance with the original spatial distribution of microcalcifications within the two clusters, and the node sequences are consistent with those in Fig. 1, which are sorted in a left-to-right and bottom-to-top direction (but alternative directions provide the same performance for the subsequent processing). As shown in Fig. 3, the connectivity of the microcalcification cluster increases from small to large scales and the corresponding microcalcification graph becomes denser and denser (more and more edges are created in the graph).

C. Multiscale Topological Feature Extraction

After generating microcalcification graphs over a range of scales, a set of graph theoretical features can be extracted to capture the topological properties of microcalcification clusters. These features will constitute the feature space for the classification of malignant and benign clusters. Before extracting the topological features of microcalcification clusters, we first provide the following definitions for general graphs. Further definitions for graphs can be found in Diestel [32]. Here, we use $G(V, E)$ to represent a graph where V is the vertex set and E is the edge set, and use $|V|$ (the cardinality of V) and $|E|$ (the cardinality of E) to denote the number of vertices and the number of edges in G, respectively. G_{conn} denotes the subgraph of G that corresponds to the largest connected component.

Definition 1. The *adjacency matrix* $A(i, j)$ of a graph

$G(V,E)$ is a $|V| \times |V|$ matrix, defined as:

$$A(i,j) = \begin{cases} 1 & \text{if } (i,j) \in E, \\ 0 & \text{otherwise.} \end{cases} \quad (1)$$

where $(i,j) \in E$ indicates (i,j) is an edge, i.e. there is an edge between vertex i and vertex j in G.

Definition 2. The *degree matrix* $D(i,j)$ of a graph $G(V,E)$ is a $|V| \times |V|$ diagonal matrix containing the degree of vertex i at entry (i,i), defined as:

$$D(i,j) = \begin{cases} d(i) & \text{if } i = j, \\ 0 & \text{otherwise.} \end{cases} \quad (2)$$

where $d(i) = \sum_{j \in V} a_{ij}$ is the number of edges incident to vertex i and $\sum_{i \in V} d(i) = 2|E|$.

Definition 3. The *Laplacian matrix* of a graph $G(V,E)$, denoted by $L(i,j)$, is defined as the difference between the *degree matrix* and the *adjacency matrix*, given by:

$$L(i,j) = D - A = \begin{cases} d(i) & \text{if } i = j, \\ -1 & \text{if } (i,j) \in E, \\ 0 & \text{otherwise.} \end{cases} \quad (3)$$

Definition 4. The *normalised Laplacian matrix* of a graph $G(V,E)$, denoted by $\mathcal{L}(i,j)$, is defined as the normalised version of the *Laplacian matrix* of G, given by:

$$\mathcal{L}(i,j) = \begin{cases} 1 & \text{if } i = j \text{ and } d(i) \neq 0, \\ -\frac{1}{\sqrt{d(i)d(j)}} & \text{if } (i,j) \in E, \\ 0 & \text{otherwise.} \end{cases} \quad (4)$$

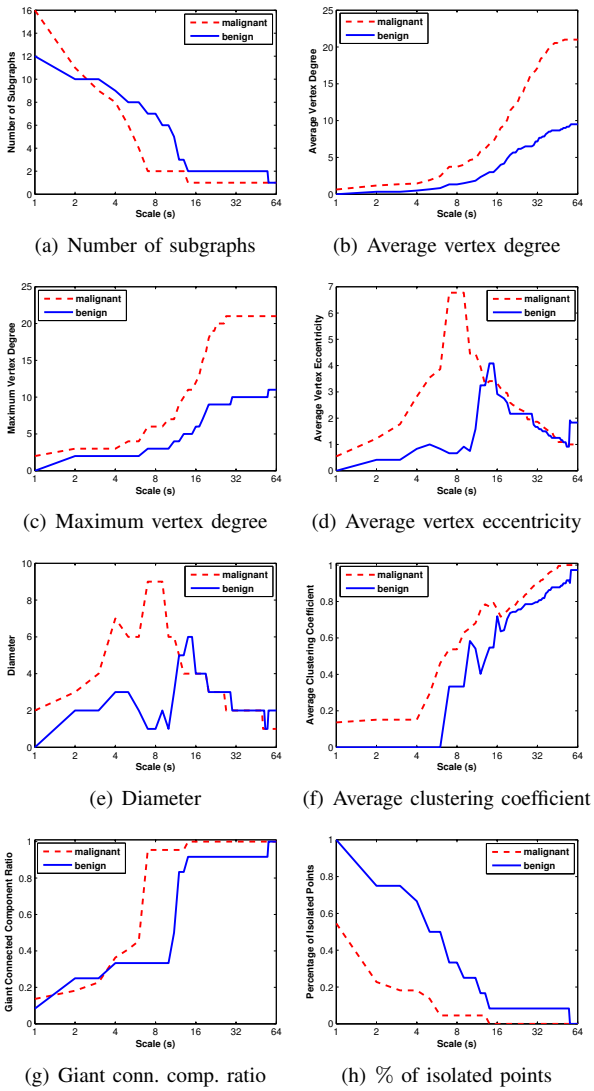
According to the above definitions, the *normalised Laplacian matrix* of G can also be given by:

$$\begin{aligned} \mathcal{L}(i,j) &= D^{-1/2} L D^{-1/2} \\ &= D^{-1/2}(D - A)D^{-1/2} \\ &= I - D^{-1/2} A D^{-1/2} \end{aligned} \quad (5)$$

with the convention that $D^{-1}(i,i) = 0$ if $d(i) = 0$ (i.e. $D(i,i) = 0$), where I is the $|V| \times |V|$ identity matrix. It should be noted that $\mathcal{L}$ is a symmetric positive semi-definite matrix and all eigenvalues of $\mathcal{L}$ are real and non-negative. In addition, it can be seen from Eq. (5) that the eigenvalues of $\mathcal{L}$ are all between 0 and 2, which are closely related to many structural properties for general graphs and play an important role in spectral graph theory [33]. The multiplicity of the eigenvalue 0, denoted for ease of notation as k, corresponds to the number of connected components in the graph, and the multiplicity of the eigenvalue 2 coincides with the number of bipartite connected components in the graph.

Definition 5. The *distance* between two vertices i and j in a graph $G(V,E)$, denoted by $dist(i,j)$, is defined as the length of the shortest path between i and j, equal to the minimum number of edges between them.

Definition 6. The *eccentricity* of a vertex i in a graph $G(V,E)$, denoted by $e(i)$, is defined as the maximum *distance* from itself to any of the reachable vetices in G, given by $e(i) = \max_{j \in V} dist(i,j)$.



(a) Number of subgraphs (b) Average vertex degree

(c) Maximum vertex degree (d) Average vertex eccentricity

(e) Diameter (f) Average clustering coefficient

(g) Giant conn. comp. ratio (h) % of isolated points

Fig. 4. The resulting eight graph feature sets extracted from the example malignant and benign microcalcification clusters in Fig. 1.

Definition 7. The *clustering coefficient* of a vertex i in a graph $G(V,E)$, denoted by $c(i)$, is the ratio of the number of actually existing edges between i's neighbours (vertices adjacent to i) and the number of all possible edges between i's neighbours, given by $c(i) = E(i)/C_k^2 = 2E(i)/(k(k-1))$, where $E(i)$ is the number of actually existing edges between i's neighbours, and k is the number of i's neighbours.

Definition 8. A vertex is considered *isolated* if it has degree equal to 0. The set X denotes the set of vertices within a graph $G(V,E)$ that are isolated, that is, $\sum_{i \in X} d(i) = 0$ and $\sum_{i \in (G \setminus X)} d(i) > 0$. Therefore, $|X|$ is equal to the total number of isolated vertices within G.

Following these definitions, we explain a set of graph metrics in Table I that will be extracted from the generated microcalcification graphs and concatenated into the feature set for the subsequent classification process.

We construct a set of microcalcification graphs based on the

TABLE I
GRAPH METRICS INVESTIGATED FOR MALIGNANCY ANALYSIS OF MICROCALCIFICATION CLUSTERS ALONG WITH THEIR DEFINITIONS AND CLINICAL INTERPRETATIONS

No.	Metric	Definition	Clinical Interpretation
1	Number of Subgraphs	k (See Def. 4)	Malignant clusters tend to have a higher number of microcalcifications and therefore their corresponding graphs contain a higher number of subgraphs at smaller scales
2	Average Vertex Degree	$\sum_{i\in V} d(i)/\|V\|$	Malignant microcalcifications appear to be more densely distributed within a cluster and therefore malignant clusters generally have a larger average vertex degree
3	Maximum Vertex Degree	$\max_{i\in V} d(i)$	Malignant microcalcifications appear to be more connected in a cluster and therefore malignant clusters tend to have a larger maximum vertex degree
4	Average Vertex Eccentricity	$\sum_{i\in V} e(i)/\|V\|$	Malignant clusters tend to have a linear distribution and therefore have larger eccentricity values
5	Diameter	$\max_{i\in V} e(i)$	Malignant clusters tend to have a linear topology and therefore have longer diameters
6	Average Clustering Coefficient	$\sum_{i\in V} c(i)/\|V\|$	Malignant microcalcifications tend to be more connected in a cluster and therefore malignant clusters have a higher average clustering coefficient
7	Giant Connected Component Ratio	$\frac{\|V\in G\|}{\|V\in G_{\text{conn}}\|}$	Malignant microcalcifications tend to be more closely grouped and therefore malignant clusters have a higher giant connected component ratio
8	Percentage of Isolated Points	$\|X\|/\|V\|$	Malignant microcalcifications tend to be more densely distributed and therefore malignant clusters tend to have a smaller percentage of isolated points

spatial connectivity relationship between microcalcifications after performing morphological dilation at multiple scales, denoted by $\mathbb{G} = G_0, G_1, \ldots, G_{S-1}$, where S is the number of scales, and $G_s (s = 0, 1, \ldots, S-1)$ denotes the microcalcification graph generated at the s^{th} scale (the 0^{th} scale corresponds to the initial microcalcification cluster without morphological dilation). We extract the eight graph metrics from each graph in $\mathbb{G}$, which produces eight graph feature sets covering S scales. We then concatenate the eight feature sets to create a feature vector, termed the multiscale topological feature vector in this paper, representing the topological characteristics of microcalcification clusters over multiple scales.

The resulting eight graph feature sets for the example malignant and benign microcalcification clusters in Fig. 1 are shown in Fig. 4, where the graph metrics are extracted from the microcalcification graphs generated at 65 scales, i.e. $\mathbb{G} = G_0, G_1, \ldots, G_{64}$ ($S = 65$). It can be seen from Fig. 4(a) that the number of subgraphs corresponding to the malignant cluster is larger than the benign cluster at the first few scales, while it decreases more drastically as the scale increases due to the fact that malignant microcalcifications are more densely distributed. When the scale increases to a certain value, the number of subgraphs remains stable and further decreases to 1 when all microcalcifications in the cluster get connected after morphological dilation. As shown in Fig. 4(b), the average vertex degree goes up continuously as the scale increases. The maximum average vertex degree is achieved when a complete microcalcification graph is formed, in which case all microcalcifications within the cluster are connected with each other. Moreover, it is shown in Fig. 4(b) that the average vertex degree values of the malignant cluster are larger than those of the benign cluster over the entire range of scales, which indicates that the malignant microcalcification cluster is more connected. Fig. 4(c) shows a set of values of the maximum vertex degree against scale which also have an increasing trend from small to large scales and tend towards stability when reaching the maximum value. Similarly, as indicated by the average vertex degree, the maximum vertex degree values for the malignant cluster are also larger than those of the benign cluster. The resulting values of the average vertex eccentricity against scale are plotted in Fig. 4(d). At the first few scales, most microcalcifications are isolated from others in the cluster, which results in small average eccentricity values (the eccentricity of isolated vertices is set to 0). When the scale increases to a specific value, the maximum average eccentricity is obtained, in which case the previously isolated microcalcifications are absorbed into a connected component with a relatively large diameter. After that, as the scale further increases, more and more microcalcifications get connected and the average vertex eccentricity starts to decrease. When all microcalcifications in the cluster are connected with each other, the average eccentricity is reduced to 1. Fig. 4(e) shows how the diameter (the maximum vertex eccentricity) of the malignant and benign clusters changes against scale, which is similar to that of the average vertex eccentricity in Fig. 4(d). In the beginning, the diameter value increases with the scale until it reaches the maximum value. After that, the diameter value gradually goes down towards the minimum value of 1 when all microcalcifications are connected. Note that the maximum diameter of the malignant cluster is larger than that of the

benign cluster. As indicated above, microcalcifications in the malignant cluster tend to present a linear topology and as such form a connected component having a longer diameter. For the resulting average clustering coefficients of the two clusters, as shown in Fig. 4(f), the malignant cluster obtains larger values at all scales than the benign cluster. Fig. 4(g) presents how the giant connected component ratio varies with scale. As the scale increases, more and more microcalcifications in the cluster are absorbed into the giant connected component until all of them are included. Thus, the resulting giant connected component ratio continuously increases until it goes up to the maximum value of 1. Note that the giant connected component ratio of the malignant cluster reaches its maximum at a much smaller scale than the benign cluster. The eighth feature set composed of the percentage of isolated points is provided in Fig. 4(h). In contrast to the giant connected component ratio, the percentage of isolated points decreases as the scale increases, which is reduced to 0 when all microcalcifications are linked together. The values for the malignant cluster are smaller than those of the benign cluster, and moreover the malignant cluster achieves 0 percentage at a much smaller scale than the benign cluster. These all indicate that the malignant cluster is more densely distributed and therefore generates a more connected microcalcification graph than the benign cluster at a specific scale.

D. Classification of Microcalcification Clusters

Four k-Nearest Neighbours (kNN) based classifiers are used for classifying microcalcification clusters into malignant and benign. The classical kNN classifier [34] is a popular and conceptually intuitive instance-based learning approach. A number of alternatives are employed which attempt to address some inherent shortcomings of the classical kNN. Fuzzy Nearest Neighbours (FNN) [35] extends the classical kNN by fuzzifying the memberships for test and training objects. Fuzzy Rough Nearest Neighbours (FRNN) [36], [37] models two different types of uncertainty: fuzziness and indiscernibility. Vaguely Quantified Nearest Neighbours (VQNN) [38] incorporates the uncertainty modelling of FRNN and also employs vague quantifiers which limit the influence that noisy data might have on the classification outcomes. These approaches offer further flexibility, improved generalisation, and retain human interpretability when compared to techniques such as ANN and SVM. In addition, it should be noted that the classical kNN is employed for the classification task such that the proposed method can be easily compared with existing work in the literature.

The kNN classification is based on a simple majority vote, unless equal class probability is indicated, in which case a Euclidean weighted approach is used. The default fuzzifier value of $m = 3.0$ is used for FNN. FRNN is stable with respect to the value of k and returns similar results but slightly different models. VQNN on the other hand results in different models when the value of k is altered. A range of values for k are employed when generating the experimental results which are documented in the following section.

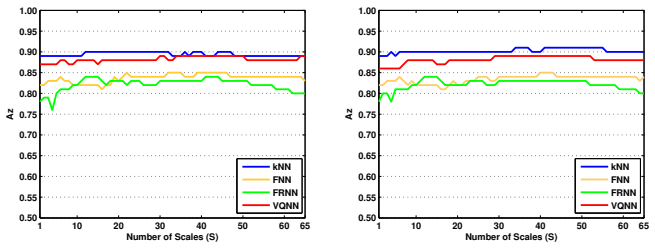
IV. Experimental Evaluation

A. Experimental Set-up

To evaluate the performance of the classifier models built using the multiscale topological feature vectors, a leave-one-out cross-validation (LOOCV) scheme was employed for all datasets, and an additional stratified 10 runs 10-fold cross-validation (10-FCV) scheme was employed for the DDSM dataset to investigate how significantly these two cross-validation schemes may affect the performance. Two evaluation metrics were used for this work. The first was overall classification accuracy (CA), which is defined as the percentage of microcalcification clusters correctly classified, providing a summary of the performance for balanced datasets (such as the datasets used here). ROC analysis was used as the second evaluation approach. An ROC curve is a plot of the true positive rate (TPR) against the false positive rate (FPR), which describes the whole range of possible operating characteristics for a binary classifier model. Here, TPR is defined as the number of correctly classified malignant cases divided by the total number of malignant cases, and FPR is defined as the number of benign cases incorrectly classified as malignant divided by the total number of benign cases. ROC analysis can be employed in order to assess the predictive ability of a classifier by using the area under the ROC curve denoted by A_z [39]. A_z is a statistically consistent measure and is equivalent to the Wilcoxon signed-ranks test, which is a non-parametric alternative to the paired t-test [40], [41]. All of the classification and evaluation aspects were completed using the WEKA data mining suite [42].

Moreover, to provide a comparison between the classification results based on manually and automatically segmented microcalcifications (and also to investigate the robustness of the proposed method to microcalcification segmentation variations), it was tested using both manual annotations and automatic detection results for the MIAS dataset.

In addition, to evaluate the stability of the proposed approach with respect to the size of the dataset, a number of subsets were randomly selected from each dataset for cross-validation. Specifically, for the MIAS dataset, two groups of random subsets were selected, consisting of 10 (5 malignant and 5 benign) and 15 (7 malignant and 8 benign) cases, respectively. For the Digital dataset, three groups of random subsets were selected, consisting of 10 (5 malignant and 5 benign), 15 (7 malignant and 8 benign) and 20 (9 malignant and 11 benign) cases, respectively. For the DDSM dataset, six groups of random subsets were selected, consisting of 10 (5 malignant and 5 benign), 15 (7 malignant and 8 benign), 20 (9 malignant and 11 benign), 40 (18 malignant and 22 benign), 80 (36 malignant and 44 benign) and 160 (72 malignant and 88 benign) cases, respectively. Each random selection was repeated 100 times, which produced 100 random subsets of each size for each dataset. The means, standard deviations, maximum and minimum values of CA and A_z were statistically analysed over each group of 100 random subsets, which are provided in the following section.



(a) A_z for DDSM (LOOCV) (b) A_z for DDSM (10-FCV)

Fig. 5. The A_z values produced by the four classifiers for the DDSM dataset using leave-one-out (a) and 10-fold cross-validation (b).

TABLE II
THE BEST CLASSIFICATION RESULTS OVER 65 SCALES. FOR 10-FCV, THE RESULTS CONTAIN MEANS AND STANDARD DEVIATIONS RESULTING FROM 100 CLASSIFIER MODELS (10 FOLDS × 10 RUNS).

(a) CA (%)

Test Data	kNN	FNN	FRNN	VQNN
MIAS (manual)	95 ($S = 10$)	95 ($S = 10$)	95 ($S = 14$)	95 ($S = 26$)
MIAS (automatic)	95 ($S = 5$)	95 ($S = 5$)	95 ($S = 8$)	95 ($S = 8$)
Digital	96 ($S = 10$)	96 ($S = 10$)	88 ($S = 24$)	88 ($S = 15$)
DDSM (LOOCV)	86.0 ($S = 40$)	85.7 ($S = 40$)	78.0($S = 12$)	84.0($S = 44$)
DDSM (10-FCV)	85.2 ± 5.7 ($S = 41$)	85.1 ± 6.5 ($S = 41$)	77.8 ± 6.8 ($S = 12$)	83.8 ± 6.1 ($S = 50$)

(b) A_z

Test Data	kNN	FNN	FRNN	VQNN
MIAS (manual)	0.96 ($S = 10$)	0.96 ($S = 10$)	0.96 ($S = 14$)	0.96 ($S = 26$)
MIAS (automatic)	0.96 ($S = 5$)	0.96 ($S = 5$)	0.96 ($S = 8$)	0.96 ($S = 8$)
Digital	0.96 ($S = 10$)	0.96 ($S = 10$)	0.90 ($S = 24$)	0.92 ($S = 15$)
DDSM (LOOCV)	0.90 ($S = 40$)	0.85 ($S = 40$)	0.84 ($S = 12$)	0.89 ($S = 44$)
DDSM (10-FCV)	0.91 ± 0.05 ($S = 41$)	0.85 ± 0.07 ($S = 41$)	0.84 ± 0.07 ($S = 12$)	0.89 ± 0.06 ($S = 50$)

B. Experimental Results

We have used two digitised and one full-field digital datasets (see Section II for details) for evaluating the performance of the proposed approach in discriminating between malignant and benign microcalcification clusters. We have investigated a range of values for S which determines the dimensionality of the feature space. As described in Section III-C, the multiscale topological feature vectors are extracted from a set of microcalcification graphs generated at scales $s = 0, 1, \ldots, S-1$, which are composed of eight graph feature sets. Thus, the dimensionality of the multiscale topological feature space is equal to $8 \times S$. The largest scale used in the experiments was set to 65, and therefore the dimensionality of the feature space was up to 520. In addition, we have used a range of values for k which determines the number of the nearest neighbours used to build the classifier models. Fig. 5 shows the results for a range of scales (S) defining the feature space for the DDSM dataset. As can be seen, the results are stable over a range of different scales. A_z results as a function of the k value show a similar stability over a range of $k = [1 \ldots 10]$ for the MIAS and Digital datasets, and $k = [12 \ldots 30]$ for the DDSM dataset. For brevity, detailed results of this are left out of the paper. Table II shows the best classification results achieved by the four classifiers over 65 scales, including CA (Table II(a)) and A_z (Table II(b)). The classification accuracy is given at the scale maximum scoring A_z value.

For the MIAS dataset, when using the manual annotations, the best CA was 95% with one benign case misclassified and the largest A_z was 0.96, produced by all the four classifiers; when using the automatic detection results, the best CA was also 95% with the same benign case misclassified, and the largest A_z of 0.96 was also obtained by all the four classifiers. For the Digital dataset, the best CA of 96% with one malignant case misclassified and the best A_z of 0.96 were achieved by kNN and FNN. For the DDSM dataset, when using leave-one-out cross-validation, kNN obtained the best CA of 86% and the largest A_z of 0.90; for 10-fold cross-validation, kNN also indicated the best performance, the CA and A_z were $85.2 \pm 5.7\%$ and 0.91 ± 0.05, respectively (standard deviations were calculated across 100 classifier models (10 folds × 10 runs)).

As described in Section IV-A, to evaluate the stability of the proposed approach with respect to the size of the dataset, a set of subsets consisting of 10, 15, 20, 40, 80, 160 cases were randomly sampled from the MIAS, DDSM and Digital datasets. 100 random subsets were generated for each number of cases. The means, standard deviations, maximum and minimum values of CA and A_z calculated over each group of 100 random subsets are provided in Table III, where the results were generated using kNN and leave-one-out cross-validation. As shown in Table III, with regard to each dataset, the random subsets containing a small number of cases produced slightly worse results with a larger standard deviation, however, the standard deviations of CA and A_z were reduced as the number of cases in the subsets was increased. Note that the random subsets containing the largest number of cases selected from the three datasets (15, 20 and 160) achieved very similar results to those based on the whole datasets.

In addition, we investigated redundancy among the defined topological feature set and explored the graph metrics which contributed most to malignancy analysis of microcalcification clusters. We performed feature selection by employing the CfsSubsetEval attribute evaluator and the GreedyStepwise search method in Weka. The CfsSubsetEval attribute evaluator evaluates the importance of a subset of features by estimating the individual predictive ability of each feature as well as the extent of redundancy between them, and as such features that are highly correlated with the class while have low intercorrelation are more likely to be selected (see [43] for more information). The GreedyStepwise search method performs a greedy forward or backward search through the feature space,

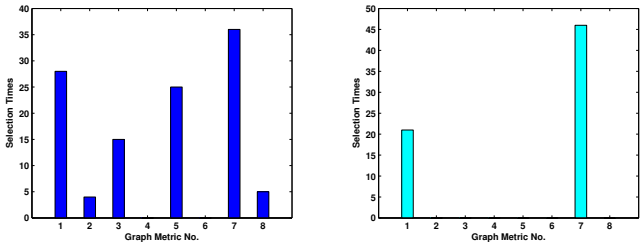
TABLE III
STATISTICAL ANALYSIS OF CA AND A_z OVER 100 RANDOM SUBSETS OF DIFFERENT NUMBERS OF CASES SELECTED FROM THE MIAS, DIGITAL AND DDSM DATASETS.

(a) CA(%)

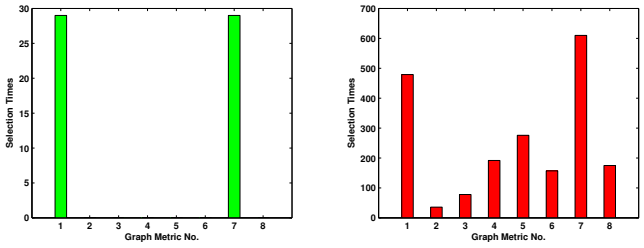
	MIAS		Digital			DDSM					
	10	15	10	15	20	10	15	20	40	80	160
Mean	93.0	95.1	93.1	94.3	96.1	88.3	90.1	89.1	88.1	87.2	86.5
Std.	6.5	4.1	7.2	4.6	3.7	9.2	7.9	7.0	4.6	3.4	2.0
Max.	100	100	100	100	100	100	100	100	97.5	95	91.3
Min.	80.0	86.7	80.0	86.7	90.0	70.0	70.0	75.0	75.0	78.8	82.5

(b) A_z

	MIAS		Digital			DDSM					
	10	15	10	15	20	10	15	20	40	80	160
Mean	0.93	0.95	0.93	0.95	0.96	0.88	0.92	0.91	0.91	0.90	0.90
Std.	0.07	0.04	0.08	0.04	0.04	0.11	0.08	0.07	0.05	0.03	0.02
Max.	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	0.96	0.95
Min.	0.74	0.88	0.76	0.87	0.90	0.62	0.67	0.70	0.73	0.80	0.86



(a) MIAS (manual annotation) $S = 10$ (b) MIAS (automatic detection) $S = 5$



(c) Digital $S = 10$ (d) DDSM $S = 40$

Fig. 6. The overview of the selection times of the eight graph metrics for the three datasets. The graph metrics from No. 1 to 8 correspond to (1) number of subgraphs, (2) average vertex degree, (3) maximum vertex degree, (4) average vertex eccentricity, (5) diameter, (6) average clustering coefficient, (7) giant connected component ratio, and (8) percentage of isolated points (see Table I).

which starts with no or all features and terminates when any addition or reduction of the currently selected feature subset results in a decrease in evaluation [44]. We used leave-one-out cross-validation for MIAS and Digital, while for DDSM we used 10 runs stratified 10-fold cross-validation. It should be noted that the feature selection is only performed on the training data and therefore it cannot overfit since there is no bias. If we did not use cross-validation then we could run into the risk of overtraining, but even with leave-one-out cross-validation the test set remains uncorrelated with the training data. The number of the resulting feature subsets was 20, 25 and 100 for MIAS, Digital and DDSM, respectively. Fig. 6 illustrates how many times each of the eight graph metrics were taken after feature selection. As can be seen from Fig. 6, graph metrics No. 1 and 7, i.e. number of subgraphs and giant connected component ratio, seemed to be the two most important graph metrics among the eight, which were most frequently selected from the multiscale topological feature set. As indicated in Table I, these two graph metrics are mainly related to the number/distribution of microcalcifications within a cluster. Their precise clinical significance for the diagnosis of malignant and benign microcalcification clusters could be further investigated.

The classification results for the three datasets after feature selection are provided in Table IV. Here, the results were generated by the kNN classifier using the resulting feature subsets, which were slightly lower when compared to those results obtained before feature selection (see Table II).

V. Discussion

As described above, good classification results have been obtained for all the three datasets. The Digital dataset provided the best results, which might be due to the more accurate detection of microcalcifications using digital mammography. As stated in Oliver et al. [16], the detection approach indicates the best performance when using the Digital dataset, and therefore more realistic detection results of microcalcifications can be provided for the classification task. The MIAS dataset produced the second best classification results, and moreover using manual annotations and automatic detections achieved the same performance. This indicates that the proposed method seems to be robust with respect to variations between manual and automatic segmentations of microcalcifications. For the DDSM dataset, very similar results were shown when using the leave-one-out and 10-fold cross-validation methods, showing a decreased performance in the results when compared to the other datasets. It might be partially explained by the fact that the detection approach performs worst for the DDSM database among the three datasets [16]. Moreover, the DDSM dataset used in our experiments contains 300 cases, which is expected to give a larger variability than the small datasets (especially as the DDSM dataset was generated using different digitisers). However, the obtained classification results are still comparable or even better than the related work reviewed by Cheng et al. [9] or Table V, where most publications used smaller databases than ours.

TABLE IV
The classification results for the three datasets after feature selection.

Test Data	Cross-Validation	CA	A_z
MIAS (manual)	leave-one-out	90%	0.91
MIAS (automatic)	leave-one-out	90%	0.93
Digital	leave-one-out	96%	0.94
DDSM	10-fold	$83.9 \pm 6.3\%$	0.90 ± 0.05

We compared the proposed method with related publications in the literature. Table V shows a summary of the comparison. Note that the various approaches use different images taken from different databases, and therefore this is a qualitative comparison. In Shen et al. [7], the 100% CA was obtained by classifying 143 individual microcalcifications from 18 biopsy proven cases based on a leave-one-microcalcification-out approach, which is different from the goal of our classification of microcalcification clusters. In Ma et al. [10], the classification of microcalcification clusters was based on the maximum feature value obtained by a selected single microcalcification rather than the whole cluster (and therefore some manual aspects were involved in the extraction process). In Ren et al. [46], the high classification performance was obtained by introducing an optimised decision making step which was performed afterwards through statistical analysis of the classifier outputs to achieve the minimum cost of error classification. As shown in Table V, the obtained classification results are comparable to the various approaches.

In addition, in order to enable a direct comparison, we extracted the features used in previously published works that showed the most promising performance [7], [10], [46], [47] and performed malignancy analysis of microcalcification clusters on our datasets. Table VI shows a summary of the

TABLE V
A QUALITATIVE COMPARISON OF OUR RESULTS WITH THOSE ACHIEVED BY RELATED WORK.

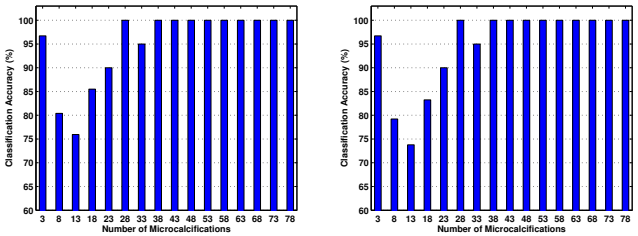
Method	Database	Cases	Feature	Classifier	Result
Shen et al. [7]	unknown	18	shape	kNN	CA = 100%
Ma et al. [10]	DDSM	183	shape	thresholding	$A_z = 0.96$
Chan et al. (1998) [21]	unknown	145	morphology	LDC	$A_z = 0.79$
Dhawan et al. [8]	unknown	191	texture&cluster	ANN	$A_z = 0.86$
Papadopoulos et al. [15]	MIAS	25	cluster	SVM	$A_z = 0.81$
Chan et al. (1997) [27]	unknown	54	texture	ANN	$A_z = 0.88$
Soltanian-Zadeh et al. [11]	Nijmegen	103	multiwavelet	kNN	$A_z = 0.89$
Betal et al [20]	Liverpool	38	shape/cluster	kNN	$A_z = 0.79, A_z = 0.84$
Rana et al. [23]	University of Chicago Hospitals	49	morphology	ANN	$A_z = 0.80$
Wei et al. [45]	University of Chicago	104	cluster&morphology	Ada-/Cas-SVM	$A_z = 0.81, A_z = 0.82$
Shao et al. [25]	Sun Yat-sen University	109	pattern factor	quantising	$A_z = 0.74$
Ren et al. (2011) [46]	DDSM	150	varied features	ANN	$A_z = 0.98$
Ren (2012) [47]	DDSM	150	varied features	ANN/SVM	$A_z = 0.93, A_z = 0.94$
Strange et al. [24]	DDSM	150	cluster	barcodes	$A_z = 0.82$
Strange et al. [24]	MIAS	20	cluster	barcodes	$A_z = 0.80$
Ours	MIAS I (manual annotation)	20	topology	kNN/FNN/FRNN/VQNN	CA = 95%, $A_z = 0.96$
Ours	MIAS I (automatic detection)	20	topology	kNN/FNN/FRNN/VQNN	CA = 95%, $A_z = 0.96$
Ours	Digital	25	topology	kNN/FNN	CA = 96%, $A_z = 0.96$
Ours	DDSM (leave-one-out CV)	300	topology	kNN	CA = 86.0%, $A_z = 0.90$
Ours	DDSM (10-fold CV)	300	topology	kNN	CA = 85.2 ± 5.7%, $A_z = 0.91 \pm 0.05$

TABLE VI
A DIRECT COMPARISON OF THE RESULTING CA AND A_z VALUES OF OUR APPROACH AND SHEN ET AL. [7], MA ET AL. [10], AND REN ET AL. [46], [47].

Test Data	Ours	Shen et al. [7]	Ma et al. [10]	Ren et al. [46], [47]
MIAS (manual annotation)	CA = 95%, $A_z = 0.96$	CA = 70%, $A_z = 0.68$	CA = 80%, $A_z = 0.76$	CA = 85%, $A_z = 0.91$
Digital	CA = 96%, $A_z = 0.96$	CA = 84%, $A_z = 0.71$	CA = 72%, $A_z = 0.68$	CA = 85%, $A_z = 0.91$
DDSM	CA = 86%, $A_z = 0.90$	CA = 73%, $A_z = 0.69$	CA = 62%, $A_z = 0.56$	CA = 82%, $A_z = 0.86$

best CA and A_z values achieved using our proposed topology based feature set and the other three feature sets [7], [10], [46], [47], where the results were all produced using kNN and leave-one-out cross-validation. As shown in Table VI, our proposed approach performed best among the four approaches and achieved the best CA and A_z results for all the three datasets. In addition, our proposed approach provided a significant improvement over the existing methods and when the results from Table VI were compared using an unpaired t-test a p value of $p < 0.01$ was obtained in all cases.

One inherent limitation of the developed method is that it cannot provide a reliable classification for the case where the cluster is structureless or few microcalcifications are segmented within the cluster. An extreme is when only a single microcalcification is detected from the cluster by the automatic detection approach, it will fail to discriminate malignant from benign based on the topology. Another concern of the proposed method is that its performance might depend on the performance of the microcalcification detection approach. False negatives or false positives may affect the global topology/connectivity of microcalcification clusters. However, the experimental results demonstrate the robustness and effectiveness of the developed method when combined with automatic microcalcification detection.



(a) DDSM (LOOCV) (b) DDSM (10-FCV)

Fig. 7. Graph showing the classification accuracy as a function of the number of microclacifications. The results are reported for the DDSM dataset using leave-one-out cross-validation (a) and 10-fold cross-validation (b) with the kNN classifier and $S = 40$ for leave-one-out cross-validation and $S = 41$ for 10-fold cross-validation.

In the experiments, the clusters with fewer microcalcifications being segmented by the automatic detection approach tend to be classified as benign since their graph metrics are more in line with those of a benign cluster. Thus, the under-detected malignant cases where the microcalcifications indicate a sparse distribution could be misclassified into benign. On the other hand, the benign cases with relatively a larger number of microcalcifications (including the false positives) being segmented from the clusters could indicate a malignant-like distribution and as such could be misclassified into malignant. Fig. 7 shows the classification accuracy as a function of the number of microcalcifications for the DDSM dataset using leave-one-out and 10-fold cross-validation. The results were obtained using the kNN classifier with $S = 40$ for leave-one-out cross-validation and $S = 41$ for 10-fold cross-validation since they correspond to the best performing scale for this classifier as displayed in Table II. For display purposes the results shown in Fig. 7 were averaged such that each bar represents the mean classification accuracy over 5 sizes of microcalcification clusters. The results show that there is a slight dip in classification accuracy when the number of microcalcifications falls in the $11 - 15$ range. This experiment was not replicated with the MIAS and Digital dataset due to their small sample size and good classification accuracy.

Due to the fact that only a few samples are misclassified, the sample size would not be big enough to show any relationship between the number of microcalcifications and the classification accuracy.

As discussed above, some highlights of this work should be noted. For the MIAS dataset, the same methodology was applied based on both manually annotated and automatically detected microcalcifications, and the same classification performance was obtained. This indicates the robustness of the proposed method to detection errors. For the DDSM dataset, we used a larger set of cases than related publications and achieved good results. In addition to these two well-known digitised databases, we evaluated our method using a full-field digital database and obtained improved classification results. This demonstrates the capability of our method in dealing with two categories of mammograms, which allows it to be applied in both film and digital mammography. We also investigated the stability of the proposed method against the size of the evaluation dataset. For each dataset, no significant difference in the classification performance was shown among the subsets of varied numbers of cases. Furthermore, we investigated the most significant microcalcification graph metrics for malignancy analysis by performing feature selection. The most frequently selected graph metrics are worth further investigation from a clinical point of view. In addition to a qualitative comparison with related publications, we implemented a direct comparison between the proposed approach and three state-of-the-art approaches, and our method demonstrated the best performance for all the three datasets used in this work. In addition, we used the CAD detection results directly instead of manual segmentation results for all the three datasets. High classification accuracies and good ROC analysis results were obtained when compared to the state-of-the-art approaches. This indicates its potential application in conjunction with automatic detection approaches in CAD systems.

As future work, other features such as shape and texture of individual microcalcifications and the whole cluster could be incorporated to build a complete framework for malignancy analysis of microcalcification clusters. A similarity measure between microcalcfication graphs can be investigated in order to discriminate between malignant and benign clusters using the graph based representation directly without generating graph feature vectors. On the other hand, alternative classifiers (e.g. random forests, ANN, and SVM) could also be investigated. In addition, we will extend the evaluation using a larger collection of digital mammograms.

VI. Conclusions

We have presented a method for classifying microcalcification clusters in mammograms based on morphological topology analysis. This is a novel approach to analyse microcalcifications in terms of the connectivity and topology for discriminating malignant from benign clusters. Unlike most features (e.g. shape/morphological features) in previous publications extracted at a single scale, a representation of microcalcification clusters covering the multiscale characteristics was developed in this paper. The topology/connectivity of microcalcification clusters was analysed using multiscale morphology. A set of microcalcification graphs were constructed to describe the topological structure of microcalcification clusters at multiple scales. When analysing the topology of microcalcification clusters, we extracted eight graph metrics from microcalcification graphs generated at multiple scales, which are number of subgraphs, average vertex degree, maximum vertex degree, average vertex eccentricity, diameter, average clustering coefficient, giant connected component ratio, and percentage of isolated points. The resulting eight graph feature sets were aggregated and constituted the multiscale topological feature vector, which has been used to classify microcalcification clusters into malignant and benign.

The proposed method has been evaluated using three datasets: MIAS, DDSM and Digital. Four k-Nearest Neighbours based algorithms (kNN, FNN, FRNN and VQNN) have been used for the classification task. Good classification results have been obtained for all the datasets. By investigating a set of S values for the number of scales and using a range of k values for the classifier, the obtained best classification accuracy was 95% for MIAS with manual annotations, 95% for MIAS with automatic detections, 96% for Digital, 86% for DDSM using leave-one-out cross validation, and $85.2 \pm 5.7\%$ for DDSM based on 10-fold cross-validation; and the largest area under the ROC curve was 0.96, 0.96, 0.96, 0.90 and 0.91 ± 0.05, respectively.

Acknowledgements

The authors would like to thank the anonymous reviewers for their comments and insight which helped to shape the final version of this paper. We would also like to thank Dr Neil Mac Parthaláin for his insight and discussions on cross-validation and feature selection. This work was funded in part by the NISCHR Biomedical Research Unit for Advanced Medical Imaging and Visualization.

References

[1] American Cancer Society, *Global Cancer Facts & Figures*, 2nd ed. Atlanta: American Cancer Society, 2011, pp. 11–12.

[2] Eurostat, *Health Statistics Atlas on Mortality in the European Union*. Office for Official Publications of the European Union, 2009, pp. 91–93.

[3] American Cancer Society, *Breast Cancer Facts & Figures 2011–2012*. Atlanta: American Cancer Society, 2011, pp. 11–15.

[4] L. Tabár, M. F. Yen, B. Vitak, H. H. T. Chen, R. A. Smith, and S. W. Duffy, "Mammography service screening and mortality in breast cancer patients: 20-year follow-up before and after introduction of screening," *The Lancet*, vol. 361, no. 9367, pp. 1405–1410, 2003.

[5] National Comprehensive Cancer Network, "NCCN Clinical Practice Guidelines in Oncology: Breast Cancer Screening and Diagnosis," 2012.

[6] Cancer Research UK, "Who is screened for breast cancer?" August 2012, http://www.cancerresearchuk.org/cancer-help/type/breast-cancer/about/screening/who-is-screened-for-breast-cancer.

[7] L. Shen, R. M. Rangayyan, and J. E. L. Desautels, "Application of shape analysis to mammographic calcifications," *IEEE Transactions on Medical Imaging*, vol. 13, no. 2, pp. 263–274, 1994.

[8] A. P. Dhawan, Y. Chitre, and C. Kaiser-Bonasso, "Analysis of mammographic microcalcifications using gray-level image structure features," *IEEE Transactions on Medical Imaging*, vol. 15, no. 3, pp. 246–259, 1996.

[9] H. D. Cheng, X. Cai, X. Chen, L. Hu, and X. Lou, "Computer-aided detection and classification of microcalcifications in mammograms: A survey," *Pattern Recognition*, vol. 36, no. 12, pp. 2967–2991, 2003.

[10] Y. Ma, P. C. Tay, R. D. Adams, and J. Z. Zhang, "A novel shape feature to classify microcalcifications," in *Proc. 17th IEEE International Conference on Image Processing*, 2010, pp. 2265–2268.

[11] H. Soltanian-Zadeh, F. Rafiee-Rad, and D. S. Pourabdollah-Nejad, "Comparison of multiwavelet, wavelet, haralick, and shape features for microcalcification classification in mammograms," *Pattern Recognition*, vol. 37, no. 10, pp. 1973–1986, 2004.

[12] E. A. Sickles, "Breast calcifications: Mammographic evaluation," *Radiology*, vol. 160, no. 2, pp. 289–293, 1986.

[13] W. Dähnert, *Radiology Review Manual*, 7th ed. Philadelphia: Lippincott Williams & Wilkins, 2011, pp. 557–558.

[14] J. Suckling, J. Parker, D. R. Dance, S. Astley, I. Hutt, C. Boggis, I. Ricketts, E. Stamatakis, N. Cerneaz, S. L. Kok, P. Taylor, D. Betal, and J. Savage, "The mammographic image analysis society digital mammogram database," in *Excerpta Medica. International Congress Series 1069*, 1994, pp. 375–378.

[15] A. Papadopoulos, D. I. Fotiadis, and A. Likas, "Characterization of clustered microcalcifications in digitized mammograms using neural networks and support vector machines," *Artificial Intelligence in Medicine*, vol. 34, no. 2, pp. 141–150, 2005.

[16] A. Oliver, A. Torrent, X. Lladó, M. Tortajada, L. Tortajada, M. Sentís, J. Freixenet, and R. Zwiggelaar, "Automatic microcalcification and cluster detection for digital and digitised mammograms," *Knowledge-Based Systems*, vol. 28, pp. 68–75, 2012.

[17] M. Elter and A. Horsch, "CADx of mammographic masses and clustered microcalcifications: A review," *Medical Physics*, vol. 36, no. 6, pp. 2052–2068, 2009.

[18] M. Lanyi, "Morphologic analysis of microcalcifications," in *Early Breast Cancer*, J. Zander and J. Baltzer, Eds. Springer Berlin Heidelberg, 1985, pp. 113–135.

[19] J. Dengler, S. Behrens, and J. F. Desaga, "Segmentation of microcalcifications in mammograms," *IEEE Transactions on Medical Imaging*, vol. 12, no. 4, pp. 634–642, 1993.

[20] D. Betal, N. Roberts, and G. H. Whitehouse, "Segmentation and numerical analysis of microcalcifications on mammograms using mathematical morphology," *British Journal of Radiology*, vol. 70, no. 837, pp. 903–917, 1997.

[21] H. P. Chan, B. Sahiner, K. L. Lam, N. Petrick, M. A. Helvie, M. M. Goodsitt, and D. D. Adler, "Computerized analysis of mammographic microcalcifications in morphological and texture feature spaces," *Medical Physics*, vol. 25, no. 10, pp. 2007–2019, 1998.

[22] Y. Jiang, R. M. Nishikawa, C. E. M. D. E. Wolverton, R. A. Schmidt, and K. Doi, "Computerized classification of malignant and benign clustered microcalcifications in mammograms," in *Proc. 19th Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, 1997, pp. 521–523.

[23] R. S. Rana, Y. Jiang, R. A. Schmidt, R. M. Nishikawa, and B. Liu, "Independent evaluation of computer classification of malignant and benign calcifications in full-field digital mammograms1," *Academic Radiology*, vol. 14, no. 3, pp. 363–370, 2007.

[24] H. Strange, Z. Chen, E. R. E. Denton, and R. Zwiggelaar, "Modelling Mammographic Microcalcification Clusters using Persistent Mereotopology," *Pattern Recognition Letters*, vol. 47, pp. 157–163, 2014.

[25] Y. Z. Shao, L. Z. Liu, M. J. Bie, C. Li, Y. Wu, X. Xie, and L. Li, "Characterizing the clustered microcalcifications on mammograms to predict the pathological classification and grading: A mathematical modeling approach," *Journal of Digital Imaging*, vol. 24, pp. 764–771, 2011.

[26] M. Yam, M. Brady, R. Highnam, C. Behrenbruch, R. English, and Y. Kita, "Three-dimensional reconstruction of microcalcification clusters from two mammographic views," *IEEE Transactions on Medical Imaging*, vol. 20, no. 6, pp. 479–489, 2001.

[27] H. P. Chan, B. Sahiner, N. Petrick, M. A. Helvie, K. L. Lam, D. D. Adler, and M. M. Goodsitt, "Computerized classification of malignant and benign microcalcifications on mammograms: Texture analysis using an artificial neural network," *Physics in Medicine and Biology*, vol. 42, pp. 549–567, 1997.

[28] S. A. Feig, B. M. Galkin, and H. D. Muir, "Evaluation of breast microcalcifications by means of optically magnified tissue specimen radiographs." *Recent Results in Cancer Research*, vol. 105, pp. 111–123, 1987.

[29] Z. Chen, H. Strange, E. R. E. Denton, and R. Zwiggelaar, "Analysis of mammographic microcalcification clusters using topological features," in *Breast Imaging*, ser. Lecture Notes in Computer Science, H. Fujita, T. Hara, and C. Muramatsu, Eds. Springer International Publishing, 2014, vol. 8539, pp. 620–627.

[30] M. Heath, K. Bowyer, D. Kopans, R. Moore, and P. J. Kegelmeyer, "The digital database for screening mammography," in *the 5th International Workshop on Digital Mammography*, 2000, pp. 212–218.

[31] P. Soille, *Morphological Image Analysis: Principles and Applications*, 2nd ed. Heidelberg: Springer-Verlag, 2003, pp. 63–103.

[32] R. Diestel, *Graph Theory*, 4th ed. Heidelberg: Springer-Verlag, 2010, pp. 1–28.

[33] F. R. K. Chung, *Spectral Graph Theory, Number 92 of CBMS Regional Conference Series in Mathematics*. American Mathematical Society, 1997, ch. 1, pp. 1–21.

[34] R. O. Duda, P. E. Hart, and D. G. Stork, *Pattern Classification*, 2nd ed. New York: John Wiley and Sons, 2000, pp. 174–188.

[35] J. M. Keller, M. R. Gray, and J. A. Givens, "A fuzzy k-nearest neighbor algorithm," *IEEE Transactions on Systems, Man, and Cybernetics*, vol. 15, no. 4, pp. 580–585, 1985.

[36] R. Jensen and C. Cornelis, "Fuzzy rough nearest neighbour classification and prediction," *Theoretical Computer Science*, vol. 412, no. 42, pp. 5871–5884, 2011.

[37] N. Mac Parthaláin, R. Jensen, Q. Shen, and R. Zwiggelaar, "Fuzzy-rough approaches for mammographic risk analysis," *Intelligent Data Analysis*, vol. 14, no. 2, pp. 225–244, 2010.

[38] C. Cornelis, M. De Cock, and A. Radzikowska, "Vaguely quantified rough sets," *Lecture Notes in Artificial Intelligence*, vol. 4482, pp. 87–94, 2007.

[39] J. R. Beck and E. K. Schultz, "The use of relative operating characteristic (ROC) curves in test performance evaluation," *Archives of pathology & laboratory medicine*, vol. 110, pp. 13–20, 1986.

[40] J. Huang and C. Ling, "Using AUC and accuracy in evaluating learning algorithms," *IEEE Transactions on Knowledge and Data Engineering*, vol. 17, no. 3, pp. 299–310, 2005.

[41] S. J. Mason and N. E. Graham, "Areas beneath the relative operating characteristics (ROC) and relative operating levels (ROL) curves: Statistical significance and interpretation," *Quarterly Journal of the Royal Meteorological Society*, vol. 128, pp. 2145–2166, 2002.

[42] M. Hall, E. Frank, G. Holmes, B. Pfahringer, P. Reutemann, and I. H. Witten, "The weka data mining software: an update," *SIGKDD Explorations*, vol. 11, no. 1, pp. 10–18, 2009.

[43] M. A. Hall, "Correlation-based feature selection for machine learning," Ph.D. dissertation, The University of Waikato, Hamilton, New Zealand, 1999.

[44] R. Caruana and D. Freitag, "Greedy attribute selection," in *Proc. 11th International Conference on Machine Learning*, 1994, pp. 28–36.

[45] L. Wei, Y. Yang, and R. M. Nishikawa, "Microcalcification classification assisted by content-based image retrieval for breast cancer diagnosis," *Pattern recognition*, vol. 42, no. 6, pp. 1126–1132, 2009.

[46] J. Ren, D. Wang, and J. Jiang, "Effective recognition of mccs in mammograms using an improved neural classifier," *Engineering Applications of Artificial Intelligence*, vol. 24, pp. 638–645, 2011.

[47] J. Ren, "ANN vs. SVM: Which one performs better in classification of MCCs in mammogram imaging," *Knowledge-Based Systems*, vol. 26, pp. 144–153, 2012.