CrossMark

Tree-based modelling for the classification of mammographic benign and malignant micro-calcification clusters

Zobia Suhail[1] · Erika R. E. Denton[2] · Reyer Zwiggelaar[1]

Abstract Computer Aided Detection (CAD) systems are being developed to assist radiologists in diagnosis. For breast cancer the emphasis is shifting from detection to classification of abnormalities. The presented work concentrates on the benign versus malignant classification of micro-calcification clusters, which are a specific type of mammographic abnormality associated with the early development of breast cancer. After segmentation (automatic or manual), tree-based representations were used to distinguish between benign and malignant clusters, which takes into account clinical criteria such as the number of micro-calcifications in the clusters and their distribution and is based on the topology of the trees and the connectivity of the micro-calcifications. The idea of using tree structure based on the distance of individual calcifications for the classification of benign and malignant micro-calcification clusters is novel and closely related to clinical perception. Tree structures used in this study are distinct from decision trees classifiers being used in many machine learning approaches. Initial evaluation on the Digital Database for Screening Mammography (DDSM) data shows promising results, with an accuracy equal to 91 %, which is comparable to state of the art CAD systems and is in line with clinical perception of the morphology and appearance of benign and malignant micro-calcification clusters.

✉ Reyer Zwiggelaar
rrz@aber.ac.uk

Zobia Suhail
zoa1@aber.ac.uk

Erika R. E. Denton
erika.denton@nnuh.nhs.uk

[1] Aberystwyth University, Aberystwyth, UK

[2] Norfolk and Norwich University Hospital, Norwich, UK

 Springer

1 Introduction

Breast cancer is one of the most common diseases found in women [42]. Early detection and assessment can increase the chances of survival [25] and Computer Aided Diagnostic (CAD) systems are being developed to provide a second opinion for diagnosis. Once detected, calcifications can be labelled as benign or malignant, which depends on several properties of the calcifications and incorrect classification can lead to inappropriate or lack of treatment [16, 22]. Chandran et al. [10] discussed several types of benign calcification and presented an algorithm for the identification of benign calcifications within the image. They used wavelet decomposition with morphological filtering for the segmentation of candidate calcifications, and then used a connected components approach to identify benign pixels. Shen et al. [41] used compactness, moments, and Fourier descriptors as a set of shape features for the classification of benign and malignant calcification. Guray et al. [20] classified common benign lesions as developmental abnormalities, inflammatory lesions, fibrocystic changes, stromal lesions, and neoplasms. Nalawade et al. [37] described various patterns of breast calcification according to their appearance and distribution from the Americal College of Radiology (ACR) Breast Imaging-Reporting and Data System (BI-RADS) publication [15]. Muttarak et al. [35] described detailed calcification types as typical benign, intermediate concern, and higher probability of malignancy, according to the types and distribution of calcifications described in BI-RADS [15].

Many researchers are using advanced machine learning techniques for the purpose of assessing calcifications in mammogram images. Dheeba et al. [14] used a Support Vector Machine (SVM) for the classification of micro-calcification clusters. They computed a feature representation for each Region of Interests (RoI), and then used that feature as input for the SVM. Chan et al. [9] used thirteen texture features from spatial grey level dependence matrices from RoIs. After sub-selecting the texture features from multidimensional feature space, they trained and tested an artificial neural network with back propagation to classify the particular calcification types.

Shao et al. [40] presented a mathematical model to characterize clustered micro-calcifications. They graded the micro-calcifications into 4 grades from 0 (benign), 1 (well-differentiated infiltrating ductal carcinoma), 2 (moderately differentiated infiltrating ductal carcinoma) to 3 (poorly differentiated infiltrating ductal carcinoma). They defined a new feature parameter θ, and coded a positive relationship between the feature parameter θ and pathological grading G of micro-calcifications. Suhail et al. [45] defined a new method for the automatic detection of abnormalities in mammogram images and introduced a new Energy Ratio function to define distinct pixel features belonging to abnormalities.

Chen et al. [11, 12] constructed a micro-calcification graph to represent the topological structure of clusters for the Mammographic Image Analysis Society (MIAS) database [44]. They analysed the topological structure by using multiscale morphology and investigated the number of independent subgraphs and the average degree of nodes as feature vectors. Chen et al. extended their work and provided additional evaluation of the proposed method [12].

Many tree-based methods have been proposed in the past for a wide range of applications like natural language processing [3, 43], road tracking [18] and biomedical science [1, 2, 4]. Many machine learning techniques like bagging [6], random forest [7] and boosting [17, 39] are based on a tree as a fundamental structure. Trees can be used to solve both classification and regression problems [8, 33]. Classification/regression trees are normally described as decision trees. In biomedical science decision trees have been used effectively to solve classification problems [1, 2, 5]. Some of the decision tree based methods have been

proposed in the literature focusing on breast-cancer research [2, 5]. One of the limitations associated with the decision tree is overfitting which indicates it can not generalize the data. Our project focuses on developing a novel computational model for the classification of malignant and benign calcifications in which we link the clinical aspects of calcifications to a tree structure, which is different from traditional decision tree concepts that have been used in the literature. The proposed tree-based model infers the classification rules after making trees instead of making rules from the data features during tree construction (Adaboost, Bagging or Random Forest classifiers). Unlike the related work presented in the literature [11, 12], in which the authors used graph structures to build the feature-set at multiple scales (the dimensionality of the feature space was equal to 520) and then used standard classifiers, we used tree-structure at a fixed scale (guided by clinical information) and the height of the trees as a standard for the benign/malignant clustering.

The overall structure of the paper is as follows: Section 2 describes the dataset. Section 3 describes the proposed algorithm. Results and discussion can be found in Section 4. Computational complexity of the algorithm is presented in Section 5. Finally conclusions are presented in Section 6.

2 Dataset

We used segmented RoIs extracted from the DDSM database [21]. The dataset has been used by other researchers for the performance evaluation of their algorithm [12]. The RoIs are of variable sizes (average size of image patches is 482 x 450 pixels), but it should be noted that our algorithm is invariant to the image size. There are 149 benign cases and 139 malignant cases in the dataset. The RoIs are probability images regarding calcification presence [38]. For evaluation of our algorithm, we used a subset of these 288 RoIs which were all classified as diffuse/scattered micro-calcification clusters according to the BI-RADS standard [15, 26]. The BI-RADS classification for the DDSM database has been provided by expert radiologist and is provided as part of the dataset. This subset contains 129 RoIs, of which 71 were malignant and 58 benign. Some example RoIs from the used database can be found in Fig. 1, where the 2nd and 4th columns are representing the annotations/segmentations.
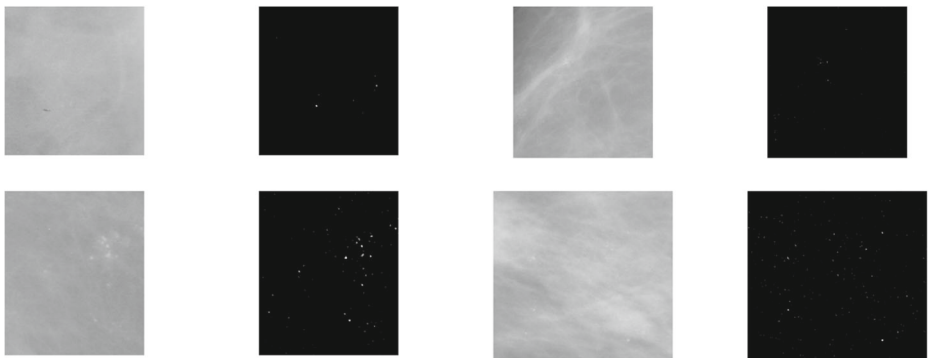

Fig. 1 Original mammographic data with segmented benign (*top row*) and malignant (*bottom row*) micro-calcifications RoIs
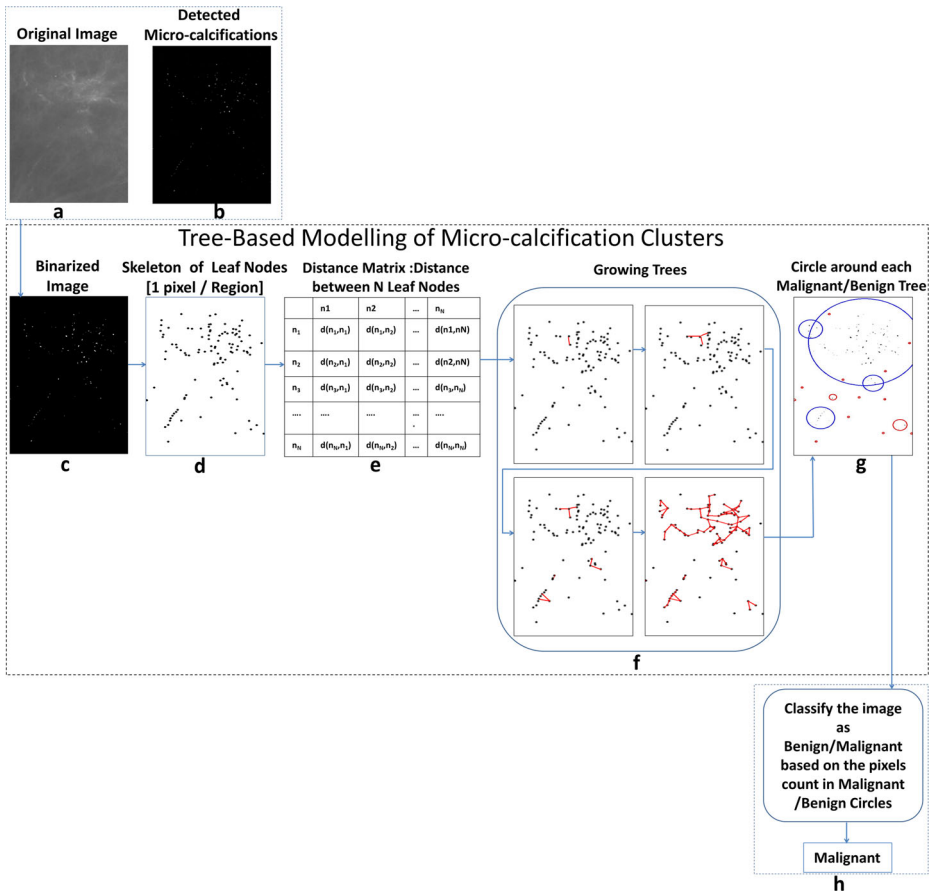
Fig. 2 Flow diagram for tree-based modelling to classify the micro-calcification clusters in mammographic RoIs: **a** mammographic RoI, **b** segmented micro-calcifications, **c** binarized image after applying specific threshold, **d** leaf node generation from 1 pixel / 7×7 connected window, **e** distance map representing the Euclidean distance between all the nodes, **f** the process of generating trees by combining the closest nodes, as can be seen the nodes from more dense area form trees, **g** drawing circles around each tree, **h** report image as benign/malignant depending on the number of pixels belonging to the benign/malignant circles (benign circles indicate trees having height ≤ 1, whereas malignant circles represent trees having height ≥ 2)

3 Proposed method

We propose a novel tree-based model for the classification of benign and malignant calcifications after detecting and segmenting the suspected abnormal regions. We exploit the properties of binary trees to classify the micro-calcification clusters. An overview of the developed methodology can be found in Fig. 2 (as well as from the pseudo-code presented in Algorithm 1) and details can be found in subsequent subsections.

Algorithm 1 Pseudocode of the proposed method

1. Take input in the form of segmented micro-calcifications RoI.
2. Binarize the image by applying a specific threshold.
3. Identify the connected components from the binarized image by considering only 1 pixel from each 7×7 connected window.
4. Construct nodes from each of the isolated pixels that we get from the last step.
5. Compute the distance map by finding out the Euclidean distance from each pixel to all other pixels.
6. Merge the closest nodes together to form binary trees.
7. Draw circle around each of the tree structures that we developed in the previous step.
8. Circles representing trees having height ≥ 2 are considered as belonging to the malignant class, whereas circles for trees having height ≤ 1 are considered as belonging to the benign class. Finally, count the pixels belonging to both the benign and malignant circles and provide results according to majority voting.

3.1 Input

The starting point is the automated detection or manual annotation of a micro-calcification cluster. An example segmented micro-calcification cluster is shown in Fig. 2b, whereas Fig. 2a is showing the original mammographic RoI. These RoIs are accurately representing mammographic RoI as a binary form, where '1' is representing a presence and '0' the absence of micro-calcification for each pixel within an image.

3.2 Binarization

For automatically detected RoIs we binarize the RoIs which removes low probability regions. We have used a threshold equal to 0.27.

$$\forall_{i,j} P(x_i, y_j) = \begin{cases} 1, & \text{if } P(x_i, y_j) \geq 0.27 \\ 0, & \text{otherwise} \end{cases}$$

We set the threshold value to create a binary image that is a balance between removing noise and retaining important image information that will be required for classification. Keeping this threshold value very high may remove important regions that may be needed in the subsequent steps and decreasing this value may add additional, non-calcification, regions, which are seen as noise. Based on the experiments, we observe that 0.27 is the optimal threshold value in terms of getting only the relevant segmented micro-calcifications.

3.3 Identifying connected components and generation of leaf node skeleton

In this step, we reduce all individual micro-calcifications to a single point, for which we have used a simple 7×7 connectivity rule. Pixel $p_2(x,y)$ is said to be 7×7 connected to pixel $p_1(x,y)$, if it appears within 7×7 surrounding pixels of $p_1(x,y)$.[1] All remaining pixels are regarded as the initial leaf nodes for the tree based modelling approach ([19] provides an excellent overview on tree based abstract data types), and the nodes

---

[1] 7×7 window size is a good approximation on the individual calcification area. Making it too large may span more than 1 calcification and making it too small may produce more than one pixel from a single RoI.

Table 1 Node structure

Node-Id :	representing Unique ID for each node
Left child :	representing left child of the node
Right child :	representing right child of the node
Connected-components :	list of connected pixels P(x,y)

(e.g. the underlying data structures) contain information about the micro-calcifications they represent. Table 1 represents the specific node structure that will be used by the current approach.

At this stage Node-Ids are assigned incrementally for all the nodes, whereas left and right children are null as there are no connections between the nodes at this stage. The connected-components are being initialized by the coordinates (x,y) of each pixel which belongs to the node.

3.4 Distance-map computation

In this step, we calculate the distance between all leaf nodes, which will be used as a basic measure of connectivity between the individual nodes. We used the Euclidean distance, which is representing the distance between pixels associated with leaf nodes N_i and N_j:

$$D_{(N_i, N_j)} = \sqrt{(N_j x - N_i x)^2 + (N_j y - N_i y)^2} \tag{1}$$

3.5 Constructing trees from closest nodes

Tree like structures are built through a recursive process which takes the distance (as computed in Section 3.4) between the initial leaf nodes into account. Algorithm 2 shows the complete procedure of constructing binary trees from the leaf nodes having minimum distance. Closest nodes have been merged recursively together to form binary trees. We used a distance equal to 40 as the upper limit, which is based on the clinical equivalent requirement of having micro-calcifications clustered within $1cm^2$.[2] The resulting representation for the RoIs is a set of trees, some of which have depth ≤ 1 (i.e. either individual micro-calcifications which are not connected or a group of 2 connected micro-calcifications), whilst others have depth ≥ 2 (i.e. represent complex clusters of micro-calcifications). In Fig. 2 we have illustrated this by representing trees as circles, where blue circles represent trees with depth ≥ 2, and red circles represent trees with depth ≤ 1 which are regarded as malignant and benign trees, respectively.

Algorithm 2 Trees construction from leaf nodes

Input: List of nodes (initially represented as leaf nodes)
    do recursively connect leaf nodes by:
        1. finding the closest pair of nodes;
        2. removing this pair from the list of nodes;
        3. merge these nodes together as a binary tree where the nodes are now represented as leaf nodes of a binary tree, and also appeared as connected-components for the root of binary tree;
        4. adding the root of the tree to the list of nodes;
    while no further pairs of nodes below a specific distance are found
return: list of nodes, in which closest nodes have been merged together as trees

---

[2] At 96 pixels/inch this is approximately equal to 40 pixels.

3.6 Output

The final stage in the classification process is to use the tree representation of the RoIs to determine if these are benign or malignant. To achieve this we count the number of pixels represented by benign (red circles) and malignant (blue circles) trees and use a simple majority vote to determine the final classification.

4 Results and discussion

The results are in line with radiologists' perceptions (see Section 3.5 for a detailed description) of micro-calcification clusters. Figure 3 shows the overall algorithm performance on a benign and malignant RoI. The benign RoI in Fig. 3 (top row) has 6 micro-calcifications. After computing the distance between all these nodes, we found not a single node that has distance ≤ 40 with any other node. Each binary tree formed from these 6 nodes has height 0, indicating that all trees belong to the benign class. The next step is to count the number of pixels belonging to the benign and malignant classes for the final image classification. There are no trees representing the malignant class, therefore the image is reported as benign using our final decision criteria. Figure 3 (bottom row) shows the algorithm flow for a malignant RoI example. After identifying the nodes, this image has 94 isolated pixels and hence 94 leaf nodes. After merging the closest nodes based on the distance criteria we end up with 13 trees with heights representing a specific cluster class for the pixels connected within a particular tree. For this malignant RoI the total number of trees with height ≤ 1 is equal to 10, whereas the total number of trees with height ≥ 2 is 3. The distribution of tree heights is shown in Table 2.

The first 7 and trees 9–11 represent the benign class, whilst trees 8, 12 and 13 represent the malignant class. After counting the number of nodes for both classes we found that 81
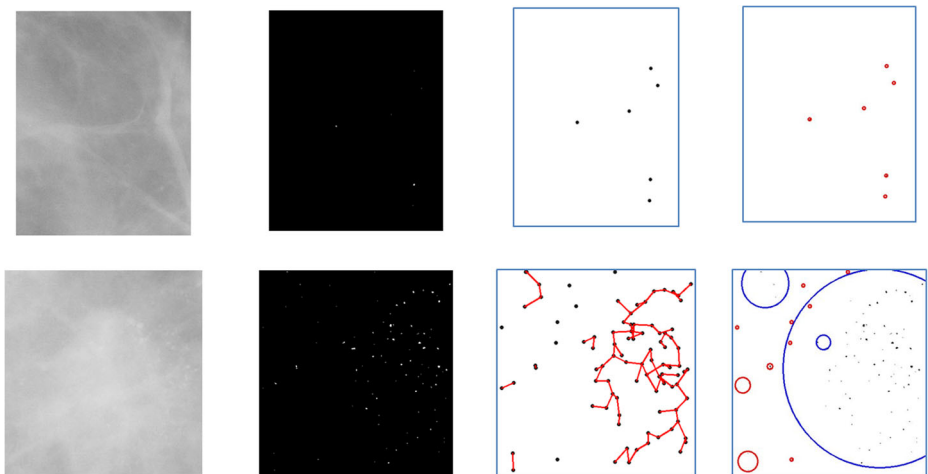

Fig. 3 Step-wise execution on benign (*top row*) and malignant (*bottom row*) example RoIs. *First column*: original mammogram RoI; *second column*: segmented and binarized micro-calcifications; *third column*: connected nodes for each binary tree; *fourth column*: boundary circles around each binary tree

Table 2 Counting the number of leaf nodes in each tree for a malignant RoI (example shown in Fig. 3, bottom row

TreeId	Height	Number of Leaf Nodes
1	0	1
2	0	1
3	0	1
4	0	1
5	0	1
6	0	1
7	0	1
8	2	3
9	1	2
10	1	2
11	1	2
12	3	6
13	7	72

nodes belong to malignant trees whereas 13 belong to benign trees, and the image is reported as malignant (based on a majority vote).

From 129 RoIs, 71 images were malignant, whereas 58 were benign. From this data set, we found 66 true positive results and 51 true negative. The performance analysis on the test dataset is given in Table 3 in terms of accuracy, sensitivity, specificity, precision and recall. The F-measure based on the precision and recall for the overall results is equal to 0.91.[3]

The algorithm also gives a good visual interpretation of the classified clusters as can be seen from the last 2 columns in Fig. 3. Results are clearly differentiating between dense and scattered micro-calcification by means of trees which are indicated by the circles around each cluster.

More results after executing the proposed algorithm on benign and malignant images can be found in Figs. 4 and 5 (RoI for the malignant and benign cases have been shown in the second column of Figs. 4 and 5) which also confirm the results discussed above. Figure 4 shows a result of the proposed algorithm's execution on two malignant RoIs, where more dense trees (height ≥ 2) are found compared to totally isolated or lesser height trees. As a result the number of pixels belonging to the malignant class are more than the benign class, so both the images were correctly reported as malignant. On the other hand in Fig. 5, there are more pixels belonging to trees having height ≤ 1 than ≥ 2 height trees. Therefore the images were successfully reported as benign. The examples in Figs. 3, 4 and 5 show the wide range of variation that occurs for malignant and benign micro-calcification clusters.

4.1 Incorrect classification results

Figure 6 shows some examples of incorrect classification results by the current approach. RoI in Fig. 6 (top row), is classified by the developed approach as malignant, although this was a benign RoI. The reason for this false positive result is that the RoI has only 1 cluster (having 3 micro-calcifications all within an area of 1 cm^2). According to the proposed

---

[3]$F - Measure = \frac{2*Precision*Recall}{Precision+Recall}$

Table 3 Performance analysis

Performance measure	Metric	Algorithm result
Accuracy	TP + TN / TP + FP + FN + TN	91 %
Sensitivity	TP / TP +FN	93 %
Specificity	TN / TN + FP	88 %
Precision	TP / TP + FP	90 %
Recall	TP / TP + FN	93 %

algorithm's result, this cluster represents a malignant RoI (height of tree associated to cluster ≥ 2). The overall image is reported as malignant as there is no other micro-calcification indicating a benign RoI. The bottom row in Fig. 6 is an example of false negative result that reported a malignant RoI as benign. Although this image has benign and malignant trees, the total count of pixels from the malignant class is less than that of the benign class. The basic reason of this misclassification is that most of the micro-calcifications are scattered and do not fall within the distance criteria (as mentioned in Section 3.4) to be merged and form dense trees.

4.2 Beyond diffused/scattered cases

For the whole dataset (288 RoIs), we get 55 % accuracy, 68 % sensitivity and 44 % specificity. Chen et al. [11, 12] used a similar approach to the proposed method by exploiting different topological structures to differentiate the benign and malignant RoIs. In their algorithm they performed evaluation on the full dataset. They presented the connectivity of individual calcification (at multiple scales) by using a graph representation. They used k-nearest neighbour (kNN)-based classifiers to classify between the benign and malignant RoIs and multi-scale topological feature vectors for classification. By examining the good performance results from their algorithm, we can expect to improve our algorithm's performance by building trees at different scales and using some multi-scale tree features (height, number of leaf nodes, etc.) to train a classifier in order to achieve improved results on the full dataset.
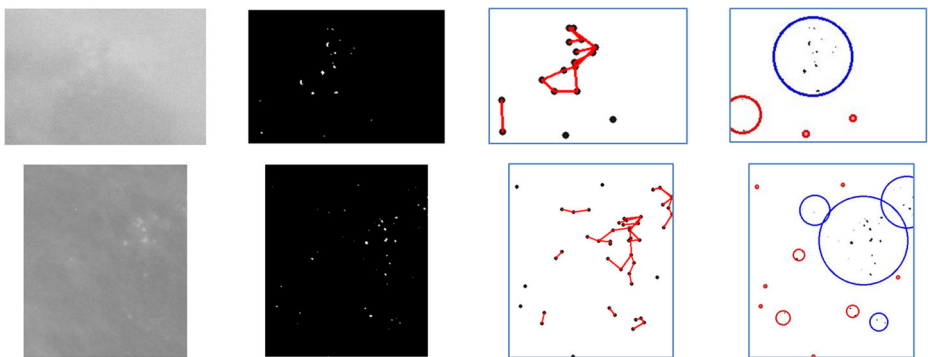

Fig. 4 Execution of the proposed algorithm on malignant RoIs (see the caption of Fig. 2 for a detailed description)
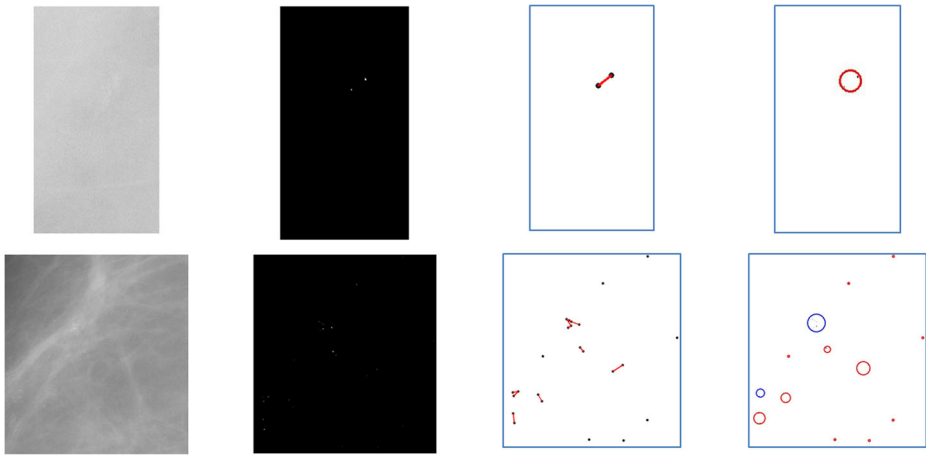
Fig. 5 Execution of the proposed algorithm on benign RoIs (see the caption of Fig. 2 for a detailed description)

5 Computational complexity analysis

Apart from the functions (*Make-Leaf-Nodes*, *Merge-closest-Nodes*, *Find-connected-component*), our program's complexity depends on the call of a recursive function *Make-Trees-Recursively*, that is used to create binary trees from leaf nodes recursively until no more closest nodes are left to be merged. If we have N leaf nodes initially and we call the function *Make-Trees-Recursively*, in a worst case it will find a closest node to a current node at the end of the list, and will execute N times. Subsequently the length of this list is reduced by 1 as one pair of pixels have been merged together and appeared as leaf nodes of a new tree whose root node has been added to the list. The second time the function *Make-Trees-Recursively* will be executed at most (N-1) times and so on. Overall execution complexity
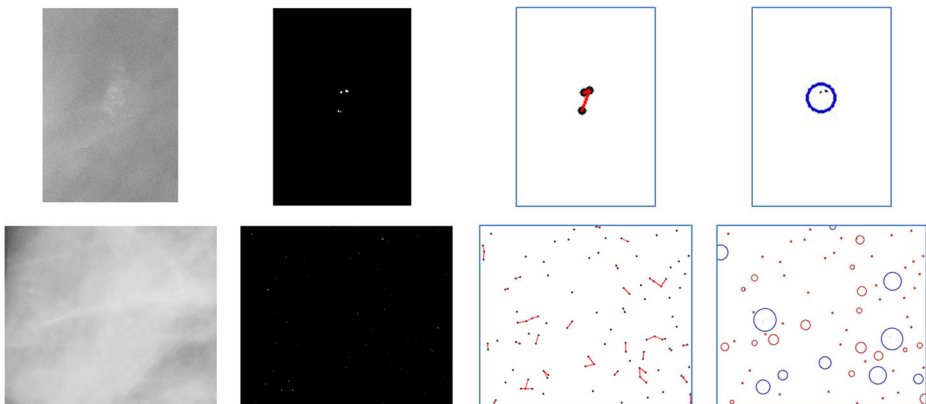

Fig. 6 Incorrect classification examples. *First row*: a benign RoI that is reported as malignant; *Second row*: malignant RoI that has been reported as benign (see the caption of Fig. 2 for a detailed description)

for *Make-Trees-Recursively* function is $N+ N-1 + N- + ... 3 + 2 + 1 = N(N+1)/2 = O(N^2)$.

6 Conclusions

Identification of benign and malignant calcification is an important part of the breast cancer diagnostic process because it can reduce incorrect treatment [16, 22]. We have introduced a novel method for the identification of benign and malignant calcifications based on a tree representation of the distribution of micro-calcifications and a distance metric. We used a conventional binary tree data structure to represent the calcification clusters. The height of trees indirectly reflects the number of pixels belonging to each cluster. Subsequently we build a relationship between the height of trees and the category of calcification (benign versus malignant). For visualisation purposes circles have been drawn around each cluster to differentiate between the benign and malignant class.

We evaluated this method on DDSM RoIs (using the diffuse/scattered cases as explained in Section 2). We obtained good results in terms of commonly used performance measures. The obtained accuracy was 91 %, whereas sensitivity and specificity were 93 % and 88 %, respectively. The results are comparable to the state of the art and reflect the radiologists' view regarding calcification type.

6.1 Translation to other application areas

In addition to malignant and benign cluster identification in mammographic images this tree-based representation can be extended to other fields. One possible use of this approach is in surveillance systems counting the number of people in a crowd in order to identify dense crowds and to be informed of abnormal situations [23, 34]. The proposed system can be translated to be used in traffic control systems, in order to identify traffic jams and to redirect traffic to appropriate routes [24]. In addition, the proposed tree-based model could be mapped to other applications like activity recognition [28, 30], motion tracking [13, 27] and multi-task classifications [29, 31, 32]. There exist a number of applications of such activity recognition in the medical domain [36, 46], the purpose of which is to identify certain postures (sitting, standing or lying) of patients/elderly people based on the data received by sensors.

6.2 Future work

The algorithm we tried is at a specific scale (see Section 3.5 for details). In future we will explore a multi-scale tree-representation for the classification of benign and malignant calcifications. After making our algorithm scale-invariant, we will investigate the effect of classifiers and compare results with state of the art algorithms developed using other topological structure representations for benign and malignant classification [12].

At the initial stage of the proposed algorithm we set a threshold of 0.27 in order to ignore the low probability regions. We set this threshold in a way that reduces noise without removing important information from the RoI. Another parameter that we set for our algorithm is the height of the trees (≤ 1 for benign RoI). In an extension of the current method, we will investigate the effect of these parameters (intensity threshold and tree-height) on the overall algorithm's accuracy.

In the current method we used a complete tree version in order to compute the height of trees. We will try to convert the current tree structure to a complete binary tree, and to make the algorithm more efficient in terms of computational complexity.

We will also compare the results of implementing other applications like activity recognition [28, 30] and human motion tracking [13, 27] to the proposed tree-based modelling and highlight the advantages and disadvantages.

References

1. Albain K, Crowley J, LeBlanc M, Livingston R (1990) Determinants of improved outcome in small-cell lung cancer: an analysis of the 2,580-patient southwest oncology group data base. J Clin Oncol 8(9):1563–1574
2. Albain K, Green S, LeBlanc M, Rivkin S, O'Sullivan J, Osborne C (1992) Proportional hazards and recursive partitioning and amalgamation analyses of the southwest oncology group node-positive adjuvant CMFVP breast cancer data base: a pilot study. Breast Cancer Res Treat 22(3):273–284
3. Bahl L, Brown P, de Souza P, Mercer R (1989) A tree-based statistical language model for natural language speech recognition. IEEE Trans Acoust Speech Signal Process 37(7):1001–1008
4. Banerjee M, Biswas D, Sakr W, Wood D (2000) Recursive partitioning for prognostic grouping of patients with clinically localized prostate carcinoma. Cancer 89(2):404–411
5. Banerjee M, George J, Song E, Roy A, Hryniuk W (2004) Tree-based model for breast cancer prognostication. J Clin Oncol 22(13):2567–2575
6. Breiman L (1996) Bagging predictors. Mach Learn 24(2):123–140
7. Breiman L (2001) Random forests. Mach Learn 45(1):5–32
8. Breiman L, Friedman J, Olshen R, Stone C (1984) Classification and regression trees. Wadsworth
9. Chan H, Sahiner B, Petrick N, Helvie M, Lam K, Adler D, Goodsitt M (1997) Computerized classification of malignant and benign microcalcifications on mammograms: texture analysis using an artificial neural network. Phys Med Biol 42(3):549
10. Chandran P, Byju N, RKumar KoshyS, Ramachandran K (2013) Benign calcification detection in mammogram images. International Journal of Electronics and Computer Science Engineering 4(2):180–189
11. Chen Z, Oliver A, Denton E, Boggis C, Zwiggelaar R (2012) Classification of microcalcification clusters using topological structure features. In: Medical image understanding and analysis, pp 37–42
12. Chen Z, Strange H, Oliver A, Denton E, Boggis C, Zwiggelaar R (2015) Topological modeling and classification of mammographic microcalcification clusters. IEEE Trans Biomed Eng 62(4):1203–14
13. Cui J, Liu Y, Xu Y, Zhao H, Zha H (2013) Tracking generic human motion via fusion of low-and high-dimensional approaches. IEEE Trans Syst Man Cybern: Systems 43(4):996–1002
14. Dheeba J, Selvi S (2011) Classification of malignant and benign microcalcification using SVM classifier. In: International conference on emerging trends in electrical and computer technology (ICETECT), IEEE, pp 686–690
15. D'orsi C, Bassett L, Berg W, Feig S, Jackson V, Kopans D (2003) Breast imaging reporting and data system: ACR BI-RADS-mammography. American College of Radiology (ACR), Reston, pp 230–234
16. Elmore J, Barton M, Moceri V, Polk S, Arena P, Fletcher S (1998) Ten-year risk of false positive screening mammograms and clinical breast examinations. N Engl J Med 338(16):1089–1096
17. Freund Y (1990) Boosting a weak learning algorithm by majority. In: COLT '90 proceedings of the third annual workshop on computational learning theory, vol 90, pp 202–216
18. Geman D, Jedynak B (1996) An active testing model for tracking roads in satellite images. IEEE Trans Pattern Anal Mach Intell 18(1):1–14
19. Goodrich MT, Tamassia R (2008) Data structures and algorithms in java. Wiley
20. Guray M, Sahin A (2006) Benign breast diseases: classification, diagnosis, and management. Oncologist 11(5):435–449
21. Heath M, Bowyer K, Kopans D, Moore R, Kegelmeyer W (2000) The digital database for screening mammography. In: Proceedings of the 5th international workshop on digital mammography, pp 212–218

22. Howard J (1987) Using mammography for cancer control: an unrealized potential. CA Cancer J Clin 37(1):33–48
23. Hussain N, Yatim H, Hussain N, Yan J, Haron F (2011) CDES: a pixel-based crowd density estimation system for Masjid Al-Haram. Saf Sci 49(6):824–833
24. Joshi M, Mishra D (2015) Review of traffic density analysis techniques. International Journal of Advanced Research in Computer and Communication Engineering 4(7)
25. Jotwani A, Gralow J (2009) Early detection of breast cancer. Mol Diagn Ther 13(6):349–357
26. Lazarus E, Mainiero M, Schepps B, Koelliker S, Livingston L (2006) BI-RADS Lexicon for US and mammography: interobserver variability and positive predictive value 1. Radiology 239(2):385–91
27. Liu Y, Cui J, Zhao H, Zha H (2012) Fusion of low-and high-dimensional approaches by trackers sampling for generic human motion tracking. In: 21st international conference on pattern recognition (ICPR), 2012, IEEE, pp 898–901
28. Liu Y, Nie L, Han L, Zhang L, Rosenblum D (2015) Action2activity: recognizing complex activities from sensor data. In: Proceedings of the international joint conference on artificial intelligence, pp 1617–1623
29. Liu Y, Liang Y, Liu S, Rosenblum D, Zheng Y (2016a) Predicting urban water quality with ubiquitous data. arXiv:161009462
30. Liu Y, Nie L, Liu L, Rosenblum D (2016b) From action to activity: sensor-based activity recognition. Neurocomputing 181:108–115
31. Liu Y, Zhang L, Nie L, Yan Y, Rosenblum D (2016c) Fortune teller: predicting your career path. In: AAAI, pp 201–207
32. Liu Y, Zheng Y, Liang Y, Liu S, Rosenblum D (2016d) Urban water quality prediction based on multi-task multi-view learning. In: Proceedings of the international joint conference on artificial intelligence
33. Loh W (2011) Classification and regression trees. Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery 1(1):14–23
34. Marana A, Velastin S, Costa L, Lotufo R (1998) Automatic estimation of crowd density using texture. Saf Sci 28(3):165–175
35. Muttarak M, Kongmebhol P, Sukhamwang N (2009) Breast calcifications: which are malignant? Singap Med J 50(9):907–13
36. Najafi B, Aminian K, Paraschiv-Ionescu A, Loew F, Bula C, Robert P (2003) Ambulatory system for human motion analysis using a kinematic sensor: monitoring of daily physical activity in the elderly. IEEE Trans Biomed Eng 50(6):711–723
37. Nalawade Y (2009) Evaluation of breast calcifications. Indian Journal of Radiology & Imaging 19(4):282–286
38. 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
39. Schapire R, Freund Y, Bartlett P, Lee W (1998) Boosting the margin: a new explanation for the effectiveness of voting methods. Ann Stat:1651–1686
40. Shao Y, Liu L, Bie M, Li C, Wu Y, Xie X, Li L (2011) Characterizing the clustered microcalcifications on mammograms to predict the pathological classification and grading: a mathematical modeling approach. J Digit Imaging 24(5):764–771
41. Shen L, Rangayyan R, Desautels J (1994) Application of shape analysis to mammographic calcifications. IEEE Trans Med Imaging 13(2):263–274
42. Siegel R, Naishadham D, Jemal A (2012) Cancer statistics. CA Cancer J Clin 62(1):10–29
43. Silva GD, Hull J (1994) Proper noun detection in document images. Pattern Recogn 27(2):311–320
44. Suckling J, Parker J, Dance D, Astley S, Hutt I, Boggis C, Ricketts I, Stamatakis E, Cerneaz N, Kok S, Taylor P (1994) The mammographic image analysis society digital mammogram database. Proc Excerpta Med Int Congr Ser 1069:375–378
45. Suhail Z, Mansoor S, Murtaza K (2015) Automatic detection of abnormalities in mammograms. BMC Med Imaging 15(1):1
46. Walker D, Heslop P, Plummer C, Essex T, Chandler S (1997) A continuous patient activity monitor: validation and relation to disability. Physiol Meas 18(1):49

Zobia Suhail received her M.Phil degree in Computer Science in 2014 from University of the Punjab, Pakistan. She is a lecturer in Computer science department of Punjab University, Pakistan since 2011. Currently she is doing her PhD in computer science from Aberystwyth University, Aberystwyth, Wales, UK. Her current research is in Computer Vision, Machine Learning and Medical image processing (specifically mammogram images). She is also first author of a journal publication 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.

Since 2005, 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.

Reyer Zwiggelaar 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, U.K., in 1993.

He is currently a Professor at the Department of Computer Science, Aberystwyth University, U.K. He is the author or coauthor of more than 200 conference and journal papers. He is an Associate Editor of the Journal of Biomedical and Health Informatics (formerly IEEE Transactions on Information Technology in Biomedicine) and Pattern Recognition. His current research interests include medical image understanding, especially focusing on mammographic and prostate data, pattern recognition, statistical methods, texture-based segmentation, and feature-detection techniques.