This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/JBHI.2014.2334491, IEEE Journal of Biomedical and Health Informatics

A CAD_X Scheme for Mammography Empowered with Topological Information from Clustered Microcalcifications' Atlases

Ioannis I. Andreadis, George M. Spyrou, and Konstantina S. Nikita

***Abstract*—A Computer Aided Diagnosis (CAD_x) framework for the diagnosis of clustered microcalcifications (MCs) has already been developed, which is based on the analysis of MCs' morphologies, the shape of the cluster they form and the texture of the surrounding tissue. In this work, we investigate the diagnostic information that the relative location of the cluster inside the breast may provide. Breast probabilistic maps are generated and adopted in the CAD_x pipeline, expecting to empower its diagnostic procedure. We propose a flowchart combining alternative classification algorithms and the aforementioned probabilistic maps in order to provide a final risk for malignancy for new considered mammograms. For the evaluation performance, a large dataset of mammograms provided from the Digital Database of Screening Mammography (DDSM) has been used. The obtained results indicate that the proposed modifications lead to the enhancement of the diagnostic process, as the classification results are further improved. Additionally, a straightforward comparison between the CAD_x pipeline and the radiologists who assessed the same mammograms, reveal that the CAD_x pipeline performs towards the right direction, as the sensitivity remains at high levels, while improving both the accuracy, from 51.4% to 69%, and the specificity, from 16.6% to 54.7%.**

***Index Terms*—Computational Intelligence, Computer Aided Diagnosis, microcalcifications, probabilistic maps, ROC analysis**

I. Introduction

MICROCALCIFICATIONS (MCs) are tiny deposits of calcium salts which can be located anywhere in breast tissue [1]. Although mammography is considered the most effective screening tool for the examination of breast MCs [2], specific inherent limitations of the method led to the development of Computer Aided Diagnosis (CAD_x) systems [3]. The role of a CAD_x system is to help the radiologists in their diagnostic process by supporting them with a reliable second opinion [4].

In a previous work [5], we have already developed a CAD_x framework for the classification of cluster of MCs. The framework implemented the main steps followed for the diagnosis, including the segmentation of MCs from the surrounding tissue, the extraction of objective features for the description of the cluster, the proper selection of the most representative features and the final classification using a machine learning algorithm [4]. A great variety of image features has been considered, containing features describing the texture of the surrounding tissue in which the MCs are embedded [6], features describing the shape of the individual MCs and their optical density [7] and finally features representing the morphology of the whole cluster and the distribution of the MCs within the cluster [8].

However, there is an additional factor which is taken under consideration by the radiologists and concerns the relative location of the findings inside the breast [9]. Despite the importance of this factor, there are few studies that have incorporated position features in their CAD_x pipeline. Veldkamp et al. [10] used the relative distance of the cluster from the pectoral muscle and the breast's skin line as measures for the discrimination of clusters of MCs. Russakoff et al [11] generated a set of probabilistic breast cancer atlases to reveal areas of higher risk of breast cancer, but took under consideration Regions of Interest (ROIs) including only cancerous masses and not cluster of MCs. The reason for this lack is probably due to the fact that the description of the relative location of a finding inside the breast requires much introductory work, since the proper segmentation of important breast landmarks (breast skin, pectoral muscle and nipple) is considered prerequisite step. In a previous work [12], we have performed the introductory steps for the use of location features in the CAD_x pipeline. Specifically, we have proposed a framework leading to the generation of breast MC occurrence and probabilistic maps. The specific maps revealed areas with higher occurrence of cluster of MCs, as well as areas that seem to be more susceptible to cancer.

In this work, we investigate whether the location of a cluster may provide quantified information to support diagnosis. We adopt the generated maps in the diagnostic procedure expecting to optimize the classification performance. For performance evaluation, we used a great number of mammograms of varying subtlety provided by the Digital Database of Screening Mammography (DDSM) [13]. To the best of our knowledge, it is the largest dataset used in CAD_x

Manuscript received June 17, 2014.

I. I. Andreadis is with the School of Electrical and Computer Engineering, National Technical University of Athens, 15573, Athens, Greece (e-mail: iandr@ biosim.ntua.gr).

G. M. Spyrou is with the Biomedical Research Foundation Academy of Athens, Athens, Greece (e-mail: gspyrou@bioacademy.gr).

K. S. Nikita is with the School of Electrical and Computer Engineering, National Technical University of Athens, 15573, Athens, Greece (corresponding author phone: +30-2107722968; fax: +30210 7722320; email: knikita@ece.ntua.gr).

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/JBHI.2014.2334491, IEEE Journal of Biomedical and Health Informatics

diagnosis for MCs, since we exploited almost all the available cases of the database containing clusters of MCs ending in a dataset consisted of 1715 different ROIs.

The rest of the paper is organized as follows: In section II we briefly present the considered dataset, we describe the basic steps of the CAD_x framework and we mainly discuss the proposed set of rules, where the contribution of the location features is implemented. In section III, we present the results achieved after ROC evaluation and we compare the reported performance with the corresponding achieved by the radiologists who assessed the same mammograms. Finally, in section IV and V, we discuss the achieved results and we proceed on the extraction of the most important conclusions, while discussing the main future work towards refining the diagnostic process.

II. Methods and Materials

A. CAD_x Framework

Each CADx system includes specific steps such as the segmentation of microcalcifications, the feature extraction process and the final classification of a considered case [4]. The framework investigated in the current study has been presented in a previous work [5], whose efficacy has been investigated using subsets of varying findings' subtlety [14]. The main steps of this CADx framework are presented in fig.1 in blue color.

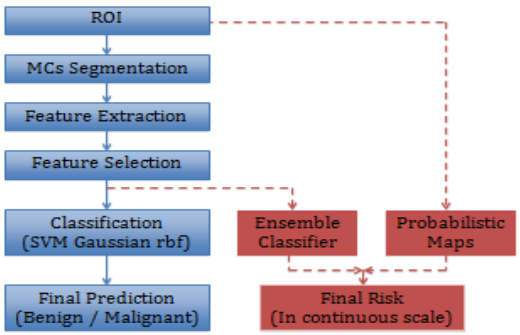

Fig. 1. Flowchart for the CAD_x framework. The main steps presented in [5] are depicted in blue color. The proposed modifications in this study are presented in red color.

A segmentation algorithm is used to isolate the detected MCs from the surrounding breast tissue. Afterwards, feature extraction methodologies are applied to extract objective image features. Specifically, a great variety of image features has been applied, including features based on shape representation of the finding (area, circularity, elongation e.t.c.), features concerning the morphology of the cluster of MCs and features representing the texture of the surrounding tissue (first-order statistics, features based on the grey-level occurrence matrices). A total of 188 different image features have been considered. Due to the great number of image features considered, a feature selection phase is appropriate, in order to remove irrelevant features and locate a satisfactory subset able to improve the classification results. To this end, the Sequential Forward Selection (SFS) method has been applied for the feature selection. Finally, a classification scheme is required to discriminate between benign and malignant cases.

In this work, we perform proper changes on the CAD_x pipeline, concerning mainly the final step, the classification scheme. The proposed changes are presented in fig. 1 in red color. Instead of using only one classification algorithm, a series of classification algorithms is exploited in order to form an ensemble classifier, which provides the estimation of the final risk assessment. Specifically, the following machine learning algorithms are utilized: Support Vector Machines, using four different kernels (gaussian rbf, sigmoid, linear and polynomial), k-Nearest Neighbors, Artificial Neural Network, Classification Tree, Random Forests and Probabilistic Neural Network using two different functions (gaussian and exponential). Therefore, we end in a set consisted of ten different classifiers.

All the aforementioned classification schemes are combined in an ensemble classifier, which estimates the risk of the malignancy of a considered ROI in a continuous scale. Specifically, each individual classifier provides a binary decision by classifying the ROI as benign or malignant. We initially apply each classification scheme and we count the number of malignant predictions (votes for malignancy). We divide then this number to the total number of votes (the total number of classifiers, herein 10). For example, since we utilized ten different classifiers, if eight of them classify the case as malignant, the final estimate of the ensemble classifier for malignancy is 80%. This is the final output of the CAD_x framework indicating the risk for malignancy for the considered case.

B. Clustered Microcalcifications' Atlases

Apart from the use of shape, cluster, distribution and textural features, we add in this work an extra feature type concerning the location of the cluster inside the breast. In a previous work [12], we have introduced the main principles for the use of location features. Specifically, we proposed a framework including segmentation and determination of important breast landmarks, computation of distances of the centre of the cluster to those landmarks and finally a pipeline leading to the generation of probabilistic maps, that are used to enhance the classification process (fig.1).

The first step in the proposed framework addresses the task of segmenting the breast area from the background in a mammogram. To this end, we applied a segmentation algorithm presented in [15]. It is a simple and fast algorithm which is mainly based on three steps, including analysis of the mammogram's histogram, application of morphological filters and smoothing of breast's border. We applied the specific algorithm to segment the breast area on all the available mammograms, both from CranioCaudal (CC) and MedioLateral (MLO) view. Apart from the breast's silhouette, another important landmark that has to be located is the pectoral muscle, which is a triangular area appeared in the left upper corner of a mammogram and it is visible exclusively on mammograms acquired from the MLO view. We applied a simple algorithm based on locating the boundary pixel between the pectoral muscle and the breast area on an initial

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/JBHI.2014.2334491, IEEE Journal of Biomedical and Health Informatics

line of the muscle. The boundary pixel is considered the pixel which appears the greatest decrease in the average brightness throughout the line. The specific process is repeated from all the lines of the image above and below the initial line, by limiting the search only in neighboring pixels of the boundary pixels defined on previous lines, since the border of the pectoral muscle is usually a straight line or at least a small curve. The criteria followed for the choice of the initial line and the choice of the boundary pixel were proposed in [16]. Finally, for the determination of the nipple, we followed basic principles of the geometry of the breast and we considered as nipple point the point of the breast border with the largest distance to the chest side for CC mammograms and the point of the breast border with the largest distance from the pectoral muscle for MLO mammograms.

After segmenting the above breast landmarks, the second step of the proposed framework concerns the extraction of position features which describe the relative location of the finding inside the breast. These features are presented in fig. 2 for a mammogram from the MLO view. To this end, using the segmented breast area, we compute the breast length and width, the distance of the centre of the cluster to the nipple of the breast (d1) and the angle that is formed between the current line and the horizontal level (θ). The distance of the centre of the cluster to the nipple is computed either relatively to the distance from the chest wall to the nipple, in case of the CC view, or relatively to the distance between the nipple and the pectoral muscle for the MLO view (d2).

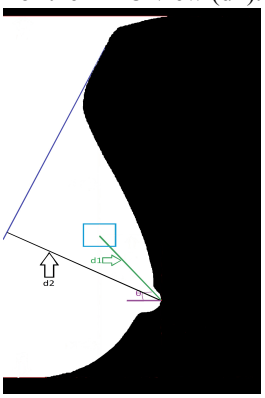

Fig. 2. Description of the relative location of the cluster inside the breast area.

Finally, using the above features, we generate breast atlases. Specifically, we project all the available ROIs on the breast silhouette of a random case which has been chosen to be the shape-container. For each different mammogram we apply the segmentation algorithms discussed earlier and we compute for each ROI the corresponding position features which provide the relative location of the cluster. Therefore, each ROI is projected on the corresponding location into the atlas shape. More specific information on the generation of breast atlases is given in our previous work [14].

Following this method, we may extract occurrence maps, as well as probabilistic maps between benignity and malignancy for each different view (CC/MLO). The achieved occurrence and discrimination maps for different parts of the breast are presented in fig. 3. The values of occurrence maps are normalized to the maximum value of occurrence. For the probabilistic maps, the red color (value=1) indicates higher probability of malignancy.

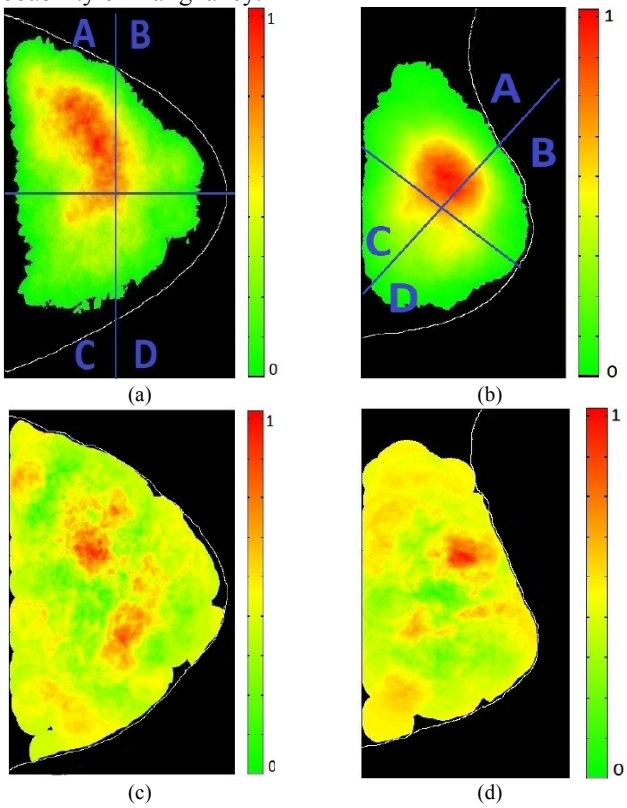

Fig. 3. Occurrence maps for (a) the CC view and (b) the MLO view and the corresponding probabilistic maps for (a) the CC view and (b) the MLO view [14].

Observing the occurrence maps, it seems that there exist indeed areas of the breast with higher occurrence of clusters of MCs. The majority of MCs clusters are located on the upper outer quadrant of the breast. Similarly, the probabilistic maps reveal areas more susceptible to cancer or the opposite, areas with increased probability to host a benign cluster. In this work, we exploit the topological information and proceed on the adoption of the specific maps in the diagnostic process of the CAD_x pipeline, as a new cluster may be projected on the probabilistic maps and therefore a priori probabilities for cancer may be estimated.

C. *Proposed Flowchart*

The classifiers implemented in the final step of the CAD_x pipeline provide a binary prediction for a new case, e.g. whether this case is benign or malignant. Applying all of them into the ensemble classifier discussed earlier, we may obtain an initial risk for the new considered case, namely $RISK_1$ hereafter. Similarly, using the probabilistic maps, we may obtain another risk, namely $RISK_2$, which is based exclusively on the location of the cluster. The main aim of this work is to propose a proper way to combine these two risks, in order to exploit the topological information that the discrimination maps offer and optimize the diagnostic process.

The proposed scenario to combine these two risks is presented in fig. 4. The main idea is to avoid the two risks to contribute equally to the final risk estimation. Instead, we

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/JBHI.2014.2334491, IEEE Journal of Biomedical and Health Informatics

want the $RISK_1$ to play the primary role for the classification of the case, since its computation is the result of several different machine learning algorithms and various feature types (morphology, distribution, texture). The role of the second risk ($RISK_2$) is to modify properly the initial risk ($RISK_1$), taking under consideration exclusively the relative location of the finding.

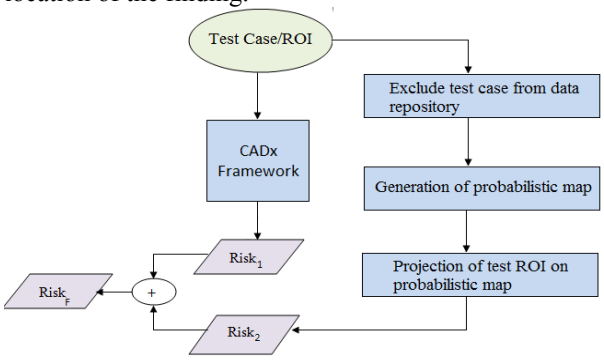

Fig. 4. Proposed flowchart for the combination of CAD_x framework and probabilistic maps.

The second risk is computed as follows: we project the new considered ROI into the corresponding atlas (CC/MLO) in order to extract a pair of probabilities, concerning benignity and malignancy, namely [P_b, P_m], where P_b = 1-P_m. The specific probabilities are computed as the mean value of the corresponding pixels of the atlas on which the current ROI is projected. The first value represents the probability for the projected cluster to be benign, while the second value represents the corresponding probability for malignancy, depending exclusively on the projection of the cluster into the probabilistic atlas. Obviously, whether the risk of benignity (P_b) is greater or smaller than the risk of malignancy (P_m), the $RISK_2$ suggests increased probability towards benignity or malignancy respectively. In order to simulate the contribution of the location of the finding, we compute the final risk ($RISK_F$) through the following equation:

$$RISK_F = RISK_1 + RISK_2 \quad (1)$$

where $RISK_2$ is computed as:

$$RISK_2 = alpha * (P_m - P_b) \quad (2)$$

The risk $RISK_F$ represents the final risk computed as indicated in the flowchart presented in fig. 4. The second term of the above relation ($RISK_2$) concerns exclusively the risk based on the location of the cluster. The parameter alpha may take only positive values and determines the grade that the second risk should contribute to the estimation of the final risk. The second risk works as a shift of the first risk, either by reducing it, when the location of the cluster recommends benignity, or by further increasing it, when the location of the cluster shows increased probability for malignancy. Therefore, the relative location of the cluster is considered as an additional component that may be analyzed independently and contribute to the initial risk estimation, which is based on the analysis of the ROI. We achieve that way to mimic radiologists' strategy, since they co-estimate for their diagnosis the location of the cluster, by splitting the mammograms into different quadrants [9].

We should note here that the classifiers of the CAD_x framework providing the $RISK_1$ have to be evaluated with cases that have not been used during their training. Similarly, a case projected on the map should not have been used previously for its generation. For this reason, the probabilistic maps are slightly differentiated from those presented in fig. 3 and reformed, in order to exclude the case considered, as shown in fig.4. In general, the validation and the evaluation of the proposed flowchart will be carried out with a totally unknown subset of cases, according to the random separation of the dataset that will be discussed in the next section.

D. Data Collection

As mentioned above, we used cases provided by the DDSM, which is the largest database currently available. Almost all cases contain images from both MLO and CC views, while a mammogram may contain more than one annotated ROI. For each ROI, the database provides accompanying files which include the assessment performed by the radiologists according to the BIRADS standard [9], information on the ROI concerning the exact location and annotation of the finding, as well as its pathology status according to the biopsy that has been preceded. We collected almost all the cases of the database that contained annotated clusters of MCs and we ended in a dataset consisted of 843 mammograms from the CC view and 846 mammograms acquired from the MLO view, containing 1715 different ROIs with both benign and malignant clusters. Only few cases have been excluded from the study, such as cases that contained clusters diffused throughout the whole breast area without a local annotation specified by radiologists, cases where segmentation of mammograms was not feasible or cases with unproven pathology results.

In order to ensure both satisfying training and evaluation of the CAD_x framework, we divided the initial dataset consisted of 1715 ROIs into two groups: the training set consisted of 1114 cases (65%) and the evaluation set consisted of 601 cases (35%). Both subsets are randomly chosen, whereas the only assumption is that the training set should be balanced by containing equal number of benign and malignant cases. As mentioned previously, the aim of this work is to propose a flowchart which fuses the CAD_x pipeline with the probabilistic maps. For this reason, we proceed on further separation of the evaluation set into two different subsets. The first subset is small enough containing 120 cases (20%) and will be used as the validation set for the design of the proposed flowchart. The second subset consisted of 481 cases (265 benign-216 malignant) will be used as the test set for the evaluation of the proposed scenario.

Previously, we have indicated the crucial role that the composition of the dataset plays on the classification performance [14]. The grade of difficulty for cluster's analysis is characterized by a value of subtlety (ranging from 1 to 5), provided in the database, where smaller values imply

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/JBHI.2014.2334491, IEEE Journal of Biomedical and Health Informatics

increasing grade of difficulty. The subtlety rating differs from the BI-RADS standard and it is a subjective measure determined by experienced radiologists, who participated in the preparation of the DDSM database. In order to achieve objective results, all three subsets (training, validation and test) have to contain cases of varying subtlety. This demand is satisfied thanks to the random separation of the initial dataset. The distribution of the cases for all three subsets based on the subtlety rating is presented in fig. 5.

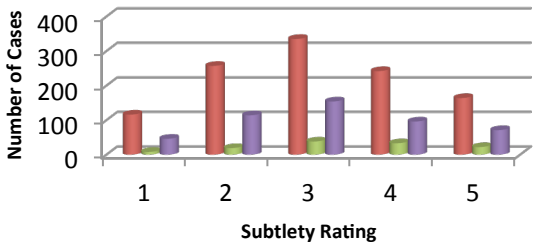

Fig. 5. Distribution of cases based on the subtlety rating for the three considered subsets. The red color represents the train set, the green color the validation set and the purple color the test set.

III. Results

According to the separation of the initial dataset into training, validation and test set we present the results achieved for each different subset. The training set has been used for the training of the classification schemes. During this phase, all classifiers are properly trained with the training set, which is consisted of 1114 cases of varying subtlety. Proper parameterization is followed for each different classifier, so as to find the optimal parameters' values for the classification task. Proper feature selection has also been applied to locate the optimal features' subset. The obtained classifiers are combined in order to form the ensemble classifier which may provide the final risk for malignancy. This final scheme is applied on the cases of both the validation and the test set to evaluate its performance.

A. Performance on the validation set

The cases of the validation set will be used for the evaluation of the general CAD_x classifier, as well as for the validation of the potential of the proposed flowchart to contribute in the optimization of the diagnostic process. The CAD_x pipeline is applied on each of the 120 cases of the validation set. For each case, we record the final risk estimated ($RISK_1$), which indicates the probability for malignancy for the considered case. Then, each case is projected on the corresponding probabilistic map, depending on the view of the mammogram, to estimate the pair of probabilities for malignancy based on the location of the cluster ($RISK_2$). The two risks are then combined following the proposed flowchart in order to estimate the final risk estimation ($RISK_F$).

Initially, we focus on the $RISK_1$ for the cases of the validation set, and we extract the corresponding ROC curve to estimate the A_z metric (area under curve), which is the best-suited metric for two-class classification problems. Considering the CAD_x framework and the initial risk $RISK_1$, we found that the achieved A_z value is equal to 0.789 (fig. 6). When applying only the Gaussian rbf SVM classifier (implemented in the previous version of the CAD_x scheme) the A_z value achieved is equal to 0.752. This fact indicates that the application of all the ten classifiers in an ensemble scheme improved the classification results.

We proceed then on the validation of the proposed flowchart. Firstly, we study the influence of the parameter alpha in the classification results. We choose different values ranging from (0, 2] with steady step 0.05. We search for the specific value which optimizes the classification performance. We observed that the maximum A_z value is achieved for the value 0.8 of the parameter alpha.

Using the specific value, we applied the flowchart to compute the final risk ($RISK_F$). We extracted the ROC curve and we observe that the probabilistic maps contribute positively in the improvement of the diagnostic performance, since the A_z value of the ROC curve is improved from 0.789 to 0.861. Both ROC curves, the former for the $RISK_1$ and the latter for the $RISK_F$, are presented in fig. 6.

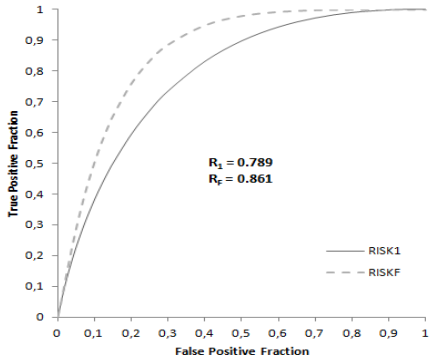

Fig. 6. ROC curves for the initial risk (RISK1) and the final risk (RISKF) using the validation set.

We may use the specific ROC curve to perform a straightforward comparison between the current CAD_x framework and the radiologists who assessed the same cases of the validation subset. We remind that the DDSM database provides also the assessment for each different case that has been performed by experience radiologists, according to the BIRADS standard [9]. Based on the final diagnosis performed by the radiologists using the BIRADS score, we follow the procedure proposed in [17] for the extraction of the ROC curve of the radiologists. In fig. 7, we provide both ROC curves for the validation set in a common graph. The difference between the two A_z values is almost 0.072, which indicates the potential of the proposed scenario to support the radiologists' diagnostic process performed in the real clinical practice.

Until now, we based the comparisons performed on the A_z value extracted from the ROC curves. In order to estimate important metrics such as the accuracy, sensitivity and specificity, we analyze further the ROC curve in order to locate the point which optimizes the discrimination between benign and malignant cases. According to [18], the optimal point of the ROC curve should be the point that has the

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/JBHI.2014.2334491, IEEE Journal of Biomedical and Health Informatics

shortest distance to the upper left of the axes, that means the point (0,1). Following this method, the optimal threshold is selected to be the percentage 27%. That means, when the final risk computed is equal or greater than 27% the case is classified as malignant, otherwise is classified as benign. However, the choice of the threshold may be slightly differentiated, in order to be in accordance to the requested behavior of the system, depending on whether we aim at high accuracy, sensitivity or specificity respectively. Concerning a CAD system, we aim in general at high sensitivity, so as to limit the number of missed cancers. For this reason, we locate the point that provides enough high sensitivity (sensitivity≥0.95), providing at the same time the greatest possible specificity. Satisfying this demand, we select as threshold the percentage 17%.

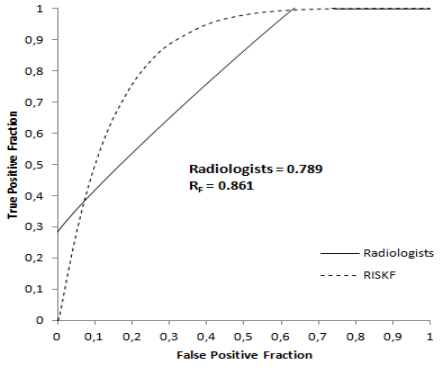

Fig. 7. ROC curves for the final risk ($RISK_F$) and the radiologists using the validation set.

Using these two different thresholds, we proceed on the computation of the accuracy, sensitivity and specificity. For reasons of comparison, we compute the same metrics for the radiologists, using the BIRADS assessment [9]. If a case is classified as BIRADS 2 or 3, the case is considered benign. On the contrary, if a case is classified as BIRADS 0, 4 or 5, the case is considered malignant. Cases classified as BIRADS 1 do not exist in the current manuscript, since we examined cases with annotated findings (clusters of MCs) and not normal mammograms. The obtained results are presented in Table I.

TABLE I
CAD_X AND RADIOLOGISTS PERFORMANCE ON THE VALIDATION SET

	Accuracy	*Sensitivity*	*Specificity*
CADx (threshold 27%)	0.817	0.850	0.783
CADx (threshold 17%)	0.808	0.950	0.667
Radiologists	0.691	1.000	0.383

We may observe that sensitivity between CAD_X and radiologists is comparable on the validation set. Especially when considering the threshold 17%, we see that both achieve very high sensitivity. However, the accuracy and the sensitivity provided by the CAD_X framework are greater than the corresponding values achieved by the radiologists.

B. Performance on the test set

Using a small validation subset at the previous section, we determined a flowchart that led to improvement of the performance (A_Z=0.845). However, the validation set is quite small to allow us to extract conclusions with high generalization ability. For this reason, we proceed on the evaluation using the test set consisted of 481 cases, which have not been used at all either during the training or the validation procedure. Therefore, we apply for each different case of the test set the CAD_X ensemble classifier, as well as the proposed flowchart presented in fig. 4, using the recommendation for the exact location of the cluster. The corresponding ROC curves are presented in fig. 8 respectively. We observe that the contribution of the probabilistic maps is also positive on the test set, since the performance is improved from 0.746 to 0.786.

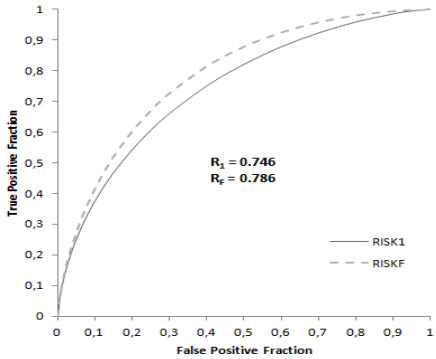

Fig. 8. ROC curves for the initial risk (RISK1) and the final risk (RISKF) using the test set.

As in the case of the validation set, we proceed on a straightforward comparison with the radiologists who assessed the mammograms of the test set. We generated the ROC curve for the radiologists presented in fig. 9 accompanied by the ROC curve of the CAD_X framework.

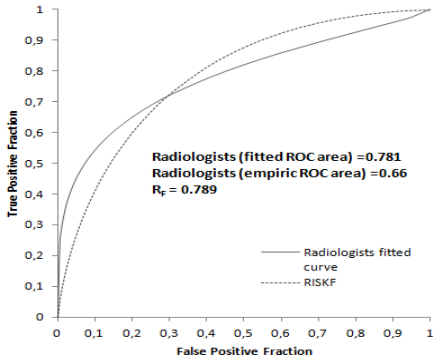

Fig. 9. ROC curves for the final risk ($RISK_F$) and the radiologists using the test set.

Due to the fact that BIRADS score is a discrete value, the ROC curve has to be fitted. However, this fitting leads to optimistic results for the classification of human experts. Specifically, we observe a quite large increase between the empiric and the fitted ROC area. The former value is the most representative evaluation criterion for the radiologists and it will be used thus as the basis for the proper comparison between the radiologists and the CAD_X. It is obvious that there

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/JBHI.2014.2334491, IEEE Journal of Biomedical and Health Informatics

is a quite importance difference (0.13) between these two A_z values.

We also compare the CAD_x system and the radiologists in terms of accuracy, sensitivity and specificity. To extract the specific measures, we use the two thresholds (27% and 17%) obtained during the validation phase and the analysis of the ROC curve. The specific results are presented in Table II. An important observation that can be made is that the CAD_x framework is able to provide better accuracy than the corresponding achieved by the radiologists. The sensitivity remains at high levels, while the specificity is enough greater than the corresponding achieved by the radiologists.

TABLE II
CAD_X AND RADIOLOGISTS PERFORMANCE ON THE TEST SET

	Accuracy	*Sensitivity*	*Specificity*
CADx (threshold 27%)	0.713	0.764	0.672
CADx (threshold 17%)	0.690	0.866	0.547
Radiologists	0.514	0.935	0.166

IV. DISCUSSION

In this work, we try to enhance the pipeline of a CAD_x framework in order to improve the diagnostic performance. Two are the main major contributions proposed in this work. The first concerns the implementation of several classification schemes and their combination into an ensemble classifier. As a result, the final output of this ensemble classifier is a continuous risk, a percentage ranging from 0 to 100, and not the binary prediction provided by each individual classifier, which consequently may be combined more easily with the probabilities that are computed afterwards from the probabilistic maps. We implemented various classification schemes, in order to ensure a satisfactory diversity in the classification rules and combined them in a powerful ensemble that exploits the individual strengths of the base classifiers, reducing at the same time their individual weaknesses. The second contribution concerns the use of topological information to support diagnosis. The novelty of this work is that we do not compute location features to feed the aforementioned classifiers, but we propose the use of such features to generate probabilistic maps. Specifically, we adopt the use of breast probabilistic maps, as a new case may be projected on them and a priori estimation for the risk of malignancy may be computed. The specific a priori risk is combined properly with the initial risk provided by the CAD_x framework to obtain the final risk estimation. Therefore, we consider the analysis of the ROI and the location of the cluster inside the breast as two independent diagnostic components. They are studied separately and the diagnostic information that each component provides is then properly combined. We manage to mimic that way the strategy that the radiologists follow when examining a new mammogram, since they analyze independently the morphology of the findings in the ROI and the relative location of the cluster inside the breast.

For the evaluation performance, we used a great variety of mammograms provided by the DDSM database. To the best of our knowledge, it is the largest dataset that has been used in CAD_x diagnosis for breast MCs. The random separation of the data into training, validation and test set serves respectively the generation of classifiers with high generalization ability, the validation of their potential and the formulation of the proposed framework and finally the objective evaluation of the proposed methodologies using a totally unknown subset.

Using the validation set, we investigate the diagnostic performance of the ensemble classifier and the achieved A_z value is 0.789. We introduce then the adoption of the generated probabilistic maps, as presented in fig. 4, following equations (1) and (2). We have noticed from the values of the probabilistic maps that the maximum difference between P_b and P_m values is about 35%. Therefore, a value of 0.8 for alpha indicates a possibility for a maximum shift in the final risk of about 0.8*35%=28%. This value is capable to shift the output to a different grade of malignancy risk. Indeed, we showed that the adoption of the probabilistic maps appeared positive contribution to the diagnostic performance, as it may provide an increase equal to 0.072 on the classification performance on the validation set. These initial results are quite encouraging, but the small size of the validation set forces us to repeat the measurements on a larger dataset.

For this reason, we exploit the final available dataset, the test set, consisted of 481 cases, which consists a very large dataset, enabling us to extract secure conclusions. The developed methodologies are sequentially applied on each case of the test set. Again, observing fig. 8, it seems that the proposed framework performs better than the initial ensemble classifier, as the improvement achieved is equal to 0.04. The final conclusion is that the adoption of probabilistic maps presents positive influence to the classification performance, both on the validation and test set.

The fact that the mammograms used in this study have been previously assessed by radiologists enabled us to perform a straightforward comparison between them and the CAD_x framework. Analyzing the ROC curves and comparing the corresponding performance in terms of sensitivity, specificity and accuracy, according to the results presented in Tables I and II, we conclude that the CAD system may retain the high levels of sensitivity by improving at the same time the specificity achieved. It is noticeable that the accuracy of the radiologists is 51.4% on the test set. In other words, almost only one out of two cases is correctly classified, mainly due to the fact that the radiologists pursuit almost perfect sensitivity so as to avoid missed cancers. This demand is satisfied by the CAD_x system, since the sensitivity obtained on the test set is slightly decreased (0.866 instead of 0.935), while the specificity achieved is about three times greater than the corresponding value achieved by the radiologists.

Despite the fact that we use DDSM data, it is difficult to perform a straightforward comparison with previous studies in the field. The reason is that the majority of the previously published studies did not use the whole set of cases but a

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/JBHI.2014.2334491, IEEE Journal of Biomedical and Health Informatics

random subset, which is not always determined. Instead, in the present study, we have incorporated almost the whole subset containing clusters of MCs (1715 cases) in order to perform large-scale evaluation. Many studies report high values of A_z (A_z>0.85), but for a small number of cases (cases<150), without discussing important factors such as the subtlety of the cases [4]. In a previous study [14], we have studied various subsets of mammograms and we have located a subset consisted of 256 cases where we reported A_z equal to 0.909. When implementing the proposed methodologies on the whole database (1715 cases) we reported a value equal to 0.75 [5]. To the best of our knowledge, few studies have used the whole database or even a well-defined subset larger than 1000 mammograms of DDSM. Hapfelmeier et al. [19] have used 1314 cases and reported A_z=0.78, while Pereira et al [20] mention that have used the whole database and reported A_z=0.607, without declaring the exact number of cases.

Despite the fact that the DDSM database is the largest publicly available database and the most popular choice for the development of CAD methodologies, it has the disadvantage that contains only cases of screen film mammograms. Unfortunately, no comparable reference database of modern full-field-digital mammograms (FFDM) is publicly available today, so as to exploit it and evaluate the proposed CAD methodologies. Taking under consideration the results of the current study and the encouraging conclusions extracted from the comparison between CAD_x and radiologists performance, we expect that the application of the proposed methodologies in FFDM along with the improved image quality will lead to an enhanced diagnostic pipeline and improved classification performance.

V. Conclusions

The main question investigated in this work is the way that we may expand CAD_x methodologies in order to optimize the diagnostic process. The proposed methods are mainly focused on the final output of the CAD_x pipeline. Instead of simply applying a trained classification scheme on new upcoming cases, we propose a scenario where extra classification schemes are applied to provide a percentage risk for the malignancy, while this initial estimated risk is empowered according to probabilistic breast maps, which provide topological information to support the diagnosis. The results provided reveal the valuable diagnostic information and the beneficiary contribution in the CAD_x pipeline that may be obtained by exploiting the probabilistic maps proposed. It is of great importance the fact that the performance achieved outperforms the corresponding achieved by the radiologists.

The specific conclusions determine the first steps towards the optimization of the diagnostic process of a CAD system for breast MCs. It seems that a set of rules which combines alternative classifiers, probabilistic breast maps and various feature types may perform better than individual classifications schemes, offering at the same time a more completed diagnostic process for the radiologists. This work will be used as the proper baseline to work towards the refinement of the diagnostic procedure in order to make it applicable in clinical routine. Our future steps include the adoption of radiologists' recommendations and data from the medical folder of the patient, in order to form a scenario for the CAD_x use, which will simulate the interaction between CAD_x and radiologists in daily clinical practice.

References

[1] M. Lanyi, " Microcalcifications in the breast—a blessing or a curse? A critical review," *Diagn. Imaging Clin. Med.*, vol. 54, pp. 126–145, 1985.

[2] H. H. Ng and M. Muttarak, "Advances in mammography have improved early detection of breast cancer," *J. HK Coll. Radiol.*, vol. 6, no. 3, pp. 126–131, 2003.

[3] M. Giger, H. Chan and J. Boone "Anniversary paper: history and status of CAD and quantitative image analysis: the role of medical physics and AAPM," *Med. Phys.*, vol. 35, pp. 5799–5820, 2008.

[4] M. Elter and A. Horsh, "CADx of mammographic masses and clustered microcalcifications: a review," *Med. Phys.*, vol. 36, pp. 2052–2068, 2009.

[5] I. Andreadis, G. Spyrou and K. Nikita, "A comparative study of image features for classification of breast microcalcifications," *IOP Meas. Sc. and Tech.*, vol. 22, no. 11, 2011.

[6] A. Karahaliou, S. Skiadopoulos, I. Boniatis, P. Sakellaropoulos, E. Likaki, G. Panayiotakis, and L. Costaridou, "Texture analysis of tissue surrounding microcalcifications on mammograms for breast cancer diagnosis," *Br. J. Radiol.*, vol. 80, pp. 648-656, 2007.

[7] M. Kallergi, "Computer-aided diagnosis of mammographic microcalcification clusters," *Med. Phys.*, vol. 31, pp. 314-326, 2004.

[8] A. Papadopoulos, D. Fotiadis, A. Likas, "Characterization of clustered microcalcifications in digitized mammograms using neural networks and support vector machines," *Artif. Intell. Med.*, vol. 34, pp. 141-150, 2005.

[9] American College of Radiology (ACR), "ACR Breast Imaging Reporting and Data System, Breast Imaging Atlas", 4th Edition, Reston, VA. USA, 2003.

[10] W.J. Veldkamp, N. Karssemeijer, J. D. Otten, and J.H. Hendricks, "Automated classification of clustered microcalcifications into malignant and benign types," *Med. Phys.*, vol. 27, pp. 2600-2608, 2000.

[11] D.B. Russakoff and A. Hasegawa, "Generation and Application of a Probabilistic Breast Cancer Atlas," In: Larsen, R., Nielsen, M., Sporring, J. (eds.) MICCAI 2006, LNCS, 4191, pp. 454-461, 2006.

[12] I. Andreadis, G. Spyrou, P. Ligomenides and K. Nikita, "Generation of clustered microcalcifications' atlases for benign and malignant cases," presented at the 13th Int. IEEE Conf. BioInformatics and BioEngineering, Chania, Greece, Nov. 10-13, 2013.

[13] M. Heath, K. Bowyer, D. Kopand, R. Moore and W. Kegelmeyer. "The Digital Database for Screening Mammography,"in Proc. of the 5th IWDM, Yaffe M. Medical Physics Publishing, 2001, pp. 212-218.

[14] I. Andreadis, G. Spyrou, P. Ligomenides and K. Nikita, "Variations on breast density and subtlety of the findings require different computational intelligence pipelines for the diagnosis of clustered microcalcifications," presented at the 13th Int. IEEE Conf. BioInformatics and BioEngineering, Chania, Greece, Nov. 10-13, 2013.

[15] T. Ojala, J. Nappi, and O. Nevalainen, "Accurate segmentation of the breast region from digitized mammograms," *Comput. Med. Imag. Graph.*, vol. 25, pp. 47-59, 2001.

[16] L. Wang, M. Zhu, L. Deng and X. Yuan, "Automatic pectoral muscle boundary detection in mammograms based on Markov chain and active contour model", *J Zhejiang Univ-Sci C*, vol. 11, no. 2, pp. 111-118, 2010.

[17] A. Karahaliou, S. Skiadopoulos, I. Boniatis, P. Sakellaropoulos, E. Likaki, G. Panayiotakis and L. Costaridou, " Texture analysis of tissue surrounding microcalcifications on mammograms for breast cancer diagnosis," *Br. J. Radiol.*, vol. 80, pp. 648-656, 2007.

[18] R. Kumar and A. Indrayan, "Receiver Operating Characteristic (ROC) Curve for Medical Researchers," *Indian Pediatrics*, vol. 48, pp. 277-287, 2011.

[19] A. Hapfelmeier and A. Horsch, "Image feature evaluation in two new mammography CAD prototypes", *Int J CARS*, vol. 6, pp. 721-735, 2011.

[20] R.R. Pereira, P. M. Azevedo Marques, M. O. Honda, S. K. Kinoshita, R. Engelmann, C. Muramatsu and K. Doi, "Usefullness of texture analysis for computerized classification of breast lesions on mammograms," *J. Dig. Imag*., vol. 20, pp. 248-255, 2007.