Real time mass classification for mammographic images: a Driven CADx scheme

Classificação de massa em tempo real para imagens mamográficas: um esquema Driven CADx

DOI:10.34119/bjhrv6n3-429

Recebimento dos originais: 23/05/2023
Aceitação para publicação: 26/06/2023

Homero Schiabel
Ph.D. in Science
Institution: Escola de Engenharia de São Carlos – Universidade de São Paulo (USP)
Address: Av. Trabalhador Sãocarlense, 400, São Carlos – SP, CEP: 13566-590
E-mail: homero@sc.usp.br

Bruno Roberto Nepomuceno Matheus
Ph.D. in Science
Institution: Escola de Engenharia de São Carlos – Universidade de São Paulo (USP)
Address: Av. Trabalhador Sãocarlense, 400, São Carlos – SP, CEP: 13566-590
E-mail: bruno.matheus@gmail.com

Fernanda Junqueira Fortes Cardoso
Graduate student in Electrical Engineering
Institution: Escola de Engenharia de São Carlos – Universidade de São Paulo (USP)
Address: Av. Trabalhador Sãocarlense, 400, São Carlos – SP, CEP: 13566-590
E-mail: fernandajunqueirafortes@usp.br

ABSTRACT

Computer-Aided Diagnosis (CADx) schemes have been proposed to serve as a supplementary image analysis tool in mammography. Experienced radiologists tend to be more assertive to such schemes in assisting their interpretation rather than solely relying on their ability to detect suspicious signals. This study focuses on a simplified version of a previously developed mammography CADx scheme, which was initially designed for digitized film, but is now specifically aimed at classifying breast nodules marked as regions of interest on digital images. This "driven" CADx scheme provides prompt indications regarding whether the selected nodule is deemed normal or suspicious. Its performance was evaluated through tests conducted on different mammograms sets – one with large number of images selected from DDSM database for training, testing and validation of classification parameters, and other comprising direct digital images from InBreast database. Remarkably, similar rates were observed for sensitivity, specificity and accuracy across these two sets (83%, 67% and 72%, respectively). The classification attributes were associated to contour, density and texture. Furthermore, a third test was conducted involving radiologists analyzing digital mammograms obtained from a specific full field digital mammography (FFDM) unit. Results showed that the Driven CADx scheme positively influenced the final diagnoses made by 3 radiologists, consistently increasing accuracy rates. This promising result allows establishing this software as a valuable tool for radiologists in the analysis of masses in digital mammography. The scheme can be implemented on any operating system, or even accessed online.

Keywords: CADx scheme, digital mammography, digital mammogram processing, mass classification in mammography.

RESUMO

Esquemas de Diagnóstico Auxiliado por Computador (CADx) foram propostos para servir como uma ferramenta suplementar de análise de imagem em mamografia. Radiologistas experientes tendem a ser mais assertivos em relação a esses esquemas para auxiliar sua interpretação, em vez de confiar apenas em sua capacidade de detectar sinais suspeitos. Este estudo se concentra em uma versão simplificada de um esquema CADx de mamografia desenvolvido anteriormente, que foi inicialmente projetado para filmes digitalizados, mas que agora se destina especificamente à classificação de nódulos mamários marcados como regiões de interesse em imagens digitais. Esse esquema CADx "orientado" fornece indicações imediatas sobre se o nódulo selecionado é considerado normal ou suspeito. Seu desempenho foi avaliado por meio de testes realizados em diferentes conjuntos de mamografias - um com grande número de imagens selecionadas do banco de dados DDSM para treinamento, teste e validação dos parâmetros de classificação e outro com imagens digitais diretas do banco de dados InBreast. Notavelmente, foram observadas taxas semelhantes de sensibilidade, especificidade e precisão nesses dois conjuntos (83%, 67% e 72%, respectivamente). Os atributos de classificação foram associados a contorno, densidade e textura. Além disso, um terceiro teste foi realizado com radiologistas analisando mamografias digitais obtidas de uma unidade específica de mamografia digital de campo total (FFDM). Os resultados mostraram que o esquema Driven CADx influenciou positivamente os diagnósticos finais feitos por três radiologistas, aumentando consistentemente as taxas de precisão. Esse resultado promissor permite estabelecer esse software como uma ferramenta valiosa para os radiologistas na análise de massas em mamografia digital. O esquema pode ser implementado em qualquer sistema operacional, ou mesmo acessado on-line.

Palavras-chave: esquema CADx, mamografia digital, processamento de mamografia digital, classificação de massas em mamografia.

1 INTRODUCTION

The most comprehensive large-scale screening method available is the mammographic exam. Unfortunately, about 20% of mammograms are false negative and up to 50% are false positive [1,2]. To minimize diagnostic errors, two radiologists usually review multiple images, and when this is not available, a Computer-Aided Detection or Diagnosis (CADe\Dx) scheme can be used [3]. Most computer-aided systems in this field are actually CADe (Computer-Aided Detection), ie., they detect and inform the location of suspicious signals on the image. They have limited use, as tests have shown that most mammography errors are caused by misinterpretation and not by missed detections [2, 4, 5]. Experienced radiologists tend to be more assertive about CAD schemes to help their analysis rather than their ability to detect suspicious signals, mainly due to the value of information about quantitative density data and and other findings [5].

The current model of our CADx scheme is based on these aspects [6]. The main characteristic is that it represents not an automatic diagnosis computer system in mammography, but a supplemental information system for the medical report. It corresponds to an integration of independent modules by using JAVA® language. The main class is called CADxLAPIMO. The ImageJ® library [7] is used to perform the opening and primary manipulation of mammography images files. Methods for selecting and cropping ROIs (regions of interest) and methods of marking the image are also included. In Pre-processing, the focus is a transformation method based on the digitizer characteristic curve ("BCC Correction" [8]), which aims at removing variations produced by the digitization process, thus reducing variations in image brightness. Previous work [9] showed the effects of this correction also in human observation, in a comparative test with other more classic pre-processing techniques for 74 images from BancoWeb [10] and DDSM [11] databases, all containing at least one unmarked signal to be used as reference. Results confirmed that the trend detected in previous CAD evaluations also happened in direct observations, with experts usually preferring images transformed by BCC correction [8].

In another previous work [12] we showed and discussed the evaluation of one of the modules of our CADx scheme – the mass segmentation step – comparing the module results with the interpretation of experienced radiologists. The evaluation essentially consisted of comparing the classification of the nodules contours given by the scheme and those considered by radiologists in order to verify not only the level of effectiveness of the automatic classification but also to show how this result can influence the radiologist evaluation. Considering the separation between benign and malignant signals in the nodule contour classification, results have indicated 82% of agreement between CADx and radiologists [12].

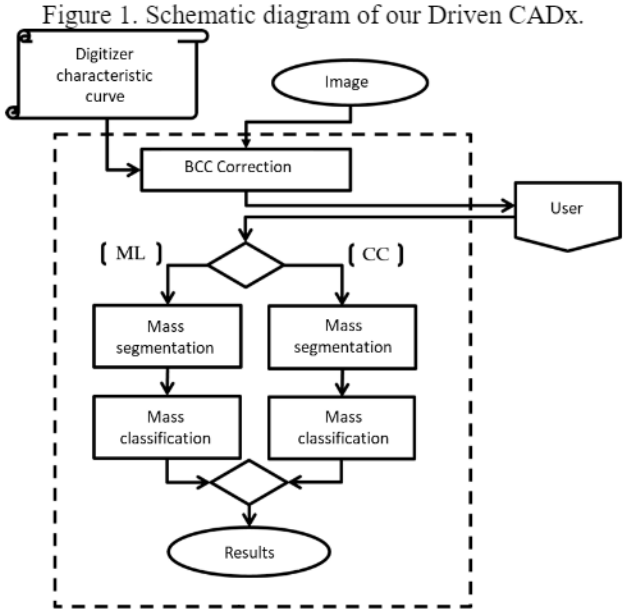
However, there are important limitations of complete CADx schemes application – ranging from the time required for the whole processing, through the high probability of dealing with false positive findings, until the classification of excessive signals. This usually causes the radiologist to lose confidence in the results and consider the computational scheme as a tool that hinders the workflow of the diagnostic analysis procedure [13]. It is considered excessive time for CADe/Dx processing when the delay is considerably longer than the time for analysis by an expert. Based on studies by Haygood [14], an experienced radiologist takes between 89 and 194 seconds to evaluate a complete mammogram. Therefore, times greater than 200 seconds (≈ 3 minutes) per exam may make it difficult to use the scheme on a large scale; times greater than 400 seconds (≈ 7 minutes) per exam are impractical, as they may represent a longer time than using another specialist for a second opinion.

Radiologists usually detect suspicious regions correctly, but misclassify them in up to 15% of cases, either false-positive or false-negative. In addition, recent researches [5, 15, 16] show that radiologists prefer to use the CAD scheme as an on-demand assistance tool, and to yield results as quickly as possible. These aspects are motivation for our implementation, aiming to allow the application of the scheme in clinical practice. This article presents a new "driven" CADx (Computer-Aided Diagnosis) scheme, providing software that could be used in hospitals or radiology clinics as a second opinion for classifying masses in mammography. This system is only aimed at classification; the radiologist is requested to select the region containing the mass in the two views of mammography and, once the region of interest (ROI) is selected, the system instantly displays the result, as a suspicious or normal signal.

2 MATERIALS AND METHODS

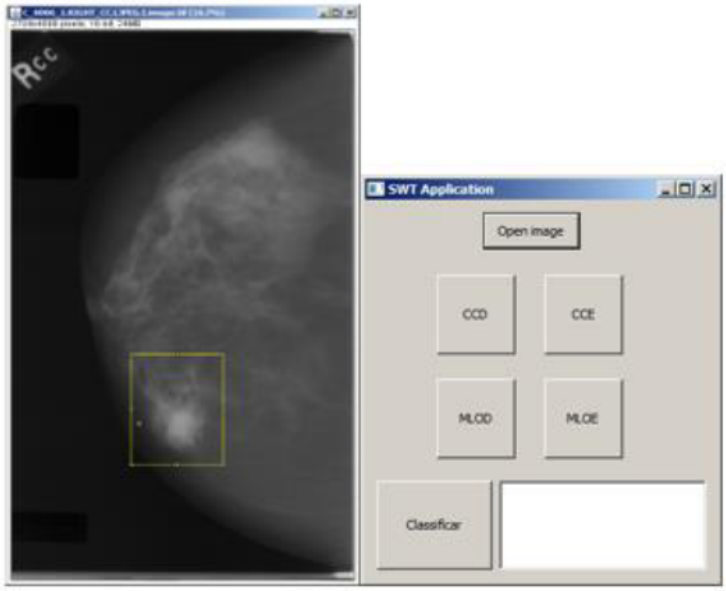
A simplified version of our CADx scheme [9, 12] was designed: the digital image is pre-processed [8] and displayed to the radiologist, who must select a particular ROI for the automated classification. The software thus classifies only the existing nodules in that ROI; it was designed to work with the same region delimited in both images corresponding to two typical projections (CC and MLO) of a given breast under investigation. Both ROIs (each relative to each projection) are ranked and a single result is shown as outcome: "suspect" or "normal". This simplified model is shown in Fig. 1. Pre-processing output images are exhibited to the radiologist through an interface extended from the subclass attribute integrated to the ImagePlus class used by the ImageJ® [7] library to display images (Fig. 2). In this window (Fig. 2a) the contrast and brightness of each image can be controlled independently. A region of the image where the nodule was detected can also be selected for automated analysis. This selected region should be delimited as well as its correspondence on the other projection of the same breast. Along with the displayed images, the software provides an independent control menu (Fig. 2b) with the function of driving the selected ROIs processing. These ROIs are then segmented using the level set method [17].

Figure 1. Schematic diagram of our Driven CADx.



Source: Authors

Figure 2. Windows displayed to the user: (a) a mammographic image for processing – the yellow rectangle represents a selected ROI; (b) control menu available together with the mammographic images.



Source: Authors

2.1 MASS SEGMENTATION

Each preselected ROI is segmented by using a direct application of level set method [17], resulting in a binary image with the white (true) values being the segmented mass and the black (false) values being the background. Two mathematical techniques of opening and closing morphology were used to overcome noise in images. Both were configured to use a disc-shaped structuring element with diameter 7 (pixels). This results in smoothing edges in the binary image and removal of noise (Figure 3). Once the segmentation is completed, the original

ROI is multiplied by the segmented image. The mass is saved in grayscale and the background is just black (value 0). This is the "Mass Only Image" (MOI), as shown in Fig. 3(c). The original ROI is also multiplied by the inverse of the segmented image, creating another one with a grayscale background, but with the mass being black (value 0). This image will be referred to here as the "Background Only Image" (BOI), as shown in Fig. 3(d).

Figure 3. Segmentation. (a) ROI; (b) binary image; (c) mass only image (MOI); (d) background only image (BOI).

(a) (b) (c) (d)

Source: Authors

2.2 FEATURES EXTRACTION

Once the image is divided into mass and background, the classification can proceed with features extraction. The scheme described here is a modification of that previously described by Ribeiro et al [18], using features as: (a) Variance [19], (b) Difference Entropy [19], (c) Compactness, (d) Mass Density and (e) Background Density. Features as Mass Density and Background Density have been combined into a single attribute called Relative Density, calculated as the ratio of Mass to Background Densities. The first two were defined by Haralick [19] to help texture characterization. They are obtained from the MOI, so that the background texture does not influence the result. Variance [18] is the statistical equivalent of standard deviation for images, evaluating the average difference between the mean value and the value of each pixel. It is calculated according to Equation 1. On the other hand, Difference Entropy [19] does not have a common statistical equivalent or a simple definition, being more easily considered only as a mathematical attribute that can help to characterize the texture. Difference Entropy can be calculated by Equations 2 and 3.

$$\sigma^2 = \sum_{i=0}^{G-1}(i-\mu)^2 p(i) \quad (1)$$

$$E_D = -\sum_{i=0}^{G-1} p_{|x-y|}(i) log\left[p_{|x-y|}(i)\right] \quad (2)$$

$$p_{|x-y|}(k) = \sum_{i=0}^{G-1}\sum_{j=0}^{G-1} p(i,j), \begin{matrix} k = |i-j| \\ k = 0,1,2..G-2 \end{matrix} \quad (3)$$

Since

i is an intensity level, μ is the average of the region and $p(i)$ is the probability of a given pixel having value i.

Compactness describes the mass contour, assessing how close to a circle a given shape is. The minimum value for Compactness is one, representing a perfect circle. As the value goes higher, the object is further from a circle. This attribute is calculated according to Equation 4.

$$C = \frac{P}{2\sqrt{\pi A}} \quad (4)$$

Where

P is the perimeter and A is the area of the mass. Both characteristics are measured using the binary segmented image, by counting the number of white (true) 4-neighbors borders as perimeter and the total number of white (true) values as the area.

Finally, Relative Density is the ratio between the mass density and the background density, as shown in Equation 5.

$$D_{Relative} = \frac{D_{Mass}}{D_{Background}} \quad (5)$$

Both densities are measured using the MOI and BOI, respectively, by calculating the average non-black value for each image. It is important to notice that this measurement is an actual calculation of radiographic densities of the mass or of the background.

3 RESULTS

3.1 CLASSIFICATION ANALYSIS AND VALIDATION

After defining the features to be considered for classifying the mass, validation tests were carried out. Firstly a set of images acquired from the DDSM database [11] was selected with some criteria: images acquisition from the same equipment and digitizer; and images containing only one mass visible in both views (craniocaudal – CC – and mediolateral oblique – MLO) for the same breast – in order to reduce interference from other findings. Thus, 568 images digitized by a Lumisys200 scanner (12 bits of contrast resolution and 50µm of spatial resolution) were chosen to compose the database for the classification tests.

Masses to be segmented and classified in these images were selected from the respective contours determined in DDSM files [11] as ROIs. Therefore, each ROI was processed as

previously described in sections 1 and 2 above. Taking as reference the malignancy information provided by the DDSM files, a ROC curve was determined for each feature in order to evaluate the best thresholds for classification, based on the cost-benefit ratio as described below and in an appendix at the end of this paper, where these curves are also illustrated.

Each feature will yield a classification such as “suspect” or “normal”. The system classifies the finding according to a set of these attributes, indicating such mass as suspicious or not. If the mass is visible in both projections of the image (CC and MLO), all 8 attributes are considered together – 4 of each segmented mass in each projection – and the classification results are used for the mass as a whole. An initial classification was obtained from each of the previously considered attributes. For a given nodule to be considered suspicious, a subset of these attributes must classify it as well; the size of this subset was defined by testing the 8 possibilities (minimum of 1 to 8 attributes classifying as *suspect*) and evaluating the sensitivity and specificity in each case.

Selected images were automatically pre-processed (using BCC Correction [8]), segmented and classified using the two projections (CC and MLO views) of the same nodule. The final classification was given through the best result obtained in the analysis of the number of matching attributes, as described in the previous section. The CADx result was then compared to the database information, and then sensitivity, specificity and accuracy rates were calculated. During the tests, the processing time of each one of 568 images was recorded, and an average of the processing time for different nodule sizes was calculated.

Although using images from the DDSM database [11] – the largest mammographic images database available to the public –, is adequate for large-scale experiments due to findings delimitations, some problems were noted mainly related to the images contrast. Thus, a study of the attributes used in the method was developed to confirm which ones should be used in the classification. Each attribute (Difference Entropy, Variance, Compactness and Relative Density) was tested independently using all 568 selected ROIs, through ROC curves (illustrated in the Appendix). The variant parameter for determining the ROC curves was the classification threshold: the value from which a certain attribute changes the result from “normal” to “suspect”. For this analysis, a true positive result (TP) was considered when the classifier pointed out a malignant nodule (according to the database report) as "suspect", and a true negative (TN) when the classification was "normal" for a benign signal. Consequently, false positive (FP) and false negative (FN) corresponded to errors related to benign and malignant cases, respectively. The ROC curves were generated using the Sensitivity x (1 - Specificity) ratio for different thresholds of the classification attributes. The closest point to the

left upper limit of the ROC curve was chosen as the best cost-benefit ratio for each attribute as illustrated in the Appendix and these ratios are shown in Table 1.

Table 1: Results for attributes evaluation considering the best cost-benefit ratio.

Attribute	**Sensitivity**	**Specificity**
Diference Entropy	76%	64%
Variance	57%	69%
Compacity	58%	52%
Relative Density	63%	54%

Source: Authors

In Table 2, Sensitivity, Specificity and Accuracy are shown for each possibility of minimum agreement between the four evaluated attributes. For example, the column with *2+ agreement* represents the results when a minimum agreement of 2 attributes indicates the signal as suspect. As the number of attributes is small, it is possible to calculate the result for all four possibilities of agreement and analyze which is the best result for the scheme.

Table 2: Results for different attributes agreements when classifying as "suspect".

	Agreement			
	1+	**2+**	**3+**	**4+**
Sensitivity	100 %	82 %	62 %	13 %
Specificity	8 %	52 %	83 %	96 %
Accuracy*	36 %	65 %	77 %	71 %

* Accuracy: represents the software-hit ratio, that is, its hit percentage, regardless of the correct classification. It can be calculated as the (TP + TN) / (number of samples)

The result for an agreement of at least 3 attributes seems to be the best, as it produced 77% accuracy, the highest value in the comparison. However, the lower sensitivity compared to *2+ agreement* raises doubt about which threshold should be used. As the purpose of this classifier is to be used for the evaluation of a possible breast cancer suspicion, reducing the possibility of a false-negative result is essential. Therefore, the classifier uses the agreement of at least two attributes to declare the case as *suspect*. It is important to emphasize that, despite this decision, the purpose of this scheme is to provide the radiologist with as much information as possible, including the number of attributes that declared the case as suspicious.

Validation was performed using the leave-one-out method, separating one case at a time and reassessing the ROC curves and results for different agreements. The results were very close to those obtained in training. To ensure that the technique was being evaluated correctly, the ten-fold technique was used, in which the set of images is divided into 10 random parts, with 9 used for training and the tenth for testing. Tests were repeated until all 10 sets were tested. The average results were also very close to the training results shown above. When the

two projections are available and selected, the classifier evaluates both together, analyzing the results of 8 attributes according to data shown in Table 3. Similarly to the case for a single projection, the best accuracy is shown in the *5+* and *6+ agreements*, but a higher Sensitivity in the *4+* case makes it the best option.

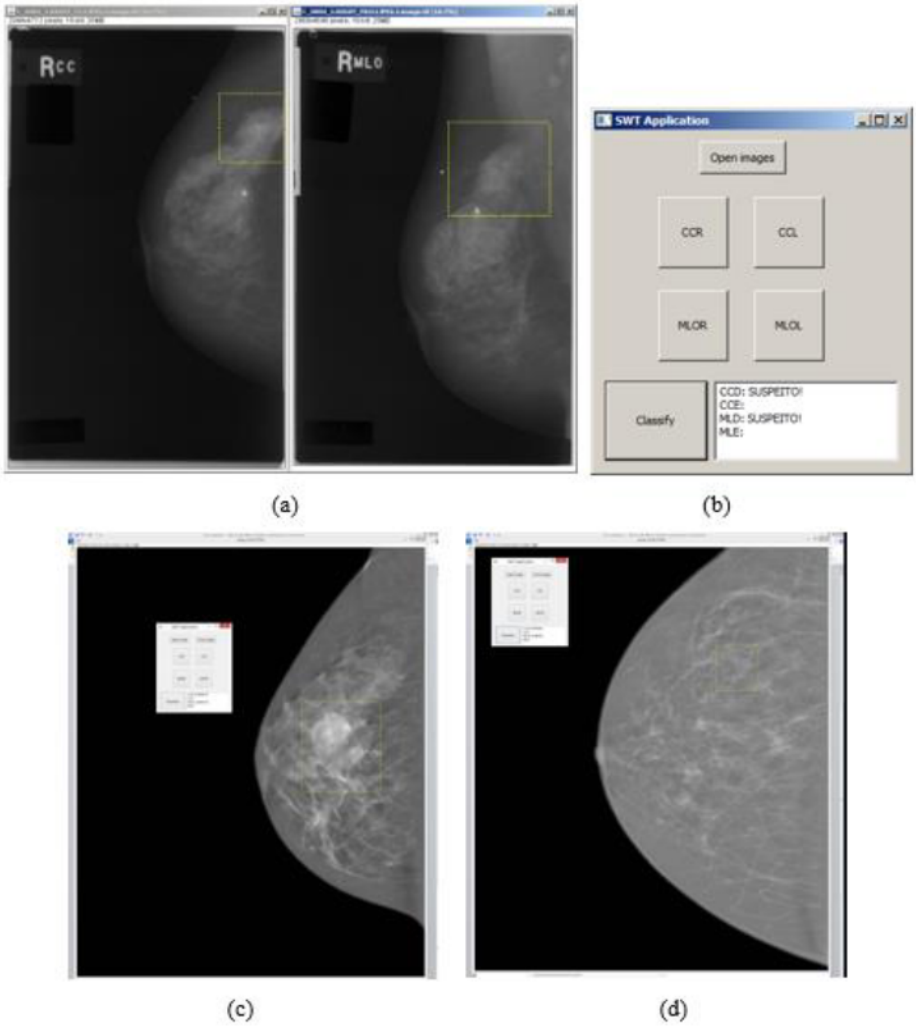
Table 3: Results for different attributes agreements when classifying as "suspect".

	Agreement							
	1+	**2+**	**3+**	**4+**	**5+**	**6+**	**7+**	**8+**
Sensitivity	100%	100%	93%	83%	65%	48%	20%	4%
Specificity	1%	10%	48%	67%	82%	90%	96%	98%
Accuracy	31%	37%	61%	72%	77%	77%	73%	70%

Source: Authors

Results of the classifier evaluation are conveyed to the radiologist in the form of a textual notification, indicating whether the selected nodule is classified as "suspect" or "normal". This information is directly presented to the physician, who has the option to save this text for future reference. Examples illustrating the display of these results are shown in Fig. 4.

Figure 4. Example of scheme outcome: (a) example of images with outlined masses; (b) example of the final classification result (the classification is displayed twice since the nodule delimitation was done in both right breast projections – MLO and CC); (c) and (d) examples of two different digital mammograms in CC projection – with indication of a suspect mass in (c) and a potential benign signal ("normal") in (d).



Source: Authors

The software interface is very simple, as it can be observed in Fig. 4(b). The "Open Images" button allows selecting the set of images to be investigate (for each case under analysis). As each image from a single exam is displayed, it must be matched according to the 4 typical projections of mammography as shown in the interface of Fig. 4: CCR, CCL, MLOR, MLOL, corresponding respectively to the CC and MLO projections, Right or Left breast. The software defines DICOM files corresponding to these images as input, although TIFF or LJPEG files could also be accessed.

Once images are selected, they are pre-processed and the software displays each one in independent simultaneous windows on the screen. Each independent image can be controlled in terms of brightness and contrast. In addition, the observer can enlarge the image (zoom effect) or return it to the original size, which is also the tool to define the limits of a ROI for

classification. Then, when selected the "Classify" button, the software will analyze and classify that ROI according to the previous descriptions.

When delimiting a ROI, information has to be provided on the following: if a single ROI is set in a single projection (CCL, for example), one should confirm whether or not that ROI is also delimited in the corresponding complementary projection (MLOL). If so, the scheme considers both ROIs as a single mass using 8 attributes; on the other hand, these ROIs will be classified individually by using 4 attributes for each. The final result is displayed in the form of text in the lower right corner, corresponding to the breast projection information as a normal or suspect nodule.

3.2 APPLYING TO IMAGES FROM FFDM SYSTEMS

A set of FFDM images from INBreast database [20] was also considered, in order to evaluate the classifier performance according to the current state-of-art of mammography exams (images obtained from DR mammography equipment). This set included 42 images matching the parameters mentioned above. Tests of this scheme with such images set yielded a similar result (72% accuracy, 83% sensitivity and 67% specificity). Even with a small set of images, similarities to test results with DDSM database [10] images show reliability when applying the scheme to direct digital images. Tests using a quad core i5 processor (3.40 GHz) with 16 GB of RAM with the images stored in a 7200 rpm HDD resulted in a delay of 3 s to pre-process and open all 4 images and, once the ROI is selected, a delay of 4 s for each pair of nodules to be processed. The time taken by this scheme is low enough to be considered a real-time processing scheme, for use by a radiologist, adding almost no diagnosis delay.

In addition, another kind of test was also achieved with images acquired from a GE Essential unit. The scheme was applied to the so-called "for presentation" images (used for visual analysis by the radiologist), obtained from the DICOM file of each case – 63 cases (4 images per case in average). But now this test has been performed requesting 3 radiologists to mark ROIs on the digital mammograms, highlighting the existing nodule before interpretation; then, the ROIs selected for each case were processed by the Driven CADx scheme, saving the corresponding results separately. Once all ROIs have been selected, the radiologists analyzed them, recording their respective opinions (such as a probable malignant or benign mass). In each mammogram analysis, the radiologist also could check the CADx result right after recording the first opinion, using the corresponding classification interface. In the end, the radiologist was invited to maintain the previous interpretation or change it depending on the "second" opinion provided by the computational scheme. None of the collaborating radiologists

had difficulty following procedures related to the Driven CADx scheme interface. Results collected from this test are presented in Table 4 which summarizes them in terms of percentage of data obtained. The "Initial" column refers to the respective rate (Accuracy, Sensitivity and Specificity) corresponding to the radiologist's first opinion when observing the image; the "CADx" column, to the result produced by the scheme; and the "Final" column, to the final assessment by the radiologist, after checking the CADx classification in each case (with the respective variation of percentages from the first to the final classification).

Table 4. Results from radiologist evaluation compared to that from the CADx scheme (percentage values as a function of the number of cases that agree or not with the original information)

RADIOLOGIST 1	**Initial**	**CADx**	**Final**	**Variation**
Accuracy	71.4%	85.7%	76.2%	5%
Sensitivity	71.4%	85.7%	71.4%	0%
Specificity	71.4%	85.7%	78.6%	7%
RADIOLOGIST 2				
Accuracy	66.7%	85.7%	71.4%	5%
Sensitivity	71.4%	85.7%	85.7%	14%
Specificity	64.3%	85.7%	64.3%	0%
RADIOLOGIST 3				
Accuracy	66.7%	85.7%	90.5%	24%
Sensitivity	57.1%	85.7%	100.0%	43%
Specificity	71.4%	85.7%	85.7%	14%

Source: Authors

Based on the results presented in Table 4, several aspects can be emphasized:

(a) The individual performance of the CADx scheme demonstrated effectiveness, successfully identifying 18 out of 21 suspected cases and 36 out of 42 benign cases – a higher rate than those achieved individually by the 3 radiologists;

(b) Radiologist 1 (the most experienced in the field) only modified the initial classification for 3 cases – originally classified as malignant – after observing the CADx result; this suggests that the system positively influenced the specificity rate for this radiologist;

(c) Radiologist 2 exhibited a similar behavior, with positive changes in rates for a highly suspicious case that was originally reported as benign;

(d) Radiologist 3 (the least experienced in the field) was the most influenced by the CADx scheme, resulting in a substantial increase in the sensitivity rate; in fact, upon reviewing the information provided by the classifier, Radiologist 3 achieved 100% accuracy in identifying suspected cases and also showed a significant improvement in the specificity rate;

(e) During the analysis, a concurrent verification was conducted to ensure that the nodules marked by the radiologists corresponded to those referenced in the acquisition of the original images (as initially classified by the CADx scheme).

Using the INBreast [20] database set, consisting of 42 FFDM images (30 malignant and 12 benign) yielded a closely comparable accuracy rate of 72%, with 83% sensitivity and 67% specificity. Due to the limited size of the image set in this particular case, it is challenging to draw statistically significant conclusions regarding the results. However, the similarities observed between the outcomes obtained from the INBreast datase [20] and the DDSM [11] images hold promise for the effectiveness of the scheme in FFDM images.

The program is accessible in two free forms: a downloadable .JAR software that functions as described earlier and a web-based version that classifies a collection of ROIs. For the web-based system, users are required to clip the ROIs, save them, and subsequently load them into the system for classification.

5 CONCLUSIONS

Despite the smaller number of nodules in comparison to previous tests conducted on cases extracted from digitized images, the results obtained in this final stage of evaluations enable us to draw important conclusions about the effectiveness and performance of our scheme under practical conditions. Two key findings deserve special attention:

(a) The individual performance of the scheme in evaluating digital radiography (DR) images surpassed its performance with digitized images of lower quality in terms of contrast, as indicated by the rates recorded in the previous stage. Although this outcome was anticipated due to the increased contrast observed in the new set of images, the necessary adaptations made to the scheme to accommodate these new characteristics had the intended effect. Upon individually analyzing each image within the newly defined set, errors were only evident in highly complex cases.

(b) Consequently, owing to the relative increase in specialists' confidence in the scheme, its positive influence on their final diagnosis was once again evident. In all sets of tests conducted, there was a demonstrable improvement in diagnostic efficacy, with at least one of the evaluated rates exhibiting an increase. This observation reinforces the conclusions drawn from the tests performed in the preceding stages of the project.

The Driven CADx scheme described herein is a fully operational breast mass classification system capable of real-time processing. Considering the limited availability of similar systems, especially those that are freely accessible, it can serve as a valuable and

important tool for obtaining second opinions in mammographic evaluations. Furthermore, the creation of a JAVA library allows future developers to utilize all the tools established for this system, facilitating potential enhancements to the scheme.

While the classification method presented in this study may not exhibit the highest accuracy reported in the literature, its distinctive structure and interface, particularly for freely downloadable software accessible via our laboratory website, make it noteworthy and useful for interested individuals.

ACKNOWLEDGMENTS

Acknowledgments to FAPESP by providing financial resources to the researches, and to Luciana Buffa Verçosa, MD., by the technical contributions.

REFERENCES

1. Kolb, T.; Lichy, J.; Newhouse, J. Comparison of the performance of screening mammography, physical examination, and breast US and evaluation of factors that influence them: an analysis of 27,825 patient evaluations. *Radiology,* 2002, v. 225, p. 165-75.

2. Abreu, M. S. de A., et al. Breast microcalcifications in screening mammography: a literature review. *Brazilian Journal of Health Review*, *4*(6), 2021, p. 27283–27289. (https://doi.org/10.34119/bjhrv4n6-292).

3. Georgian-Smith, D; Moore, R.H.; Halpern, E. Blinded Comparison of Computer-Aided Detection with Human Second Reading in Screening Mammography, *AJR Women's Imaging,* 2007, v. 189, p. 1135-1141.

4. Meyer, J.E.; Eberlein, T.J.; Stomper, P.C.; Sonnenfeld, M.R. Biopsy of occult breast lesions. Analysis of 1261 abnormalities, *Journal of the American Medical Association,* 1990, v. 263, p. 2341-2343.

5. Karssemeijer, N. CAD: more than a perception aid. *Medical Physics Web*, Winter 2011. p. 6.

6. Schiabel, H. et al. A prototype of mammography CADx scheme integrated to imaging quality evaluation techniques. In Proc. of SPIE Medical Imaging: Computer-aided Diagnosis, Orlando, FL, USA, v. 7963, p. 796323-1 - 796323-12, 12-17 Feb 2011.

7. ImageJ. Available online: https://imagej.nih.gov/ij/index.html (accessed on 12 Dec 2022).

8. Goes, R.F.; Schiabel, H.; Sousa, M.A.Z. Automatic scanning software based on the characteristic curve of mammograms digitizers. *Journal of Electronic Imaging*, 2013, v. 22, p. 1 – 9.

9. Matheus, B.R.N.; Marcomini, K.D.; Schiabel, H. Segmentation techniques evaluation based on a single compact breast mass classification scheme. In Proc. of SPIE Medical Imaging: Image Processing, San Diego, CA, USA, v. 9787, p. 978426-1 – 978426-9, 27 Feb – 02 Mar 2016.

10. Matheus, B.R.N.; Schiabel, H. Online mammographic images database for development and comparison of CAD schemes. *Journal of Digital Imaging*, 2011, v. 24, p. 500–506.

11. Heath, M.; Bowyer, K.; Kopans, D.; Moore, R.; Kegelmeyer, W.P. The digital database for screening mammography; In Proc. of the Fifth International Workshop on Digital Mammography, Medical Physics Publishing. 2001. p. 212-218.

12. Matheus, B.R.N.; Gonçalves, S.; Schiabel, H. Automated mass classification in CADx mammography scheme using density parameters – In Proc of 17th International Workshop on Computer-Aided Diagnosis (29th International Congress on Computer Assisted Radiology and Surgery – CARS2015), IJCARS, v. 10, suppl. 1, Barcelona, Spain, 24-27 Jun 2015, , p. S286.

13. Siegel, E.L. (Un. Maryland Med Ctr, USA). Re-thinking CAD for the future: a clinical perspective. Personal communication, SPIE Medical Imaging: Computer-aided Diagnosis, San Diego, CA, USA, 19 Feb 2014.

14. Haygood, T.M. et al. Timed efficiency of interpretation of digital and film-screen screening mammograms. *American Journal of Roentgenology*, 2009, v. 1, p. 216-220.

15. Kooi, T.; Mordang, J-J.; Karssemeijer, N. Conditional random field modelling of interactions between findings in mammography. In Proc. of SPIE Medical Imaging: Computer-aided Diagnosis, Orlando, FL, USA, v. 10134, p. 101341E-1 - 101341E-8, 11-16 Feb 2017.

16. Holland, K. et al. Quantification of mammographic masking risk with volumetric breast density maps: how to select women for supplemental screening. In Proc. of SPIE Medical Imaging: Computer-aided Diagnosis, San Diego, CA, USA, v. 9785, p. 97850I, 27 Feb – 03 Mar 2016.

17. S. Osher, J. A. Sethian, "Fronts propagating with curvature dependent speed: algorithms based on Hamilton-Jacobi formulations". J. of Computational Physics, vol. 79, pp. 12-49, 1988.

18. Ribeiro, P.B.; Romero, R.A.; Oliveira, P.R.; Schiabel, H.; Verçosa, L.B. Automatic segmentation of breast masses using enhanced ICA mixture model. *Neurocomputing,* 2013, v. 120, p. 61-71.

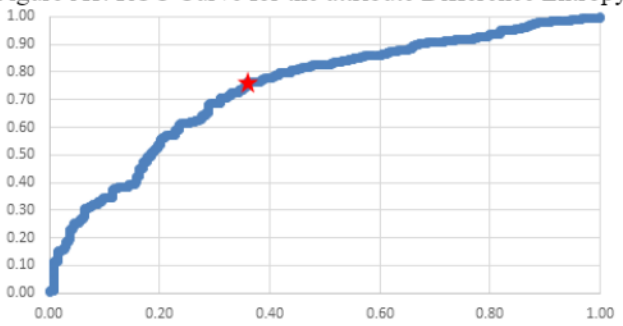
19. Haralick, R.; Shanmugam, K.; Distein, I. Textural features for image classification, *IEEE Trans. on Systems, Man, and Cybernetics - SMC,* 1973, v. 3, p. 610-621.

20. Moreira, I. et al. INbreast: toward a full-field digital mammographic database, *Academic radiology,* 2012, v. 19, p. 236-48.

APPENDIX

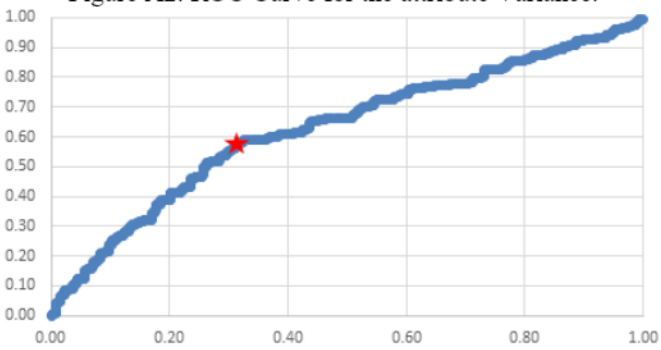
This section illustrates the ROC curves generated by testing independently each attribute defined in section 2 (Difference Entropy, Variance, Compactness and Relative Density) with 568 previously selected ROIs from DDSM database [11]. They were determined by changing the parameter corresponding to the classification threshold (the value from which a given attribute changes the result from "non-suspect" to "suspect"). The highlighted point in each Figure shows always the best cost-benefit for that respective test.

Figure A1. ROC Curve for the attribute Difference Entropy.



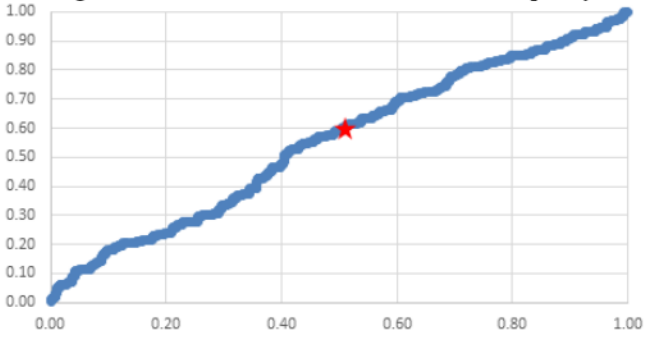
Source: Authors

Figure A2. ROC Curve for the attribute Variance.



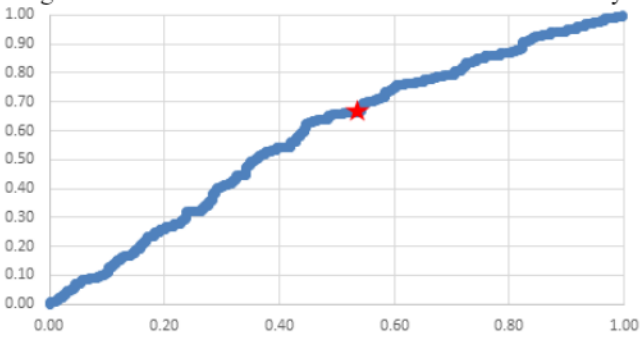
Source: Authors

Figure A3. ROC Curve for the attribute Compacity.



Source: Authors

Figure A4. ROC Curve for the attribute Relative Density.



Source: Authors