Educational software as a tool for teaching & learning of digital mammography*

Uso de software como ferramenta pedagógica no processo de ensino-aprendizagem da mamografia digital

Simone Elias[1], Silvio Ricardo Pires[2], Ana Claudia Patrocinio[1], Regina Bitelli Medeiros[3]

Abstract
OBJECTIVE: The present study was aimed at evaluating the performance of residents trained in the reading of conventional and digital mammography images with a specific computational tool. MATERIALS AND METHODS: The training was accomplished in the Laboratory of Medical Images Qualification (QualIM – Laboratório de Qualificação de Imagens Médicas). Residents in radiology performed approximately 1,000 readings of a set of 60 images acquired from a statistical phantom (Alvim®) presenting microcalcifications and fibers with different sizes. The analysis of the residents' performance in the detection of these structures was based on statistical parameters. RESULTS: Values for detection probability were respectively 0.789 and 0.818 for conventional and digital systems. False-positive rates were 8% and 6%, and true-positive rates, 66% and 70% respectively. The total kappa value was 0.553 for readings on the negatoscope (hard-copy readings), and 0.615 on the monitor (soft-copy readings). The area under the ROC curve was 0.716 for hard-copy readings and 0.810 for soft-copy readings. CONCLUSION: The training has showed to be effective, with a positive impact on the residents' performance, representing an interesting educational tool. The results of the present study suggest that this method of training based on the reading of images from phantoms can improve the practitioners' performance in the interpretation of mammographic images.
Keywords: Digital image; Teaching & learning; Mammography; Educational tool; Soft-copy reading; Hard-copy reading.

Resumo
OBJETIVO: Avaliar o impacto sobre o treinamento de residentes utilizando uma ferramenta computacional dedicada à avaliação do desempenho da leitura de imagens radiológicas convencionais e digitais. MATERIAIS E MÉTODOS: O treinamento foi realizado no Laboratório de Qualificação de Imagens Médicas (QualIM). Os residentes de radiologia efetuaram cerca de 1.000 leituras de um total de 60 imagens obtidas de um simulador estatístico (Alvim®) que apresenta fibras e microcalcificações de dimensões variadas. O desempenho dos residentes na detecção dessas estruturas foi avaliado por meio de parâmetros estatísticos. RESULTADOS: Os resultados da probabilidade de detectabilidade foram de 0,789 e 0,818 para os sistemas convencional e digital, respectivamente. As taxas de falso-positivos foram de 8% e 6% e os valores de verdadeiro-positivos, de 66% e 70%, respectivamente. O valor de kappa total foi 0,553 para as leituras em negatoscópio e 0,615 em monitor. A área sob a curva ROC foi de 0,716 para leitura em filme e 0,810 para monitor. CONCLUSÃO: O treinamento proposto mostrou ser efetivo e apresentou impacto positivo sobre o desempenho dos residentes, constituindo-se em interessante ferramenta pedagógica. Os resultados sugerem que o método de treinamento baseado na leitura de simuladores pode produzir um melhor desempenho dos profissionais na interpretação das imagens mamográficas.
Unitermos: Imagem digital; Ensino-aprendizagem; Mamografia; Ferramenta pedagógica; Leitura em filme; Leitura em monitor.

INTRODUCTION

Professionals in the medical area are constantly required to acquire new skills in order to get updated with the technological progress within their specialities. Education is also affected by these changes, requiring new pedagogical models.

Information technology has contributed in this process, promoting professional capacitation, thus representing an alternative means to transform the pedagogical practice into an accessible, flexible and dynamic structure.

Mammography is one of the most demanding imaging methods in terms of professional knowledge and experience in the interpretation of lesions. However, it is not a certainty that the benefits from this expe-

rience can be extended to digital mammography, a new version of a well established diagnostic modality [1].

Breast images interpretation, in particular, is a topic of such a relevance that the Society of Breast Imaging runs a teaching & learning program dedicated to this speciality. This program is designed with three objectives: providing guidance for teachers and program leaders; evaluating and refining the training of the specialist in breast diagnosis; guiding specialized practitioners to keep themselves updated[2].

The Food and Drug Administration (FDA) requires a minimum eight-hour training in digital mammography for a practitioner to be able to interpret this new image standard[3]. Some authors, however, highlight the necessity of further, wider studies for determining if such training workload is sufficient to assure the control over the technology, maintaining the diagnostic accuracy[4].

Individual preferences and time of experience have a direct influence on the radiologist's option for softcopy or hardcopy reading. Generally, radiologists with more than 30 years of experience prefer hardcopy reading. In terms of sensitivity, there is no significant difference between the two modalities, but it seems there be a difference in the specificity and positive predictive value for three in every four practitioners who prefer softcopy reading. However, it is still to be established if the specialist presents better performance and less subjectivity with this reading modality[5].

However, differences in the monitor's luminance, utilization of softwares as tools for images manipulation and other characteristics involving the display of the image on the monitor screen can directly influence the practitioner's performance[6,7].

It is assumed that softcopy reading can qualify a professional to take advantage of the digital system, considering that personal preferences and subjectivity in the images interpretation seem to affect the specificity and sensitivity for lesions detection[5].

The adaptation of the professional plays a significant role in this period of transition from the conventional to the digital technology, where the recognition

of the new image standard (particularly in terms of brightness and contrast resolution) and the accurate differentiation between actual structures and artifacts have been gradually integrated in order to maintain the potential benefits of the populational screening, namely, early detection and fast reading[5].

A method supported by a dedicated software is proposed to foster this transition. The present study is aimed at evaluating the impact of the utilization of this computational tool for training residents in the reading of digital images.

MATERIALS AND METHODS

In 2006, a pilot project approved by the Committee for Ethics in Research was developed with residents of the Department of Imaging Diagnosis at Universidade Federal de São Paulo/Escola Paulista de Medicina (Unifesp/EPM), São Paulo, SP, Brazil. The residents were given a training developed for digital technology applied to mammography, besides the official content of the medical residency program. The training consists of theoretical activities (classes about digital systems, digital imaging and pre-processing; monitors; quality-control; BI-RADS® classification) and practical activities (on negatoscopes and monitors) developed in a dedicated laboratory within the institution.

The laboratory infrastructure encompasses software and hardware for training, including high-resolution Clinton® 3 Mpixel LCD monitors (Clinton Electronics Corporation; Rockford, USA) and LCD

Barco® 5 Mpixel monitors (BarcoView; Kortrijk, Belgium), specific Planilux® negatoscope (Planilux; Warstein, Germany) and environment with controlled luminance, in compliance with the FDA recommendations[3]. Each resident performance in terms of detectability is measured by means of images reading on films (hardcopy reading), and digitized images reading on monitors (softcopy reading). These images were acquired from a statistical mammographic phantom, and stored in a data/image bank.

The mammographic phantom (Figure 1) model Alvim Statistical Phantom 18-209 (R&D Ltd.; Jerusalem, Israel) is composed of a main acrylic 1.5 cm-thick plate and other three, secondary 1.0 cm-thick plates which, as a whole, simulate a typically compressed breast (4.5 cm). Within the internal structure of the main plate, there are 100 cylinders which may be randomly distributed, 25 of them including objects simulating microcalcifications and fibers of different sizes[8]. Generally, this type of phantom is utilized only for research purposes, because its utilization as a routine in quality control programs implies a method of difficult application.

In the first phase of the program, the residents perform the reading of phantom images on the dedicated negatoscope and classify their findings into five confidence levels, according to the following scoring system: 100 (certainty about the presence of the object), 75 (probable presence of the object), 50 (uncertainty about the presence of the object), 25 (improbable presence of the object), 0 (definitely absent object). In
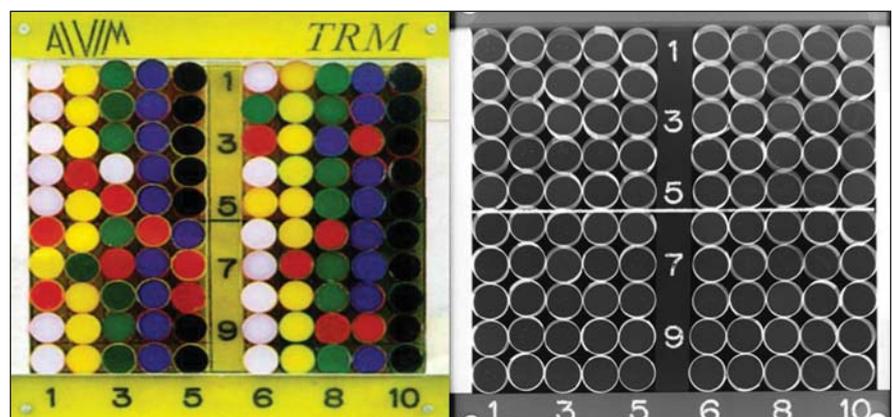

Figure 1. Alvim® statistical phantom (at left) and respective radiographic image (at right).

the second phase, the training is developed with high-resolution monitors. The software GQM® version 2[9], developed by the laboratory dedicated to the training process, allows the images manipulation by means of digital tools for brightness and contrast settings, magnification and Grayscale inversion (Figure 2).

Additionally, this program allows the automatized analysis of the residents' performance in the detection of signs present on the images. The statistical analysis defines the observer's profile, calculating the following evaluation indices: detection probability (Pdet), kappa (κ) values[10], true-positive results, false-negative results, and area under the ROC curve[11] for microcalcifications and fibers of different sizes.

Other capabilities include comparison between softcopy and hardcopy readings, evaluation of intra- and interobserver agreement in the different training phases, and monitoring of the time spent for each digital image reading.

Once these phases are completed, the professionals are given access to an image bank with mammograms acquired in different equipment of several radiological units and classified in accordance with different levels of complexity. A golden-standard data set (stored in the data bank after double-reading by specialists) registers the diagnostic impression based on these images. This images bank is continuously updated for utilization in the third phase of the training. In this phase, the professional undergoing the training gathers the information required for preparing his/her own report that is stored in a specific interface. A golden-standard report is available for self-training and comparison of the diagnostic impressions. A tutorial is also available for instruction in this phase.

The results of the analysis of 2,000 phantom images interpreted by experienced specialists with a minimum eight-hour training in digital systems, were adopted as a parameter for evaluating the residents' performance[9], besides the values established by regulation of the Brazilian Ministry of Health [12]. These parameters are shown on Tables 1 and 2.

The performance indices utilized for evaluating the residents were: κ value,
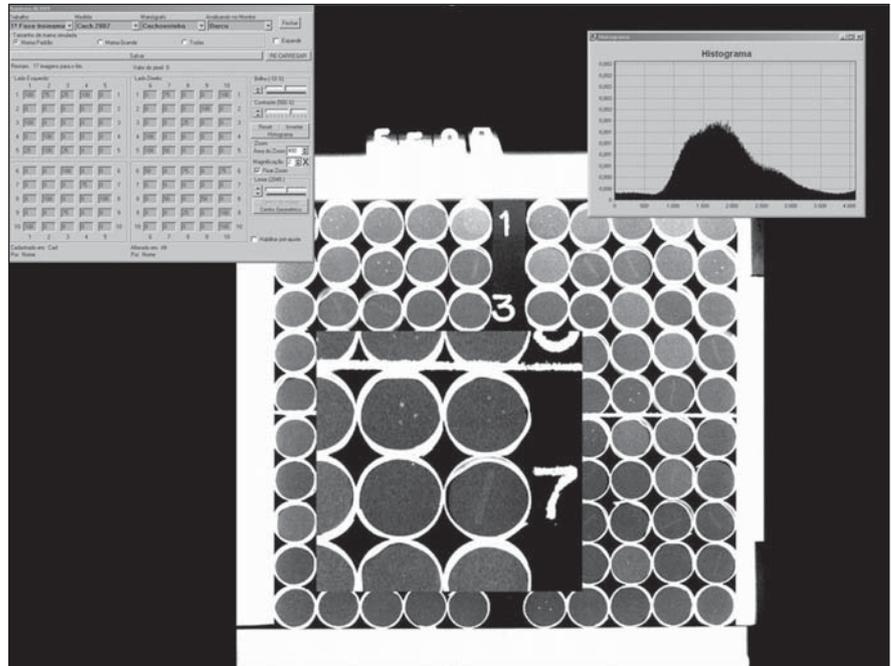
Figure 2. Interface for training in phantom images interpretation. The software provides digital tools similar to the ones of a digital mammography workstation. On the right upper corner, there is a window to be fulfilled with the user's rate of detection probability, and at right, the visualization of the image histogram.

Table 1 Reference values for the reading of about 2,000 phantom images interpreted by specialists[9] and respective parameters of the Brazilian Ministry of Health[12].

Structure	Dimension	Minimum required identification rate
Microcalcifications	0.20 mm	10% to 15%
	0.25 mm	40% to 45%
	0.32 mm*	100%
Fibers	0.45 mm	20% to 30%
	0.60 mm	50% to 60%
	0.75 mm*	100%

* Values established by technical regulation of the Brazilian Ministry of Health, Order (Portaria) No. 453, dated June 2, 1998.

Pdet, area under the ROC curve, besides true-positive and false-positive results.

The κ value measures the degree of agreement between answers (presence or absence of structures); Pdet indicates the chance of structures identification with a determined degree of certainty either with true-positive or false-positive results. The data evaluation based on a ROC curve[11] allows a graphic and fast analysis of the relation between sensitivity and specificity both individually and in group.

RESULTS

A total of 1,003 Alvim®-type phantom image readings were performed by the resi-

Table 2 Reference values for reading of about 2,000 phantom images interpreted by specialists[9].

Index	Reference values
Kappa	> 0,7
Area under the ROC curve	> 0,8
True-positive	> 85%
False-positive	Up to 8%

dents as follows: 588 readings on negatoscope, and 415 readings on high-resolution monitors.

The results were based on the average number of readings performed by all residents, with a comparison of their performance in hardcopy and softcopy readings.

Figure 3 presents the mean values of performance indices among hardcopy and softcopy readings.

Figure 4 shows the histograms with the distribution of κ values with different reading methods.

Figure 5 shows ROC curves for conventional and digital readings.

Table 3 presents the true-positive and false-positive rates, besides the area under the ROC curve in both reading methods.

The mean time for reading an image on the monitor at the beginning of the training was 7.5 minutes, and, at the end, 4.0 minutes.

DISCUSSION

Users can benefit from the utilization of the digital technology in the interpretation of images; however, the professional must be trained for adaptation to the new visual patterns.
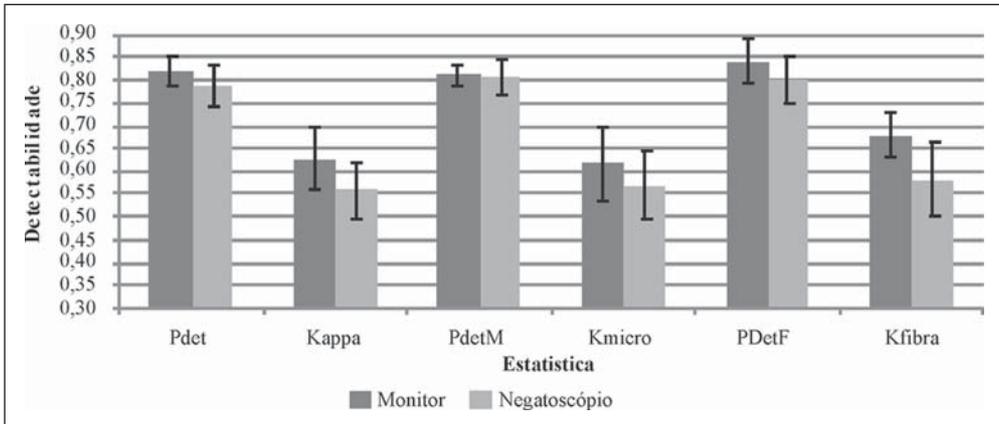
Figure 3. Average performance indices in the different reading systems (negatoscope and monitor). Pdet, detection probability; K, kappa values; M, microcalcifications; F, fibers.
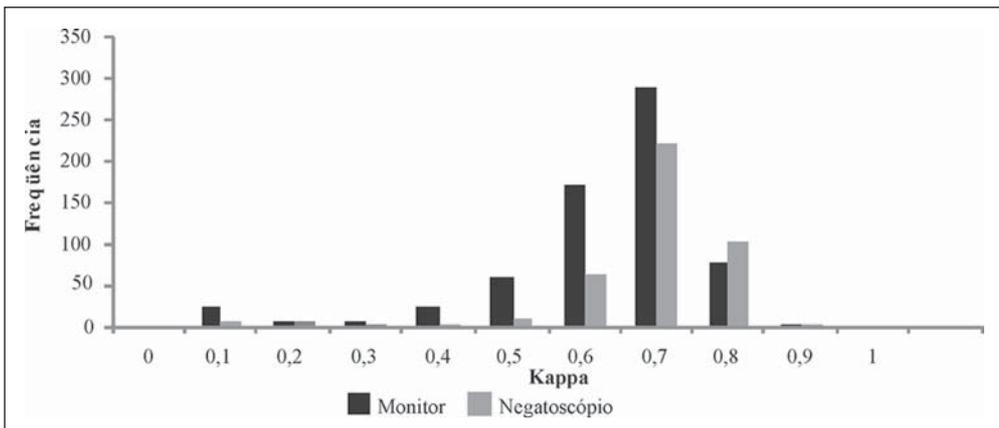
Figure 4. Histograms of kappa values distribution in different reading systems.
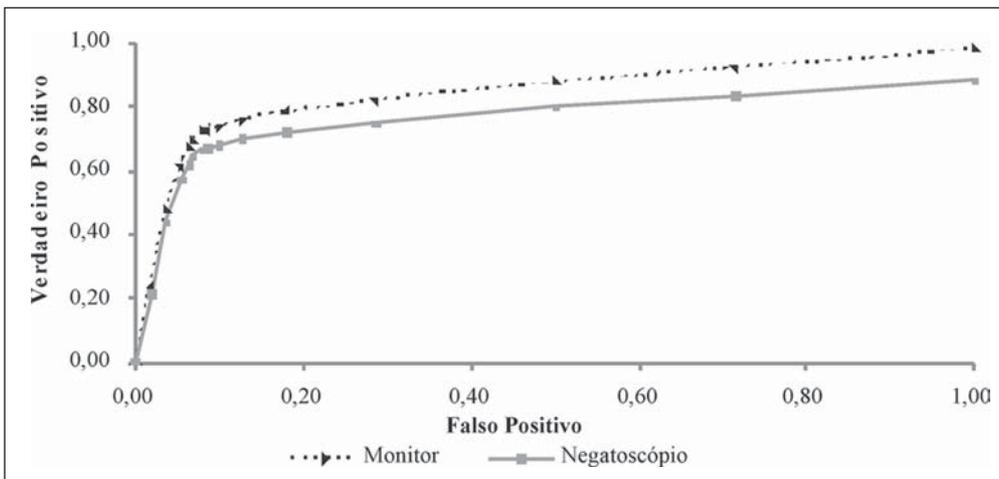
Figure 5. Area under the ROC curve in conventional ($A_Z = 0.761$) and digital systems ($A_Z = 0.810$).

Table 3 Values of the area under the ROC curve, true-positive and false-positive results on the different reading methods.

Reading method	True-positive	False-positive	Area under the ROC curve
Negatoscope	66%	8%	0.761
Monitor	70%	6%	0.810

Reference values (desirable values)[9]: true-positive > 85%; false-positive up to 8%; area under the ROC curve > 0.8.

Advantages can be clearly observed when the practitioners are appropriately trained and qualified[13].

The present study is intended to aid in the training of professionals during the period of introduction of the digital technology and contributing to the reduction of variations or disagreements in the images interpretation, minimizing false-positive and false-negative results.

Images interpretation on monitor screens (softcopy reading) have been widely studied to evaluate the practitioners' performance as compared with the conventional system (hardcopy reading)[14,15]. The majority of these studies have concluded that the images interpretation on monitor screens is widely accepted, the adaptation of the specialist is rapid, and the accuracy and time spent in the interpretation compare to the conventional system, provided the professional is duly trained[16,17].

An overall analysis demonstrated that the Pdet by the residents during the learning process was higher in the digital system. However, considering the characteristic of this indicator, false-positive results should be evaluated for a correct interpretation. As far as the reference values are considered, a false-positive results rate of up to 8% is acceptable.

The higher Pdet (= 0.818), associated with the lower rate of false-positive results (6%) confirms a better identification of structures on the monitor screen (softcopy reading). These results indicate that the digital reading presents both higher sensitivity and specificity.

Also, κ results corroborate a higher detectability of structures at digital reading for both microcalcifications and fibers. But, it was in the detection of fibers that a clear improvement could be observed, suggesting that the manipulation of images brightness and contrast allowed by the digital system may have influenced this result.

Based on the results of κ values, a histogram with the distribution of these data could be constructed, allowing the evaluation of the interobserver reading variability (Figure 4). On this Figure, a narrower base can be observed with a symmetrical distribution (similar to a typical distribution curve), demonstrating a lower rate of subjectivity in the softcopy readings.

The evaluation of the time spent in the digital reading has shown that, with the progression of the training, this time decreased without, however, affecting the residents' performance, which demonstrates the effectiveness of the training.

The overall analysis demonstrated that, for a less experienced professional, false-positive results have a negative influence on the final accuracy of the reading[13]. This lower accuracy is primarily a result of the lack of perception and experience during the training in mammography, which restricts the resident's capability to recognize "objects", complicating the recognition of real lesions and artifacts. The solution proposed is a systematic training tutorial associating the image perception with the feedback on reasons that aid in the decision making process[18].

Reading velocity and improved performance of experienced observers in the interpretation of mammographic images involve a shift in the mechanism of image perception from a "search-to-find" process[18] to a relatively faster holistic mode[19]. This shift depends, among others, on the observer's expertise in the method and the presence of an actual or virtual tutorial.

Even though the results of the present study indicate a better performance in this reading mode (digital phantom images allowing manipulation), some inherent aspects may have contributed to this conclusion. The training with films (hardcopy) preceded the training with the digital system; therefore the Professional, although short, have acquired some experience in the reading of phantom images. The personal motivation may also have influenced the observers' concentration at the moment of the softcopy reading. Additionally, the initial contact with the technological innovation may have positively influenced the professionals' interest in this phase of the training.

It is important to observe that the results of the present study were based on phantom images reading, with the objective of quantifying data and evaluating the influence of the training on the overall professionals' performance. Additionally, it should be noted that, in the reading of mammographic images, where the fibroglandular tissue and lesion present with different patterns, the detection process becomes even more complex, requiring more experience of the professional.

CONCLUSION

The method of training proposed in the present study has shown to be effective, with a positive impact on the residents' performance, constituting an interesting pedagogical tool.

The results suggest that the method of training based on phantom reading can produce a better performance of the practitioners in the interpretation of mammographic images.

It is the authors' opinion that this training method with a dedicated software plays a significant role as an ancillary tool in the complex, multifactorial process of learning in mammography.

REFERENCES

1. Elmore JG, Wells CK, Howard DH. Does diagnostic accuracy in mammography depend on radiologists' experience? J Womens Health. 1998;7: 443–9.

2. The Society of Breast Imaging – SBI (2005-2008). ACR – Breast imaging training. [cited 2008 Sep 25]. Available from: http://sbi-online.org

3. Radiological Health. MQSA: Mammographic Quality Standards Act – final regulation program.

[cited 2008 Sep 25]. Available from: http://www.fda.gov/cdrh/mammography/digital.html

4. Monsees B. The Mammography Quality Standards Act. An overview of the regulations and guidance. Radiol Clin North Am. 2000;38:759–72.

5. Obenauer S, Hermann KP, Marten K, et al. Soft copy versus hard copy reading in digital mammography. J Digit Imaging. 2003;16:341–4.

6. Pisano ED, Chandramouli J, Hemminger BM, et al. Does intensity windowing improve the detection of simulated calcifications in dense mammograms? J Digit Imaging. 1997;10:79–84.

7. Krupinski E, Roehrig H, Furukawa T. Influence of film and monitor display luminance on observer performance and visual search. Acad Radiol. 1999;6:411–8.

8. Gurvich VA. Statistical approach for image quality evaluation in daily medical practice. Med Phys. 2000;27:94–100.

9. Pires SR. "Software" gerenciador de uma base de dados e de imagens mamográficas classificadas segundo um índice de qualidade [tese de mestrado]. São Paulo: Universidade Federal de São Paulo; 2003.

10. Altman DG. Practical statistics for medical research. 1st ed. London: Chapman & Hall; 1991.

11. Metz CE. Receiver operating characteristic analysis: a tool for the quantitative evaluation of observer performance and imaging systems. J Am Coll Radiol. 2006;3:413–22.

12. Brasil. Ministério da Saúde. Secretaria de Vigilância Sanitária. Diretrizes de proteção radiológica em radiodiagnóstico médico e odontológico. Portaria nº 453, de 2/6/1998. Brasília: Diário Oficial da União; 1998.

13. Nodine CF, Kundel HL, Mello-Thoms C, et al. How experience and training influence mammography expertise. Acad Radiol. 1999;6:575–85.

14. Pisano ED, Gatsonis C, Hendrick E, et al. Diagnostic performance of digital versus film mammography for breast-cancer screening. N Engl J Med. 2005;353:1773–83.

15. Krug KB, Stützer H, Girnus R, et al. Image quality of digital direct flat-panel mammography versus an analog screen-film technique using a phantom model. AJR Am J Roentgenol. 2007;188:399–407.

16. Kundel HL, Polansky M. Measurement of observer agreement. Radiology. 2003;228:303–8.

17. Hemminger BM. Soft copy display requirements for digital mammography. J Digit Imaging. 2003;16:292–305.

18. Kundel HL, Nodine CF, Conant EF, et al. Holistic component of image perception in mammogram interpretation: gaze-tracking study. Radiology. 2007;242:396–402.

19. Mello-Thoms C, Hardesty L, Sumkin J, et al. Effects of lesion conspicuity on visual search in mammogram reading. Acad Radiol. 2005;12:830–40.