

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

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

Simone Elias¹, Silvio Ricardo Pires², Ana Claudia Patrocínio¹, Regina Bitelli Medeiros³

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.

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.

Elias S, Pires SR, Patrocínio AC, Medeiros RB. Uso de *software* como ferramenta pedagógica no processo de ensino-aprendizagem da mamografia digital. Radiol Bras. 2009;42(2):115–120.

INTRODUÇÃO

Os profissionais da área médica são constantemente exigidos a alcançar novas habilidades para que possam acompanhar

a evolução da tecnologia dentro de suas especialidades. Paralelamente, essas mudanças atingem também a educação, exigindo novos modelos pedagógicos.

A informática tem contribuído neste processo, propiciando a capacitação profissional. Constitui-se, assim, em uma alternativa para transformar a prática pedagógica em uma estrutura acessível, flexível e dinâmica.

* Trabalho realizado no Departamento de Diagnóstico por Imagem da Universidade Federal de São Paulo/Escola Paulista de Medicina (Unifesp/EPM), São Paulo, SP, Brasil.

1. Doutoras, Pós-doutorandas do Departamento de Diagnóstico por Imagem da Universidade Federal de São Paulo/Escola Paulista de Medicina (Unifesp/EPM), São Paulo, SP, Brasil.

2. Doutor, Professor Adjunto da Pró-Reitoria de Graduação da Universidade Federal de São Paulo/Escola Paulista de Medicina (Unifesp/EPM), São Paulo, SP, Brasil.

3. Pós-doutorado, Professora Associada do Departamento de Diagnóstico por Imagem da Universidade Federal de São Paulo/Escola Paulista de Medicina (Unifesp/EPM), São Paulo, SP, Brasil.

Endereço para correspondência: Dra. Simone Elias, Rua Mirassol, 313, Vila Clementino, São Paulo, SP, Brasil, 04044-010. E-mail: simone.elias@cfmr.epm.br

Recebido para publicação em 12/11/2008. Aceito, após revisão, em 5/1/2009.

A mamografia é um dos exames de imagem que mais exige do profissional conhecimento e experiência na interpretação das lesões. Todavia, não se tem certeza se os benefícios dessa experiência podem ser estendidos para a mamografia digital, uma nova versão de uma modalidade diagnóstica bem estabelecida⁽¹⁾.

Particularmente, a interpretação da imagem mamária constitui-se em um tópico de tamanha relevância que a Society of Breast Imaging possui um programa de ensino-aprendizagem dedicado a esta especialidade. Esse programa serve a três propósitos: prover um guia para os professores e chefes de programas; avaliar e aperfeiçoar o treinamento do especialista em diagnóstico mamário; orientar os especialistas práticos a manterem-se atualizados⁽²⁾.

Paralelamente, a Food and Drug Administration (FDA) requer do profissional um mínimo de oito horas de treinamento em mamografia digital antes que ele possa interpretar esse novo padrão de imagem⁽³⁾. Alguns autores, no entanto, sugerem estudos mais amplos para saber se este tempo é suficiente para assegurar o domínio sobre a tecnologia e manter a acurácia diagnóstica⁽⁴⁾.

Preferências individuais e tempo de experiência influenciam diretamente as opções do radiologista para leitura em filme (*hard-copy*) ou em monitor (*soft-copy*). Radiologistas com mais de 30 anos de experiência preferem, em geral, a leitura em filme. A avaliação da sensibilidade não evidencia diferenças relevantes entre as duas modalidades, mas parece haver diferença na especificidade e no valor preditivo positivo para três em cada quatro leitores que preferem leitura em monitor. No entanto, ainda não está bem estabelecido se nessa modalidade de leitura o especialista apresenta melhor desempenho e menor subjetividade⁽⁵⁾.

Contudo, diferenças de luminância dos monitores, a utilização de ferramentas de *software* para manipulação das imagens e outras características que envolvem a apresentação da imagem no monitor podem ter influência direta no desempenho do profissional^(6,7).

Supõe-se que a leitura em monitores possa capacitar o profissional a extrair as vantagens do sistema digital, uma vez que

as preferências individuais ou subjetividade na interpretação parecem atuar na especificidade e sensibilidade para a detecção de lesões⁽⁵⁾.

É importante a adaptação do profissional durante esse período de transição, da tecnologia convencional para a digital, em que o reconhecimento do novo padrão de imagem (principalmente diferenças de brilho e contraste) e a identificação precisa entre estruturas reais e artefatos estão sendo gradativamente incorporados para que os benefícios potenciais do rastreamento populacional sejam mantidos: detecção precoce e leitura rápida⁽⁵⁾.

Para promover essa transição propõe-se um método que conta com o apoio de um *software* dedicado. O objetivo do presente trabalho é avaliar o impacto do uso desta ferramenta computacional no treinamento de médicos residentes nas leituras de imagens digitais.

MATERIAIS E MÉTODOS

Em 2006, iniciou-se um projeto piloto aprovado pelo Comitê de Ética em Pesquisa, no qual os residentes do Departamento de Diagnóstico por Imagem da Universidade Federal de São Paulo/Escola Paulista de Medicina (Unifesp/EPM) recebem um treinamento desenvolvido para tecnologia digital aplicada à mamografia, além do conteúdo oficial do programa da residência médica. O treinamento consiste em atividades teóricas (aulas referentes ao tema: sistemas digitais; imagem digital e pré-processamento; monitores; controle de qualidade; sistema BI-RADS[®]) e atividades práticas (em negatoscópio e monitores) em la-

boratório dedicado na instituição de desenvolvimento.

O laboratório possui infraestrutura em *software* e *hardware* preparados para o treinamento, incluindo monitores de alta resolução LCD Clinton[®] 3 Mpixel (Clinton Electronics Corporation; Rockford, EUA) e LCD Barco[®] 5 Mpixel (BarcoView; Kortrijk, Bélgica), negatoscópio específico Planilux[®] (Planilux; Warstein, Alemanha) e ambiente com luminância controlada, conforme recomendado pela FDA⁽³⁾. O desempenho da detectabilidade de cada residente é mensurado por meio da leitura de imagens geradas em filme, digitalizadas e apresentadas em monitor. Essas imagens foram obtidas de um simulador de objetos, estatístico (provenientes de um único equipamento mamográfico) e então armazenadas no banco de dados e imagens.

O simulador mamográfico representado na Figura 1, modelo Alvim Statistical Phantom 18-209 (R&D Ltda.; Jerusalém, Israel) é composto por uma placa principal de acrílico de 1,5 cm de espessura e outras três, secundárias, com 1,0 cm. Estas, quando utilizadas em conjunto, simulam uma mama comprimida de padrão normal (4,5 cm). Na estrutura interna da placa principal há 100 cilindros, que podem ser distribuídos aleatoriamente, dentre os quais 25 cilindros contêm objetos que simulam microcalcificações e fibras de diferentes tamanhos⁽⁸⁾. Em geral, este tipo de simulador é utilizado em pesquisa, porque seu uso na rotina do programa de garantia de qualidade implica método de difícil aplicação.

Na primeira etapa, os residentes realizam a leitura do simulador no negatoscópio dedicado e classificam seus achados em

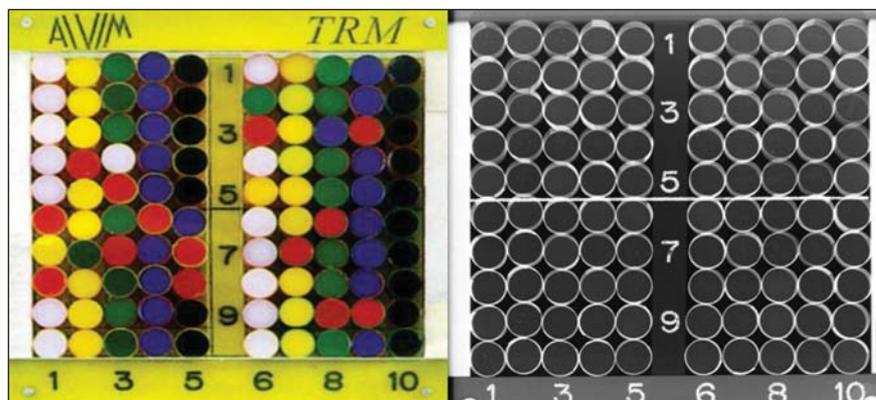


Figura 1. Simulador (*phantom*) estatístico Alvim[®] (à esquerda) e sua imagem radiográfica (à direita).

cinco níveis de confiança, conforme a seguinte pontuação: 100 (quando tem certeza da presença do objeto), 75 (quando julgam provável a presença do objeto), 50 (quando julgam incerta a presença do objeto), 25 (quando julgam improvável a presença do objeto), 0 (quando julgam o objeto definitivamente ausente). Na segunda etapa, o treinamento é realizado em monitores de alta resolução. Nesta etapa, o *software* GQM[®] versão.2⁽⁹⁾, desenvolvido no laboratório e dedicado ao treinamento, permite a manipulação da imagem, pois disponibiliza ferramentas digitais que possibilitam o ajuste de contraste e brilho, ampliação e inversão dos níveis de cinza (Figura 2).

Esse programa permite ainda analisar de forma automatizada o desempenho do residente na detecção dos sinais presentes na imagem. A análise estatística dos dados permite conhecer o perfil do leitor, em que índices de avaliação podem ser calculados: probabilidade de detectabilidade (Pdet), valores de kappa (κ)⁽¹⁰⁾, verdadeiro-positivos, falso-negativos e área sob a curva ROC⁽¹¹⁾ para microcalcificações e fibras de diferentes dimensões.

Também é possível efetuar as comparações entre os sistemas de leituras (filme e monitor), intra e interobservador nas diversas fases de treinamento e monitorar o intervalo de tempo despendido na leitura de cada imagem digital.

Finalizadas essas etapas, um banco de imagens é disponibilizado, no qual mamografias provenientes de diversos serviços de radiologia, realizadas em diferentes equipamentos, foram classificadas de acordo com diferentes graus de complexidade. Um laudo padrão-ouro (armazenado no banco após dupla leitura por especialistas) registra a impressão diagnóstica dessas imagens. Este banco de imagens está em contínua atualização para o uso na terceira etapa do treinamento.

Na terceira etapa, o profissional em treinamento inclui informações para a confecção do seu próprio laudo, que são armazenadas em uma interface específica. O laudo padrão-ouro é disponibilizado para auto-treinamento e comparação da impressão diagnóstica. Nesta fase, também é possível contar com a orientação de um tutor.

Os resultados da análise de 2.000 imagens de simuladores, lidas por especialis-

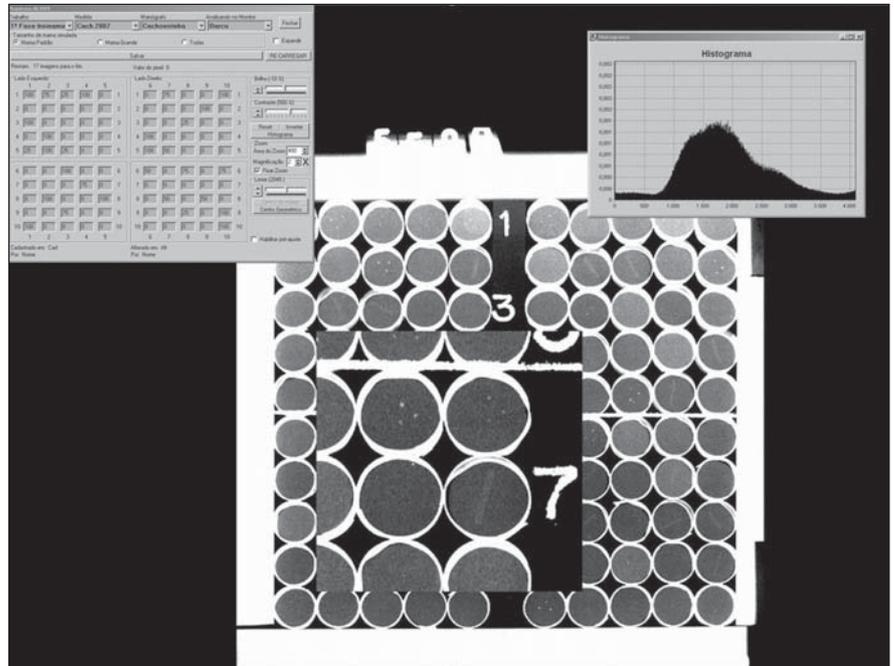


Figura 2. Interface de treinamento para a interpretação das imagens de simuladores. O programa disponibiliza ferramentas digitais à semelhança da *workstation* de uma mamografia digital. No canto superior esquerdo observa-se a janela para o preenchimento das probabilidades de detectabilidade de leitura do usuário, e à direita, a visualização do histograma da imagem.

Tabela 1 Valores de referência extraídos da leitura de cerca de 2.000 imagens de simuladores interpretadas por especialistas⁽⁹⁾ e do Ministério da Saúde⁽¹²⁾.

Estrutura	Dimensão	Porcentagem mínima que as estruturas devem ser identificadas
Microcalcificações	0,20 mm	10% a 15%
	0,25 mm	40% a 45%
	0,32 mm*	100%
Fibras	0,45 mm	20% a 30%
	0,60 mm	50% a 60%
	0,75 mm*	100%

* Valores estabelecidos por regulamento técnico do Ministério da Saúde, Portaria nº 453, de 2/6/1998.

tas experientes, que possuíam no mínimo oito horas de treinamento em sistemas digitais, foram tomados como referência para fins de avaliação do desempenho dos médicos residentes⁽⁹⁾, além dos valores estabelecidos por regulamento do Ministério da Saúde⁽¹²⁾. Esses parâmetros estão dispostos nas Tabelas 1 e 2.

Os índices de desempenho utilizados para a avaliação dos residentes foram valor de κ , Pdet, área sob a curva ROC, além dos resultados verdadeiro-positivos e falso-positivos.

O valor de κ mede o grau de concordância das respostas (presença ou ausência das estruturas). A Pdet indica a chance de o leitor identificar as estruturas com certo grau

Tabela 2 Valores de referência extraídos da leitura de cerca de 2.000 imagens de simuladores interpretadas por especialistas⁽⁹⁾.

Índice	Valores de referência
Kappa	Acima de 0,7
Área sob a curva ROC	Acima de 0,8
Verdadeiro-positivos	Acima de 85%
Falso-positivos	Até 8%

de certeza, sejam resultados verdadeiro-positivos ou falso-positivos.

A avaliação dos dados por meio da construção da curva ROC⁽¹¹⁾ permite uma análise gráfica e rápida dos resultados da relação entre sensibilidade e especificidade individual e do grupo.

RESULTADOS

Foram realizadas 1.003 leituras de simuladores do tipo Alvim[®] pelos residentes, sendo 588 imagens lidas em negatoscópio e 415, em monitores de alta resolução.

Os resultados foram baseados na média da leitura da totalidade desses profissio-

nais, comparando-se o desempenho em negatoscópio *versus* monitor.

A Figura 3 apresenta os valores médios dos índices de desempenho entre leitura em negatoscópio e monitor.

A Figura 4 mostra os histogramas da distribuição dos valores de κ nos diferentes sistemas de leitura.

A Figura 5 mostra as curvas ROC nos sistemas de leitura convencional e digital.

A Tabela 3 mostra os resultados de verdadeiro-positivos, falso-positivos e área sob a curva ROC nos dois sistemas de leitura.

O tempo médio de leitura de cada imagem em monitor no início do treinamento foi de 7,5 min, e ao final deste, de 4,0 min.

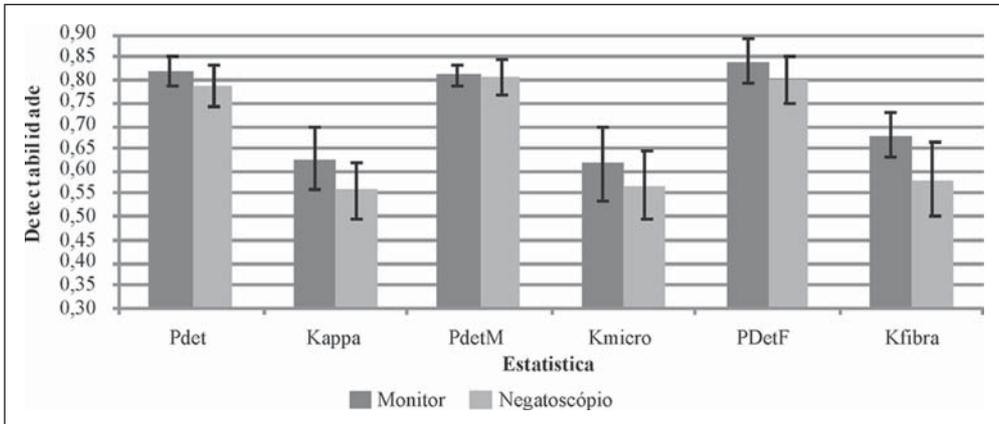


Figura 3. Médias dos índices de desempenho nos diferentes sistemas de leitura (monitor e negatoscópio). Pdet, probabilidade de detectabilidade; K, valores de kappa; M, microcalcificações; F, fibras.

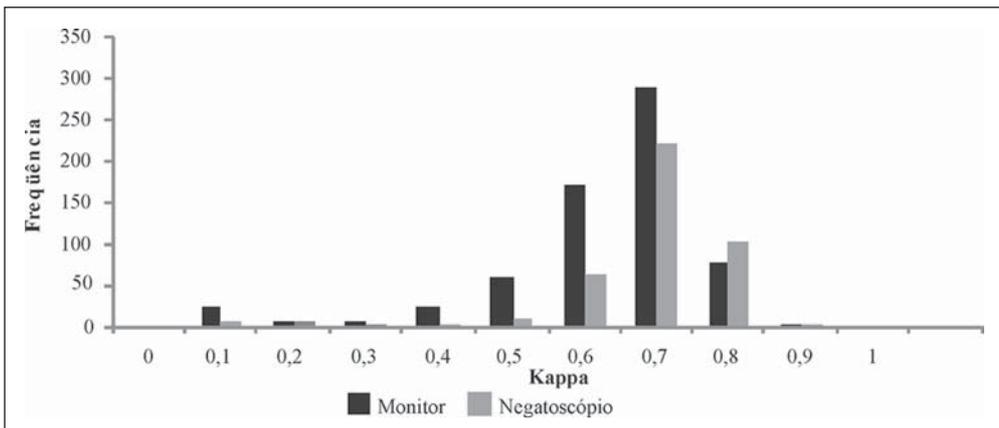


Figura 4. Histogramas da distribuição dos valores de kappa nos diferentes sistemas de leitura.

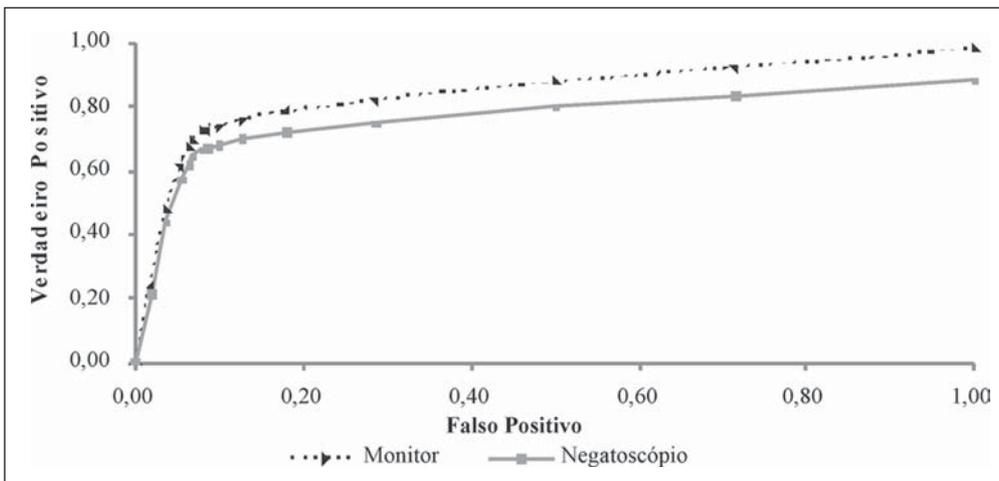


Figura 5. Área sob a curva ROC nos sistemas convencional ($A_z = 0,761$) e digital ($A_z = 0,810$).

Tabela 3 Valores de verdadeiro-positivos, falso-positivos e da área sob a curva ROC nos diferentes sistemas de leitura.

Sistema de leitura	Verdadeiro-positivos	Falso-positivos	Área sob a curva ROC
Negatoscópio	66%	8%	0,761
Monitor	70%	6%	0,810

Valores de referência (desejáveis)⁽⁹⁾: verdadeiro-positivos > 85%; falso-positivos até 8%; área sob a curva ROC superior a 0,8.

DISCUSSÃO

O uso da tecnologia digital na interpretação de imagens poderá beneficiar seus usuários, porém, é necessário que o profissional seja treinado para a adaptação aos novos padrões visuais. As vantagens são evidentes quando os profissionais são treinados e qualificados de forma adequada⁽¹³⁾.

O presente trabalho pretende auxiliar no treinamento dos profissionais no período de incorporação desta tecnologia e contribuir para redução das variações ou discordâncias entre as interpretações de profissionais da área, minimizando os resultados falso-positivos e falso-negativos.

A interpretação de imagens em monitores tem sido amplamente estudada, com o objetivo de avaliar o desempenho dos profissionais na interpretação de imagens em comparação ao sistema convencional^(14,15). A maioria desses estudos conclui que a interpretação das imagens em monitores é bem aceita, a adaptação do especialista é rápida e a precisão e o tempo gasto na interpretação são comparáveis aos do sistema convencional, desde que o profissional esteja devidamente treinado^(16,17).

Uma análise geral mostrou que a Pdet dos residentes durante o processo de aprendizagem foi superior no sistema digital. No entanto, devido à característica deste indicador, é necessário que se avalie os resultados falso-positivos para sua correta interpretação. Segundo os valores de referência estipulados, são aceitáveis resultados falso-positivos de até 8%.

A Pdet superior (= 0,818), associada ao menor índice de falso-positivos (6%), confirma a melhor identificação de estruturas no monitor. Esses resultados indicam que a leitura digital apresentou maior sensibilidade e especificidade.

Além disso, os resultados de κ reforçam que houve uma melhor detectabilidade de estruturas na leitura digital, tanto para a

identificação de microcalcificações como para fibras. Porém, foi a leitura de fibras que apresentou melhora evidente, sugerindo que a manipulação do brilho e contraste propiciada pelo sistema digital pode ter influenciado nesse resultado.

Com os resultados dos valores de κ foi possível construir um histograma da distribuição desses dados, cuja base permite avaliar a variabilidade da leitura intraobservador (Figura 4). Nessa figura observa-se uma base mais estreita e com distribuição simétrica (semelhante a uma curva de distribuição normal), revelando uma menor subjetividade das leituras efetuadas nos monitores.

A avaliação do tempo despendido na leitura digital mostrou que, com o evoluir do treinamento, os residentes diminuíram este tempo, sem, no entanto, comprometer o desempenho. Isto mostra a efetividade do treinamento.

A análise geral mostra que para o profissional menos experiente os resultados falso-positivos representam uma grande influência sobre a acurácia final⁽¹³⁾. Esta menor acurácia deve-se, principalmente, à falta de experiência da percepção durante o treinamento com mamografia, fator que limita a habilidade em reconhecer “objetos” e dificulta o reconhecimento de lesões reais e artefatos. A solução proposta é um tutor de treinamento, sistemático, que associe a percepção da imagem com o retorno sobre as razões que auxiliam a tomada da decisão⁽¹⁸⁾.

A velocidade da leitura e o melhor desempenho dos observadores experientes na interpretação de mamografias envolvem uma mudança no mecanismo de percepção da imagem, do processo de “ponto-a-ponto”⁽¹⁸⁾ para o modelo global, relativamente mais rápido⁽¹⁹⁾. Esse processo de mudança é dependente, entre outros fatores, da experiência no método e da presença de um “tutor” real ou virtual.

Ainda que nossos resultados indiquem um desempenho superior neste tipo de leitura (imagem digital de simuladores com possibilidade de manipulação), alguns aspectos inerentes do método podem ter colaborado nessa conclusão. O treinamento em filmes foi realizado anteriormente ao treinamento em sistema digital; assim, o profissional obteve, ainda que pequena, uma experiência na leitura dos simuladores. A motivação pessoal também pode ter influenciado a concentração no momento da leitura em monitor. O contato inicial com a inovação tecnológica pode também ter colaborado com maior interesse nesta etapa do treinamento.

Convém advertir que os resultados aqui apresentados foram baseados em leitura de simuladores, com o objetivo de quantificar os dados e analisar se o treinamento interferiu no desempenho final. Deve-se ressaltar que na leitura da mamografia, em que o tecido fibroglandular e as lesões apresentam-se sob diferentes padrões, o processo de detecção torna-se ainda mais complexo e exige muita experiência do profissional.

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.

Acreditamos que esse método de treinamento utilizando um *software* dedicado constitui-se em uma importante ferramenta auxiliar no complexo processo multifatorial do aprendizado da mamografia.

Agradecimentos

À Fundação de Amparo à Pesquisa do Estado de São Paulo (Fapesp), pelo apoio financeiro.

REFERÊNCIAS

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.