

Avaliação de desempenhos de esquema de diagnóstico auxiliado por computador (CAD) para diferentes grupos de imagens mamográficas

Computer-aided diagnosis (CAD) performances evaluation to different mammographic image groups

Ana C. Patrocínio¹, Michele F. Angelo², Simone Elias³, Leandro P. Freitas⁴, Homero Schiabel⁴ e Regina B. Medeiros³

¹Faculdade de Engenharia Elétrica da Universidade Federal de Uberlândia (UFU) – Uberlândia (MG), Brasil.

²Departamento de Tecnologia da Universidade Estadual de Feira de Santana (UEFS) – Feira de Santana (BA), Brasil.

³Departamento de Diagnóstico por Imagem da Universidade Federal de São Paulo (Unifesp) – São Paulo (SP), Brasil.

⁴Departamento de Engenharia Elétrica da Universidade de São Paulo (USP) – São Carlos (SP) Brasil.

Resumo

Este trabalho descreveu os testes de um esquema de diagnóstico auxiliado por computador (CAD) com dois diferentes grupos de imagens mamográficas e comparou com o desempenho das respostas dos especialistas. Foram utilizadas imagens com comprovações patológicas com regiões de interesse (RIs) com nódulos benignos e malignos. O grupo 1 de imagens foi composto por 102 RIs apenas com nódulos malignos, e o grupo 2 por 50 RIs, contendo nódulos benignos e malignos. As imagens do grupo 1 passaram por dupla leitura de especialistas e suas respostas foram comparadas com as do CAD. O CAD apresentou área sob a curva ROC (A_z) de 0,94 e 0,84 para os grupos 1 e 2 respectivamente. Enquanto os especialistas apresentaram A_z de 0,85 para o grupo 1.

Palavras-chave: diagnóstico auxiliado por computador; diagnóstico por imagem; mamografia.

Abstract

This work described the tests of computer-aided diagnosis (CAD) scheme with two different groups of mammographic images and to compare with the performance of specialists. Images were used with pathological proofs with regions of interest (ROI) containing benign and malign nodules. The images of group 1 were composed by 102 ROI with malign nodules only, and the group 2 composed by 50 ROI with benign and malign nodules. The images of the group 1 were submitted by double reading of specialists and your answers were compared with the one of CAD. CAD presented area under the curve ROC (A_z) of 0.94 and 0.84 for the groups 1 and 2 respectively. While the specialists presented A_z 0.85 for the group 1.

Keywords: computer-aided diagnosis; diagnostic imaging; mammography.

Introdução

Esforços têm sido realizados a fim de se detectar precocemente o câncer de mama, por ser essa a forma mais efetiva de diminuir a mortalidade por essa doença. Isso é possível por meio da mamografia, que é considerada o exame de maior sensibilidade para o rastreamento do câncer de mama ainda impalpável.

A fim de colaborar na avaliação e análise de mamografias digitais, diversos trabalhos vêm sendo desenvolvidos utilizando esquemas computacionais (CAD – computer-aided diagnosis), que têm por objetivo auxiliar no

diagnóstico de achados mamografias dando ao especialista uma “segunda opinião”. Entre as técnicas empregadas, há um particular interesse nas que buscam detectar e/ou classificar microcalcificações¹⁻⁷ e nódulos⁸⁻¹³.

Estudos indicam que o desempenho de um CAD está diretamente relacionado com a base de imagens que é utilizada para a sua validação. Segundo Wirth¹⁴, o uso extensivo de coleções de mamografias com características variadas pode levar ao desenvolvimento de algoritmos mais robustos.

Dessa forma, o objetivo deste trabalho foi analisar o resultado de um CAD para detecção e classificação de massas nodulares quando aplicado a grupos diferentes de imagens.

Correspondência: Ana Claudia Patrocínio – Faculdade de Engenharia Elétrica - Universidade Federal de Uberlândia – Av. João Naves de Ávila, 2160 - Bloco 3N – Campus Santa Mônica – CEP: 38400-902 – Uberlândia – MG – E-mail: acpatroc@yahoo.com.br

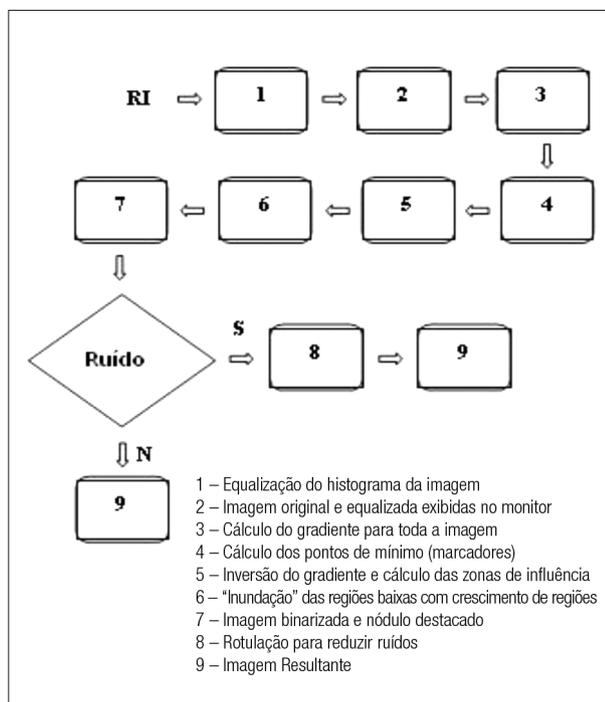


Figura 1. Diagrama esquemático do sistema para detecção de nódulos¹³

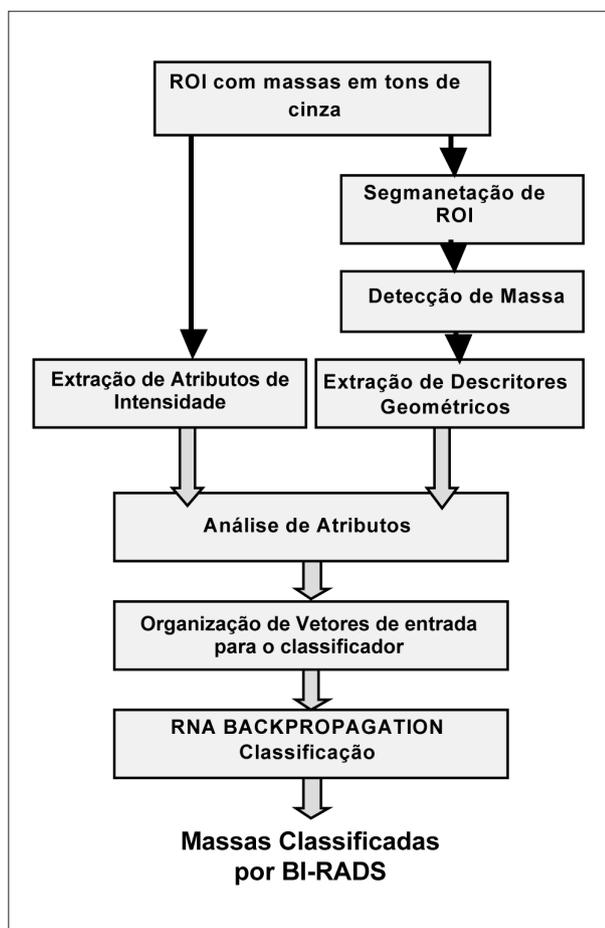


Figura 2. Diagrama esquemático do sistema de classificação de massas nodulares¹⁷

Material e métodos

Foram comparadas respostas do CAD para dois grupos distintos de imagens, o grupo 1 corresponde a 102 imagens com nódulos malignos adquiridas de um mesmo equipamento mamográfico analógico. O grupo 2 correspondeu a 50 imagens sendo 34 contendo nódulos malignos e 16 nódulos benignos, porém no grupo 2 as imagens foram adquiridas em equipamentos diferentes, incluindo imagens de equipamentos digitais e analógicos, os quais passaram por testes de qualidade, verificando-se que estavam em conformidade com o exigido pela portaria 453/1998.

As imagens adquiridas em equipamentos analógicos foram digitalizadas em um scanner Lumiscan 75 com 12 bits de resolução de contraste e 70 µm de resolução espacial.

As imagens do grupo 1 foram submetidas à dupla leitura de especialistas e tanto as imagens do grupo 1 quanto as do grupo 2 possuem confirmação patológica.

A técnica de segmentação¹³ desenvolvida no Laboratório de Análise e Processamento de Imagens Médicas e Odontológicas (LAPIMO) da Escola de Engenharia de São Carlos da Universidade de São Paulo (EESC-USP) foi utilizada visando à detecção de massas suspeitas em regiões de interesse (RIs) extraídas de imagens mamográficas digitalizadas. Originalmente, implementada para casos de mamas densas, onde a principal característica da RI é a alta densidade e, portanto, contraste reduzido, uma técnica de pré-processamento, equalização do histograma, foi implementada a fim de melhorar o contraste da imagem.

Em seguida, a segmentação foi aplicada com base na Transformada Watershed^{15,16}. Essa transformada investiga os níveis de cinza do histograma a partir da frequência desses níveis, em que uma mesma frequência corresponde às estruturas de interesse e altas frequências correspondem a gradiente e bordas.

Para corrigir os defeitos residuais, uma técnica de pós-processamento foi aplicada. Para esta fase utilizou-se a técnica de rotulação, que atribui uma única cor para pixels pertencentes a uma mesma região conectada. Após a rotulação, a imagem resultante é alinhada à imagem original em níveis de cinza, com o intuito de comparar as duas imagens e destacar o nódulo.

A Figura 1 apresenta um diagrama esquemático do sistema de detecção é apresentado.

Após a detecção o próximo passo a ser realizado pelo CAD é classificar o nódulo através do padrão BI-RADS®. Para isso, foi utilizado o esquema computacional desenvolvido por Patrocínio¹⁷, que consiste na classificação automática de achados mamográficos, através de uma abordagem híbrida de classificadores e extração de atributos (Figura 2). Foi implementada uma rede Backpropagation que classifica nódulos utilizando descritores geométricos e uma entrada com informação extraída de atributos de intensidade, produzindo na saída classificação da região de Interesse entre as categorias BI-RADS®.

Resultados

A análise do desempenho do CAD considerou as categorias BI-RADS 1, 2 e 3 como sendo lesões negativas e as 0, 4 e 5 como sendo lesões positivas, para que a análise seguindo a teoria das curvas ROC¹⁸ fosse facilmente empregada na avaliação dos resultados. Uma vez que as curvas ROC quantificam o desempenho do sistema considerando acertos e erros em relação aos casos positivos e negativos apenas.

A Tabela 1 apresenta as respostas do CAD para o grupo 1 de imagens, para o grupo 2 e para os especialistas em relação ao grupo 1, segundo a teoria de detecção de sinais utilizada em curvas ROC, tais como: probabilidades de acertos [verdadeiras positivas (VP) e verdadeiras negativas (VN)]; probabilidades de erros [falsos positivos (FP) e falsos negativos (FN)]; e área sob a curva da característica operativa do receptor (curva ROC) (A_z).

Tanto para os resultados quantificados dos especialistas quanto do CAD para o grupo 1 de imagens não foram computados resultados verdadeiros negativos e nem resultados falsos positivos pois neste grupo de imagens não existiam lesões benignas.

As Figuras 3, 4 e 5 correspondem às curvas ROC dos sistemas apresentados na Tabela 1 – CAD (grupo 1), CAD (grupo 2) e especialistas.

Discussão e conclusões

O CAD apresentou um alto desempenho ($A_z = 0,94$) nos testes com as imagens do grupo 1, bem como nos testes das imagens do grupo 2 ($A_z = 0,84$), principalmente considerando que no grupo 2 foi possível computar tanto as taxas erros FP e FN e de acertos VP e VN.

Para ambos os grupos de imagens, 1 e 2, o CAD teve alta sensibilidade com taxas de acertos VP (0,84 e 0,82 respectivamente), e VN de 0,81 para o grupo 2. Embora a sensibilidade do CAD tenha sido próxima para os dois grupos de imagens, o desempenho do sistema nos dois grupos teve uma diferença significativa. Isso pode ser justificado pelas taxas de erro FP e FN que puderam ser computadas no grupo 2, uma vez que este grupo apresentou maior variedade de lesões incluído os nódulos benignos. Ainda considerando o desempenho do CAD para os grupos 1 e 2 de imagens, a taxa FN para ambos, embora abaixo de 0,2 sugere que as rotinas de segmentação e classificação devem ser aprimoradas com treinamento de redes neurais com maior número de casos variados na tentativa de diminuir o erro, pois se pensarmos que este número representa casos que deixarão de ser tratados, as respostas FN do CAD são significativas.

A taxa FP do CAD para o grupo 2, embora não tenha sido muito baixa, ficou abaixo de 0,2, o que mostra ganhos em relação à literatura nas taxas apresentadas nas avaliações feitas por especialistas.

Tabela 1. Resultados de respostas do CAD para os grupos 1 e 2 e dos especialistas para o grupo 1

	VP	VN	FP	FN	A_z
CAD (grupo 1)	0,84	-	-	0,16	0,94
CAD (grupo 2)	0,82	0,81	0,19	0,18	0,84
Especialistas	0,74	-	-	0,26	0,85

VP: verdadeiras positivas; VN: verdadeiras negativas; FP: falsos positivos; FN: falsos negativos; A_z : área sob a curva ROC

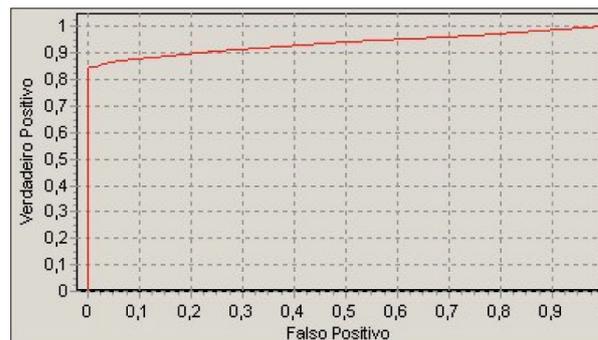


Figura 3. Curva ROC do desempenho do CAD para o grupo 1

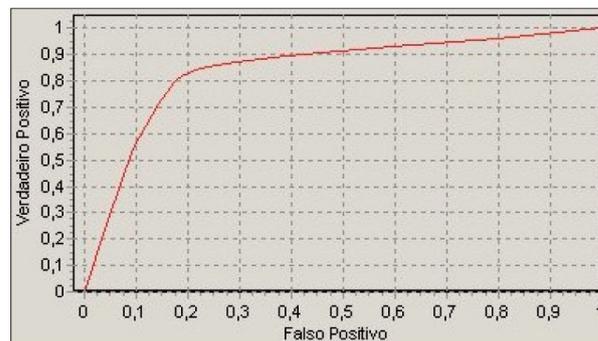


Figura 4. Curva ROC do desempenho do CAD para o grupo 2

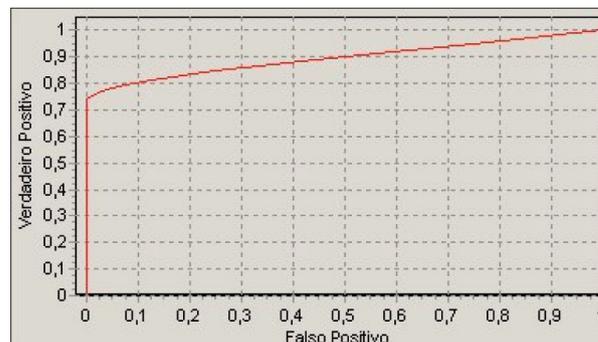


Figura 5. Curva ROC do desempenho dos especialistas para o grupo 1

Com a análise dos resultados dos especialistas na avaliação das imagens do grupo 1, o CAD mostra superioridade tanto na sensibilidade (VP), quanto no desempenho (A_z), porém na taxa de erro (FN) os especialistas tiveram uma taxa bastante alta conforme mostrado na Tabela 1.

Foram computados para o grupo 1 os acertos e erros caso a caso tanto dos especialistas quanto do CAD e foi possível comparar pontualmente os erros de ambos os sistemas. Dessa forma observou-se que os dois sistemas (especialistas e CAD) são complementares, pois em 92% dos casos indicados como FN pelos especialistas na avaliação das imagens do grupo 1, os mesmos casos foram classificados corretamente pelo CAD. E em 87% dos casos classificados como FN pelo CAD, os mesmos casos foram indicados corretamente pelos especialistas.

Considerando todas as respostas (VP e FN) na avaliação do grupo 1, por ambos sistemas (especialistas e CAD), houve 61% de concordância entre eles.

Os resultados das avaliações dos grupos 1 e 2 apresentados mostram uma flexibilidade do CAD quando aplicado a imagens adquiridas em equipamentos diferentes com lesões classificadas em diversas categorias BI-RADS e consequentemente de caracterização de formatos e tamanhos variados, isso mostra que as técnicas são robustas e o classificador apresenta boa generalização.

A qualidade das imagens, assegurada pelos testes de controle de qualidade dos equipamentos mamográficos, os quais as imagens foram adquiridas, contribuíram para que o desempenho do CAD fosse satisfatório em ambos os grupos aqui testados.

Agradecimentos

Os autores agradecem à Fundação de Amparo à Pesquisa do Estado de São Paulo (FAPESP) pelo apoio financeiro.

Referências

- Goes CE, Schiabel H, Nunes FLS. Evaluation of microcalcifications segmentation techniques for dense breast digitized images. *J Dig Imag*. 2002;15(1):231-33.
- Cheng HD, Cai X, Chen X, Hu L, Lou X. Computer-aided detection and classification of microcalcification in mammo-grams: a survey. *Pattern Recognition*. 2003; 36:2937-91.
- Patrocínio AC, Schiabel H, Romero RAF. Evaluation of bayesian network to classify clustered microcalcifications. In: *Proceedings of SPIE. San Diego (CA) – USA*; 2004;5370:1026-33.
- Papadopoulos A, Fotiadis DI, Likas A. Characterization of clustered microcalcifications in digitized mammograms using neural networks and support vector machines. *Artif Intell Med*. 2005;34:141-50.
- Peng Y, Yao B, Jiang J. Knowledge-discovery incorporated evolutionary search for microcalcification detection in breast cancer diagnosis. *Artif Intell Med*. 2006;37(1):43-53.
- Jiang J, Yao B, Wason AM. A genetic algorithm design for microcalcification detection and classification in digital mammograms. *Comput Med Imaging Graph*. 2007;31(1):49-61.
- Halkiotis S, Botsis T, Rangoussi M. Automatic detection of clustered microcalcifications in digital mammograms using mathematical morphology and neural networks. *Signal Processing*. 2007;87:1559-68.
- Hadijski L, Chan HP, Sahiner B, Helvie MA, Roubidoux MA, Blance C, et al. Improvement in radiologists' characterization of malignant and benign breast masses on serial mammograms with computer-aided diagnosis: an ROC study. *Radiology*. 2004;233(1):255-65.
- Ribeiro PB, Schiabel H, Patrocínio AC, Romero RAF. Selection of the Best Features for the Classification of Contour of Mammary Nodules Using Artificial Neural Networks. *Proceedings of Annual Meeting of the Society for Imaging Informatics in Medicine (SIIM)*; Providence; 2007;92-3.
- Varela C, Timp S, Karssemeijer N. Use of border information in the classification of mammographic masses. *Phys Med Biol*. 2006;51:425-41.
- Pasquale D, Fantacci ME, Kasae P, Retico A. Characterization of mammographic masses using a gradient-based segmentation algorithm and a neural classifier. *Comp Biol Med*. 2007;37(10):1479-91.
- Younesi F, Alam N, Zoroofi RA, Ahmadian A, Guiti M. Computer-aided mass detection on digitized mammograms using adaptive thresholding and fuzzy entropy. *Conf Proc IEEE Eng Med Biol Soc*. 2007;2007:5638-41.
- Schiabel H, Santos VT, Angelo MF. Segmentation technique for detecting suspect masses in dense breast digitized images as a tool for mammography CAD schemes. In: *23rd Annual ACM Symposium on Applied Computing*. 2008;1333-37.
- Wirth M, Lyon J, Fraschini M, Nikitenko D. The effect of mammogram databases on algorithm performance. *Proceedings of 17th IEEE Symposium on Computer-Based Medical Systems*. 2004;15-20.
- Beucher S, Lantuéjoul C. Use of Watersheds in contour detection. In: *International Workshop on Image Processing. Real Time Edge and Motion Detection/Estimation, Rennes, France*; 1979. p. 2.1-2.12.
- Roerdink Jos BTM, Meijster A. The watershed transform: definitions, algorithms and parallelization strategies. *Fundamenta Informaticae*. 2000;41:187-228.
- Patrocínio AC. Classificador automático de achados mamográficos em imagens digitais de mamas densas utilizando técnicas híbridas. [tese]. São Carlos: Departamento de Engenharia Elétrica da Universidade de São Paulo (EESC/USP); 2004.
- Metz CE. Receiver operating characteristic analysis: a tool for the quantitative evaluation of observer performance and imaging systems. *J Am Coll Radiol*. 2006;3(6):413-22.