

## Detecção de regiões assimétricas em mamografias digitais utilizando índice de Getis-Ord.

Antônio Marcos V. Sales<sup>1</sup>, Aristófanés Corrêa Silva<sup>1</sup>, Anselmo Cardoso de Paiva<sup>1</sup>

<sup>1</sup>Núcleo de Computação Aplicada (NCA), Universidade Federal do Maranhão (UFMA)

marco-sales@nca.ufma.br, ari@dee.ufma.br, paiva@deinf.ufma.br

**Abstract.** *Breast cancer accounts for 22% of new cancer cases worldwide each year, according to the National Cancer Institute. Early detection greatly increases the chances of cure. This paper presents the use of the Getis-Ord index in its general form in the search for structural asymmetries between pairs of regions of left and right mammograms, and subsequent training and placement in areas asymmetric and symmetric with a Support Vector Machine (SVM). Regions with lesions were classified as asymmetric with performance ranging from 52% to 100% efficacy. The cases that did not obtain the desired results suggest improvement in feature extraction and increasing training base of SVM.*

**Resumo.** *O câncer de mama representa 22% dos novos casos de câncer no mundo a cada ano, segundo o Instituto Nacional do Câncer. O diagnóstico precoce aumenta bastante as chances de cura. Este trabalho apresenta o uso do Índice de Getis-Ord em sua forma geral na busca de assimetrias estruturais entre pares de regiões de mamografias esquerdas e direitas, e posteriores treinamento e classificação em regiões assimétricas e simétricas com uma Máquina de Vetores de Suporte (MVS). Regiões com lesão foram classificadas como assimétricas com desempenho variando de 52% a 100% de acerto. Os casos que não obtiveram o resultado desejado sugerem a melhoria na extração de características e o aumento da base de treinamento da MVS.*

### 1. Introdução

Câncer é o nome dado ao crescimento desordenado de células que ataca tecidos e órgãos, podendo se tornar muito agressivo e causando o surgimento de tumores malignos. O câncer de mama, segundo tipo mais frequente no mundo, corresponde a 22% dos novos casos a cada ano e é o mais comum no sexo feminino, de acordo com o Instituto Nacional do Câncer (INCA) [Instituto Nacional do Câncer 2010]. O diagnóstico precoce aumenta as chances de cura. A mamografia, o exame mais usado na prevenção e diagnóstico do câncer de mama, permite a descoberta de lesões em fase inicial. A análise de imagens mamográficas pode ser melhorada com o auxílio de processamento digital de imagens e visão computacional.

Estudos anteriores mostraram que imagens de mamografia esquerda e direita da mesma paciente tendem a ter elevado grau de simetria em sua estrutura, sendo observada uma ligação entre o alto grau de assimetria e o risco de câncer [Scutt et al. 2006]. A análise de textura em mamografias [Braz Junior et al. 2006] pode ser auxiliada pelo uso de descritores espaciais [Braz Junior 2008]. Além disso, vêm sendo propostos métodos de análise de simetria entre as mamas em conjunto com a estatística espacial

para uma melhor extração das características [Ericeira et al. 2010].

Este trabalho apresenta o uso do descritor espacial conhecido como Índice de Getis-Ord, em sua forma geral, na detecção de regiões assimétricas em mamografias digitalizadas. Com a detecção das assimetrias alerta-se aos médicos para regiões que merecem melhor observação.

## 2. Índice de Getis-Ord

O índice de Getis-Ord é uma medida de associação espacial que se apresenta em duas formas, local e geral. A sua forma local mede a associação a partir da concentração dos pontos que se encontram dentro de uma distância  $d$  do ponto original. A sua forma geral considera todos os pares de pontos possíveis a uma distância  $d$  entre si [Getis and Ord 1992]. Esta última forma é utilizada neste trabalho e se apresenta como segue:

$$G(d) = \frac{\sum_{i=1}^n \sum_{j=1}^n w_{ij}(d) x_i x_j}{\sum_{i=1}^n \sum_{j=1}^n x_i x_j}, j \neq i \quad (1)$$

Na Equação 1,  $G(d)$  representa o quociente do somatório de todos os pares de valores  $(x_i, x_j)$  que estejam a uma distância menor ou igual a  $d$  um do outro pelo somatório de todos os pares  $(x_i, x_j)$  possíveis.  $w_{ij}$  é a matriz de vizinhança e assume o valor 1 caso o par esteja dentro da distância  $d$  e 0 caso contrário.

Esta medida fornece um índice que indica o grau de associação entre os pixels para toda a região considerada, segundo a distância  $d$ . Isto indica  $G(d)$  como uma medida a ser usada na verificação de simetria entre regiões esquerda e direita, pois a comparação é feita pelo grau de associação dos pixels pertencentes a cada região, o que é diferente de, por exemplo, uma simples subtração entre os valores de intensidade dos pixels. Sendo assim pode-se medir o grau de simetria entre as regiões através da diferença entre seus  $G(d)$ s.

## 3. Metodologia

Foram utilizadas imagens do banco DDSM (*Digital Database for Screening Mammography*) que contém 2620 casos adquiridos a partir de hospitais e instituições [Heath et al. 2001].

Os pares de imagens selecionados foram submetidos à etapa de pré-processamento: redimensionamento para 512x512 pixels para tornar o processo do cálculo de  $G(d)$  mais viável computacionalmente; segmentação das mamas (utilizando Watershed); e registros rígido e deformável (utilizando o algoritmo “Demons”), no intuito de tornar as imagens esquerda e direita de cada par mais próximas uma da outra. Maiores detalhes podem ser encontrados em [Ericeira et al. 2010].

Essas imagens foram divididas em 256 regiões de tamanho 32x32 pixels para calcular os valores de  $G(d)$  para cada região selecionada correspondente nas mamas esquerda e direita do par. Foram calculados 31 valores de  $G(d)$  para cada região,  $d$  variando de 1 a 31 (31 é a distância máxima). Em seguida, para cada distância  $d$ , calculou-se o módulo da diferença entre os valores de  $G(d)$  da mama esquerda e da

direita, ou seja,  $|G(d)_{esquerda} - G(d)_{direita}|$ . Assim obteve-se, para cada par de regiões, 31 valores de  $|G(d)_{esquerda} - G(d)_{direita}|$ . Foram utilizados os pares sem lesão para gerar vetores (de 31 valores cada) pertencentes a regiões simétricas e, a partir dos demais, foram gerados vetores com valores de regiões assimétricas. Mantiveram-se os vetores de regiões simétricas e assimétricas em uma proporção aproximada de 2 para 1, respectivamente. Estes vetores foram usados para treinamento de uma Máquina de Vetores de Suporte (MVS) e posterior classificação de regiões em simétrica e assimétrica. O software utilizado para treinamento e classificação foi o LIBSVM [Chih-Chung and Chih-Jen 2001].

#### 4. Testes e Resultados

Para validar a metodologia proposta selecionaram-se 62 pares (esquerda e direita) de mamografias, sendo 18 pares sem lesões, dos quais foram gerados 704 vetores referentes a regiões simétricas, e 44 pares apresentando algum tipo de lesão, dos quais foram gerados 352 vetores referentes a regiões assimétricas. Estes vetores foram usados exclusivamente no treinamento da MVS. Outros 12 novos pares de imagens foram selecionados para teste. Nesta fase foi usada a técnica *leave-one-out*.

O resultado de um dos testes realizados é mostrado na Figura 1. Corresponde ao par de imagens A1468 do banco DDSM. A MVS treinada classificou como assimétricas, neste teste, 100% das regiões marcadas como lesão no banco DDSM (destacadas em vermelho). Além destas, 35 regiões que não estavam marcadas no banco DDSM foram classificadas como assimétricas (destacadas em verde). A Figura 2 mostra uma comparação entre as médias de  $|G(d)_{esquerda} - G(d)_{direita}|$  de todos os vetores calculados de regiões simétricas e assimétricas da amostra completa, para cada distância.

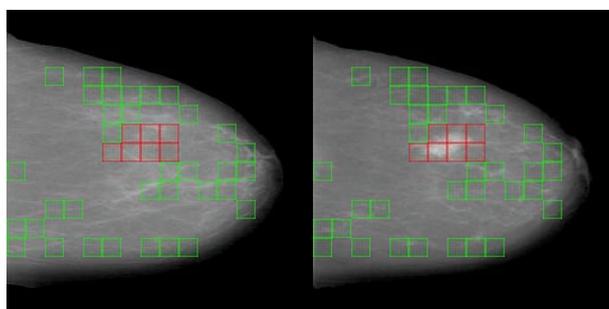


Figura 1. Par de mamogramas de uma paciente com regiões assimétricas detectadas marcadas em verde e vermelho. Regiões em vermelho correspondem à lesão.

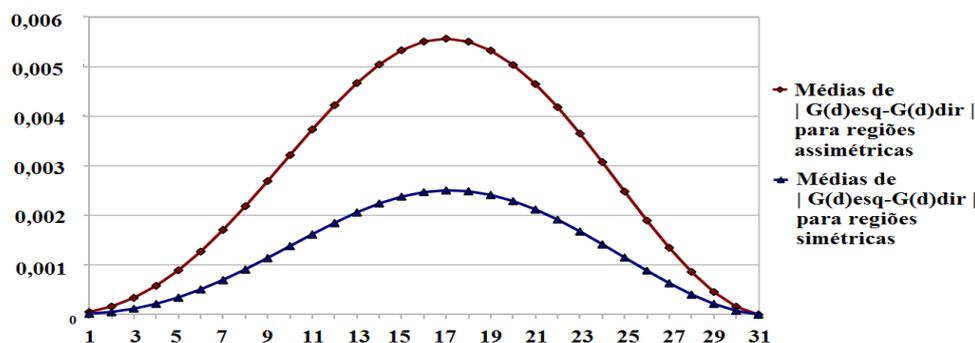


Figura 2. Gráfico comparando as médias de  $|G(d)_{esquerda} - G(d)_{direita}|$  entre todas as regiões simétricas e assimétricas da amostra completa.

Em 6 dos 12 casos testados, 100% das regiões com algum tipo de lesão foram classificadas como assimétricas, e nos outros 6 casos, essa eficácia variou de 52% a 96%, isto é, a metodologia proposta conseguiu detectar apenas parte da lesão. É importante salientar que houve outras assimetrias detectadas, mas que não estavam marcadas como lesão no banco DDSM. Estas assimetrias merecem melhor observação.

## 5. Conclusão

Este trabalho apresentou a utilização do Índice de Getis-Ord na detecção de regiões assimétricas em pares de imagens de mamografia dentro de uma metodologia de extração de características, treino e classificação com uma MVS em regiões assimétricas e simétricas. Regiões indicadas no banco DDSM como possuidoras de lesão foram classificadas como assimétricas com um desempenho variando de 52% a 100% de acerto, além das assimetrias detectadas não marcadas como lesão no banco DDSM. Trabalhos futuros pretendem aumentar a base de treinamento e considerar o uso de outras medidas estatísticas, assim como o aperfeiçoamento da extração de características das regiões em estudo para uma classificação mais eficaz das mesmas e a distinção entre lesão e não-lesão nas regiões assimétricas.

## Referências

- Braz Junior, Geraldo (2008), “Classificação de Regiões de Mamografias em Massa e Não Massa usando Estatística Espacial e Máquina de Vetores de Suporte”, Dissertação (Mestrado em Engenharia de Eletricidade), Centro de Ciências Exatas e Tecnologia, Universidade Federal do Maranhão.
- Braz Junior, Geraldo., Corrêa Silva, Aristófanés., Cardoso de Paiva, Anselmo., Cesar Muniz de Oliveira, Alexandre. (2006), “Identificação de Massas em Mamografias usando Textura, Geometria e Algoritmos de Agrupamento e Classificação”, VI Workshop de Informática Médica - WIM2006.
- Chih-Chung, C., Chih-Jen L. (2001), “LIBSVM: a library for support vector machines”, <http://www.csie.ntu.edu.tw/~cjlin/libsvm/>, Março de 2011.
- Ericeira, D. R., Corrêa Silva, Aristófanés., Cardoso de Paiva, Anselmo. (2010), “Detecção de Regiões Suspeitas em Mamografias Digitais Utilizando Descrição Espacial com Variograma Cruzado”, XII Congresso Brasileiro de Informática em Saúde – CBIS2010.
- Getis, A., and Ord, JK. (1992), “The Analysis of Spatial Association by Use of Distance Statistics.” *Geographical Analysis* 24 (July).
- Heath, M., Bowyer, K., Kopans, D., Moore, R., and Kegelmeyer, WP., (2001), “The Digital Database for Screening Mammography,” in *Proceedings of the Fifth International Workshop on Digital Mammography*, M.J. Yaffe, ed., 212-218, Medical Physics Publishing. ISBN 1-930524-00-5.
- Instituto Nacional do Câncer (2010), “Estimativa 2010 - Incidência de Câncer no Brasil”, <http://www.inca.gov.br/estimativa/2010/>, Março de 2011.
- Scutt D., Lancaster G. A.& Manning J.T. *Breast Cancer Res.*, 8. R14 (2006), “Breast Asymmetry and Predisposition to Breast Cancer”, *Breast Cancer Research*, licensee BioMed Central Ltd.