

Classificação de Mamografias pela densidade do tecido mamário

Welber Ribeiro da Silva

Universidade Federal de Ouro Preto
Campus Universitário, Ouro Preto – MG – Brazil

19 de Abril de 2012

Definição

Definição

Câncer de mama é o desenvolvimento anormal das células do seio. Estas células crescem e substituem o tecido saudável.

Estatísticas

- Segunda principal causa de morte por câncer em mulheres
- Um milhão de novos casos diagnosticados anualmente
- Uma em cada quatorze mulheres desenvolvem o câncer

Tratamento

- O diagnóstico precoce é fundamental para o sucesso do tratamento
- 75-80% das pacientes estão em estágios avançados ao serem diagnosticadas

Mamografia

Definição

Forma particular de radiografia que utiliza níveis de radiação mais baixos que os de uma radiografia convencional.

Influencia do tecido mamário na mamografia

- Composição do tecido mamário pode dificultar a detecção de lesões
- O tecido fibroglandular tem alta densidade e atenua muito a energia efetiva do feixe de raio X
- O tecido gorduroso absorve pouco, refletindo bem o feixe de raio X

Exemplos de Mamografias

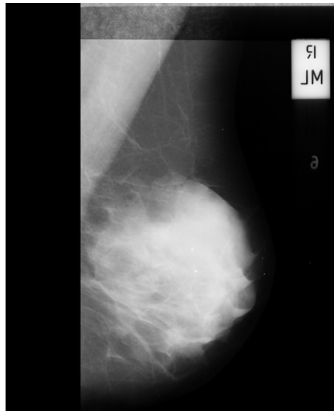


Figura: Mamografias com diferentes proporções de composição do tecido.

Riscos relacionados a composição do tecido mamário

- Mamas densas têm um risco de 4 a 6 vezes maior de desenvolver câncer de mama
- Tecido denso em mais de 50 % da mama pode ser responsável por cerca de um terço dos casos de câncer

Análise da densidade das mamas em mamografias

- Diante desses riscos a densidade é incorporada ao laudo das mamografias
- Para padronizar os laudos é criado pelo Colégio Americano de Radiologia a tabela BI-RADS

BI-RADS I

- predominantemente gordurosas
- até 25% de componente fibroglandular

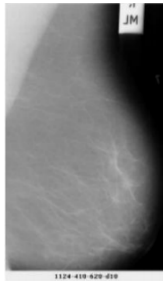


Figura: Mamografia BI-RADS I

BI-RADS II

- parcialmente gordurosas
- 26% a 50% do volume da mama é tecido fibroglandular

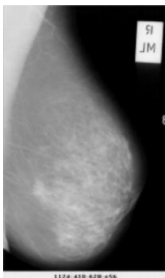


Figura: Mamografia BI-RADS II

BI-RADS III

- densa de forma heterogênea
- 51% a 75% de tecido fibroglandular

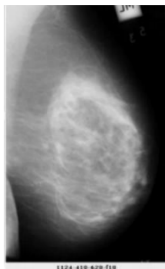


Figura: Mamografia BI-RADS III

BI-RADS IV

- extremamente densa
- mais de 75% de tecido fibroglandular

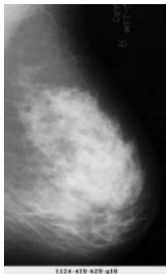


Figura: Mamografia BI-RADS IV

Sistemas CAD para nomografias

- Objetivam reduzir a variabilidade na análise e melhorar a precisão na interpretação
- Aumentam a detecção precoce do câncer de mama principalmente em mamas densas
- Alguns são baseados em CBIR-Content-based image retrieval system

Principais tarefas do sistema CAD para mamografias

- Pré-processamento automático
- Classificação da mamografia de acordo com a densidade do tecido mamário
- Classificação da mamografia de acordo com a lesão
- Segmentação da lesão

Objetivo do trabalho

- Classificação da mamografia de acordo com a densidade do tecido mamário

Hipóteses trabalhadas

- Técnicas de representação através da análise de componentes principais são superiores a outras para representação do tecido da mama;
- A combinação de descritores de textura e outras técnicas representam melhor do que individualmente o tecido da mama;
- A extração de características de apenas uma parte da mamografia é suficiente para a classificação tecido;
- Modelos gerados partir de uma pequena quantidade de mamografias podem ser suficientemente generalistas;

Sheshadri et al

- 6 medidas estatísticas baseadas no histograma
- 320 mamografias do MIAS Database
- 80% de acurácia
- não relata outras medidas de avaliação como a acurácia média da validação cruzada

Subashini et al

- 9 medidas estatísticas do histograma
- 43 imagens do *MIAS Database*
- acurácia média de 95,44% com validação cruzada de 3 folds.

Oliveira et al

- 2DPCA em duas dimensões sobre ROIs de 300 x 300 pixels
- 5.024 mamografias de um total de 10.605 do IRMA database
- porém apenas 3.168 são mamografias originais, o restante são cópias escaneadas em modelos diferentes de scanners
- acurácia média de 83% a 97% com validação cruzada de 10 folds.

Kinoshita et al

- 88 características relacionadas a contorno, textura, momentos, transformada de random, granulometria e histograma
- O vetor de características é reduzido utilizando a técnica PCA
- 1080 mamografias não disponíveis publicamente
- acurácias de 87% a 93% utilizando a validação *leave one out*.

Mamografias utilizadas

- 320 de 322 mamografias de 1024 x 1024 pixels do MIAS Database
- rotuladas de acordo com a densidade em três classes:

Classe	Instâncias
Gorduroso	104
Gorduroso-Glandular	104
Denso-Glandular	112

Tabela: Distribuição das instâncias da base MIAS

Pré Processamento

- Remoção de ruídos, como o músculo peitoral e etiquetas
- Abordagens manuais, semi-automáticas, e automáticas
- Métodos analisados não consideram a existência de tecido mamário sobre o músculo peitoral a ser removido

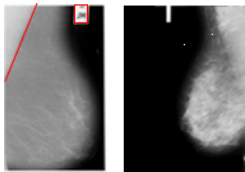


Figura: Pré-processamento (a) Exemplo de regiões a serem removidas.
(b) Exemplo de mamografia com sobreposição do músculo peitoral.

Abordagem utilizada

- Pré processamento manual
- Remoção de ruídos
- Extração de ROIs de 300×300 pixels
- Publicadas em https://github.com/welber/cad_mammography/tree/master/MIAS_PRE_PROC

Características experimentadas

- Descritores de textura estatísticos a partir do histograma;
- Descritores de textura estatísticos a partir da matriz de co-ocorrência;
- Descritores de textura a partir do espectro de Fourier;
- Momentos invariantes de Hu;
- PCA sobre a matriz da imagem vetorizada;
- 2DPCA sobre a matriz da imagem;

Normalização das características

- Todas as características foram normalizadas entre -1 e 1 através do método min-max

Treinamento e classificação

- Classificador SVM por apresentar grande capacidade de generalização
- kernel RBF
- Parâmetros calibrados por meio da técnica grid search para os diferentes agrupamentos de características
- Parâmetros para o melhor conjunto de características $c = 8.192$ e $\gamma = 0,03125$.
- 10-fold cross validation

Resultados dos experimentos

Características	Mamografia	ROI
Text. estatísticas co-ocorrência	72.18	70.93
Text. estatísticas histograma	73.43	71.87
Text. estatísticas histograma e co-ocorrência	77.18	73.75
Text. spectral	61.56	71.56
Momentos invariantes de Hu	62.25	69.06
PCA 5 primeiro componentes	50.62	55.62
PCA 10 primeiros componentes	50.31	57.81
Text. estatísticas histograma, co-ocorrência, e mom. inv.	75.00	74.68
Text. estatísticas histogram, co-ocorrência, e spectral	65.31	71.25

Resultados dos experimentos com 44 e 320 instâncias

Tabela:

Características	44 inst.	320 inst.
Text. estatísticas histograma	74.41	73.43
Text. estatísticas histograma e co-ocorrência	88.37	77.18

Resultados dos experimentos com análise de componentes principais

Tabela:

# PCs	PCA	2DPCA	2DPCA & Text.
Primeiros 5	55.62	59.68	72.18
Primeiros 6	56.56	59.68	72.50
Primeiros 7	57.18	59.37	72.81
Primeiros 8	58.75	59.68	72.50
Primeiros 9	60.62	58.75	72.50
Primeiros 10	57.81	58.43	72.18
Primeiros 11	57.50	58.75	72.81
Primeiros 20	56.25	55.93	73.43

Matriz de confusão para o melhor grupo de características

Tabela: Matriz de confusão para os descritores estatísticos do histograma e da matriz de co-ocorrência

-	D	F	G
D	97	0	10
F	3	91	8
G	12	13	86

Conclusão

- Melhor acurácia obtida foi 77.18% utilizando a combinação dos descritores estatísticos extraídos do histograma aos extraídos da matriz de co-ocorrência
- As técnicas de análise de componentes principais não foram superiores as estatísticas
- A utilização de pequenas ROIs não se provaram inferiores a utilização de toda a mamografia.
- É preciso diferenciar melhor a densidade intermediária

Trabalhos futuros

- Utilizar classificação hierárquica para a identificação e classificação de lesões
- No primeiro nível classificar pela densidade
- No segundo pela existência ou não de lesões
- No terceiro pelo tipo de lesão

Referências

- T. Subashini, V. Ramalingam, and S. Palanivel, "Automated assessment of breast tissue density in digital mammograms," *Computer Vision and Image Understanding*, vol. 114, no. 1, pp. 33–43, 2010.
- J. E. E. D. Oliveira, A. M. C. Machado, G. C. Chavez, A. P. B. Lopes, T. M. Deserno, and A. de A. Araújo, "Mammosys: A content-based image retrieval system using breast density patterns," *Computer Methods and Programs in Biomedicine*, vol. 99, no. 03, pp. 289–297, 2010.
- N. F. Boyd, H. Guo, L. J. Martin, L. Sun, J. Stone, E. Fishell, R. A. Jong, G. Hislop, A. Chiarelli, S. Minkin, and M. J. Yaffe, "Mammographic density and the risk and detection of breast cancer," *The New England Journal of Medicine*, vol. 356, no. 3, pp. 227–236, 2007.
- A. Oliver, X. Llad, E. Prez, J. Pont, E. R. E. Denton, J. Freixenet, and J. Mart, "A statistical approach for breast density segmentation," *Journal of Digital Imaging*, vol. 23, no. 5, pp. 527–537, 2010.
- S. K. Kinoshita, P. M. de Azevedo-Marques, R. R. Pereira, J. A. H. Rodrigues, and R. M. Rangayyan, "Content-based retrieval of mammograms using visual features related to breast density patterns," *Journal of Digital Imaging*, vol. 20, no. 2, pp. 172–190, 2007.
- H. S. Sheshadri and A. Kandaswamy, "Experimental investigation on breast tissue classification based on statistical feature extraction of mammograms," *Computerized Medical Imaging and Graphics*, vol. 31, no. 1, pp. 46–48, 2007.
- J. Suckling, "The mammographic image analysis society digital mammogram database exerpta medica," *International Congress Series*, no. 1069, pp. 375–378, 1994.
- R. C. Gonzalez and R. E. Woods, *Digital Image Processing*, 3rd ed. Prentice Hall, 2008.
- J. Yang, D. Zhang, A. F. Frangi, and J. yu Yang, "Two-dimensional pca: A new approach to appearance-based face representation and recognition," *IEEE Transactions on Pattern Analysis Machine and Intelligence*, vol. 26, no. 1, pp. 131–137, 2004.
- C. Cortes and V. Vapnik, "Support-vector networks," *Machine Learning*, vol. 20, pp. 273–297, 1995.
- S. Theodoridis and K. Koutrambas, *Pattern Recognition*, 4th ed. Academic Press, 2009.
- C.-C. Chang and C.-J. Lin, "Libsvm: A library for support vector machines," *ACM Trans. Intell. Syst.*