

Classificação Automática de Tecido Estromal em Imagens da Próstata Baseada em Descritores de Textura

Domingos L. L. Oliveira¹, Marcelo Z. Nascimento², Leandro A. Neves³,
Moarcir F. Godoy⁴, Yan A. S. Duarte¹, Pedro F. F. Arruda⁵, Dalisio S. Neto⁶

¹Centro de Matemática, Computação e Cognição
Universidade Federal do ABC (UFABC) - Santo André, SP, Brasil

²Faculdade de Computação - FACOM
Universidade Federal de Uberlândia - Uberlândia - MG

³Departamento de Ciência de Computação e Estatística (DCCE)
Universidade Estadual Paulista (UNESP) - São José do Rio Preto, SP, Brasil

⁴Núcleo Transdisciplinar para Estudo do Caos e da Complexidade (NUTTECC)
Faculdade de Medicina de São José do Rio Preto (FAMERP)
São José do Rio Preto, SP, Brasil

⁵Serviço de Cirurgia do Transplante Renal do Hospital de Base
Fundação Faculdade Regional de Medicina (FUNFARME)
São José do Rio Preto, SP, Brasil.

⁶Serviço de Anatomia Patológica do Hospital de Base
Fundação Faculdade Regional de Medicina (FUNFARME)
São José do Rio Preto, SP, Brasil.

domingos.latorre@gmail.com

Abstract. *This paper presents an automatic classification system of stromal tissue in the prostate images based on texture descriptors. Normal, hyperplasia and cancer tissue class were evaluated. Initially, the segmentation of stromal tissue and extraction of regions of interest (ROI's) was performed. Then, wavelet decomposition was applied to the ROI's. Next, the Haralick and SVD descriptors were extracted. Feature selection based on genetic algorithm was applied. Finally, classification was performed using Random Forest algorithm. The area under the ROC curve (AUC) was calculated for normal versus cancer, normal versus hyperplasia and hyperplasia versus cancer, were achieved 0.962, 0.836 to 0.886, respectively.*

Resumo. *Esse artigo apresenta um sistema automático de classificação de tecido estromal em imagens da próstata baseado em descritores de textura. Foram avaliados tecidos das classes normal, hiperplasia e câncer. Inicialmente, foi realizada a segmentação do tecido estromal e extração de regiões de interesse (ROI's). Então, foi aplicada a decomposição wavelet nas ROI's. Em seguida, foram extraídos os descritores de Haralick e SVD. Um método de seleção de características baseado em algoritmo genético foi aplicado. Por fim, a classificação foi realizada utilizando algoritmo Random Forest. A área sob a curva ROC (AUC) foi calculada. Para as classes normal versus câncer, normal versus hiperplasia e hiperplasia versus câncer, foram alcançadas 0.962, 0.836 a 0.886, respectivamente.*

versus hiperplasia e hiperplasia versus câncer, foram alcançados os valores de AUC 0,962, 0,836 e 0,886, respectivamente.

1. Introdução

As doenças da próstata são consideradas um grave problema de saúde, considerando o comprometimento que essas provocam na qualidade de vida do paciente. Dentre essas doenças estão o câncer de próstata e a hiperplasia. Com relação ao câncer de próstata, o Instituto Nacional de Câncer (INCA) [INCA 2011] estimou o número de 60.180 novos casos dessa neoplasia para o ano de 2012. Esse valor corresponde a um risco estimado de 62 casos novos a cada 100 mil homens. Para a hiperplasia, o valor estimado de incidência é de 50% entre homens com idade superior a 60 anos [Paolone 2010].

A análise histológica realizada por meio de microscópio é considerada o procedimento que possibilita avaliar apropriadamente regiões de interesse de uma próstata e, conseqüentemente, contribuindo com o diagnóstico e prognóstico de vários tipos de câncer [Gurcan et al. 2009]. No câncer de próstata essa análise é executada por um patologista que avalia visualmente as características de cada componente histológicos da amostra (lúmen glandular, tecido epitelial e tecido estromal). O padrão de forma, tamanho e conteúdo interno de cada componente histológico são os atributos comumente avaliados em diferentes níveis de magnificação [Epstein and Netto 2007].

Sistemas de apoio ao diagnóstico (CAD) vem sendo desenvolvidos e podem produzir resultados consistentes e reprodutíveis para detecção e estadiamento de doenças. Dentre os sistemas CAD propostos para aplicação em doenças da próstata estão os trabalhos de Diamond *et al.* [Diamond et al. 2004], Doyle *et al.* [Doyle et al. 2006] e Nguyen *et al.* [Nguyen et al. 2012]. No estudo de Diamond *et al.* [Diamond et al. 2004] foi proposto um método de classificação entre as classes normal e câncer utilizando características de textura e área do lumen. Para os testes foram utilizadas 8 imagens de tecido de 100x100 pixels com 40x de magnificação. O método obteve acurácia de 79,3%. Já no trabalho de Doyle *et al.* [Doyle et al. 2006] foi alcançada a acurácia de 88% para classificação entre as classes normal e câncer, utilizando medidas estatísticas, textura e características wavelet. Em outro estudo, proposto por Nguyen *et al.* [Nguyen et al. 2012], os autores utilizaram características da estrutura glândular e informação de texturas, para classificação de imagem saudáveis e câncer. O valor de 85,6% de acurácia foi alcançado com avaliações realizadas em 82 regiões de interesse obtidas de imagem de tecido com 10x de magnificação. É importante considerar que apenas o método de Diamond *et al.* [Diamond et al. 2004] utilizou imagens do tecido estromal. Além disso, nenhum dos estudos realizou testes com imagens de tecido hiperplásico. Em contraste com essas abordagens, é proposto um método de classificação de imagens de tecido da próstata utilizando regiões de interesse (ROI) do tecido estromal das classes normal, hiperplasia e câncer. Esse método é baseado em decomposição wavelet [Mallat 1999], combinações de descritores de texturas de Haralick [Haralick et al. 1973] e da técnica de decomposição em valores singulares (*Singular Value Decomposition* - SVD). O número de características foi reduzido utilizando um método de seleção baseado em algoritmo genético e análise de correlação de atributos e a classificação foi realizada utilizando o método Random Forest.

2. Metodologia

O método proposto consiste na aplicação de cinco etapas: (1) segmentação do tecido estromal, extração de Regiões de Interesse (*Regions of Interest* - ROI's) e componentes de cor *Red* (R), *Green* (G) e *Blue* (B); (2) decomposição wavelet das imagens dos componentes de cor; (3) extração de características de textura utilizando descritores de Haralick, SVD e criação dos vetores; (4) redução de dimensionalidade dos vetores utilizando um método de seleção baseado em algoritmo genético e análise de correlação de atributos; (5) classificação utilizando algoritmo baseado em árvore de decisão Random Forest. A Figura 1 apresenta o fluxograma do método proposto. Nesse estudo foram utilizados os softwares Matlab® R2011a e Weka 3.6.

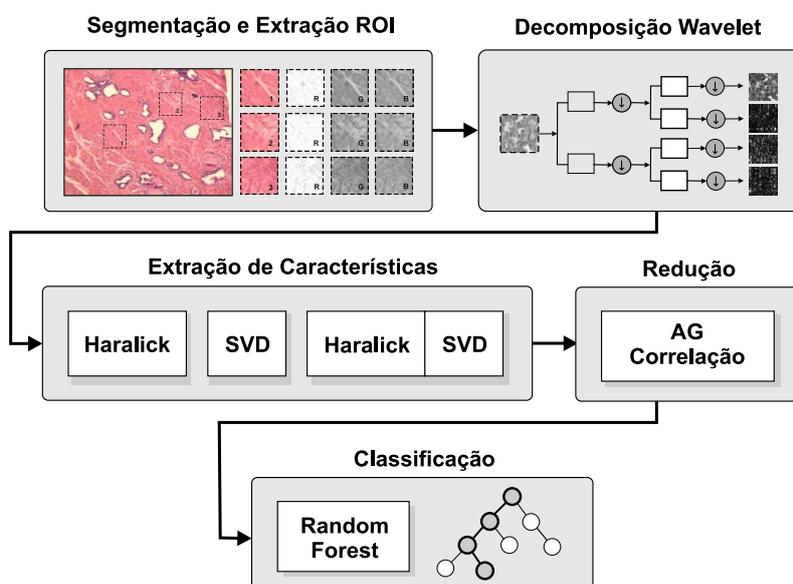


Figura 1. Fluxograma do método proposto.

2.1. Banco de Imagens

O banco de imagens foi composto de lâminas histológicas produzidas a partir das próstatas de 35 pacientes, com idade entre 50 e 75 anos ($64,4 \pm 5,9$ anos), independente de raça, portadores de câncer de próstata. Os pacientes foram atendidos pela equipe do Serviço de Urologia do Hospital de Base de São José do Rio Preto, SP, no período de 2007 a 2008. De cada lâmina, foram selecionadas regiões por dois patologistas e fotografadas digitalmente com magnificação de 40x, utilizando-se um microscópio trinocular Olympus BX41, com objetiva planacromática de 10x acoplada a uma câmera digital Color Samsung SCC-131, com adaptador Olympus U-TV1X-2. As imagens constituíram três grupos (normal, hiperplásico e tumoral) com 35 casos cada. Portanto, foram avaliadas 105 imagens com magnificação de 40x. As imagens foram salvas no padrão RGB, formato tiff, 1200x900 pixels, com 24 bits de quantização.

2.2. Segmentação

Com objetivo de extrair ROI's do tecido estromal foi aplicado um processo de segmentação, dividido em três etapas: (1) segmentação de núcleos de células cubóides; (2) segmentação de lumens glandulares; (3) segmentação do tecido estromal.

A segmentação de núcleos de células cubóides foi realizada utilizando o método descrito em [Oliveira et al. 2012b]. Nesse método, inicialmente, foi extraído o componente de cor R da imagem original. Em seguida, a técnica de alargamento de contraste foi aplicada para otimizar o contraste entre núcleos e regiões adjacentes da imagem. Por fim, foi aplicada a técnica de limiarização global com limiar definido manualmente com valor de 135. As regiões falsos positivos foram removidas por meio da análise de área das regiões candidatas a núcleo. Como resultado da aplicação na imagem original I (Figura 2 (a)), foi gerada a imagem binária I_{ncc} , que representa regiões de núcleos de células cubóides, definidas em preto na Figura 2 (b).

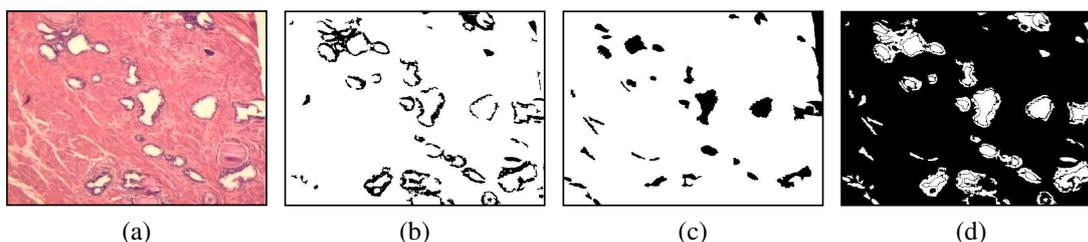


Figura 2. Amostra de tecido da próstata diagnosticada como normal (H&E, 40x) (a). Imagens binárias com regiões de núcleos de células cubóides (b) e lúmens glandulares (c) segmentadas em preto. Imagem binária com regiões de tecido estromal segmentadas em preto (d).

Para segmentar as regiões de lúmens glandulares foi aplicado o método proposto por Oliveira *et al.* [Oliveira et al. 2012a]. Esse método considera as técnicas de alargamento de contraste e o filtro de difusão anisotrópica, ambos aplicados sobre a imagem em níveis de cinza. Em seguida, foi aplicada a técnica de limiarização global com limiar definido manualmente com valor 190. Por fim, os pixels definidos como falsos positivos foram removidos por meio da análise de área de regiões candidatas a lúmen. O resultado foi a imagem binária I_{lg} , que representa as regiões de lúmens glandulares, definidas em preto na Figura 2 (c).

A segmentação do tecido estromal foi obtida por meio de uma operação lógica *OR* entre as duas imagens binárias I_{ncc} e I_{lg} , seguida de uma operação lógica *NOT*. Essas operações foram definidas pela seguinte equação:

$$I_{est}(x, y) = \neg(I_{ncc}(x, y) \vee I_{lg}(x, y)), \quad (1)$$

onde, o valor de cada pixel da imagem I_{est} , nas coordenadas x e y é calculado por meio de operações lógicas utilizando as imagens binárias I_{ncc} e I_{lg} .

Um exemplo de resultado dessa segmentação é apresentado na Figura 2 (d), com regiões de tecido estromal definidas em preto. Utilizando a imagem I_{est} foram extraídas automaticamente três ROI's quadradas de 64x64 pixels das regiões segmentadas do tecido estromal de cada imagem original I do conjunto. Após extração, foram obtidos os componentes de cor R, G e B do modelo de cor RGB de cada ROI. Um exemplo dessa etapa é ilustrado na Figura 3.

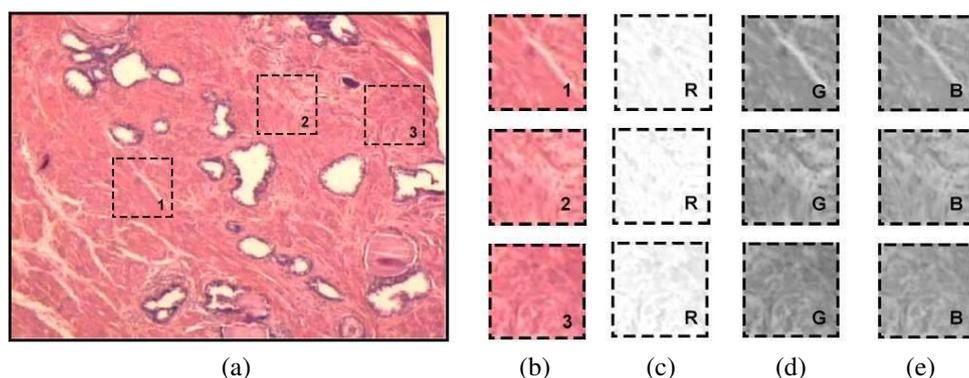


Figura 3. Exemplo de extração de ROI's e imagens de cada componente de cor do modelo de cores RGB. Amostra de tecido da próstata diagnosticada como normal (H&E, 40x) (a), ROI's obtidas do tecido estromal (b), componentes de cor R (c), G (d) e B (e) de cada ROI.

2.3. Extração de Características

Para cada componente de cor de cada ROI, foram aplicadas a etapa de decomposição wavelet, extração dos descritores de Haralick e aplicação da técnica SVD.

2.3.1. Decomposição Wavelet

A abordagem multiescala de imagens permite a análise de dados e detalhes da imagem em diferentes resoluções. Uma das técnicas utilizadas para se obter a representação multiescala é a Transformada Wavelet (TW). Nessa técnica a transformação de um sinal no domínio do espaço para os domínios da frequência e do tempo é obtida pela equação:

$$Wf(a, b) = \int_{-\infty}^{\infty} f(t)\psi_{a,b}(t)dt \quad (2)$$

onde, $f(t)$ é o produto interno de um sinal com uma função wavelet-mãe escalonada e transladada no eixo temporal $\psi_{a,b}(t)$, a representa a escala (dilatação) e b representa a transladação (deslocamento no tempo) da *wavelet-mãe* $\psi_{a,b}(t)$ definida pela equação 3:

$$\psi_{a,b}(t) = \frac{1}{\sqrt{a}}\psi\left(\frac{x-b}{a}\right) \quad (3)$$

A TW apresenta duas variações: Transformada Wavelet Contínua (TWC) e Transformada Wavelet Discreta (TWD). Nesse estudo foi aplicada a TWD por meio de *Quadrature Mirrors Filters* (QFM) e de operações de amostragem [Mallat 1999]. Para o processo de obtenção dos coeficientes wavelets a função wavelet-mãe Daubechies 8. Foram definidos 3 níveis de decomposição e a aplicação da TWD em cada nível gerou 4 subimagens de coeficientes wavelet, sendo elas: aproximação (cA), detalhe horizontal (cH), detalhe vertical (cV) e detalhe diagonal (cD). A imagem de aproximação foi descartada, utilizando-se apenas as subimagens de detalhe, já que essas apresentam descritores de texturas mais relevantes [Arivazhagan and Ganesan 2003]. A Figura 4 apresenta um es-

quema ilustrativo da decomposição wavelet em 3 níveis e um exemplo da aplicação em componente de cor de uma ROI.

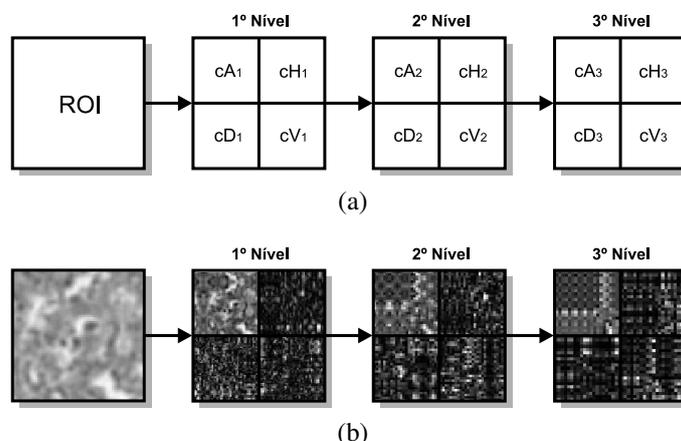


Figura 4. Exemplo de decomposição wavelet de 3 níveis: exemplo ilustrativo (a) e resultado da decomposição de uma ROI (b)

Utilizando as três subimagens dos coeficientes de detalhes dos três níveis de decomposição wavelet, foram totalizadas nove imagens por componente de cor de cada ROI ($cH_1, cV_1, cD_1, cH_2, cV_2, cD_2, cH_3, cV_3, cD_3$). Dessas imagens foram extraídos os descritores de Haralick e aplicada a técnica de SVD.

2.3.2. Descritores de Haralick

Os descritores de Haralick [Haralick et al. 1973] são medidas que utilizam abordagem estatística de segunda ordem e são obtidos por meio do cálculo de matrizes de coocorrência. Essa matriz foi construída para cada subimagem com N coeficientes *wavelets*. Foram extraídos 13 descritores de Haralick, com ângulos θ de $0^\circ, 45^\circ, 90^\circ$ e 135° e distância $d = 1$: energia, correlação, inércia, entropia, momento da diferença inversa, média da soma, variância da soma, entropia da soma, média da diferença, variância da diferença, entropia da diferença, informação da medida de correlação 1 e 2.

2.3.3. SVD

O método da decomposição em valores singulares (*Singular Value Decomposition* - SVD) possui aplicações em diversas áreas: estatística, processamento de sinais e imagens, entre outras. Esse método consiste em decompor uma matriz, cujos elementos podem ser compostos pela intensidade dos *pixels* pertencentes à textura da imagem em uma multiplicação de matrizes. Os valores singulares obtidos como resultado dessa decomposição e suas distribuições fornecem informações úteis sobre textura [Pedrini and Schwartz 2008].

Uma imagem representada pela matriz $A_{M \times N}$, sendo M e N os índices da matriz A e $M > N$, pode-se expressá-la pela equação:

$$A = USV^T \quad (4)$$

Na decomposição da matriz A , denominada decomposição em valores singulares, as matrizes U e V , são matrizes ortogonais e apresentam, respectivamente, $M \times N$ e $N \times N$. Os valores da matriz S são os valores singulares utilizados na construção do vetor que descreve a textura. A extração dos valores singulares das ROI's foram obtidos por meio da aplicação do método descrito em [Simon et al. 2004].

2.4. Redução de Dimensionalidade

Experimentos realizados em algoritmos de classificação baseado em árvore de decisão, tem mostrado que a adição de um atributo randômico ao conjunto de dados causa impacto direto no desempenho da classificação, reduzindo sua acurácia em 5 a 10% [Witten and Frank 2005]. Então, foi aplicada uma técnica de redução baseada em Algoritmos Genéticos (AG), disponível no software Weka. Como função de avaliação do AG foi utilizada a técnica de seleção baseada em correlação [Hall 1999].

2.5. Classificação

Para a etapa de classificação foi utilizado o método Random Forest proposto por Breiman [Breiman 2001]. Apesar de relativamente novo, o Random Forest é utilizado em diversos estudos de classificação de imagens médicas [Ramos et al. 2012, Ko et al. 2011], Durante a execução do método o resultado de vários classificadores são combinados, sendo que cada classificador é uma árvore de decisão construída com profundidade máxima, sem poda.

Para etapa de treinamento e validação do classificador foi utilizada a técnica de validação cruzada (cross-validation). Nessa técnica, o conjunto de dados foi dividido aleatoriamente em n subconjuntos disjuntos de mesmo tamanho. Um desses subconjuntos foi reservado para teste e os demais para o treinamento. Esse procedimento foi repetido n vezes, utilizando em cada uma delas um subconjunto diferente para teste e os restantes para treinamento. Após procedimento, foi calculada a área sob a curva ROC (AUC).

3. Resultados e Discussão

Os resultados dos cálculos de AUC para os vetores construídos com os descritores de Haralick, SVD, Haralick com SVD, para cada combinação de classes de tecido e componentes de cor são apresentados nas Tabelas 1, 2 e 3, respectivamente. As curvas ROC dos melhores resultados de AUC para as combinações de classes Normal versus Câncer, Normal versus Hiperplasia e Hiperplasia versus Câncer são apresentadas na Figura 5 (a), Figura 5 (b) e Figura 5 (c), respectivamente.

Nos resultados da classificação entre classes de tecido normal e câncer, nota-se valores relevantes de AUC de 0,962 (Tabela 1, grupo 1), 0,906 (Tabela 2, grupo 1) e 0,957 (Tabela 3, grupo 1). Isso ocorre pois uma das características apresentadas em imagens do tecido da próstata com câncer é a invasão do tecido estromal por células cubóides do tecido glândular. Essa invasão altera os padrões de forma e intensidade de cor da região, aumentando o seu poder discriminativo por meio de informações de textura.

Os resultados de classificação entre classes normal versus hiperplasia e hiperplasia versus câncer apresentaram AUC de 0,836 (Tabela 3, grupo 2) e 0,886 (Tabela 3, grupo 3). Esses resultados sugerem que a combinação de características de texturas dos coeficientes wavelet podem otimizar os resultados de classificação de imagens de tecido com

Tabela 1. Valores da área sob a curva ROC (AUC) das combinações e componentes de cor utilizando descritores de Haralick.

Combinação	Componente de Cor	AUC	Desvio Padrão
Normal e Câncer	R	0,962	0,0438
	G	0,948	0,0600
	B	0,932	0,0705
Normal e Hiperplasia	R	0,769	0,199
	G	0,729	0,166
	B	0,692	0,226
Hiperplasia e Câncer	R	0,878	0,109
	G	0,840	0,123
	B	0,822	0,151

Tabela 2. Valores da área sob a curva ROC (AUC) das combinações e componentes de cor utilizando SVD.

Combinação	Componente de Cor	AUC	Desvio Padrão
Normal e Câncer	R	0,906	0,092
	G	0,874	0,121
	B	0,901	0,0961
Normal e Hiperplasia	R	0,704	0,181
	G	0,686	0,189
	B	0,747	0,161
Hiperplasia e Câncer	R	0,807	0,148
	G	0,788	0,139
	B	0,741	0,167

Tabela 3. Valores da área sob a curva ROC (AUC) das combinações e componentes de cor utilizando a agregação dos descritores de Haralick e SVD.

Combinação	Componente de Cor	AUC	Desvio Padrão
Normal e Câncer	R	0,939	0,057
	G	0,951	0,089
	B	0,957	0,061
Normal e Hiperplasia	R	0,836	0,177
	G	0,758	0,178
	B	0,775	0,177
Hiperplasia e Câncer	R	0,886	0,0897
	G	0,851	0,0890
	B	0,804	0,134

hiperplasia. Entretanto, os experimentos sugerem que somente informações texturais do tecido estromal não são suficientes para classificação relevante de tecidos hiperplásicos. Para se obter melhores resultados de AUC é necessário investigar combinações de diferentes características do tecido estromal e características de outras regiões da próstata, como lumens e células cubóides.

Com relação a análise de componentes de cor, nota-se na Tabela 1, Tabela 2 e Tabela 3, que o componente de cor R apresentou melhores resultados em 7 das 9 combinações de testes. Isso ocorre devido as informações de contraste de intensidade

entre os diferentes componentes como núcleos e citoplasma das células. Esse contraste está relacionado as propriedades dos corantes Hematoxilina&Eosina, utilizados no processo de coloração do tecido. O corante hematoxilina é atraído pelos ácidos nucléicos presentes nos núcleos celulares, corando-os de azul púrpura, enquanto o corante eosina é atraído pelos elementos básicos do citoplasma das células, corando-o de rosa a vermelho [Junqueira and Carneiro 2008].

Analisando as curvas ROC (Figura 5 (a), Figura 5 (b) e Figura 5 (c)) nota-se que para os vetores de Haralick, SVD e agregação Haralick com SVD, o método proposto apresenta distinções relevantes entre classes normal versus câncer.

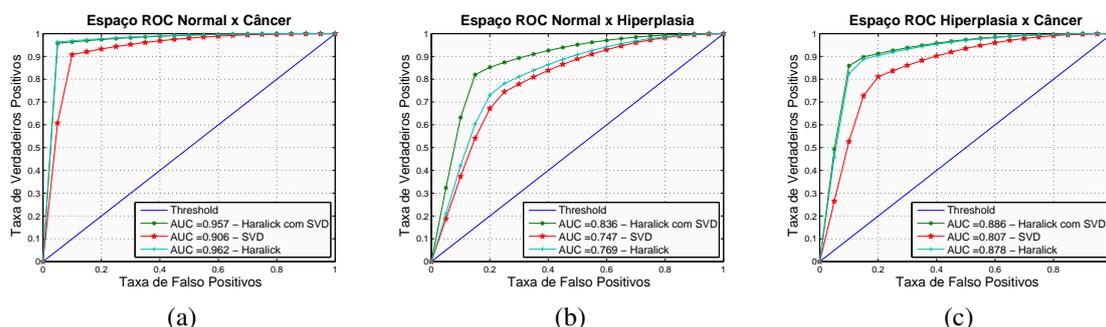


Figura 5. Espaço ROC das combinações de classes Normal versus Câncer (a), Normal versus Hiperplasia (b) e Hiperplasia versus Câncer (c).

4. Conclusão

Nesse artigo foi apresentado um método automático de classificação de imagens de tecido estromal entre as classes normal, hiperplasia e câncer. As principais características desse método são: (1) utilização de características de textura dos coeficientes wavelet; (2) resultados relevantes para classificação entre imagens da classe normal e câncer. (3) utilização de pequenas ROIs de resolução espacial de 64x64, diminuindo o custo computacional e tempo de aplicação. O projeto está em fase de prototipagem. Para trabalhos futuros, novas características do tecido estromal e de componentes como lúmen glandular e células cubóides serão investigados com objetivo de aumentar AUC das regiões com hiperplasia.

5. Agradecimentos

Ao CNPq pelo apoio financeiro.

References

- Arivazhagan, S. and Ganesan, L. (2003). Texture segmentation using wavelet transform. *Pattern Recognition Letters*, 24(16):3197–3203.
- Breiman, L. (2001). Random forests. *Machine learning*, 45(1):5–32.
- Diamond, J., Anderson, N. H., Bartels, P. H., Montironi, R., and Hamilton, P. W. (2004). The use of morphological characteristics and texture analysis in the identification of tissue composition in prostatic neoplasia. *Human Pathology*, 35(9):1121–1131.

- Doyle, S., Madabhushi, A., Feldman, M., and Tomaszewski, J. (2006). A boosting cascade for automated detection of prostate cancer from digitized histology. In *Medical Image Computing and Computer-Assisted Intervention—MICCAI 2006*, pages 504–511. Springer.
- Epstein, J. and Netto, G. (2007). *Biopsy interpretation of the prostate*. Lippincott Williams & Wilkins.
- Gurcan, M., Boucheron, L., Can, A., Madabhushi, A., Rajpoot, N., and Yener, B. (2009). Histopathological image analysis: A review. *Biomedical Engineering, IEEE Reviews in*, 2:147–171.
- Hall, M. A. (1999). *Correlation-based feature selection for machine learning*. PhD thesis, The University of Waikato.
- Haralick, R. M., Shanmugam, K., and Dinstein, I. H. (1973). Textural features for image classification. *Systems, Man and Cybernetics, IEEE Transactions on*, (6):610–621.
- INCA (2011). Estimativa 2012: Incidência de Câncer no Brasil. Technical report, Instituto Nacional de Câncer.
- Junqueira, L. and Carneiro, J. (2008). *Junqueira’s basic histology: text & atlas*. Guanabara Koogan, 11 edition.
- Ko, B. C., Kim, S. H., and Nam, J.-Y. (2011). X-ray image classification using random forests with local wavelet-based cs-local binary patterns. *Journal of Digital Imaging*, 24(6):1141–1151.
- Mallat, S. (1999). A wavelet tour of signal processing.
- Nguyen, K., Sabata, B., and Jain, A. K. (2012). Prostate cancer grading: Gland segmentation and structural features. *Pattern Recognition Letters*, 33(7):951–961.
- Oliveira, D., Nascimento, M., Neves, L., Godoy, M., and Arruda, PFF, N. D. (2012a). Algoritmo para segmentação de lumens glandulares em tecidos da próstata. In *Anais do XXIII Congresso Brasileiro em Engenharia Biomédica XXIII CBEB*.
- Oliveira, D., Nascimento, M., Neves, L., Godoy, M., and Arruda, PFF, N. D. (2012b). Segmentation of cell nuclei regions in epithelium of prostate glands. *Journal of Modelling and Simulation of Systems*, (3):21.
- Paolone, D. R. (2010). Benign prostatic hyperplasia. *Clinics in geriatric medicine*, 26(2).
- Pedrini, H. and Schwartz, W. R. (2008). *Análise de imagens digitais: princípios, algoritmos e aplicações*. Thomson Learning.
- Ramos, R. P., Nascimento, M. Z. d., and Pereira, D. C. (2012). Texture extraction: An evaluation of ridgelet, wavelet and co-occurrence based methods applied to mammograms. *Expert Systems with Applications*.
- Simon, C. P., Blume, L., and Doering, C. I. (2004). *Matemática para economistas*. Bookman.
- Witten, I. H. and Frank, E. (2005). *Data Mining: Practical machine learning tools and techniques*. Morgan Kaufmann.