

Desenvolvimento de descritores baseado em análise de forma para diagnóstico de lesões pulmonares

Antonino Calisto dos S. Neto¹, Alexandre Ribeiro C. Ramos¹, Antonio Oséas de C. Filho¹, Alcilene D. de Sousa¹, Patricia Medyna L. de L. Drumond¹

¹ Laboratório de Investigação em Pesquisa e Poética Digital, Universidade Federal do Piauí, rua Cicero Duarte, bairro Junco, Picos - PI, Brasil

{antuninosantos, alexandre.cajamos, antoniooseas, alcileneeluzsousa}@gmail.com, patriciamedyna@ufpi.edu.br

Abstract. Lung cancer is one of the most common types of cancer and the one with the highest mortality rate in the world. The automation of diagnostic computer vision systems, through the analysis of medical images, provides an interpretation regarding the pathology. The idea is to use the features of the form from the images of lung nodules, then sort on malignant or benign. This paper presents the development of descriptors based on shape analysis for characterization of nodule. The tests have promising results with an accuracy of 92 %, specificity of 89.2%, sensitivity of 91.5 % and a área under the ROC curve of 0.920.

Resumo. O câncer de pulmão é um dos tipos mais comuns de câncer e o que possui maior índice de mortalidade no mundo. A automatização do diagnóstico por sistemas de visão computacional, através da análise das imagens médicas, proporciona uma interpretação a respeito da patologia. A ideia é utilizar as características de forma extraídas das imagens dos nódulos pulmonares, para classificar em maligno ou benigno. Este trabalho apresenta o desenvolvimento de descritores baseado na análise de forma, para caracterização do nódulo. Os testes apresentaram resultados promissores com uma acurácia de 92%, uma especificidade de 89,2%, uma sensibilidade 91,5% e uma área sob a curva ROC de 0,920.

1. Introdução

Com o avanço da tecnologia, diversas áreas têm sido beneficiadas na simplificação de soluções para seus problemas. Uma dessas áreas é a área da saúde, destacando as pesquisas relacionadas ao diagnóstico do câncer. O câncer é caracterizado pelo crescimento desordenado das células, que invadem tecidos e células vizinhas, tornando-se, às vezes, agressivas e incontroláveis [NCI 2017].

Atualmente, o câncer de pulmão, é o câncer com maior taxa de mortalidade mundial no homem e o segundo na mulher, ficando atrás apenas para o câncer de mama [IARC 2017].

Diferentes ferramentas foram desenvolvidas para ajudar no diagnóstico do câncer de pulmão, sendo uma das mais conhecidas as análises de imagens obtidas por meio de Tomografia Computadorizada (TC). A TC surgiu como uma maneira eficaz para auxiliar na detecção de nódulos pulmonares, tornando-se bastante usual para o diagnóstico do câncer.

No entanto, na análise de um exame pode ocasionar alguns problemas no diagnóstico, por ser realizada de forma subjetiva, além da complexidade dos diferentes formatos que um nódulo pode assumir, ocasionando erros em sua análise, causada por fadiga visual, baixa qualidade de imagem, distração, entre outros.

Assim, pode-se utilizar sistemas CADx (“*Computer-Aided Diagnosis*”) com a finalidade de melhorar a acurácia do diagnóstico (maligno e benigno), servindo de uma “segunda opinião” para a tomada de decisão do especialista.

2. Trabalhos Relacionados

Na literatura, muitos trabalhos estão relacionados com o desenvolvimento de sistemas automáticos para classificação dos nódulos pulmonares. A seguir, tem-se os resumos dos trabalhos mais recentes relacionados ao método proposto.

Orozco et al. (2015) propõe um método que avalia os descritores mais significativos utilizando uma *Support Virtual Machine* (SVM). A pesquisa conta com a validação de imagens de 45 TC das bases *Lung Image Database Consortium – (LIDC-IRDI)*, onde obtiveram resultados promissores, com taxas de acertos de sensibilidade de 90,9%, especificidade de 73,91% e acurácia de 82%.

Froner (2015) utilizou em seu trabalho atributos morfológicos, de intensidade e textura para uma melhor interpretação no diagnóstico auxiliado por computador. Ela apresentou uma avaliação para a utilização de dados de pacientes e atributos quantitativos de nódulos pulmonares em imagens de Tomografias Computadorizadas (TC) de pulmão para a elaboração de um modelo que melhor classificasse a malignidade dos nódulos, onde obteve um melhor resultado sob curva ROC de 0.923 de precisão e no melhor caso, uma acurácia de 85,1 %.

No trabalho de Dhara et al. (2016) utilizou-se uma estrutura de segmentação para todos os tipos de nódulos pulmonares (benignos e malignos), baseado na textura interna (sólida / parte sólida e não sólida) e externa (pleural e vascular). Foram utilizados descritores de forma e textura para a representação dos nódulos provenientes da base LIDC-IDRI. Como resultado obteve-se uma sensibilidade de 89,73%, uma especificidade de 86,36% e uma área sob a curva ROC de 0,882.

Em Wang et al (2016) é proposto a metodologia baseada no método *Radiomic* quantitativo para predição de tumor maligno e benigno. Utiliza-se uma análise radiográfica de 150 características quantificando a intensidade, forma e textura da image do tumor pulmonary utilizando a base LIDC-IDRI. Obteve-se como resultados uma acurácia de 86%, uma sensibilidade de 82,5 % e uma especificidade de 89,5 %.

Neste trabalho foi desenvolvido uma metodologia utilizando apenas descritores de forma aplicados em uma quantidade maior de casos, além de possuir uma menor discrepância entre os valores de sensibilidade e especificidade, quando comparados com a maioria dos trabalhos relacionados.

3. Metodologia

A metodologia desenvolvida para a classificação dos nódulos pulmonares em maligno ou benigno seguiu algumas etapas: Inicialmente houve a aquisição das imagens; posteriormente, foram extraídas as características a partir dos índices de forma; em seguida, utilizou-se a ferramenta WEKA para classificação em maligno e benigno.

3.1. Aquisição de Imagens

Para aplicação dos algoritmos desenvolvidos foi utilizada neste trabalho a base de imagens pública LIDC-IDRI, disponibilizada pelo *National Cancer Institute of EUA* (NCI). Esse banco de dados de imagens foi criado a partir da combinação das bases *Lung Image Database Consortium* (LIDC) e o *Image Database Resource Initiative* (IDRI). Essa base foi utilizada para treinamento dos algoritmos estudados, otimização e validação dos resultados adquiridos [Armato III et al. 2011].

A criação da LIDC-IDRI resultou em uma base composta por 1018 exames, de acordo com as informações e marcações da base LIDC. Cada exame dessa base possui um arquivo XML (*eXtensible Markup Language*), com marcações das coordenadas dos nódulos e da quantidade de fatias do exame, descritas por 4 especialistas [Armato III et al. 2011]. Foram utilizados testes de eficiência nessa base, utilizando 1403 nódulos, sendo 1009 benignos e 394 malignos [Carvalho Filho 2016].

3.2. Descritores de Características

A fase de extração foi desenvolvida a partir da obtenção de características baseadas na forma dos nódulos pulmonares, pois estas podem fornecer informações de sua malignidade, como: volume, esfericidade, entre outros.

A fase de classificação utilizou como entradas, as características obtidas na fase de extração, as quais são submetidas à avaliação de um classificador e a partir de um treinamento prévio, informar a natureza maligna ou benigna.

3.3.1. Desproporção Esférica

A desproporção esférica é medida por meio da comparação entre a área de um objeto com a área que este objeto viria a possuir se fosse totalmente esférico, podendo assim, obter uma diferenciação quantitativa de sua estrutura morfológica [Sousa 2007].

A Equação 1 define como é calculada a desproporção esférica de um objeto, onde A é a área da superfície do objeto, R é o raio estimado da esfera como o mesmo volume do objeto. O raio é obtido através da equação 2, onde V é o volume do objeto.

$$DespEsf = \frac{A}{4\pi R^2} \quad (1)$$

$$R = \frac{\sqrt[3]{3V}}{4\pi} \quad (2)$$

3.3.2. Compacidade

Essa medida geométrica é calculada através da medição da densidade em relação a uma figura perfeitamente densa, ou seja, uma esfera.

A Equação 3 apresenta a fórmula para calcular a compacidade [Sampaio 2015], onde p representa área da superfície do objeto e A representa o volume interno do objeto.

$$C = \frac{V}{\frac{4\pi R^3}{3}} \quad (3)$$

3.3.3. Medidas Estatísticas

Tendo em vista essa variação e com o intuito de produzir informações mais consistentes quanto ao diagnóstico, foi calculado o desvio padrão e a média da variância da distância

entre o centro de massa do nódulo e a sua borda. As Equações 4 e 5 apresentam os cálculos dessa medida, onde M representa a média da distância entre o centro do candidato a nódulo e a sua borda, sendo n a quantidade de pontos identificados no decorrer de toda a borda. X_i indica a distância entre o centro do candidato a nódulo e um ponto i da borda.

$$M = \frac{\sum_i^n X_i}{n} \quad (4)$$

$$DP = \sqrt{\frac{\sum_i^n (X_i - M)^2}{n}} \quad (5)$$

3.3.4. Diâmetros de Feret

O Diâmetro de Feret pode ser considerado a distância entre dois planos que definem o limite de um objeto por meio de retas perpendiculares a esse objeto e em uma direção específica. A Figura 1 mostra um exemplo.

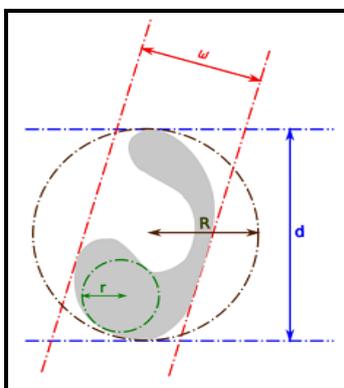


Figura 1. Demonstração de Feret. Adaptado de Sampaio (2015).

Por meio da aplicação dessa técnica sobre uma imagem 3D, é possível extrair informações como volume interno do objeto (A), área da superfície (P), raio maior (R) e menor (r) de um circunferência interna da forma, e os diâmetros mínimo (w) e máximo (d) de Feret.

3.3.5. Esqueleto

Os descritores que utilizam o esqueleto como base, podem ser usados para obter as propriedades do esqueleto, como o tamanho do esqueleto, definido a partir do centro de massa de alguns círculos até as bordas de um limite definido (No caso da Figura 2, utilizou-se um retângulo, para que através deste método, calcular o tamanho do esqueleto), e utilizá-lo para extrair características à respeito da forma [Sebastian 2001]. Como exemplo, pode-se citar um descritor proposto por Klein *et. al* (2000), em que comparam grafos a partir da utilização de um algoritmo de distância de edição.

Considera-se que um ponto pertence ao esqueleto se ele for o centro de um círculo máximo e se este círculo tocar a borda em pelo menos dois pontos diferentes [Braz Júnior 2014]. Essa técnica foi utilizada em conjunto, através do tamanho do esqueleto (E) com os diâmetros mínimo (w) e máximo (d) de Feret para criar novos descritores, como pode ser observado nos índices 15 e 16 da Tabela 1.

Na Figura 2 é ilustrado os pontos pertencentes ao esqueleto e a ilustração dos círculos que tocam os limites das bordas.

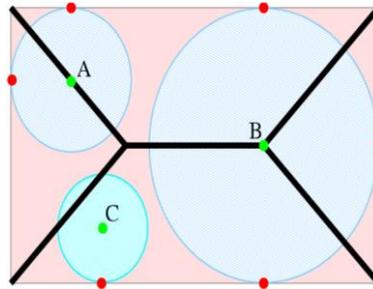


Figura 2. Ilustração do retângulo do esqueleto, onde *A* e *B* pertencem ao esqueleto, e *C* não faz parte do mesmo. Adaptado de [Sampaio 2015].

3.3.6 Descritores combinados com o esqueleto e o Diâmetro de Feret

Para melhorar a precisão das informações obtidas através da forma nódulos pulmonares, foram implementados alguns descritores de forma, descritos no trabalho de Sampaio (2015).

Esses descritores podem ser observados na Tabela 1 onde *A* é o volume interno do objeto, *P* é a área da superfície, *R* é o raio maior e *r* é o raio menor de uma circunferência interna da forma, *w* é o diâmetro mínimo e *d* é o diâmetro máximo de Feret, e *Eq* é o tamanho do esqueleto.

Tabela 1 - Descritores extraídos de acordo com o trabalho de Sampaio (2015).

Descritor	Fórmula	Descritor	Fórmula
1	$\frac{r}{R}$	8	$\frac{\pi w}{P}$
2	$\frac{w}{2R}$	9	$\frac{4\pi A}{p^2}$
3	$\frac{A}{2\pi R}$	10	$\frac{2d}{P}$
4	$\frac{2r}{d}$	11	$\frac{4R}{P}$
5	$\frac{4A}{\pi d^2}$	12	$\frac{2r}{w}$
6	$\frac{R\sqrt{3}}{d}$	13	$\frac{d}{Eq}$
7	$\frac{2\pi r}{P}$	14	$\frac{w}{Eq}$

3.3.7. Esfericidade

O cálculo da taxa de esfericidade (*E*) serve para medir o comportamento do nódulo, levando em consideração um objeto esférico [Silva 2007]. Assim, a esfericidade tem como objetivo medir o quão “perfeito” seria um nódulo, já que um nódulo perfeito, seria equivalente a uma esfera “perfeita” [Silva 2004]. A Equação 6 representa tal conceito.

$$E = \frac{6V\sqrt{\pi}}{A^2} \quad (6)$$

onde V é o volume do nódulo do nódulo, e A é a área da superfície do nódulo.

3.3.8. Densidade Esférica

Com a densidade esférica (*DensEsf*) é possível medir o quão compacto é um objeto, comparando o seu volume em relação ao volume de um *bounding box* [Sousa 2008]. O cálculo da densidade esférica é definida pela Equação 7:

$$DensEsf = \frac{100 NV}{V} \quad (7)$$

onde NV representa o número de *voxels* (*pixel* em 3D) do nódulo e V é o volume do nódulo.

3.3.9. Medidas de Irregularidade

A irregularidade de um nódulo é dado pela razão entre o volume e o quadrado da área de um nódulo [Froner 2015]. O cálculo da medida de irregularidade é definida pela Equação 8:

$$I = \frac{A^2}{V} \quad (8)$$

onde V representa o volume do nódulo e A representa a área da superfície do nódulo.

3.3.10. Circularidade

A circularidade é bastante eficaz na definição de formas, pois representa a medida específica do contorno dos nódulos que estão sendo estudados [Braz Júnior 2014]. O cálculo da Circularidade é definida pela Equação 9:

$$C = \frac{A^2}{4\pi V} \quad (9)$$

onde A é a área da superfície do nódulo e V é o volume do nódulo. Assim, quanto mais próximo o objeto estiver de um círculo, mais próximo de 1 será o valor de sua circularidade [Braz Júnior 2014].

3.4. Classificação

A classificação se utiliza da descoberta de elementos e a relação entre eles, com o intuito de chegar a uma possível conclusão [Giger 1999]. Essa técnica é muito utilizada para o treinamento de algoritmos em processamento de imagens [Landwehr et al. 2003].

Os 4 classificadores aos quais esse trabalho foi submetido (Simple logic [Frank 2014], LMT [Frank 2014], *Bagging* [Frank 2014] e o *Radial Basis Function Classifier* [WEKA 2017]) usaram o método *k-fold cross validation*, em que utilizam 10 conjuntos de características ($k=10$) onde 9 são para treinamento e 1 para testes. Esse processo é repetido 10 vezes, sendo que o conjunto no momento escolhido para o teste, é diferente do anterior, para que no final seja gerada uma média dos resultados.

3.5. Métricas de Avaliação

As métricas utilizadas para validação dos resultados foram especificidade [Zhu et al. 1998], sensibilidade [Zhu et al. 1998], acurácia [Provost e Domingos 2000], índice Kappa [Cook 1998] e a curva ROC [Martinez et al. 2003].

A sensibilidade é conceituada como a probabilidade do teste em fornecer um resultado positivo, desde que o indivíduo seja portador da enfermidade. Este teste define o número de nódulos que são detectados corretamente. [Zhu et al. 1998].

A especificidade, por sua vez, se caracteriza como a probabilidade do fornecimento de um resultado negativo pelo teste. Ele mede a proporção de pessoas sem a doença, indicando a qualidade do teste em identificar os indivíduos não doentes, caracterizando o número de casos que foram identificados como não-nódulos [Zhu et al. 1998].

Já acurácia, calcula o total de acertos baseado em todas as instâncias classificadas, calculado a partir de todos os casos (classificados corretamente e incorretamente) [Provost e Domingos 2000].

O índice Kappa, é considerada como uma das métrica mais importantes na etapa de classificação, representando uma totalidade na matriz de confusão. Ele mede a concordância formulada para medir o desacordo do conjunto de respostas baseando-se em observadores, para assim, chegar em um consenso [Cook 1998]. A qualidade do índice Kappa é definida de acordo com o seu valor em uma escala de precisão, como pode ser observado na Tabela 2.

Tabela 2. Níveis de exatidão do índice Kappa. Adaptado de [Martinez et al. 2003].

Índice Kappa (K)	Qualidade
$K \leq 0,2$	Ruim
$0,2 < K \leq 0,4$	Razoável
$0,4 < K \leq 0,6$	Bom
$0,6 < K < 0,8$	Muito Bom
$K \geq 0,8$	Excelente

E por último, a área sob a curva ROC (*Receiver Operating Characteristic*) descreve, uma maneira aceitável e de grande utilização na literatura médica e na comunidade científica, de avaliar o desempenho quantitativo do método proposto, auxiliando na validação do teste [Martinez et al. 2003].

4. Resultados

Para validação da metodologia desenvolvida foram realizados testes de eficiência com base na LIDC-IDRI utilizando 1403 nódulos, sendo 1009 benignos e 394 malignos, proveniente do resultado da detecção feito por Carvalho Filho (2016). Os resultados e a eficiência das características para cada classificador utilizado podem ser observados na Tabela 3.

A Tabela 3 mostra que o classificador *SimpleLogistic* obteve o melhor resultado com um valor de 89,2% de especificidade (*E*), um valor de 91,5% de sensibilidade (*S*) e 92% de Acurácia (*A*) com um índice Kappa 0,6874 e uma área sob curva ROC de 0,920.

Considerando os termos qualitativos do trabalho, a metodologia apresenta um resultado para o diagnóstico de nódulos pulmonares em maligno ou benigno, que condiz com as opiniões dos especialistas.

Tabela 3. Resultados para os múltiplos classificadores.

Classificador	KAPPA	Curva ROC	S(%)	E(%)	A(%)
<i>Bagging</i>	0,6856	0,909	92,25	88	88
LMT	0,6874	0,920	91,5	89,2	91
<i>RBFClassifier</i>	0,6779	0,916	92,5	87,7	90
<i>SimpleLogistic</i>	0,6874	0,920	91,5	89,2	92

A Tabela 4 mostra alguns trabalhos com resultados aceitos na literatura, comparados com a metodologia proposta.

Tabela 4. Comparação entre os trabalhos relacionados e a metodologia proposta.

TRABALHOS	E(%)	A(%)	S(%)
[Froner 2015]	-	85,1	-
[Orozco et al. 2015]	73,91	82	90,9
[Dhara et al. 2016]	86,36	-	89,73
[Wang et al. 2016]	82,5	86	89,5
Metodologia	89,2	92	91,5

De acordo com a Tabela 4, em comparação com os trabalhos relacionados, a metodologia proposta apresenta um resultado aceitável no que diz respeito às métricas, possuindo um erro aproximado de 8% devido alguns nódulos malignos, às vezes, possuírem forma semelhante aos nódulos benignos, fazendo com que alguns dos descritores não obtivessem uma descrição relevante. Além disso, a base LIDC-IDRI é bastante diversificada quanto a sua composição, com marcações de diferentes especialistas provenientes de laboratórios diferentes [Armato III et al. 2011].

5. Conclusão

É notável que o diagnóstico de câncer de pulmão auxiliado por computador contribui de forma significativa para um diagnóstico correto e em menor tempo, pois representa um considerável aumento na probabilidade de sobrevivência dos pacientes. Esse trabalho apresentou uma metodologia automática para classificação de nódulos pulmonares baseada na análise de forma com múltiplos classificadores, com a intenção de diferenciar os nódulos pulmonares em maligno e benigno, servindo como uma segunda opinião para o especialista.

Os resultados obtidos confirmam um resultado promissor das técnicas de extração da forma, com uma taxa de acerto de 92%, proporcionando um tratamento precoce e com maiores chances de um prognóstico mais favorável ao paciente.

Para trabalhos futuros, pode-se verificar e selecionar alguns dos descritores citados que obtenham melhor relevância quanto à classificação (e/ou implementar outros descritores de forma, combinando-os com atributos funcionais e/ou atributos de textura), utilizando estes como parâmetros para a etapa de classificação em maligno e benigno.

Além disso, pode-se utilizar na etapa de classificação uma rede neural baseada em uma abordagem *Deep Learning* (aprendizagem profunda), onde a rede irá ser treinada, podendo fazê-la reconhecer de forma mais consistente os padrões de características dos nódulos benignos e malignos, obtendo uma classificação mais eficaz.

Referências

- Armato III, S. G.; McLennan G.; Bidaut L.; McNitt-Gray M.F.; Meyer C.R.; Reeves A.P.; Zhao B.; Aberle D.R.; Henschke C.I.; Hoffman E.A.; et al. "The Lung Image Database Consortium (LIDC) and Image Database Resource Initiative (IDRI): a completed reference database of lung nodules on CT scans", *Med. Phys.*, 38 (2011), pp. 915–931.
- Braz Júnior, G. B.. "Detecção de regiões de massas em mamografias usando índices de diversidade geoestatística e geometria côncava". Univ. Federal do Maranhão, Maranhão, Centro de Ciências e Tecnologias 2014.
- Carvalho Filho, A. O de.. Métodos para sistemas CAD e CADx de nódulo pulmonar baseada em tomografia computadorizada usando análise de forma e textura, tese de doutorado, Área de Engenharia de Eletricidade, Universidade Federal do Maranhão, Centro de Ciências Exatas e Tecnologia, São Luís, Maranhão, 2016.
- Chen, W.; Li, Z.; Bai, L.; e Lin, Y.. "NF- kappaB, a mediator for lung carcinogenesis and a target for lung cancer prevention and therapy." *Frontiers in bioscience: a journal and virtual library*, 2011, v. 16, pp. 1172-1185.
- Cook, R. J.. "Kappa and its Dependence on Marginal Rates", *Encyclopedia of BioStatistics*, P. Armitage and T. Colton, eds., John Wiley & Sons, 1998, pp. 2166–2168.
- Dandil, E.; Cakiroglu, M.; Eksi, Z.; Ozkan, M.; Kurt, O.; Canan, A., "Artificial neural network-based classification system for lung nodules on computed tomography scans", *Soft Computing and Pattern Recognition (SoCPaR) 2014 6th International Conference of*, pp. 382-386, pp. 11– 14 Aug. 2014.
- Dhara, A. K., Mukhopadhyay, S., Dutta, A., Garg, M., & Khandelwal, N. A combination of shape and texture features for classification of pulmonary nodules in lung ct images. *Journal of digital imaging*, 2016, vol. 29, n. 4, pp. 466-475.
- Frank, E. "Fully supervised training of Gaussian radial basis function networks in WEKA". Department of Computer Science, The University of Waikato, 2014.
- Frazão, A., "Que tipo de nódulo pode ser câncer," *Tua saúde*, <https://www.tuasaude.com/que-tipo-de-nodulo-no-pulmao-pode-ser-cancer/> (acesso em Nov. 15, 2016).
- Froner, A. P. P.. Caracterização de nódulos pulmonares em imagens de tomografia computadorizada para fins de auxílio ao diagnóstico, Dissertação de mestrado, Área de Engenharia Elétrica, Pontifícia Universidade Católica do Rio Grande Do Sul, Faculdade de Engenharia Elétrica, Troy, Porto Alegre, 2015.
- Giger, M. L.. "Computer-aided diagnosis". *RSNA Categorical Course in Breast Imaging*, 1999, pp. 783-792.
- International Agency for Research on Cancer - IARC. "About ANCErModial". Disponível em: <<http://www-dep.iarc.fr>>. Acessado em 15 Fev. 2017).

- Klein, P. N.; Tirthapura, S.; Sharvit, D.; e Kimia, B. B.. A tree-edit-distance algorithm for comparing simple, closed shapes. In Symposium on Discrete Algorithms, pp. 696–704, 2000.
- Landwehr, N.; Hall, M. ; e Frank, E.. “Logistic model trees,” in Proc. 14th Eur. Conf. Machine Learning, 2003, vol. 2837, pp. 241–252.
- Manzano-Mancho, D.; e Gómez-Pérez, A..” An overview of methods and tools forontology learning from texts”, The Knowledge Engineering Review, vol 19, n.3, pp. 187-212, 2005.
- Martinez, E. Z.. Louzada, N. F.; e Pereira, B. B.. “A curva ROC para tests diagnósticos”. Cadernos Saúde Coletiva 11, 2003, pp. 7-31.
- National Cancer Institute - NCI. “Lung Cancer. U.S. National Institute of Health”. Disponível em: <www.cancer.gov/about-cancer/what-is-cancer --> What Is Cancer?r>. Acessado em 15 Fev. 2017).
- Orozco, H.M.; Villegas, O. O. V.; Sanchez, V. G. C.; Dominguez, M. D. J. N. “Automated system for lung nodules classification based on wavelet feature descriptor and support vector machine”, Biomedical Engineering Online, v.14, no. 9 ,2015.
- Provost, F. e Domingos, P.. “Well-Trained Pets: Improving Probability Estimation Trees”, 2000, Stern School of Business, New York Univ.
- Sampaio, W. B. . “Mass Detection in Mammography Images using a Methodology Adapted to Breast Density”, Univ. Federal of the Maranhão, Maranhão, Center for Science and Technology, 2015.
- Sebastian, T. B.; Klein, P. N.; e Kimia, B. B.. "Recognition of shapes by editing shock graphs," in IEEE ICCV, 2001.
- Silva, A. C.. “Algoritmos para Diagnóstico Assistido de Nódulos Pulmonares Solitários em Imagens de Tomografia Computadorizada”. Tese de Doutorado do Programa de Pós-graduação em Informática, Departament de Informática, Pontificia Universidade Católica do Rio de Janeiro, 2004.
- Silva, A. C.. “Medidas globais em 3d para diagnóstico de nódulo pulmonar”. Disse, Instituto de Matemática Pura e Aplicada–IMPA, 2007.
- Sousa, J. R. F. S.. “Metodologia para detecção automática de nódulos pulmonares”. Dissertação de Mestrado do Programa de Pós-graduação em Engenharia da Eletricidade, Universidade Federal do Maranhão 2007.
- Sousa, J. R. F. S.; Silva, A. C.; e Paiva, A. C.. “Lung structure classification using 3D geometric measurements and SVM. Progress in Pattern Recognition, Image Analysis and Applications.”. Lecture Notes in Computer Science.,2008, v. 4756, pp. 783-792.
- Wang, Jun, et al. "Prediction of malignant and benign of lung tumor using a quantitative radiomic method." Engineering in Medicine and Biology Society (EMBC), 2016 IEEE 38th Annual International Conference of the. IEEE, 2016.
- WEKA – Machine Learning Group at the University of Waikato. “Weka 3: Data Mining Software in Java”. Disponível em: <<http://www.cs.waikato.ac.nz/ml/weka/>>. Acessado 20 Mar. 2017)
- Zhu, W.; Zeng, N.; e Wang, N.. “Sensitivity, specificity, accuracy associated confidence interval and roc analysis with practical sas implementations” in In Proceedings of the NorthEast SAS Users Group. Conference on Evolutionary Programming, San Diego, CA, 1998, pp. 201-208.