

Recuperação de Nódulos Pulmonares por Conteúdo: uma abordagem *Radiomics* em Pesquisa Reprodutível

Marcelo Costa Oliveira¹, David Jones Ferreira de Lucena¹, Ailton Felix¹

¹Instituto de Computação – Universidade Federal de Alagoas (UFAL)
Campus A.C. Simões - Av. Lourival Melo Mota, s/n, 57072-970 – Maceió – AL – Brasil

{oliveiramc,davidjones,aflf}@ic.ufal.br

Abstract. *Early diagnosis and treatment of lung cancer are effective forms of ensure the patient's life. However, the detection and classification of lung nodules are challenging tasks for the specialists, because the nodules are small and have low contrast. The purpose of this work was to evaluate the precision of the 3D Shape and 3D Intensity attributes associated with Margin Sharpness and 3D Texture attributes available in a reproducible research context. The attributes created a 66-dimensional Radiomics vector that was applied to the content-based image retrieval. The algorithm showed precision of the 0.841 and 0.803 in the retrieval of benign and malignant nodules, respectively.*

Resumo. *O diagnóstico e o tratamento precoce do câncer de pulmão são formas efetivas de garantir a vida dos pacientes. Contudo, a detecção e classificação dos nódulos pulmonares são tarefas desafiadoras aos especialistas, pois os nódulos são pequenos e apresentam baixo contraste. O objetivo deste trabalho foi avaliar a precisão dos atributos de Forma 3D e Intensidade 3D associados a atributos de Textura 3D e Nitidez de Borda disponíveis no contexto de pesquisa reprodutível. Os atributos formaram um vetor Radiomics de 66 dimensões e foram aplicados à recuperação de imagens por conteúdo. O resultado da precisão do algoritmo foi de 0,841 e 0,803 na recuperação de nódulos benignos e malignos, respectivamente.*

1. Introdução

O câncer é caracterizado como um crescimento anormal de células que invadem e destroem os seus tecidos vizinhos. Diante da estimativa de mais de 600 mil novos casos estimados para 2016, o câncer compreende a segunda maior causa de morte da população brasileira. Segundo dados do Instituto Nacional do Câncer (INCA) [INCA 2016], são estimados 17.330 novos casos de câncer de pulmão (CP) para o ano de 2016 .

O CP é uma doença agressiva em que geralmente a presença de sintomas é um sinal de prognóstico ruim. A taxa de sobrevivência de um paciente com CP diagnosticado após cinco anos é de apenas 15%. Porém, a identificação de nódulos malignos logo no início da doença garante ao paciente uma taxa de sobrevivência de 49%. Entretanto, o diagnóstico do câncer de pulmão é uma tarefa complexa e apresenta três grandes desafios: interpretar as imagens, identificar as lesões e classificar os nódulos. Ainda, imagens médicas são extremamente complexas por natureza, e o processo de interpretação de imagens é um desafio reconhecido mesmo para radiologistas experientes. Logo, é importante o uso

de ferramentas de auxílio computadorizado ao diagnóstico (CAD) com o objetivo de melhorar tanto a acurácia diagnóstica quanto a consistência na interpretação do diagnóstico por imagens do CP [Kuruvilla and Gunavathi 2014].

A recuperação de imagens baseada em conteúdo (CBIR) tem sido descrita como uma potencial ferramenta CAD, pois é capaz de auxiliar o processo de decisão clínica, recuperando em grandes bases de dados, casos similares já diagnosticados pertencentes à mesma modalidade, região anatômica e com alterações estruturais provocadas por determinadas doenças. Sistemas CAD baseados em CBIR podem ser muito úteis em um ambiente clínico real, pois podem auxiliar o radiologista na tomada de decisão diagnóstica, ou aumentar sua certeza, valendo-se de um modelo de decisão baseada em exemplo. Por exemplo, se um caso desconhecido for considerado muito similar a um caso já diagnosticado como maligno, o radiologista terá mais confiança na decisão diagnóstica do caso desconhecido em maligno [Oliveira and Ferreira 2013]. Um dos desafios no desenvolvimento das aplicações de CBIR é definir métodos de extração de atributos efetivos e eficientes para representar as imagens.

Radiomics é um conceito recente para a extração e análise em larga escala de atributos quantitativos obtidos a partir de imagens médicas usando avançados algoritmos matemáticos. O objetivo é revelar as características das lesões que podem passar despercebidas se analisadas a olho nú [Gillies et al. 2016]. O uso da abordagem *radiomics* permite avaliar e diagnosticar o câncer de forma objetiva, pois atribui às características das imagens valores quantitativos, diminuindo significativamente o problema da variabilidade intra/inter-observador e melhorando a acurácia no diagnóstico quando comparado às estratégias qualitativas de avaliação do câncer [Kumar et al. 2015].

Em [Ferreira Junior et al. 2016a] os autores propuseram uma base para a pesquisa reprodutível de nódulos malignos e benignos composta por 1944 nódulos associados a atributos *radiomics* de Textura e Nitidez de Borda. Entretanto, a base proposta pelos autores não possui atributos capazes de caracterizar a intensidade e a forma, que é uma das principais características utilizadas pelo radiologista durante o diagnóstico de nódulos pulmonares. Logo, o objetivo deste trabalho foi desenvolver os atributos de forma 3D e intensidade 3D, e avaliar a precisão destes algoritmos quando inseridos ao vetor de atributos *radiomics* proposto em [Ferreira Junior et al. 2016a] no contexto de pesquisa reprodutível.

2. Metodologia

Uma visão geral da arquitetura usada neste trabalho pode ser vista na Figura 1. Neste trabalho foi utilizado o banco de dados de nódulos pulmonares público desenvolvido por Ferreira Junior [Ferreira Junior et al. 2016a], este banco é voltado para a pesquisa reprodutível e contém imagens do Lung Image Database Consortium (LIDC) (Seção 2.1).

A segmentação manual dos nódulos foi realizada pelos especialistas do LIDC. Deses nódulos segmentados foram extraídos os atributos *radiomics* de Textura 3D, Nitidez de Borda, Intensidade 3D e Forma 3D (Seção 2.2). Após a extração de atributos, foi realizada a análise de similaridade dos nódulos pulmonares dentro do conceito da CBIR.

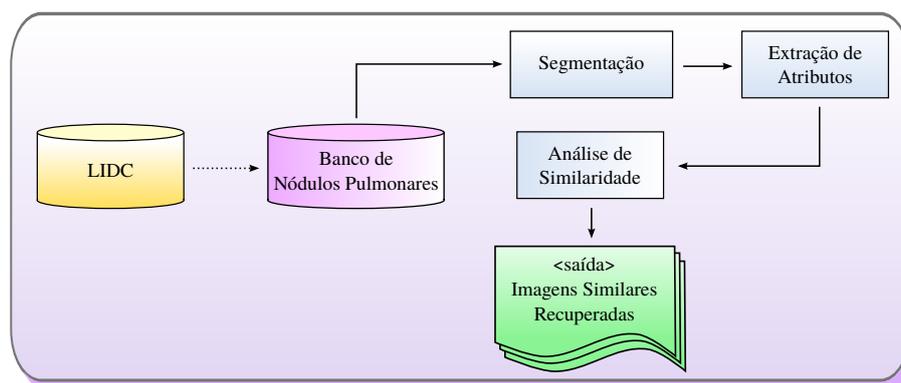


Figura 1. Visão geral da arquitetura CBIR.

2.1. Banco de Nódulos Pulmonares

O LIDC [Armato et al. 2011] é um repositório formado por imagens de Tomografia Computarizada (TC) de câncer de pulmão com lesões identificadas e classificadas por quatro radiologistas em um processo de interpretação de imagem que requereu dos especialistas a análise das imagens e marcação das lesões usando uma interface gráfica. Os nódulos identificados foram classificados pelos radiologistas de acordo com algumas características subjetivas, entre elas a probabilidade de malignidade, de acordo com as seguintes condições:

- Malignidade 1: probabilidade alta para ser benigno;
- Malignidade 2: probabilidade moderada para ser benigno;
- Malignidade 3: probabilidade indeterminada;
- Malignidade 4: probabilidade moderada para ser maligno;
- Malignidade 5: probabilidade alta para ser maligno.

O LIDC é uma coleção não organizada em esquema de banco de dados, assim, não existe correlação entre imagens, dados dos exames e classificação dos nódulos pelos radiologistas. Além disso, o LIDC não contém informações a respeito do tamanho do nódulo ou atributos das imagens. Logo, optamos por usar o Banco de Nódulos Pulmonares (BNP) desenvolvido por Ferreira Junior et al. [Ferreira Junior et al. 2016a]. Este banco utiliza uma abordagem NoSQL orientado a documentos com o MongoDB, onde todas as imagens das lesões foram manualmente segmentadas usando as marcações realizadas pelos radiologistas e então armazenadas no banco de dados. Esta característica permitiu fácil acesso as imagens contendo apenas as lesões na etapa de extração de atributos. O BNP possui 752 exames e 1944 nódulos pulmonares do LIDC nas cinco classificações de malignidade.

Neste trabalho nódulos com probabilidades de malignidade 1 e 2 foram considerados benignos, e nódulos com probabilidades de malignidade 4 e 5, malignos. Porém, nódulos com probabilidade de malignidade 3 foram desconsiderados, pois possuem probabilidade de malignidade indeterminada. Ao término destes recortes, a quantidade de nódulos benignos superou de forma excessiva a quantidade de nódulos benignos. Assim, foi realizada uma distribuição equânime da quantidade de nódulos benignos e malignos com o objetivo de avaliar os resultados da CBIR utilizando uma base balanceada (Tabela 1).

2.2. Extração de Atributos

A combinação de diferentes categorias de atributos é bastante comum em aplicações reais para se alcançar uma melhor acurácia na classificação [Wu and He 2015]. Portanto, neste trabalho foram usadas quatro categorias de atributos de imagem: Análise de Textura 3D (AT) e Análise de Nitidez de Borda (ANB), disponível no BNP; e os atributos desenvolvidos neste trabalho, Atributos de Intensidade 3D e Forma 3D.

No BNB cada nódulo possui um vetor de AT 3D de dimensão 32 obtido a partir da matriz de coocorrência [Ferreira Junior et al. 2016a]. Os atributos desenvolvidos pelo autor foram: Energia, Entropia, Momento da Diferença Inverso, Matiz, Contraste, Proeminência, Correlação e Variância, considerando quatro orientações angulares. Todos os nódulos também estão associados aos atributos do ANB 3D. Para obter estes atributos os autores desenharam linhas perpendiculares sobre todas as bordas das fatias dos nódulos e armazenaram a intensidade dos pixels dessas linhas em um simples array ordenado. Os atributos utilizados pelo autor foram: Diferença entre os extremos, Soma dos valores, Soma dos quadrados, Soma dos logs, Média geométrica, Variância da população, Variância da amostra, Desvio padrão, Medida de kurtosis, Medida de skewness e Segundo momento central. Cada nódulo foi caracterizado com um vetor de atributos ANB com 12 dimensões.

Os atributos de Análise de Intensidade 3D (AI) e Análise de Forma 3D (AF) foram desenvolvidos neste trabalho usando linguagem de programação JAVA versão 1.8. Os atributos de Análise de Intensidade 3D desenvolvidos neste trabalho foram propostos por [Nowik 2013]. Para cada nódulo foi associado um vetor de atributos com dimensão 14 obtido a partir das fatias do nódulo. Os atributos de Análise de Intensidade 3D desenvolvidos estão listados nas equações de 1 a 14.

	Benigno		Maligno		
Probabilidade de Malignidade	1	2	4	5	Total
Número de Nódulos	32	32	32	32	128
Soma	64		64		

Tabela 1. Distribuição balanceada entre nódulos benignos e malignos após desconsiderar a probabilidade de malignidade 3.

Aerts [Aerts et al. 2014] propôs oito atributos de análise de forma 3D que foram adaptados e implementados neste trabalho. Segundo [Brandman 2011], o tamanho do nódulo é um importante indicativo da sua malignância. Logo, consideramos neste trabalho o diâmetro do nódulo como atributo de forma, assim cada nódulo foi associado a um vetor de AF de nove dimensões. Além do diâmetro, os atributos implementados foram Compacidade (1 e 2), Desproporção esférica, Esfericidade, Área, Área da superfície, Relação superfície-volume e Volume, listados nas equações 15 a 23.

$$\text{Energia} = \sum_i^n x_i^2, \quad (1)$$

$$\text{Intensidade média } (\bar{x}) = \frac{1}{n} \sum_{i=1}^n x_i, \quad (2)$$

$$\text{Intensidade mediana}, \quad (3)$$

$$\text{Intensidade mínima } (I_m), \quad (4)$$

$$\text{Intensidade máxima } (I_M), \quad (5)$$

$$\text{Entropia} = - \sum_{k=1}^N p(x_k) \log_2(p(x_k)), \quad (6)$$

$$\text{Kurtosis} = \frac{\sum_{i=1}^n (x_i - \bar{x})^4}{(n-1)s^4}, \quad (7)$$

$$\text{Skewness} = \frac{\sum_{i=1}^n (x_i - \bar{x})^3}{(n-1)s^3}, \quad (8)$$

$$\text{Desvio médio absoluto} = \frac{1}{n} \sum_{i=1}^n |x_i - \bar{x}|, \quad (9)$$

$$\text{Range} = |I_M - I_m|, \quad (10)$$

$$\text{Raiz quadrada média} = \sqrt{\frac{\sum_{i=1}^n x_i^2}{n}}, \quad (11)$$

$$\text{Desvio padrão} = \sqrt{\frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2}, \quad (12)$$

$$\text{Uniformidade} = \sum_{k=1}^N p(x_k)^2, \quad (13)$$

$$\text{Variância} = \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2, \quad (14)$$

onde x_i é o i -ésimo valor de intensidade da imagem; s é o desvio padrão da intensidade; n é o número de pixels na região; e $p(x_k)$ a probabilidade de ocorrência do k -ésimo valor de intensidade em um conjunto de N intensidades.

$$\text{Compacidade 1} = \frac{V}{\sqrt{\pi} A^{\frac{2}{3}}}, \quad (15)$$

$$\text{Compacidade 2} = 36\pi \frac{V^2}{A^3}, \quad (16)$$

$$\text{Desproporção esférica} = \frac{A}{4\pi R^2}, \quad (17)$$

$$\text{Esfericidade} = \frac{\pi^{\frac{1}{3}}(6V)^{\frac{2}{3}}}{A}, \quad (18)$$

$$\text{Área } (A_1) = \sum_{i \in f} p_i, \quad (19)$$

$$\text{Área da Superfície } (A_2) = 4\pi R^2, \quad (20)$$

$$\text{Ralação superfície-volume} = \frac{A}{V}, \quad (21)$$

$$\text{Volume } (V) = \left(\sum_{f \in F} \sum_{i \in f} p_i \right) \cdot \text{Esp}, \quad (22)$$

onde f representa o conjunto de pixels (p) delimitados pelas coordenadas cartesianas da segmentação do nódulo; F é o conjunto de todas as fatias do nódulo; Esp corresponde ao valor da espessura das fatias; para os atributos em função da área A , tanto A_1 quanto A_2 pode ser usado, devido a abordagem do autor; para os atributos em função de um valor de raio (R), o mesmo foi definido como segue:

$$\text{Raio } (R) = \sqrt[3]{\frac{3V}{4\pi}} \quad (23)$$

2.3. Métrica de similaridade e Avaliação da Precisão

Cada atributo extraído possui o seu próprio intervalo de valores e que não são necessariamente coincidentes. Para utilizar métricas de similaridades baseadas em distância, faz-se necessária a normalização da base para que todos os dados se localizem em um intervalo de valores específico [Ferreira de Lucena et al. 2016]. O método de normalização aplicado neste trabalho foi a Normalização Estatística (Transformação Z).

Segundo [Akgül et al. 2010], o método mais comum para avaliar a eficiência de um sistema de recuperação da informação é analisar a Precisão e Revocação. Precisão é a proporção de objetos recuperados que são relevantes para a consulta. Revocação indica a proporção de objetos relevantes na base de dados que foram recuperados ao responder uma consulta. Entretanto, é de grande valiar mensurar a precisão de um sistema de CBIR também na recuperação dos primeiros casos retornados, ou seja, aqueles que estão nas posições iniciais do ranking de similaridade. Dado que os primeiros casos recuperados

(mais similares) são os que terão maior influência no diagnóstico do especialista. Para isso, a métrica utilizada é conhecida como Precisão(N), onde N é o número de casos retornados. [Müller et al. 2004].

Para avaliar a Precisão(N) da recuperação feita com com nódulos representados através dos atributos citados anteriormente, foi utilizada a técnica LOO (Leave-One-Out) sobre a base de nódulos. Esta técnica consiste em utilizar todos os nódulos da base como referência para recuperação selecionando um por vez para obter o nódulo de referência. Ou seja, no caso da base de nódulos selecionada neste trabalho, foram feitas 128 iterações de recuperação e calculada a Precisão(N) da recuperação. A cada iteração, um nódulo diferente da base foi selecionado como referência e os demais foram utilizados para medir a distância para a referência e, com isso, constituir o ranking de similaridade. A similaridade entre os nódulos da base em relação ao nódulo de referência foi obtida através do cálculo da Distância Euclidiana medida entre os seus vetores de atributos radiomics. Assim, quanto menor a distância, ou seja, quanto mais próximo de zero, mais similar são os nódulos.

Os testes foram realizados segundo os seguintes critérios: (1) foram executadas 128 rodadas de recuperação de imagens, considerando os nódulos da base (64 benignos e 64 malignos); (2) a cada rodada foi selecionado um nódulo de referência por meio da técnica LOO; (3) os nódulos recuperados foram ordenados de acordo com a menor Distância Euclidiana em relação ao nódulo de referência da rodada; (4) a precisão da CBIR foi avaliada levando em consideração a ordem dos nódulos resultantes do processo de recuperação, para isso foi calculada a média da Precisão(N) das 64 iterações de recuperação na recuperação dos nódulos benignos e, da mesma forma, a média da Precisão(N) das 64 iterações de recuperação na recuperação dos nódulos malignos considerando, sempre considerando os 10 primeiros resultados (Precisão(10)).

3. Resultados e Discussão

A Figura 2 apresenta os resultados da recuperação P(10) dos nódulos benignos e malignos. Os resultados de precisão foram obtidos a partir da integração dos atributos desenvolvidos neste trabalho com os atributos disponibilizados no BNP (AT e ANB). A média de precisão para o dez primeiros nódulos malignos foi de $0,803 \pm 0,23$ e para os benignos de $0,841 \pm 0,19$. Ao avaliarmos o potencial dos atributos integrados no processo de decisão do especialista, considerando os três primeiros casos, a precisão obtida foi de $0,890 \pm 0,20$ na recuperação de nódulos malignos e $0,911 \pm 0,18$ para nódulos benignos.

A diferença de 6,87% da precisão da CBIR entre os nódulos benignos e malignos sugere que alguns atributos foram capazes de caracterizar de maneira mais eficiente o potencial benigno. Em nossos testes preliminares verificamos que devido a forma espicular do nódulo maligno, a caracterização da forma é um desafio, principalmente em nódulos precoces (<10mm), inclusive dificultando o processo de segmentação manual. Ainda, segundo [Ferreira Junior et al. 2016b], o atributo de proeminência da textura é um forte atributo de caracterização do nódulo benigno.

A tabela 2 apresenta a precisão dos resultados entre P(1) a P(10) dos nódulos malignos obtidos neste trabalho (segunda coluna) e os confronta com os resultados de precisão de outros atributos publicados na literatura. Para garantir uma comparação justa com as outras avaliações, o primeiro nódulo recuperado P(1) de outros trabalhos foi com-

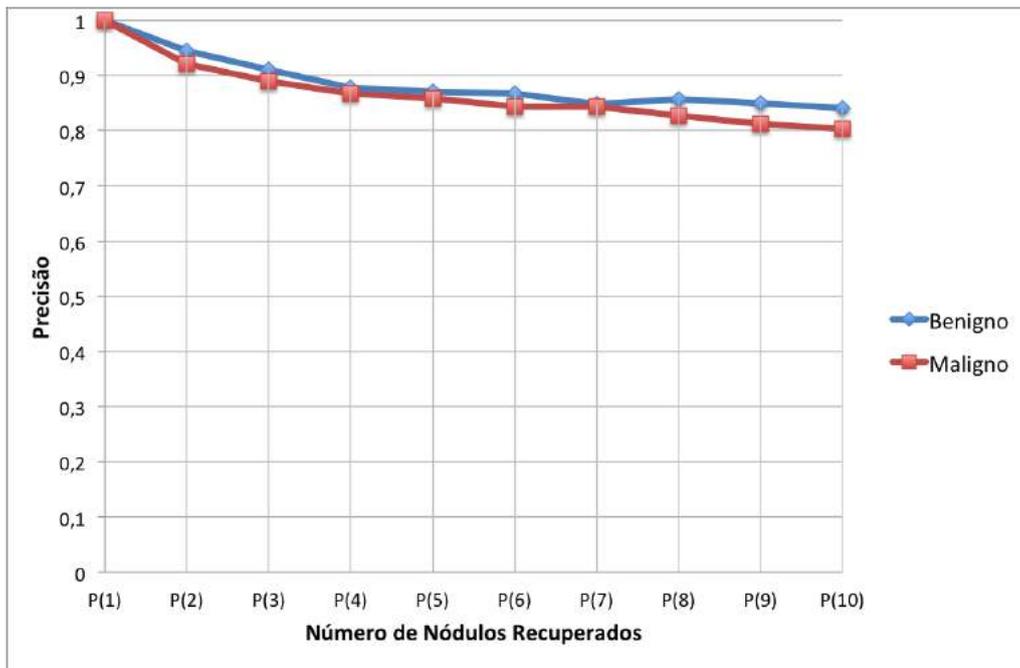


Figura 2. Precisão dos atributos integrados na recuperação dos primeiros dez nódulos similares.

parado com o segundo nódulo recuperado P(2) do nosso trabalho, visto que o primeiro nódulo recuperado neste trabalho é sempre o caso de referência. Ainda, comparamos os resultados da literatura com a nossa menor precisão média P(10), que foi a obtida na recuperação dos nódulos malignos.

Ao comparamos os nossos resultados com [Ferreira Junior et al. 2016b] verificamos que os atributos desenvolvidos neste trabalho (AI e AF) aumentaram a precisão (P10) em 8,0 pontos percentuais quando integrados aos atributos propostos pelo autor. Nossos resultados em P(10) também foram superiores quando comparados ao Atributo de Markov (43%), Filtro de Gabor (38%) e Registro de Imagem (10%). Esta comparação reforça a importância em utilizar os atributos de forma integrada, segundo o conceito *radiomics*.

4. Conclusão

Este trabalho apresentou a avaliação da precisão de um vetor de atributos composto por 66 atributos *radiomics* obtidos a partir da textura, borda, forma e intensidade do nódulo. A precisão foi avaliada na recuperação dos dez primeiros casos, pois é a forma como a CBIR tem maior potencial de auxiliar o especialista no diagnóstico.

Os resultados evidenciaram a importância em integrar os atributos, mesmo sem realizar uma seleção de características obtivemos precisão de 0,841 e 0,803 na recuperação de nódulos benignos e malignos, respectivamente.

Neste trabalho evidenciamos a importância da pesquisa reprodutível, pois ao integrarmos os atributos desenvolvidos por [Ferreira Junior et al. 2016a] (AT e ANB) aos atributos desenvolvidos neste trabalho conseguimos aumentar a precisão da recuperação de nódulos pulmonares malignos em 8%. Também confrontamos os nossos resultados com atributos publicados na literatura para a recuperação de nódulos pulmonares. A pre-

COMPARAÇÃO DE EFICIÊNCIA						
<i>Nódulos Recuperados</i>	Proposto	Atributos AT e ANB	<i>Atributos Haralick</i>	<i>Atributos Markov</i>	<i>Filtros Gabor</i>	<i>Registro de Imagem</i>
1	0,92	0,85	0,29	0,89	0,88	–
2	0,89	0,80	0,23	0,82	0,81	–
3	0,86	0,78	0,21	0,72	0,73	–
4	0,85	0,76	–	–	–	–
5	0,85	0,75	0,18	0,52	0,57	–
6	0,84	0,74	–	–	–	–
7	0,84	0,73	–	–	–	–
8	0,82	0,72	–	–	–	–
9	0,81	0,72	–	–	–	–
10	0,80	0,72	0,15	0,37	0,42	0,7

Tabela 2. Comparação da precisão média P(10) dos atributos integrados proposto neste trabalho com atributos publicados na literatura [Ferreira Junior et al. 2016b].

cisão dos atributos integrados superou os atributos de Markov em 43,9%, Filtro de Gabor em 35,66% e Registro de Imagem em 10%. Contudo, os atributos desenvolvidos (AF e AI) não promoveram um aumento da precisão na recuperação dos nódulos benignos ao compararmos com [Ferreira Junior et al. 2016b]. Ainda, acreditamos que o desvio padrão de $\pm 0,17$ e $\pm 0,19$ na precisão de nódulos benignos e malignos, respectivamente, seja resultado da dificuldade do especialista em segmentar manualmente nódulos menores que 10mm, o que compromete a precisão dos atributos de forma na caracterização do nódulo. Uma outra problemática da base, são os nódulos não sólidos, onde aplicar os atributos de Intensidade para a caracterização dos nódulos é um desafio. Logo, é necessário realizar futuros experimentos para avaliar a eficiência dos atributos na caracterização dos nódulos.

A precisão dos atributos desenvolvidos neste trabalho foi obtida usando um banco de dados de nódulos pulmonares disponível publicamente, o que permitiu a pesquisa reprodutível e a validação cruzada entre atributos. Utilizar uma mesma base de imagens permitiu comparar os resultados de maneira mais fidedigna do que utilizando uma outra base de imagens, garantindo assim consistência na comparação dos resultados. Portanto, encorajamos que outros autores utilizem o BNP para testes e publiquem os seus atributos dentro do conceito de pesquisa reprodutível, avançando o estado-da-arte da aplicação *radiomics* na classificação de nódulos pulmonares.

Referências

- Aerts, H. J. W. L., Velazquez, E. R., Leijenaar, R. T. H., Parmar, C., Grossmann, P., Cavalho, S., Bussink, J., Monshouwer, R., Haibe-Kains, B., Rietveld, D., Hoebbers, F., Rietbergen, M. M., Leemans, C. R., Dekker, A., Quackenbush, J., Gillies, R. J., and Lambin, P. (2014). Decoding tumour phenotype by noninvasive imaging using a quantitative radiomics approach. In *Nature communications*, volume 5, n. 4006, pages 1–8.
- Akgül, C. B., Rubin, D. L., Napel, S., Beaulieu, C. F., Greenspan, H., and Acar, B. (2010).

- Content-based image retrieval in radiology: Current status and future directions. *Journal of Digital Imaging*, 24(2):208–222.
- Armato, S. G., McLennan, G., Bidaut, L., et al. (2011). The lung image database consortium (lidc) and image database resource initiative (idri): A completed reference database of lung nodules on ct scans. *Medical Physics*, 38(2):915–931.
- Brandman, Scott MD; Ko, J. P. M. (2011). Pulmonary nodule detection, characterization, and management with multidetector computed tomography. *Journal of Thoracic Imaging*, 26(2):90–105.
- Ferreira de Lucena, D. J., Ferreira Junior, J. R., Machado, A. P., and Oliveira, M. C. (2016). Automatic weighing attribute to retrieve similar lung cancer nodules. *BMC Medical Informatics and Decision Making*, 16(2):79.
- Ferreira Junior, J. R., Oliveira, M. C., and de Azevedo-Marques, P. M. (2016a). Cloud-based nosql open database of pulmonary nodules for computer-aided lung cancer diagnosis and reproducible research. *Journal of Digital Imaging*, pages 1–14.
- Ferreira Junior, J. R., Oliveira, M. C., and de Azevedo-Marques, P. M. (2016b). Integrating 3d image descriptors of margin sharpness and texture on a gpu-optimized similar pulmonary nodule retrieval engine. *The Journal of Supercomputing*, pages 1–17.
- Gillies, R. J., Kinahan, P. E., and Hricak, H. (2016). Radiomics: Images are more than pictures, they are data. *Radiology*, 278(2):563–577. PMID: 26579733.
- INCA (2016). Instituto nacional de câncer. last accessed 22-09-2016.
- Kumar, D., Shafiee, M. J., Chung, A. G., Khalvati, F., Haider, M. A., and Wong, A. (2015). Discovery radiomics for computed tomography cancer detection. *CoRR*, abs/1509.00117.
- Kuruvilla, J. and Gunavathi, K. (2014). Lung cancer classification using neural networks for {CT} images. *Computer Methods and Programs in Biomedicine*, 113(1):202 – 209.
- Müller, H., Michoux, N., Bandon, D., and Geissbuhler, A. (2004). A review of content-based image retrieval systems in medical applications—clinical benefits and future directions. *International Journal of Medical Informatics*, 73(1):1 – 23.
- Nowik, S. D. K. (2013). *The use of surrounding lung parenchyma for the automated classification of pulmonary nodules*. Ms (master of science) thesis, University of Iowa.
- Oliveira, M. C. and Ferreira, J. R. (2013). A bag-of-tasks approach to speed up the lung nodules retrieval in the bigdata age. In *e-Health Networking, Applications Services (Healthcom), 2013 IEEE 15th International Conference on*, pages 632–636.
- Wu, H. and He, L. (2015). Combining visual and textual features for medical image modality classification with ℓ_p -norm multiple kernel learning. *Neurocomputing*, 147:387 – 394. Advances in Self-Organizing Maps Subtitle of the special issue: Selected Papers from the Workshop on Self-Organizing Maps 2012 (WSOM 2012).