

# Os efeitos da utilização de atributos perinodulares na classificação de nódulos pulmonares

José L. L. Calheiros<sup>1</sup>, Lucas B. V. Amorim<sup>1</sup>, Lucas L. Lima<sup>1</sup>, Marcelo C. Oliveira<sup>1</sup>

<sup>1</sup>Instituto de Computação  
Universidade Federal de Alagoas (UFAL) – Maceió, AL – Brasil

{lucaslc, lucas, lll, oliveiramc}@ic.ufal.br

**Abstract.** *Currently, most CADx systems have been using only descriptors from the pulmonary nodule region. Recent studies indicate that there is a significant interaction between the pulmonary nodule and its surroundings, the parenchyma, however, this region has been little used for the lung cancer diagnosis process. The objective of this work was to investigate the performance of image descriptors extracted from the regions of the nodule, border and parenchyma (intranodular and perinodular attributes), in the identification of their potential for malignancy. In this work, 897 pulmonary nodules were evaluated with 121 image descriptors extracted from the tumor region. The descriptors were selected by genetic algorithm and the performance evaluation was done through the area under the ROC curve (AUC) with 10 folds cross-validation and 5 repetitions. Our best-evaluated model obtained an average AUC of 0.916, accuracy of 84.26%, sensitivity of 84.45% and specificity of 83.84%. The results obtained sustain that the use of perinodular features effectively improves the classification performance of pulmonary nodules.*

**Resumo.** *Atualmente grande parte dos sistemas CADx vêm utilizando apenas descritores oriundos da região do nódulo pulmonar. Trabalhos recentes indicam que há uma interação significativa entre o nódulo pulmonar e seu entorno, o parênquima, entretanto essa região tem sido pouco utilizada para o processo de diagnóstico do câncer pulmonar. O objetivo deste trabalho foi investigar o desempenho de descritores de imagem extraídos das regiões do nódulo, borda e parênquima (atributos intranodulares e perinodulares), na identificação de seu potencial para malignidade. Neste trabalho foram avaliados 897 nódulos pulmonares com 121 descritores de imagem extraídos da região tumoral. Os descritores foram selecionados por algoritmo genético e a avaliação de desempenho foi feita através da área sob a curva ROC (AUC) com validação cruzada de 10 folds e 5 repetições. Nosso melhor modelo avaliado obteve AUC média de 0,916, acurácia de 84,26%, sensibilidade de 84,45% e especificidade de 83,84%. Os resultados obtidos sustentam que a utilização de atributos perinodulares melhoram efetivamente o desempenho de classificação de nódulos pulmonares.*

## 1. Introdução

O câncer pode ser descrito como um conjunto de doenças caracterizadas pelo crescimento desordenado de células que invadem e destroem tecidos e órgãos. Dentre os

tipos de câncer o mais letal e comum é o câncer pulmonar, sendo responsável por 2,09 milhões de mortes em 2018, 18,4% do total de mortes por câncer no mesmo ano [World Health Organisation 2019, Bray et al. 2018]. O principal fator relacionado à sobrevivência do paciente é o momento da detecção do câncer. Se detectado em estágios iniciais a taxa de sobrevivência em um período de 1 ano é de 81-84%, e em estágios mais avançados essa taxa é reduzida para 15-19% [Knight et al. 2017, Bannister and Broggio 2016].

O diagnóstico do câncer pulmonar através da tomografia computadorizada (TC) se dá pela avaliação visual das imagens por um radiologista, um processo complexo e desafiador mesmo para profissionais experientes [Akgül et al. 2010]. O exame de TC é composto por uma série de imagens que representam cortes da caixa torácica do paciente, cada corte é analisado pelo radiologista em busca de irregularidades que possam identificar um nódulo pulmonar. Após a identificação do nódulo são avaliadas características morfológicas, tamanho, localização e intensidade do tumor. Fatores como fadiga, iluminação, e nível de experiência do profissional podem afetar a avaliação do radiologista tornando mais provável a ocorrência de um erro de diagnóstico [Stec et al. 2018].

Com o propósito de melhorar a acurácia e consistência no diagnóstico de imagens médicas as ferramentas de diagnóstico auxiliado por computador (CADx - Computer Aided Diagnosis) fornecem apoio à decisão do radiologista, agindo como uma segunda opinião [Firmino et al. 2016]. Uma possível abordagem para uma ferramenta CADx baseia-se na segmentação e extração de atributos quantitativos de uma região de interesse (ROI - Region of Interest) em cada imagem de um conjunto de exames nos quais é conhecida a natureza do tumor. Esses atributos são então empregados em algoritmos de aprendizagem de máquina para realização de classificação de novas amostras de tumor.

Diversos esforços vêm sendo aplicados com o objetivo de gerar ferramentas CADx mais robustas e precisas, em geral, esses esforços são concentrados na análise e desenvolvimento de atributos que melhor representam a região do nódulo pulmonar [Paul et al. 2018, Zhou et al. 2018, Chen et al. 2018, Coroller et al. 2015, Wei et al. 2018, Dhara et al. 2016, Felix et al. , Ferreira et al. 2017]. Atributos volumétricos, de forma, intensidade, borda e textura do nódulo já foram determinados relevantes para o processo de diagnóstico do câncer pulmonar [Wei et al. 2018, Dhara et al. 2016, Ferreira et al. 2017, Way et al. 2009]. Alguns estudos sugerem que nódulos malignos e benignos interagem de formas diferentes com o seu entorno, o parênquima pulmonar, indicando que esta região pode trazer informações quanto à natureza do tumor [Dilger et al. 2015, Uthoff et al. 2019]. A utilização dessa região apresentou bons resultados na classificação de nódulos pulmonares pequenos, que possuem limitada quantidade de *voxels* disponíveis para sua representação [Huang et al. 2018]. Entretanto, estes estudos foram realizados em condições limitadas, seja por número de nódulos avaliados, ou pela falta de um teste estatístico que comprove a significância desses resultados. Além disso, são escassos os trabalhos a respeito da importância de determinados tipos de atributos dessa região no desempenho de classificação do câncer pulmonar.

O objetivo deste trabalho foi investigar o desempenho de descritores de imagem extraídos das regiões do nódulo, borda, e parênquima pulmonar na identificação de seu potencial para malignidade. Especificamente este trabalho se propõe a verificar o impacto da utilização das regiões de borda e parênquima tumoral na análise e diagnóstico

do câncer pulmonar.

## 2. Materiais e métodos

Este trabalho apresenta um estudo retrospectivo que incluiu 897 nódulos pulmonares pertencentes a exames de TC do repositório público do Lung Image Database Consortium (LIDC). A Figura 1 apresenta o esquema geral da metodologia aplicada.

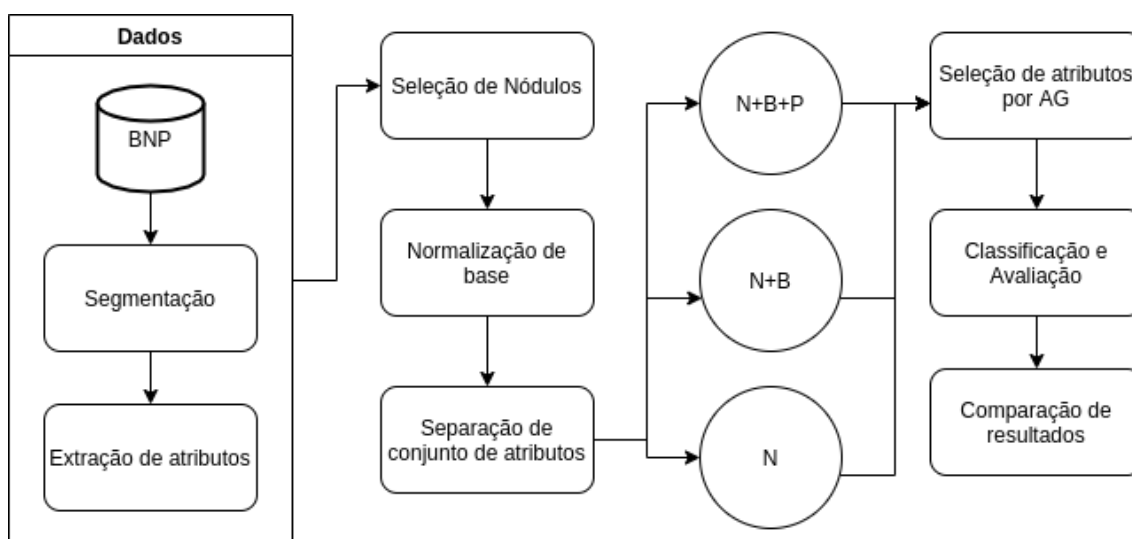


Figura 1. Esquema da metodologia aplicada no trabalho.

Todos os experimentos deste trabalho foram realizados em um PC GNU/Linux Ubuntu 18.04 LTS, Intel Core i5 CPU 3.30GHz, com 8GB de RAM. Os softwares utilizados foram: python (versão 3.7.3), sklearn (versão 0.21.2), imblearn (versão 0.5.1).

### 2.1. Banco de nódulos pulmonares

O LIDC contém 1.018 exames de TC de 1.010 pacientes com 244.559 imagens, assim como marcações e classificações de nódulos pulmonares identificados por quatro radiologistas experientes. As classificações dos nódulos foram realizadas segundo algumas características subjetivas, dentre elas a probabilidade de malignidade em cinco níveis [Armato et al. 2011].

Os dados do LIDC não estão organizados em um esquema de banco de dados e possuem diferentes formatos de dados, dificultando sua utilização. Portanto, optou-se pela utilização do Banco de Nódulos Pulmonares (BNP) [Junior et al. 2016]. O BNP foi desenvolvido no Laboratório de Telemedicina e Informática Médica (LaTIM) com o objetivo de organizar os dados presentes no LIDC. O BNP possui 1.944 nódulos pulmonares com marcações, segmentações dos nódulos e seus respectivos parênquimas, e um conjunto de 121 atributos radiômicos referentes a cada par de segmentação.

De acordo com a classificação de cinco níveis de probabilidade de malignidade, e para os propósitos deste trabalho, os nódulos com níveis 1 e 2 foram considerados benignos, e os nódulos com níveis de malignidade 4 e 5, malignos. Nódulos com nível de malignidade igual a 3 foram descartados devido à incerteza definida pelo especialista, restando 1.171 nódulos pulmonares. Devido à complexidade em segmentar o parênquima

no entorno de nódulos não sólidos, foram selecionados apenas nódulos de natureza sólida, restando após esta etapa 897 nódulos. A Tabela 1 sumariza a distribuição de classes dos nódulos utilizados.

	Benigno		Maligno		
Probabilidade de malignidade	1	2	4	5	Total
Número de nódulos	252	364	154	127	897
Somatório	616		281		

Tabela 1. Distribuição de classes dos nódulos pulmonares.

## 2.2. Conjuntos de atributos

Os 121 atributos presentes no BNP estão divididos em 4 categorias: intensidade, forma, textura e nitidez de borda. A Tabela 2 apresenta a região de extração de cada categoria de atributo.

	Nódulo	Parênquima
Atributos de Intensidade	14	14
Atributos de Forma	9	-
Atributos de Textura	36	36
Atributos de Nitidez de Borda	12	

Tabela 2. Categorias de atributos e regiões de extração.

Os **atributos de intensidade** (AI) seguiram a proposta de [Dilger 2013], foram extraídos de segmentações das regiões do nódulo, e parênquima pulmonar e são compostos por informações a respeito do histograma do valor de intensidade dos *voxels* da área segmentada. Os seguintes atributos foram extraídos: energia, intensidade máxima, mínima, média e mediana, entropia, kurtosis, skewness, desvio médio absoluto, intervalo de intensidade, média da raiz quadrada, desvio padrão, uniformidade e variância.

Os **atributos de forma** (AF) foram extraídos apenas da região do nódulo e representam suas características geométricas, seguindo a proposta de [Aerts et al. 2014]. Os atributos extraídos foram: desproporção esférica, esfericidade, área dos cortes, área da superfície, relação superfície-volume, volume, diâmetro, e duas medidas de compacidade.

Os **atributos de textura** (AT) contém informações a respeito do aspecto e repetição de padrões de imagem e suas variações na área segmentada. Os AT foram extraídos das regiões do nódulo e parênquima seguindo a proposta de [Haralick et al. 1973]. Os AT foram obtidos através da aplicação de uma matriz de co-ocorrência em 4 orientações ( $0^\circ$ ,  $45^\circ$ ,  $90^\circ$  e  $135^\circ$ ) às equações dos seguintes atributos: energia, entropia, momento de diferença inversa, sombra, inércia, variância, proeminência, correlação e homogeneidade.

Os **atributos de nitidez de borda** (ANB) foram baseados em uma análise estatística, seguindo o modelo parcialmente proposto por [Xu et al. 2012]. A análise foi feita com base na geração de linhas ortogonais à borda do nódulo, os valores dos pixels contidos nestas linhas foram guardados em um array de níveis de cinza, este array foi ordenado em ordem crescente, e os atributos estatísticos foram calculados em razão de seu

conteúdo. As seguintes medidas foram calculadas: diferença de extremidades, soma de valores, soma de quadrados, soma dos logaritmos, média aritmética, média geométrica, variância da população, variância de amostra, desvio padrão, kurtosis, skewness e segundo momento central.

Para avaliação das diferentes regiões do ambiente tumoral foram criados três conjuntos de dados. Os conjuntos de dados foram formulados para investigar o impacto da adição dos atributos de determinada região no processo de classificação. O primeiro conjunto identificado como N contém os atributos extraídos apenas da região do nódulo pulmonar. O segundo conjunto, N+B, corresponde aos atributos do conjunto N acrescidos dos atributos de nitidez de borda. E por fim, o conjunto N+B+P é resultante do acréscimo dos atributos oriundos da região do parênquima pulmonar.

### 2.3. Seleção de atributos e classificação

Para selecionar os atributos mais relevantes para o processo de classificação foi utilizado o algoritmo genético. A métrica otimizada pelo algoritmo foi a área sob a curva ROC (AUC - Area Under the ROC Curve). A população foi composta por 50 indivíduos, onde apenas os 4 melhores de cada geração foram selecionados para compor a nova geração através de um cruzamento uniforme com taxa de 60% e probabilidade de mutação de 5%. O critério de parada foi um máximo de 100 gerações ou 10 gerações sem mudança de melhor indivíduo.

Foram utilizados ao todo 6 algoritmos de aprendizagem de máquina para a realização da classificação: *Decision Tree* (DT), *K-Nearest Neighbors* (KNN), *Logistic Regression* (LR), *Random Forest* (RF) *Support Vectors Machine* (SVM), e *XGBoost*. Foram avaliadas duas diferentes configurações do KNN,  $k = 10$  e  $k = 20$ . Três configurações do RF com 100, 500 e 1000 árvores. Três *kernels* no SVM, linear, RBF, polinomial. E cinco configurações no XGBoost com diferentes números de árvores, 15, 18, 20, 25, e 50. Cada modelo foi aplicado aos três conjuntos de dados e individualmente teve seus atributos selecionados pelo algoritmo genético.

### 2.4. Avaliação

A avaliação da classificação dos modelos foi feita através de validação cruzada de 10 *folds* com 5 repetições. Nesta forma de avaliação as métricas foram calculadas em 5 diferentes configurações da validação cruzada de 10 *folds*, e o resultado foi obtido pela média aritmética dessas métricas dentre o total de *folds*. Para cada *fold* foram calculadas 6 métricas: acurácia, *F1-score*, precisão, especificidade, sensibilidade e AUC.

Em modelos baseados em árvore, a importância dos atributos foi calculada utilizando a importância de Gini. Também utilizamos a frequência dos atributos nos melhores avaliados modelos preditivos como um indicativo de sua importância.

Durante a etapa de treinamento dos modelos, as bases da etapa de treinamento foram balanceadas com a utilização do algoritmo SMOTE (Synthetic Minority Over-sampling Technique). Optou-se pela utilização deste algoritmo através de testes preliminares.

Para avaliar a significância estatística da diferença de desempenho obtido pelos conjuntos de atributos testados optamos pela utilização do teste *Wilcoxon Signed-Rank*.

Na comparação de duas populações de performance de conjuntos, um  $p$ -valor  $< 0.05$  indica que essa diferença possui significância estatística.

### 3. Trabalhos relacionados

[Beig et al. 2019] combinou atributos perinodulares e intranodulares com objetivo de gerar um modelo classificação para diferenciar adenocarcinomas e granulomas. Foram utilizados atributos de forma, textura, e *wavelet* em um conjunto com 290 nódulos pulmonares, dos quais 145 foram utilizados como conjunto independente para teste. A combinação de atributos obteve AUC de 0,80, comparativamente, com apenas atributos da região do nódulo o desempenho foi de 0,75, e na utilização de uma rede convolucional para esta tarefa o desempenho foi de 0,76.

[Dilger et al. 2015] utilizou um conjunto de 358 atributos radiômicos (nódulo, parênquima e global) de 50 nódulos pulmonares. O objetivo de seu estudo foi investigar o desempenho na classificação de nódulos pulmonares ao utilizar atributos radiômicos oriundos do parênquima. Seu modelo preditivo com os atributos do nódulo e parênquima obteve uma AUC média de 0,938, superando em 0,20 o modelo composto por atributos apenas do nódulo. Seus achados indicaram que a presença destes atributos influenciam positivamente a performance da classificação, entretanto, devido ao baixo número de amostras disponíveis, a significância estatística de seus resultados não pôde ser verificada.

[Uthoff et al. 2019] realizou um estudo sobre os efeitos da inclusão do parênquima pulmonar no processo de classificação. Foi utilizado um conjunto de 363 nódulos pulmonares (74 malignos, 289 benignos) para treinamento do modelo preditivo e um conjunto com 100 nódulos (50 malignos, 50 benignos) foram utilizados para validação do modelo de forma exclusiva. Foi realizada uma análise comparativa do desempenho do modelo ao adicionar diversas quantidades do parênquima, seus resultados indicaram que o tamanho do parênquima incluído que otimiza o ganho de informação da natureza do tumor é de 75% do diâmetro do nódulo. A AUC média obtida pelo modelo avaliado no conjunto independente foi 0,965, comprovada significância estatística.

### 4. Resultados e discussão

Os modelos de classificação deste trabalho foram avaliados com 5 repetições de validação cruzada com 10 *folds*. Os 10 *folds* foram calculados de forma estratificada, representando a real porcentagem de classes na base de dados. Cada *fold* foi composto de aproximadamente 61 nódulos benignos e 28 malignos. Durante cada etapa de treinamento 9 *folds* foram selecionados resultando num conjunto de aproximadamente 550 nódulos benignos e 252 malignos. O algoritmo SMOTE foi utilizado para o balanceamento desse conjunto de treinamento gerando aproximadamente 298 amostras sintéticas de nódulos malignos.

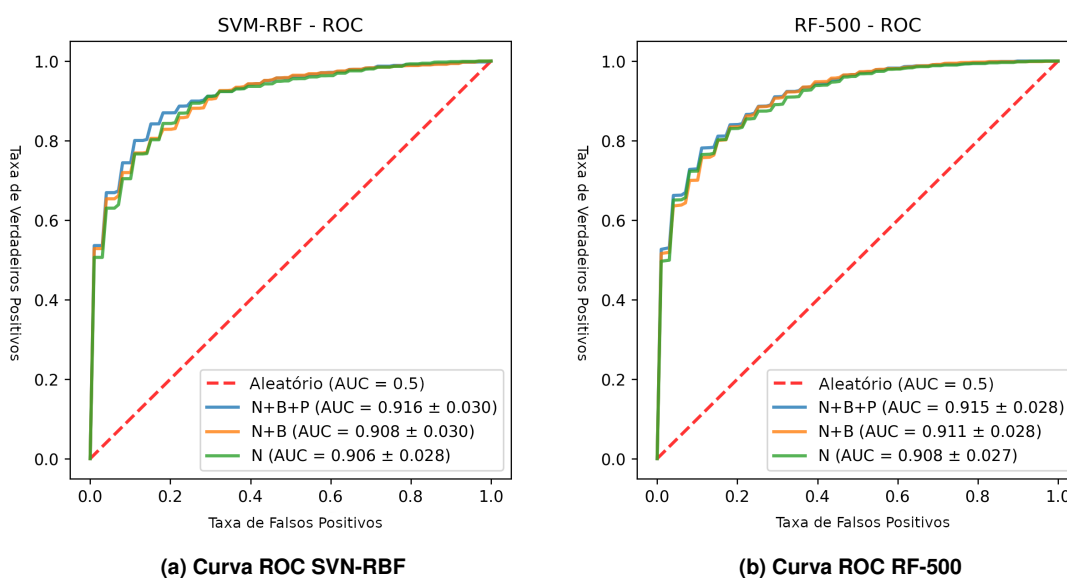
A Tabela 3 apresenta a média dos resultados das classificações obtidos pelas validações cruzadas para cada modelo de aprendizagem de máquina e a média geral destes para cada conjunto de atributos. Os melhores resultados médios absolutos de cada modelo, para todas as métricas, foram obtidos no conjunto N+B+P, a média geral dos modelos deste conjunto também foi superior a dos demais conjuntos em todos os casos. A AUC média geral do conjunto N+B+P foi de  $0,901 \pm 0,032$ , superior em 0,006 em relação ao seu valor no conjunto N+B e 0,008 no conjunto N. O teste de hipótese aplicado

Conjunto	Modelo	Acurácia	F1-score	Precisão	Sensitividade	Especificidade	AUC
N+B+P	DT	80,24 ± 3,20	85,32 ± 2,48	87,13 ± 3,32	83,76 ± 4,21	72,53 ± 8,19	0,781 ± 0,040
	KNN-10	80,95 ± 4,77	84,92 ± 4,01	92,63 ± 3,66	78,56 ± 5,51	86,19 ± 7,04	0,909 ± 0,034
	KNN-20	80,31 ± 3,94	84,15 ± 3,45	<b>93,71 ± 3,09</b>	76,52 ± 5,05	<b>88,60 ± 5,76</b>	0,910 ± 0,033
	LR	83,38 ± 3,62	87,21 ± 2,95	92,35 ± 3,53	82,79 ± 4,53	84,69 ± 7,36	0,912 ± 0,033
	RF-100	83,26 ± 2,97	87,51 ± 2,33	89,71 ± 3,20	85,58 ± 4,03	78,15 ± 7,69	0,913 ± 0,028
	RF-500	83,48 ± 3,25	87,67 ± 2,54	89,89 ± 3,22	85,71 ± 4,17	78,58 ± 7,46	0,915 ± 0,029
	RF-1000	83,17 ± 3,39	87,35 ± 2,75	90,06 ± 3,13	85,00 ± 4,78	79,14 ± 7,23	0,914 ± 0,027
	SVM-Linear	83,28 ± 3,54	87,24 ± 2,83	91,50 ± 3,10	83,47 ± 4,09	82,84 ± 6,66	0,911 ± 0,033
	SVM-Poly	83,86 ± 3,45	<b>88,10 ± 2,66</b>	89,13 ± 3,49	<b>87,34 ± 4,80</b>	76,23 ± 8,81	0,903 ± 0,033
	SVM-RBF	<b>84,26 ± 3,46</b>	88,01 ± 2,80	92,07 ± 3,06	84,45 ± 4,48	83,84 ± 6,60	<b>0,916 ± 0,031</b>
	XGBoost-15	81,58 ± 3,62	85,78 ± 2,97	91,16 ± 3,39	81,17 ± 4,55	82,50 ± 7,15	0,904 ± 0,031
	XGBoost-18	81,65 ± 3,47	85,83 ± 2,92	91,19 ± 3,51	81,30 ± 4,96	82,43 ± 7,72	0,906 ± 0,031
	XGBoost-20	81,54 ± 3,83	85,68 ± 3,21	91,34 ± 3,15	80,87 ± 5,11	82,99 ± 6,63	0,907 ± 0,032
	XGBoost-25	81,94 ± 3,42	86,04 ± 2,86	91,48 ± 3,09	81,39 ± 4,76	83,14 ± 6,73	0,908 ± 0,030
	XGBoost-50	82,47 ± 3,47	86,58 ± 2,82	91,20 ± 3,24	82,56 ± 4,45	82,29 ± 7,09	0,911 ± 0,030
		<b>Média</b>	<b>82,36 ± 3,56</b>	<b>86,49 ± 2,91</b>	<b>90,97 ± 3,28</b>	<b>82,70 ± 4,63</b>	<b>81,61 ± 7,21</b>
N+B	DT	78,99 ± 4,00	84,35 ± 3,02	86,42 ± 3,73	82,52 ± 4,22	71,25 ± 8,99	0,769 ± 0,049
	KNN-10	80,74 ± 3,56	84,64 ± 3,23	93,16 ± 2,80	77,73 ± 5,01	87,32 ± 5,45	0,906 ± 0,030
	KNN-20	80,39 ± 4,16	84,39 ± 3,62	92,78 ± 3,45	77,59 ± 5,37	86,54 ± 6,81	0,907 ± 0,035
	LR	82,16 ± 4,39	86,15 ± 3,71	91,80 ± 3,25	81,35 ± 5,73	83,92 ± 6,75	0,907 ± 0,033
	RF-100	82,79 ± 3,30	87,22 ± 2,58	88,97 ± 3,22	85,72 ± 4,31	76,38 ± 7,82	0,909 ± 0,029
	RF-500	83,06 ± 3,40	87,33 ± 2,74	89,59 ± 3,12	85,39 ± 4,91	77,95 ± 7,50	0,911 ± 0,028
	RF-1000	82,77 ± 3,34	87,10 ± 2,66	89,55 ± 3,42	85,00 ± 4,74	77,87 ± 8,04	0,911 ± 0,029
	SVM-Linear	81,76 ± 3,96	85,81 ± 3,35	91,77 ± 3,21	80,77 ± 5,24	83,92 ± 6,71	0,905 ± 0,032
	SVM-Poly	80,93 ± 3,64	87,22 ± 2,15	81,23 ± 4,17	94,48 ± 3,65	51,24 ± 14,23	0,900 ± 0,032
	SVM-RBF	82,50 ± 3,78	86,71 ± 3,07	90,42 ± 3,34	83,47 ± 4,76	80,35 ± 7,53	0,908 ± 0,031
	XGBoost-15	80,55 ± 3,85	84,94 ± 3,32	90,39 ± 3,25	80,35 ± 5,59	81,00 ± 6,92	0,897 ± 0,031
	XGBoost-18	80,62 ± 3,66	84,99 ± 3,10	90,58 ± 3,57	80,29 ± 5,19	81,35 ± 7,64	0,898 ± 0,030
	XGBoost-20	80,49 ± 3,94	84,87 ± 3,36	90,50 ± 3,47	80,12 ± 5,36	81,28 ± 7,31	0,898 ± 0,031
	XGBoost-25	80,93 ± 3,72	85,28 ± 3,23	90,46 ± 3,42	80,94 ± 5,63	80,94 ± 7,69	0,897 ± 0,031
	XGBoost-50	81,16 ± 3,32	85,59 ± 2,72	90,06 ± 3,49	81,75 ± 4,53	79,87 ± 7,87	0,902 ± 0,029
		<b>Média</b>	<b>81,32 ± 3,73</b>	<b>85,77 ± 3,06</b>	<b>89,85 ± 3,39</b>	<b>82,50 ± 4,95</b>	<b>78,75 ± 7,82</b>
N	DT	79,29 ± 4,02	84,49 ± 3,29	86,78 ± 3,42	82,60 ± 5,81	72,02 ± 8,72	0,773 ± 0,044
	KNN-10	80,99 ± 4,17	84,96 ± 3,68	92,55 ± 3,01	78,72 ± 5,56	85,97 ± 5,87	0,901 ± 0,030
	KNN-20	79,70 ± 4,15	83,79 ± 3,76	92,31 ± 3,39	76,97 ± 5,84	85,69 ± 6,74	0,905 ± 0,033
	LR	81,49 ± 4,08	85,49 ± 3,53	92,19 ± 3,42	79,93 ± 5,66	84,90 ± 6,92	0,904 ± 0,032
	RF-100	83,30 ± 3,26	87,60 ± 2,55	89,34 ± 3,44	86,13 ± 4,54	77,09 ± 8,48	0,905 ± 0,027
	RF-500	82,99 ± 3,47	87,35 ± 2,70	89,23 ± 3,48	85,74 ± 4,52	76,95 ± 8,47	0,908 ± 0,028
	RF-1000	82,83 ± 3,00	87,14 ± 2,45	89,62 ± 3,24	85,03 ± 4,76	78,00 ± 7,96	0,907 ± 0,028
	SVM-Linear	81,15 ± 3,60	85,27 ± 3,14	91,72 ± 3,00	79,86 ± 5,08	83,98 ± 6,23	0,904 ± 0,032
	SVM-Poly	82,27 ± 3,25	87,37 ± 2,32	85,73 ± 3,91	89,35 ± 4,42	66,76 ± 10,86	0,890 ± 0,038
	SVM-RBF	83,18 ± 3,59	87,14 ± 2,95	91,52 ± 3,10	83,34 ± 4,77	82,84 ± 6,71	0,906 ± 0,028
	XGBoost-15	79,53 ± 3,60	83,99 ± 3,17	90,46 ± 3,46	78,67 ± 5,61	81,42 ± 7,52	0,898 ± 0,032
	XGBoost-18	80,62 ± 3,62	85,02 ± 3,05	90,43 ± 3,60	80,48 ± 5,30	80,93 ± 7,87	0,898 ± 0,032
	XGBoost-20	79,88 ± 3,49	84,26 ± 3,10	90,80 ± 3,32	78,86 ± 5,48	82,13 ± 7,22	0,900 ± 0,031
	XGBoost-25	81,14 ± 3,40	85,48 ± 2,96	90,44 ± 3,49	81,33 ± 5,45	80,72 ± 7,94	0,899 ± 0,031
	XGBoost-50	81,58 ± 3,34	85,89 ± 2,77	90,47 ± 3,31	81,94 ± 4,63	80,79 ± 7,38	0,900 ± 0,030
		<b>Média</b>	<b>81,33 ± 3,60</b>	<b>85,68 ± 3,03</b>	<b>90,24 ± 3,37</b>	<b>81,93 ± 5,16</b>	<b>80,01 ± 7,66</b>

**Tabela 3. Média das métricas das validações cruzadas dos modelos de classificação.**

aos resultados gerais, na relação dos conjuntos N+B+P e N+B, e, N+B+P e N para cada métrica comprovam a significância estatística ( $p < 0,05$ ) em todos os casos. Na relação entre os conjuntos N+B e N para as métricas acurácia e *F1-score* o teste de hipótese mostrou-se inconclusivo ( $p > 0,05$ ).

O modelo melhor avaliado no conjunto N+B+P utilizou o algoritmo SVM com *kernel* RBF, e obteve AUC média de  $0,916 \pm 0,031$ , em relação ao desempenho obtido pelo mesmo algoritmo no conjunto N+B o resultado foi superior em um valor de AUC média de  $0,008$ , em relação ao conjunto N a superioridade foi de  $0,010$  (Figura 2a). O número de atributos selecionados para o modelo de SVM-RBF nos conjuntos N+B+P, N+B e N foram, respectivamente, 30 (nódulo: 9 AI, 7 AT, 3 AF; borda: 2 ANB; parênquima: 2 AI, 7 AT), 13 (nódulo: 1 AI, 8 AT, 3 AF; borda: 1 ANB) e 14 (nódulo: 5 AI, 6 AT, 3 AF). O teste de hipótese aplicado aos resultados da métrica AUC obtidos por este modelo mostrou significância estatística na relação dos conjuntos N+B+P e N ( $p = 0,0003 < 0,05$ ), e N+B+P e N+B ( $p = 0,002 < 0,05$ ). A relação do conjunto N+B e N mostrou-se inconclusiva ( $p = 0,34 > 0,05$ ).



**Figura 2. Curvas ROC dos dois melhor avaliados modelos de classificação.**

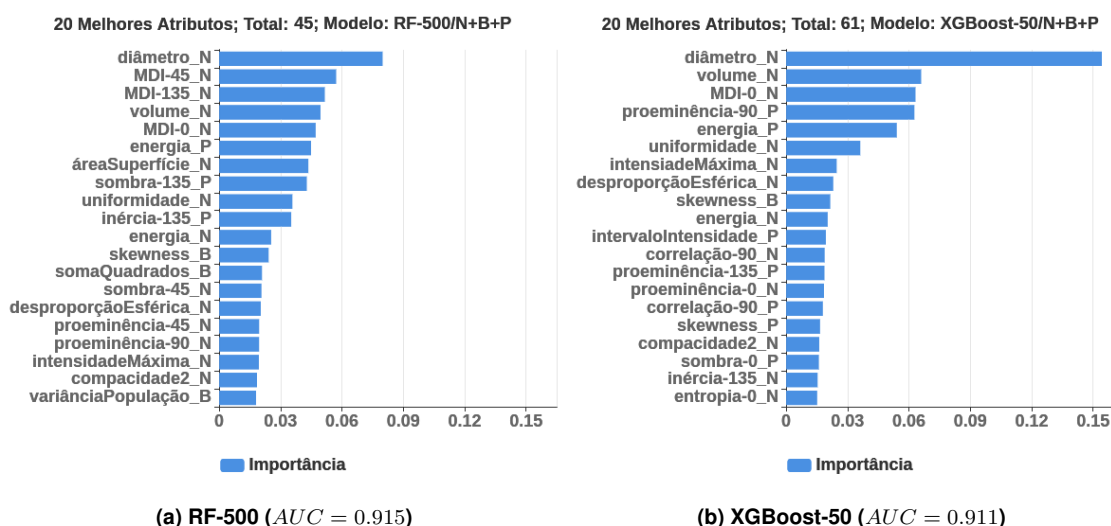
Avaliamos a frequência de ocorrência de um atributo na seleção como um indicativo de sua importância. Para cada algoritmo avaliado apenas a sua melhor configuração foi utilizada no processo de análise da frequência dos atributos, com exceção do algoritmo *Decision Tree* que foi descartado por seu baixo desempenho. Portanto, o resultado apresentado abaixo é referente aos modelos SVM-RBF, RF-500, KNN-20, LR, e XGBoost-50. Foram considerados de maior relevância os atributos que tenham sido utilizados em mais da metade desses cinco modelos, ou seja, aqueles com frequência  $\geq 3$ :

- Frequência 5: Sombra do nódulo em  $45^\circ$ , skewness da borda;
- Frequência 4: Intensidade máxima, média e desproporção esférica do nódulo, skewness e proeminência em  $135^\circ$  do parênquima, e soma de quadrados da borda;
- Frequência 3: Entropia, uniformidade, entropia em  $0^\circ$ , homogeneidade em  $0^\circ$  e  $45^\circ$ , correlação em  $90^\circ$  e  $135^\circ$ , sombra em  $135^\circ$ , diâmetro, compacidade2, proe-



minência em 45° e 90°, e esfericidade do nódulo, energia, entropia, correlação em 90°, energia em 45°, entropia em 135°, sombra em 90° e 135° do parênquima, e segundo momento central da borda.

Para os melhores avaliados modelos baseados em árvore, foi calculada a importância de Gini para a avaliação dos atributos. A Figura 3 mostra os 20 mais importantes atributos selecionados pelos modelos RF-500 e XGBoost-50 no conjunto N+B+P. A utilização dessa métrica pode nos indicar o quanto os atributos oriundos do parênquima influenciaram no processo de classificação. Uma análise na Figura 3a nos mostra que 3 dos 20 mais importantes atributos é originado da região do parênquima (6 ao contabilizar atributos de borda), indicando sua relevância no processo de classificação. Na Figura 3b esse número chega a 7 (8 ao contabilizar os atributos de borda). Ambas figuras indicam a alta importância de atributos de forma como diâmetro, volume, e desproporção esférica, da energia do parênquima e do skewness da borda.



**Figura 3. Importância de Gini dos 20 melhores atributos dos modelos RF-500 e XGBoost-50 no conjunto N+B+P. MDI = Momento de Diferença Inverso. \*\_N = Atributo de Nódulo. \*\_B = Atributo de Borda. \*\_P = Atributo de Parênquima.**

Os resultados apresentados na Tabela 3 evidenciam a superioridade do conjunto N+B+P sobre os conjuntos N+B e N, para todas as métricas avaliadas o desempenho médio do conjunto N+B+P foi superior ao desempenho médio dos demais conjuntos. E considerando a métrica AUC, a performance obtida no conjunto N+B+P, para cada modelo avaliado, foi superior aos seus respectivos pares nos conjuntos N+B e N, o teste de hipótese na relação dos conjuntos N+B+P e N nestes casos apresentou significância estatística ( $p < 0,05$ ) para todos os modelos avaliados, com exceção do modelo DT ( $p = 0,282 > 0,05$ ). Tais resultados corroboram resultados já encontrados nos trabalhos de Uthoff, Dilger e Beig et al. [Uthoff et al. 2019, Dilger et al. 2015, Beig et al. 2019], indicando ganhos de performance na utilização dos conjuntos de atributos citados.

## 5. Conclusão

Neste trabalho investigamos a utilização de atributos extraídos das regiões do nódulo, borda, e parênquima do ambiente tumoral, no desenvolvimento de um modelo para

classificação de nódulos pulmonares. Os resultados obtidos através da aplicação de algoritmos de aprendizagem de máquina nos diferentes conjuntos de atributos evidenciam que a utilização das regiões perinodulares de borda e parênquima possibilitam uma melhora efetiva no desempenho de classificação do câncer pulmonar. A superioridade de desempenho do conjunto N+B+P tanto em termos de média geral quando de valores absolutos nos modelos avaliados nos indica um comportamento estável da melhora da classificação, o teste de hipótese nos atesta a significância estatística destes resultados. Nosso melhor modelo de classificação foi obtido pelo algoritmo SVM com *kernel* RBF no conjunto N+B+P, o modelo obteve AUC média de 0,916, acurácia de 84,26%, sensibilidade de 84,45% e especificidade de 83,84%. Destaca-se também a alta taxa de importância dos atributos de borda e parênquima, que chegaram a ocupar 6 posições na lista dos 20 atributos mais importantes para o melhor avaliado modelo de classificação baseado em árvore, com o algoritmo *Random Forest*. Ressalta-se portanto a importância da utilização da região do entorno do nódulo no processo de classificação do câncer pulmonar.

## Referências

- Aerts, H., Rios Velazquez, E., Leijenaar, R., Parmar, C., Grossmann, P., Cavalho, S., Bussink, J., Monshouwer, R., Haibe-Kains, B., Rietveld, D., Hoebers, F., M Rietbergen, M., Leemans, C., Dekker, A., Quackenbush, J., Gillies, R., and Lambin, P. (2014). Decoding tumour phenotype by noninvasive imaging using a quantitative radiomics approach. *Nature communications*, 5:4006.
- Akgül, C., Rubin, D., Napel, S., Beaulieu, C., Greenspan, H., and Acar, B. (2010). Content-based image retrieval in radiology: Current status and future directions. *Journal of digital imaging : the official journal of the Society for Computer Applications in Radiology*, 24:208–22.
- Armato, 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. (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.
- Bannister, N. and Broggio, J. (2016). Cancer survival by stage at diagnosis for england (experimental statistics): Adults diagnosed 2012, 2013, 2014 and followed up to 2015. *Produced in collaboration with Public Health England*.
- Beig, N., Khorrami, M., Alilou, M., Prasanna, P., Braman, N., Orooji, M., Rakshit, S., Bera, K., Rajiah, P., Ginsberg, J., Donatelli, C., Thawani, R., Yang, M., Jacono, F., Tiwari, P., Velcheti, V., Gilkeson, R., Linden, P., and Madabhushi, A. (2019). Perinodular and Intranodular Radiomic Features on Lung CT Images Distinguish Adenocarcinomas from Granulomas. *Radiology*, 290(3):783–792.
- Bray, F., Ferlay, J., Soerjomataram, I., L. Siegel, R., Torre, L., and Jemal, A. (2018). Global cancer statistics 2018: Globocan estimates of incidence and mortality worldwide for 36 cancers in 185 countries: Global cancer statistics 2018. *CA: A Cancer Journal for Clinicians*, 68.
- Chen, C. H., Chang, C. K., Tu, C. Y., Liao, W. C., Wu, B. R., Chou, K. T., Chiou, Y. R., Yang, S. N., Zhang, G., and Huang, T. C. (2018). Radiomic features analysis in computed tomography images of lung nodule classification. *PLoS ONE*, 13(2).

- Coroller, T. P., Grossmann, P., Hou, Y., Rios Velazquez, E., Leijenaar, R. T., Hermann, G., Lambin, P., Haibe-Kains, B., Mak, R. H., and Aerts, H. J. (2015). CT-based radiomic signature predicts distant metastasis in lung adenocarcinoma. *Radiotherapy and Oncology*, 114(3):345–350.
- Dhara, A. K., Mukhopadhyay, S., Dutta, A., Garg, M., and Khandelwal, N. (2016). A Combination of Shape and Texture Features for Classification of Pulmonary Nodules in Lung CT Images. *Journal of Digital Imaging*, 29(4):466–475.
- Dilger, S., Uthoff, J., Judisch, A., Hammond, E., Mott, S., Smith, B., Newell, Jr, J., Hoffman, E., and (de Ryk) Sieren, J. (2015). Improved pulmonary nodule classification utilizing quantitative lung parenchyma features. *Journal of Medical Imaging*, 2:041004.
- Dilger, S. K. N. (2013). The use of surrounding lung parenchyma for the automated classification of pulmonary nodules.
- Felix, A., Oliveira, M., Machado, A., and Raniery, J. Using 3D Texture and Margin Sharpness Features on Classification of Small Pulmonary Nodules. Technical report.
- Ferreira, J. R., de Azevedo-Marques, P. M., and Oliveira, M. C. (2017). Selecting relevant 3d image features of margin sharpness and texture for lung nodule retrieval. *International Journal of Computer Assisted Radiology and Surgery*, 12(3):509–517.
- Firmino, M., Angelo, G., Morais, H., Dantas, M. R., and Valentim, R. (2016). Computer-aided detection (CADe) and diagnosis (CADx) system for lung cancer with likelihood of malignancy. *BioMedical Engineering OnLine*, 15(1):2.
- Haralick, R. M., Shanmugam, K., and Dinstein, I. (1973). Textural features for image classification. *IEEE Transactions on Systems, Man, and Cybernetics*, SMC-3(6):610–621.
- Huang, P., Park, S., Yan, R., Lee, J., Chu, L. C., Lin, C. T., Hussien, A., Rathmell, J., Thomas, B., Chen, C., HalS, R., Ettinger, D. S., Brock, M., Hu, P., Fishman, E. K., Gabrielson, E., and Lam, S. (2018). Added value of computer-aided CT image features for early lung cancer diagnosis with small pulmonary nodules: A matched case-control study. *Radiology*, 286(1):286–295.
- Junior, J. R. F., Oliveira, M. C., and de Azevedo-Marques, P. M. (2016). Cloud-based nosql open database of pulmonary nodules for computer-aided lung cancer diagnosis and reproducible research. *Journal of digital imaging*, 29(6):716–729.
- Knight, S., Crosbie, P., Balata, H., Chudziak, J., Hussell, T., and Dive, C. (2017). Progress and prospects of early detection in lung cancer. *Open Biology*, 7:170070.
- Paul, R., Hawkins, S. H., Schabath, M. B., Gillies, R. J., Hall, L. O., and Goldgof, D. B. (2018). Predicting malignant nodules by fusing deep features with classical radiomics features. *Journal of Medical Imaging*, 5(01):1.
- Stec, N., Arje, D., Moody, A. R., Krupinski, E. A., and Tyrrell, P. N. (2018). A systematic review of fatigue in radiology: Is it a problem?
- Uthoff, J., Stephens, M. J., Newell, J. D., Hoffman, E. A., Larson, J., Koehn, N., De Stefano, F. A., Lusk, C. M., Wenzlaff, A. S., Watza, D., Neslund-Dudas, C., Carr, L. L., Lynch, D. A., Schwartz, A. G., and Sieren, J. C. (2019). Machine learning

- approach for distinguishing malignant and benign lung nodules utilizing standardized perinodular parenchymal features from CT. *Medical Physics*, page mp.13592.
- Way, T. W., Sahiner, B., Chan, H.-P., Hadjiiski, L., Cascade, P. N., Chughtai, A., Boggot, N., and Kazerooni, E. (2009). Computer-aided diagnosis of pulmonary nodules on ct scans: Improvement of classification performance with nodule surface features. *Medical Physics*, 36(7):3086–3098.
- Wei, G., Cao, H., Ma, H., Qi, S., Qian, W., and Ma, Z. (2018). Content-based image retrieval for Lung Nodule Classification Using Texture Features and Learned Distance Metric. *Journal of Medical Systems*, 42(1).
- World Health Organisation (2019). Cancer. <http://www.who.int/mediacentre/factsheets/fs282/fr/>, Last accessed on 2019-08-22.
- Xu, J., Napel, S., Greenspan, H., Beaulieu, C., Agrawal, N., and Rubin, D. (2012). Quantifying the margin sharpness of lesions on radiological images for content-based image retrieval. *Medical physics*, 39:5405–18.
- Zhou, H., Dong, D., Chen, B., Fang, M., Cheng, Y., Gan, Y., Zhang, R., Zhang, L., Zang, Y., Liu, Z., Zheng, H., Li, W., and Tian, J. (2018). Diagnosis of Distant Metastasis of Lung Cancer: Based on Clinical and Radiomic Features. *Translational Oncology*, 11(1):31–36.