

Pulmonary Nodule Classification with 3D Convolutional Neural Networks

Anthony E. A. Jatobá*, Lucas L. Lima†, Marcelo C. Oliveira*

*Programa de Pós-Graduação em Informática (PPGI)

Instituto de Computação, Universidade Federal de Alagoas

†Programa de Pós-graduação Interunidades em Bioengenharia, Universidade de São Paulo

*Maceió, AL, Brasil, †Ribeirão Preto, SP, Brasil

Email: {aeaj, oliveiramc}@ic.ufal.br*, lucaslima7@usp.br†

Resumo—Lung cancer is a leading cause of death worldwide and its early detection is critical for patient survival. However, the diagnosis is still a challenging task, in which computer-aided diagnosis (CADx) systems try to assist by providing a second opinion to a radiologist. In this work, we propose a 3D Convolutional Neural Network for classification of solid pulmonary nodules into benign and malignant. We evaluated different approaches for the nodule volume assembling and tuned our models in an automated fashion. Our models achieved satisfactory results, with AUC of 0.89, accuracy of 81.37% and a sensibility of 84.83%. Moreover, our results have shown that the first slices of a nodule provide the best results and only five nodule slices are enough for a 3D CNN achieve its best results.

Index Terms—Lung nodule classification, 3D convolutional neural networks

I. INTRODUÇÃO

O câncer de pulmão é a forma mais frequentemente diagnosticada de câncer e uma das principais causas de mortalidade no mundo [1]. A detecção precoce do câncer de pulmão é crucial, possibilitando taxas de sobrevivência de 70-90%, contra até 15% quando o diagnóstico é feito nos últimos estágios da doença [2]. A tomografia computadorizada (TC) é o método preferencial para a detecção precoce do câncer de pulmão, produzindo um volume de cortes em alta resolução e contraste que permitem uma melhor caracterização dos tumores [2], [3].

Mas o diagnóstico ainda apresenta desafios, pois cada imagem gerada pelo exame deve ser examinada cuidadosamente, tarefa sujeita a fatores como fadiga e experiência do radiologista [4], [5]. As ferramentas de auxílio computadorizado ao processo de diagnóstico (do inglês *Computer-Aided Diagnosis* - CADx) surgem para ajudar neste trabalho. Sistemas CADx empregam técnicas de identificação e classificação de imagens para auxiliar o diagnóstico do radiologista, tornando esta tarefa mais rápida e precisa [3].

Sistemas CADx tradicionais funcionam através da extração de descritores da imagem e posterior aplicação de técnicas de Aprendizagem de Máquina (AM) para a classificação [6]. Nos últimos anos houve um crescimento na aplicação de Aprendizagem Profunda (AP), um conjunto de técnicas de AM que se propõe a extrair representações dos dados envolvidos no problema, dispensando a extração prévia de atributos [7]. Uma Rede Neural Convolutiva (RNC) é uma arquitetura

de AP que revolucionou diversos ramos de visão computacional, como classificação de imagens, detecção de objetos e segmentação semântica [8]–[10].

Motivado pelo sucesso das RNC, desenvolveu-se uma série de trabalhos que emprega a técnica para diagnóstico médico, incluindo a classificação de nódulos pulmonares em TC [11]. No entanto, grande parte destes trabalhos utilizam RNC 2D, que ignoram a informação tridimensional do exame de TC. Nos últimos anos, foram propostos modelos que utilizam a terceira dimensão em análises médicas. Porém, os resultados dessas abordagens ainda são preliminares, mas apontam ganhos no desempenho das tarefas de detecção e classificação de nódulos pulmonares em relação aos métodos tradicionais.

O desenvolvimento de redes neurais de AP é complexo e laborioso, por se tratarem de modelos muito sensíveis às suas configurações [12]. Trabalhos recentes reforçam a importância de se investir esforços na etapa de configuração dos modelos como forma de se obter melhores resultados [13], [14]. Estes problemas podem ser mitigados por meio de técnicas automatizadas de Otimização de Hiperparâmetros (OH), que têm apresentado resultados competitivos com a configuração manual feita por especialistas [15].

O objetivo principal deste trabalho consistiu em avaliar uma RNC 3D configurada por meio de OH para a classificação dos nódulos pulmonares entre benignos e malignos. Como objetivo secundário, serão apresentadas diferentes metodologias para a composição do volume nodular a ser utilizado por uma RNC 3D, um problema em aberto na literatura ao qual este artigo traz contribuições ao estado-da-arte.

II. TRABALHOS RELACIONADOS

Os primeiros trabalhos para classificação automática de nódulos pulmonares consistiram na extração de descritores de imagem para posterior uso por classificadores como *Support Vector Machine* (SVM) [16], *Random Forest* (RF) [17] e Redes Neurais Artificiais (RNA) [18]. Com a difusão do uso de AP nos últimos anos, esta família de técnicas foi rapidamente empregada na classificação de nódulos pulmonares, trazendo avanços significativos na área [7], [19].

Hua et al. [3] propõe *Deep Belief Network* (DBN) e CNN para classificação de nódulos pulmonares entre benignos e malignos. Um conjunto de 2.545 nódulos de diâmetro superior

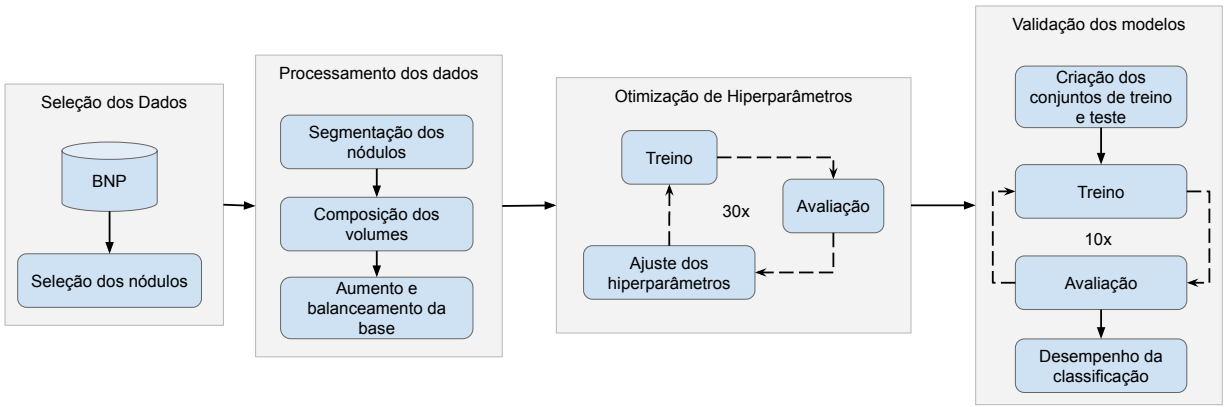


Figura 1. Esquema geral da metodologia empregada.

a 3mm foi selecionada da base LIDC-IDRI. Os resultados foram obtidos por meio de validação *leave-one-out* para a DBN e CNN, além de KNN e SVM servindo como *baseline*. A DBN obteve uma sensibilidade de 73,4% e especificidade de 82,2%, enquanto a CNN obteve 73,3% e 78,7%, apontando a efetividade de AP na classificação de nódulos pulmonares em imagens de TC. Sendo um dos primeiros estudos a aplicar AP neste problema, os modelos utilizados foram rapidamente superados pelos trabalhos seguintes.

Shen et al. [20] propõe CNN multi-escala (MCNN), em que recortes de diferentes dimensões são extraídos de cada nódulo, são dados como entrada para uma CNN 3D e combinados para extrair atributos para classificação usando SVM ou RF. A abordagem foi validada em 1.375 nódulos (880 benignos, 495 malignos) do LIDC-IDRI usando validação cruzada 5-fold. O melhor modelo obteve acurácia de 86,64%. 8 diferentes configurações para a CNN foram avaliadas e os classificadores SVM e RF foram otimizados por meio de *grid-search*. A otimização dos modelos poderia ter sido melhor explorada como forma de atingir melhores resultados.

Kang et al. [4] avaliou MCNNs 3D de diferentes arquiteturas para classificação binária (benignos e malignos) e ternária (benignos, malignos e maligno metastáticos) de nódulos pulmonares. Os modelos foram treinados com 776 nódulos (186 benignos, 590 malignos) da base LIDC-IDRI. Os modelos utilizados foram CNN *multi-view*, uma CNN 3D *feedforward* e uma CNN 3D com arquitetura de grafo acíclico dirigido (do inglês *Direct Acyclic Graph*, DAG). Os melhores modelos obtiveram taxas de erro de 4,59% na classificação binária e 7,70% na classificação ternária. Os resultados foram obtidos por meio de validação cruzada 10-fold. No entanto, os autores usaram aumento de base na base de testes da validação cruzada, prática que não é comum.

Dey et al. [21] propôs quatro CNNs 3D com arquitetura *multi-view* para classificação de nódulos em benignos e malignos. Os modelos foram avaliados em uma base de dados privada de 147 nódulos (37% benignos, 63% malignos). Dado o pequeno número de amostras, as redes foram pré-treinadas com 686 nódulos (46% benignos, 54% malignos) do LIDC-

IDRI. Os melhores modelos obtiveram uma AUC de 0,86 sem a transferência de aprendizagem e 0,90 com o pré-treino. Os resultados foram obtidos por meio de validação cruzada 5-fold. As dimensões da base de dados e sua exclusividade tornam difícil a comparação de resultados com este trabalho.

Lima et al. [22] utilizou CNN 2D para classificação de nódulos pulmonares precoces (com diâmetro entre 5 e 10mm) em benignos e malignos. Os autores avaliaram diferentes estratégias para classificar um nódulo com base na classificação de cada um de seus cortes. As redes foram otimizadas por meio do algoritmo *Tree-structured Parzen Estimator* (TPE), atingindo acurácia de 88% no melhor modelo. O uso de validação cruzada poderia gerar resultados mais generalizáveis e cabem ressalvas à aplicação deste modelo a nódulos não-precoces.

III. MATERIAIS E MÉTODOS

A metodologia empregada neste trabalho consiste na seleção de casos de interesse, processamento dos dados, Otimização de Hiperparâmetros e validação dos modelos obtidos. A figura 1 resume os passos da metodologia empregada.

Todos os experimentos deste trabalho foram realizados em uma GPU Nvidia Titan X com 12 GB de RAM, 3.584 núcleos, velocidade de 1,5 GHz e arquitetura Pascal. Os frameworks utilizados foram: Hyperas [23] (versão 0.4) com o Hyperopt [24] (versão 0.1.1) como backend e Keras [25] (versão 2.2.4) com o Tensorflow [26] (versão 1.11.0).

A. LIDC-IDRI e BNP

O LIDC-IDRI [27] é um repositório público de exames de TC, contendo 1.018 exames de 1.010 pacientes e 244.529 imagens de TC de tórax. As lesões foram identificadas e classificadas por quatro radiologistas experientes, que avaliaram os nódulos acerca de uma série de características subjetivas, dentre elas a probabilidade de malignância em uma escala de 1 a 5, onde:

- **Malignância 1:** alta probabilidade de ser benigno;
- **Malignância 2:** probabilidade moderada de ser benigno;
- **Malignância 3:** malignância indefinida;

- **Malignância 4:** probabilidade moderada de ser maligno;
- **Malignância 5:** alta probabilidade de ser maligno.

Apesar de ser uma das bases mais completas disponíveis, o LIDC-IDRI é composto por uma série de arquivos em formatos distintos e não relacionados, o que dificulta a manipulação das informações contidas na base. Diante deste problema, optamos por utilizar o Banco de Nódulos Pulmonares (BNP) [28]. O BNP foi desenvolvido em MongoDB e disponibiliza exames do LIDC-IDRI de forma organizada e de fácil acesso. Foram selecionados os nódulos do radiologista que identificou o maior número de lesões [28].

Atualmente o BNP contém 752 exames e 1.944 nódulos segmentados corte a corte manualmente pelo especialista, processo ilustrado na figura III-A. As imagens originais de TC do tórax foram armazenadas no padrão de imagens médicas DICOM. De maneira a uniformizar as imagens da base do BNP, foi aplicado o janelamento de escalas de cinza do pulmão fixando a janela em 1600 e o nível em -600 unidades Hounsfield.

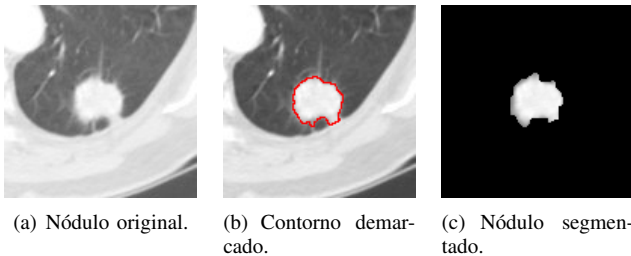


Figura 2. Processo de segmentação de um nódulo.

B. Seleção dos Nódulos

Os nódulos presentes no BNP tiveram suas dimensões medidas por Lima Filho [29]. Foram calculadas as distâncias entre os pontos mínimo e máximo ao longo dos eixos x e y de todos os cortes de cada nódulo. A distância de maior valor foi armazenada no BNP. Um nódulo pulmonar é definido como uma opacidade focal arredondada de diâmetro entre 3mm e 30mm, logo, para este trabalho, foram considerados os nódulos destas dimensões [30]. Além disso, foram selecionados apenas os nódulos sólidos, pois sua demarcação é feita com maior precisão pelos radiologistas. Ao fim desta etapa de seleção, obtivemos um total de 1.006 nódulos, distribuídos como mostrado na Tabela I.

Probabilidade de malignância	Benignos		Malignos		Total
	1	2	4	5	
Número de nódulos	304	394	171	137	1.006
Por categoria	698		308		

Tabela I
DISTRIBUIÇÃO DOS NÓDULOS SÓLIDOS DE 3-30MM NO BNP.

C. Composição dos Volumes

As RNCs precisam que suas entradas sejam fornecidas em uma resolução padronizada, portanto, os volumes a serem

usados por estas redes devem ter suas dimensões de altura, largura e profundidade em um formato predeterminado. Temos um número variável de cortes por nódulo, cuja distribuição é apresentada na figura 3. Para tratar deste problema empregamos estratégias de normalização do número de cortes de cada nódulo de forma a criar volumes de dimensões uniformes.

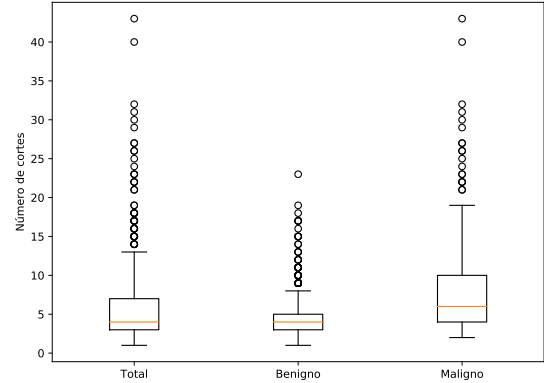


Figura 3. Distribuição do número de cortes na base.

As metodologias empregadas são listadas a seguir e resumidas na figura 4. Sendo n o número de cortes de um nódulo e k o número de cortes que compõem o volume:

- **Primeiros cortes:** consiste em selecionar os k primeiros cortes de um nódulo na ordem em que estão dispostos.
- **Cortes alternados:** são selecionados os cortes 1, corte n e $1 + \frac{n-1}{k-1} \cdot i$, onde $i = \{1, 2, \dots, k-2\}$ e \cdot é a função de arredondamento inteiro. Este cálculo fornece um volume onde os cortes são alternados. Esta estratégia é usada em [4].
- **Corte principal centralizado:** o corte de maior diâmetro é selecionado, junto com os $\lceil \frac{k-1}{2} \rceil$ cortes anteriores e os $\lfloor \frac{k-1}{2} \rfloor$ cortes posteriores.

Os cortes selecionados são organizados num volume de dimensão $64 \times 64 \times k$ voxels. Nas situações em que $n < k$, os k cortes são atingidos adicionando imagens completamente pretas após os cortes selecionados.

Os valores de k avaliados foram 5, 6 e 7 cortes. Escolhemos o número de 6 cortes por se tratar da mediana do número de cortes dos nódulos malignos. Valores menores podem descartar informação de muitos nódulos e valores muito maiores podem gerar um grande número de volumes com cortes na cor preta. O valor de 6 cortes foi usado também por Kang et al [4].

D. Balanceamento e Aumento de Base

Como apresentado na tabela I, os dados se encontram desbalanceados, com 698 nódulos benignos para 308 malignos. Esse desequilíbrio pode tornar os modelos enviesados. O problema foi solucionado efetuando rotações nos volumes usados para o treino das redes, de forma a balancear a base e aumentá-la. Esta estratégia é comum na literatura [4]. Como na etapa de otimização são selecionados 100 nódulos para validação e na de classificação os *folds* de validação possuem entre 60 e 62 nódulos, as rotações foram feitas de forma distinta

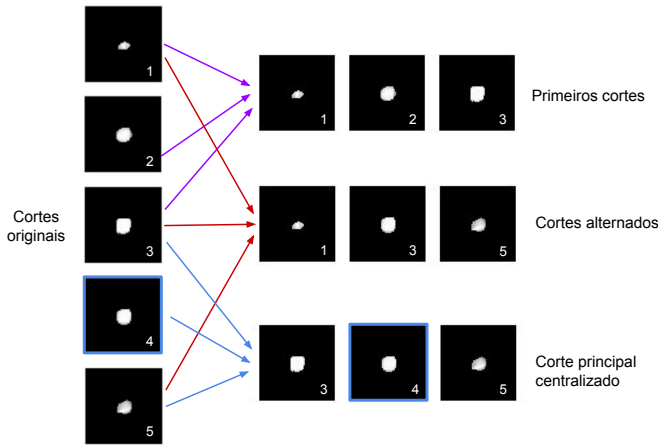


Figura 4. Resumo das estratégias de composição de volumes empregadas.

para manter os conjuntos balanceados. A seguir é explicado como foi efetuado o aumento da base nas etapas OH e de classificação. A tabela II resume as operações.

1) Balanceamento na Otimização de Hiperparâmetros:

Para esta etapa, foram selecionados 50 nódulos malignos e 50 nódulos benignos para constituir o conjunto de validação, enquanto o restante foi usado como conjunto de treino. Na base de treino, aos 648 nódulos benignos foram efetuadas rotações em intervalos de 90°, totalizando 2.592 nódulos benignos; aos 258 malignos foram feitas rotações em intervalos de 36°, totalizando 2.580 nódulos.

2) Balanceamento na Etapa de Classificação: A validação foi feita por meio de validação cruzada 10-fold com nódulos balanceados, gerando conjuntos de teste com 30-31 nódulos de cada classe, restando cerca de 667 nódulos benignos e 277 nódulos malignos para compor o conjunto de treino. Como estas são quantidades distintas das usadas na etapa de OH, as rotações foram efetuadas de maneira diferente para manter os dados balanceados: os 667 nódulos benignos são rotacionados em intervalos de 72°, totalizando 3.335 nódulos; os 277 malignos são rotacionados em intervalos de 30°, totalizando 3.324 nódulos.

Tabela II
BALANCEAMENTO E AUMENTO DE BASE NAS DIFERENTES ETAPAS.

Etapa	Nódulos	Teste	Resto	Rotação	Treino
OH	698	50	648	90°	2592
	308	50	258	36°	2580
Validação	698	~31	~667	72°	3335
	308	~31	~277	30°	3324

E. Otimização de Hiperparâmetros

A configuração dos modelos foi feita por uso do algoritmo TPE [31]. Como temos um grande número de modelos a ser avaliado, optamos por trabalhar com um espaço de busca reduzido. Desta forma podemos executar menos passos da busca e atingir bons resultados mais rapidamente. Os valores avaliados foram tomados com base em experimentos manuais

com topologias de rede e são apresentados na tabela III e tabela IV.

1) *Tree-structured Parzen Estimator*: O TPE é uma técnica para OH proposta por Bergstra et al. [31]. Diferente da busca em grade e aleatória, o TPE usa um modelo probabilístico para, dado o histórico da busca, encontrar a região no espaço de hiperparâmetros que possivelmente contém o conjunto ótimo de valores. Foi demonstrado que a técnica TPE atinge taxas de erros menores que a busca manual e aleatória [31].

As arquiteturas avaliadas são chamadas RNC1 e RNC2:

- **RNC1**: arquitetura composta por uma camada convolucional, uma camada max-pooling e duas camadas densas com dropout.
- **RNC2**: arquitetura composta por duas camadas convolucionais e max-pooling intercaladas e duas camadas densas com dropout.

A tabela III descreve o espaço de busca para a arquitetura RNC1.

Tabela III
ESPAÇO DE BUSCA PARA A ARQUITETURA RNC1.

Camada	Hiperparâmetro	Espaço de busca
Convolutacional	Número de filtros	{32, 48, 64, 96}
	Tamanho do Kernel	(3x3x3)
Max-pooling	Tamanho do Kernel	(2x2x2)
Densa 1	Número de unidades	{64, 96, 128}
	Dropout	[0, 0.5]
Densa 2	Número de unidades	{16, 24, 32}
	Dropout	[0, 0.5]

De maneira similar, para a arquitetura RNC2 temos o espaço de busca apresentado na tabela IV. Neste caso, foi aplicado um *padding* nas camadas convolucionais, para evitar que os *feature maps* tivessem suas dimensões muito reduzidas para as operações das camadas seguintes.

Tabela IV
ESPAÇO DE BUSCA PARA A ARQUITETURA RNC2.

Camada	Hiperparâmetro	Espaço de busca
Convolutacional 1	Número de filtros	{16, 32, 48}
	Tamanho do Kernel	(3x3x3)
Convolutacional 2	Número de filtros	{64, 96, 128}
	Tamanho do Kernel	(3x3x3)
Max-pooling	Tamanho do Kernel	(2x2x2)
Densa 1	Número de unidades	{64, 96, 128}
	Dropout	[0, 0.5]
Densa 2	Número de unidades	{16, 24, 32}
	Dropout	[0, 0.5]

Em ambas as arquiteturas, o algoritmo de otimização de aprendizagem escolhido foi o *RMSprop* e a taxa de aprendizagem foi fixada em 10^{-4} . As funções de ativação escolhidas foram *ReLU* para as unidades das camadas convolucionais e densas e *sigmoide* na camada de saída. Esses valores apresentaram melhores resultados nos estudos iniciais e são amplamente usados na literatura [4]. A função de perda utilizada foi *binary crossentropy*.

A busca de hiperparâmetros foi configurada para executar 30 avaliações de modelos e o critério de aptidão usado foi a

acurácia do modelo na classificação dos exemplos da base de validação.

F. Validação

A validação dos modelos foi feita usando validação cruzada 10-fold. Os nódulos que compõem os conjuntos de treino e teste foram os mesmos para todas as estratégias, permitindo uma comparação justa entre os modelos. As configurações da rede foram as encontradas na etapa de OH.

Para cada modelo foram calculadas a acurácia, especificidade, sensibilidade, F1-score e AUC. Foram geradas as curvas ROC para cada modelo. A curva ROC e a AUC são técnicas bem estabelecidas na literatura de sistemas CADx, sendo portanto a métrica que usaremos na escolha do melhor modelo. Este conjunto de métricas é frequente em trabalhos de classificação na área da saúde e são bastante descritivas do comportamento dos modelos [4], [21].

IV. RESULTADOS E DISCUSSÃO

Os resultados da classificação para os 10 melhores modelos obtidos dentre as combinações de arquitetura (RNC1 e RNC2) e estratégias de composição de volume são sintetizados na tabela V do melhor para o pior em relação à AUC e acurácia, nesta ordem. A coluna *Arq.* indica a arquitetura usada no modelo em questão (RNC1 ou RNC2), a coluna *estratégia* indica como os volumes foram montados e a coluna *cortes*, o número de cortes utilizado nesta composição.

Tabela V
RESULTADO DA CLASSIFICAÇÃO PARA AS DIFERENTES ESTRATÉGIAS.

Arq.	Estratégia	Cortes	Acc.	Sens.	Spec.	F1	AUC
RNC1	Primeiros	5	81.37%	84.83%	77.91%	81.65	0.89
RNC1	Primeiros	6	80.73%	83.77%	77.69%	79.77	0.89
RNC2	Alternados	5	80.55%	78.90%	82.19%	80.46	0.89
RNC2	Principal	5	80.05%	80.19%	79.90%	79.64	0.89
RNC2	Principal	6	79.73%	75.96%	83.49%	80.12	0.89
RNC2	Alternados	6	79.72%	78.56%	80.88%	79.66	0.89
RNC2	Alternados	7	79.72%	78.56%	80.88%	79.50	0.89
RNC2	Principal	7	79.40%	78.25%	80.56%	79.13	0.89
RNC1	Alternados	5	81.20%	81.18%	81.22%	81.13	0.88
RNC1	Principal	6	81.05%	76.97%	85.13%	81.74	0.88

A maioria dos modelos apresentados atingiu AUC de 0.89. Os melhores valores de acurácia e sensibilidade foram, respectivamente, 81,37% e 84,83%, atingidos pelo primeiro modelo da lista. Este modelo apresenta a melhor capacidade de classificar os casos positivos (nódulos malignos). O último modelo da lista apresentou os melhores valores de Especificidade e F1-measure, de 85,13% e 81,74, respectivamente. Este modelo tem a melhor capacidade de distinguir casos negativos (nódulos benignos).

Pela tabela, observamos que as redes de arquitetura RNC2 aparecem em maior quantidade, com 6 modelos, embora os primeiros lugares sejam das redes RNC1. Apenas dois modelos utilizaram 7 cortes na composição dos nódulos, com a maioria utilizando 5 e 6 cortes.

A figura 5 apresenta a curva ROC para o melhor modelo da tabela V de acordo com nossos critérios. Este modelo utiliza arquitetura RNC1 e volumes com os 5 primeiros cortes de um

Tabela VI
COMPARAÇÃO DOS RESULTADOS OBTIDOS COM OUTROS MODELOS.

Modelo	Acc.	Sens.	Spec.	F1	AUC
Random Forest	76.00%	78.00%	74.00%	-	0.85
RNC 2D	80.22%	79.23%	81.20%	79.90	0.89
RNC 3D	81.37%	84.83%	77.91%	81.65	0.89

nódulo. A linha azul é a curva média das iterações da validação cruzada. A área em cinza representa o desvio padrão destas curvas, sendo um indicativo da estabilidade do modelo.

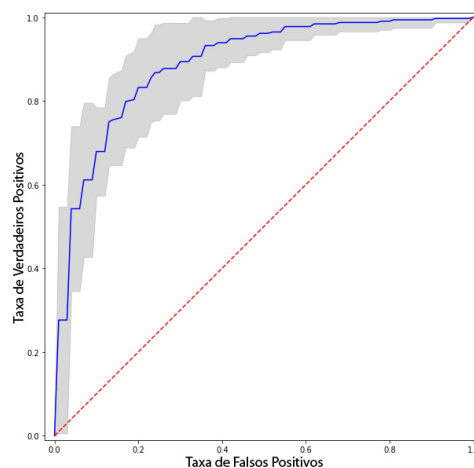


Figura 5. Curva ROC do modelo de melhor desempenho.

A tabela VI apresenta o modelo de melhor desempenho em comparação com outras técnicas. O modelo *Random Forest* consiste na aplicação desta técnica sobre 48 atributos 3D de textura e borda extraídos dos nódulos [6]. O modelo *RNC 2D* denota uma Rede Neural Convolutiva 2D desenvolvida com a mesma metodologia proposta para as RNC 3D, mas utilizando apenas o corte principal de cada nódulo.

Ambas as RNC utilizadas apresentaram um desempenho superior às técnicas tradicionais, com a vantagem de não se fazer necessária a etapa de extração de atributos. A RNC 3D apresentou uma pequena vantagem sobre a RNC 2D.

V. CONCLUSÃO

Este trabalho apresentou um modelo de classificação de nódulos pulmonares utilizando redes neurais convolucionais 3D. O modelo proposto atingiu AUC de 0,89, acurácia de 81,37% e sensibilidade de 84,83%, superiores às técnicas tradicionais e à RNC 2D. Evidenciamos que, dentre as estratégias avaliadas, o melhor desempenho foi obtido pelo uso dos primeiros cortes dos nódulos, embora a vantagem seja tênue. Observamos também que volumes de 5 cortes obtiveram desempenho igual ou superior aos volumes de 6 e 7 cortes, o que indica que 5 cortes são o suficiente para a obtenção de bons resultados, com a vantagem de gerar modelos menos exigentes computacionalmente. Estes resultados empíricos estão atrelados à metodologia empregada neste trabalho, em especial ao uso de imagens de nódulos segmentados. Como

trabalhos futuros, planejamos avaliar os passos empregados com imagens não segmentadas de nódulos, bem como utilizar diferentes abordagens para completar os volumes de nódulos com poucos cortes.

REFERÊNCIAS

- [1] F. Bray, J. Ferlay, I. Soerjomataram, R. L. Siegel, L. A. Torre, and A. Jemal, "Global cancer statistics 2018: Globocan estimates of incidence and mortality worldwide for 36 cancers in 185 countries," *CA: a cancer journal for clinicians*, vol. 68, no. 6, pp. 394–424, 2018.
- [2] S. Blandin Knight, P. A. Crosbie, H. Balata, J. Chudziak, T. Hussell, and C. Dive, "Progress and prospects of early detection in lung cancer," *Open biology*, vol. 7, no. 9, p. 170070, 2017.
- [3] K.-L. Hua, C.-H. Hsu, S. C. Hidayati, W.-H. Cheng, and Y.-J. Chen, "Computer-aided classification of lung nodules on computed tomography images via deep learning technique," *OncoTargets and therapy*, vol. 8, 2015.
- [4] G. Kang, K. Liu, B. Hou, and N. Zhang, "3d multi-view convolutional neural networks for lung nodule classification," *PloS one*, vol. 12, no. 11, p. e0188290, 2017.
- [5] M. J. Chuquicusma, S. Hussein, J. Burt, and U. Bagci, "How to fool radiologists with generative adversarial networks? a visual Turing test for lung cancer diagnosis," in *2018 IEEE 15th International Symposium on Biomedical Imaging (ISBI 2018)*. IEEE, 2018, pp. 240–244.
- [6] J. R. Ferreira, M. C. Oliveira, and P. M. de Azevedo-Marques, "Characterization of pulmonary nodules based on features of margin sharpness and texture," *Journal of digital imaging*, vol. 31, no. 4, pp. 451–463, 2018.
- [7] G. Litjens, T. Kooi, B. E. Bejnordi, A. A. A. Setio, F. Ciompi, M. Ghahouani, J. A. Van Der Laak, B. Van Ginneken, and C. I. Sánchez, "A survey on deep learning in medical image analysis," *Medical image analysis*, vol. 42, pp. 60–88, 2017.
- [8] C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov, D. Erhan, V. Vanhoucke, and A. Rabinovich, "Going deeper with convolutions," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2015, pp. 1–9.
- [9] S. Ren, K. He, R. Girshick, and J. Sun, "Faster r-cnn: Towards real-time object detection with region proposal networks," in *Advances in neural information processing systems*, 2015, pp. 91–99.
- [10] L.-C. Chen, G. Papandreou, I. Kokkinos, K. Murphy, and A. L. Yuille, "Semantic image segmentation with deep convolutional nets and fully connected crfs," *arXiv preprint arXiv:1412.7062*, 2014.
- [11] D. Ravi, C. Wong, F. Deligianni, M. Berthelot, J. Andreu-Perez, B. Lo, and G.-Z. Yang, "Deep learning for health informatics," *IEEE journal of biomedical and health informatics*, vol. 21, no. 1, pp. 4–21, 2017.
- [12] G. Montavon, G. Orr, and K.-R. Müller, *Neural networks: tricks of the trade*. Springer, 2012, vol. 7700.
- [13] A. Coates and A. Y. Ng, "The importance of encoding versus training with sparse coding and vector quantization," in *Proceedings of the 28th international conference on machine learning (ICML-11)*, 2011, pp. 921–928.
- [14] N. Pinto, D. Doukhan, J. J. DiCarlo, and D. D. Cox, "A high-throughput screening approach to discovering good forms of biologically inspired visual representation," *PLoS computational biology*, vol. 5, no. 11, p. e1000579, 2009.
- [15] J. Bergstra and Y. Bengio, "Random search for hyper-parameter optimization," *Journal of Machine Learning Research*, vol. 13, no. Feb, pp. 281–305, 2012.
- [16] A. Tartar, A. Akan, and N. Kilic, "A novel approach to malignant-benign classification of pulmonary nodules by using ensemble learning classifiers," in *2014 36th Annual International Conference of the IEEE Engineering in Medicine and Biology Society*. IEEE, 2014, pp. 4651–4654.
- [17] F. Han, H. Wang, G. Zhang, H. Han, B. Song, L. Li, W. Moore, H. Lu, H. Zhao, and Z. Liang, "Texture feature analysis for computer-aided diagnosis on pulmonary nodules," *Journal of digital imaging*, vol. 28, no. 1, pp. 99–115, 2015.
- [18] A. Felix, M. C. Oliveira, A. Machado, and J. R. Ferreira, "Using 3D Texture and Margin Sharpness Features on Classification of Small Pulmonary Nodules," in *Proceedings...*, Conference on Graphics, Patterns and Images, 29. (SIBGRAPI). IEEE Computer Society's Conference Publishing Services, 2016.
- [19] Y. Yang, X. Feng, W. Chi, Z. Li, W. Duan, H. Liu, W. Liang, W. Wang, P. Chen, J. He *et al.*, "Deep learning aided decision support for pulmonary nodules diagnosing: a review," *Journal of thoracic disease*, vol. 10, no. Suppl 7, p. S867, 2018.
- [20] W. Shen, M. Zhou, F. Yang, C. Yang, and J. Tian, "Multi-scale convolutional neural networks for lung nodule classification," in *International Conference on Information Processing in Medical Imaging*. Springer, 2015, pp. 588–599.
- [21] R. Dey, Z. Lu, and Y. Hong, "Diagnostic classification of lung nodules using 3d neural networks," in *2018 IEEE 15th International Symposium on Biomedical Imaging (ISBI 2018)*. IEEE, 2018, pp. 774–778.
- [22] J. J. F. O. M. C. Lima, Lucas Lins de, "Efficient hyperparameter optimization of convolutional neural networks on classification of early pulmonary nodules," in *32nd IEEE International Symposium on Computer-Based Medical Systems (IEEE CBMS2019)*, 2019.
- [23] M. Pumperla. (2019) Hyperas. [Online]. Available: <http://maxpumperla.com/hyperas/>
- [24] J. Bergstra, D. Yamins, and D. D. Cox, "Hyperopt: A python library for optimizing the hyperparameters of machine learning algorithms," in *Proceedings of the 12th Python in science conference*. Citeseer, 2013, pp. 13–20.
- [25] F. Chollet *et al.*, "Keras," <https://github.com/fchollet/keras>, 2015.
- [26] M. Abadi, P. Barham, J. Chen, Z. Chen, A. Davis, J. Dean, M. Devin, S. Ghemawat, G. Irving, M. Isard *et al.*, "Tensorflow: A system for large-scale machine learning," in *12th {USENIX} Symposium on Operating Systems Design and Implementation ({OSDI} 16)*, 2016, pp. 265–283.
- [27] S. G. Armato, G. McLennan, L. Bidaut, M. F. McNitt-Gray, C. R. Meyer, A. P. Reeves, B. Zhao, D. R. Aberler, C. I. Henschke, E. A. Hoffman *et al.*, "The lung image database consortium (lidc) and image database resource initiative (idri): a completed reference database of lung nodules on ct scans," *Medical physics*, vol. 38, no. 2, pp. 915–931, 2011.
- [28] J. R. F. Junior, M. C. Oliveira, and P. M. de Azevedo-Marques, "Cloud-based nosql open database of pulmonary nodules for computer-aided lung cancer diagnosis and reproducible research," *Journal of digital imaging*, vol. 29, no. 6, pp. 716–729, 2016.
- [29] A. Lima Filho, A. P. Machado, and M. Oliveira, "Modelo para Classificação de Nódulos Pulmonares Pequenos usando Descritores Radiomics," Master's thesis, University of Alagoas - (UFAL), 2016.
- [30] C. I. S. Silva, E. Marchiori, A. S. Souza Júnior, and N. L. Müller, "Illustrated brazilian consensus of terms and fundamental patterns in chest ct scans," *Jornal Brasileiro de Pneumologia*, vol. 36, no. 1, pp. 99–123, 2010.
- [31] J. S. Bergstra, R. Bardenet, Y. Bengio, and B. Kégl, "Algorithms for hyper-parameter optimization," in *Advances in neural information processing systems*, 2011, pp. 2546–2554.