

Classificação de Nódulos Pulmonares Utilizando Redes Neurais Convolucionais 3D

Thiago J. B. Lima¹, Daniel S. Luz^{1,2}, Rodrigo M. S. Veras³, Flávio H. D. Araújo^{1,4}

¹Programa de Pós Graduação em Engenharia Elétrica,
Universidade Federal do Piauí - Teresina, Brasil

²Departamento de Informática,
Instituto Federal do Piauí - Picos, Brasil

³Departamento de Computação,
Universidade Federal do Piauí - Teresina, Brasil

⁴Departamento de Sistemas de Informação,
Universidade Federal do Piauí - Picos, Brasil

{thiagobjb12, rveras, flavio86}@ufpi.edu.br, daniel.luz@ifpi.edu.br

Abstract. Lung cancer is the second most common in Brazil. The early detection of pulmonary nodules is essential for patient survival. In this work we propose an algorithm based on the 3D Convolutional Neural Network to classify pulmonary nodules as benign or malignant in computed tomography images. The proposed architecture has two blocks of convolutional layers followed by a pooling layer, two fully connected layers and a softmax. The use of data augmentation to balance the training set produced promising results, with accuracy of 0.9077, kappa of 0.7774, sensitivity of 0.8481, specificity of 0.9322 and AUC of 0.8901, in the classification of pulmonary nodules.

Resumo. O câncer de pulmão é o segundo mais comum no Brasil. A detecção precoce de nódulos pulmonares é essencial para a sobrevivência do paciente. Neste trabalho propomos um algoritmo baseado na Rede Neural Convolutiva 3D para classificar nódulos pulmonares como benignos ou malignos em imagens da tomografia computadorizada. A arquitetura proposta possui dois blocos de camadas convolucionais seguidos por uma camada de pooling, duas camadas totalmente conectadas e uma softmax. O uso de aumento de dados para balanceamento do conjunto de treino produziu resultados promissores, com acurácia de 0,9077, kappa de 0,7774, sensibilidade de 0,8481, especificidade de 0,9322 e AUC de 0,8901, na classificação de nódulos pulmonares.

1. Introdução

No final do século 20, o câncer de pulmão se tornou uma das causas de mortes mais evitáveis desde que a sua detecção seja realizada precocemente. No Brasil, o câncer pulmonar é o segundo mais comum e o primeiro em todo o mundo [Inca 2019]. O exame de tomografia computadorizada (TC) é, segundo [Yan et al. 2016], o método preferido por especialistas para realização da triagem não invasiva de pacientes com a doença.

A análise das imagens da TC ainda é uma tarefa desafiadora, pois a densidade dos nódulos pode ser semelhante a outras estruturas pulmonares e, além disso, os especialistas possuem um grande número de imagens para analisar em um processo repetitivo. Um nódulo pulmonar é definido como uma opacidade radiológica única, esférica, de até 3 cm de diâmetro, circundada pelo parênquima pulmonar. Lesões acima de 3 cm são frequentemente classificadas como malignas [de Carvalho Filho et al. 2017].

Com intuito de aumentar a precisão na detecção e diagnósticos de lesões para auxiliar o especialista com uma segunda opinião, ferramentas de cunho computacional denominadas *Computer-Aided Detection* (CADe) e *Computer-Aided Diagnosis* (CADx) têm sido amplamente exploradas [Masood et al. 2018]. Na detecção de nódulos pulmonares, a grande preocupação é descobrir se um nódulo representa ou não o câncer. Neste contexto, é importante focar na detecção de anormalidades malignas.

A metodologia utilizada nesse trabalho para a detecção de nódulos malignos emprega as seguintes etapas: aquisição das imagens, segmentação, extração de características/classificação e por fim a avaliação dos resultados. A base de imagens utilizada foi a *Lung Image Database Consortium (LIDC) and Image Database Resource Initiative (IDRI)* (LIDC-IDRI) [Armato III et al. 2011]; para a segmentação de regiões de nódulos foi utilizado um algoritmo baseado em *quality threshold clustering* proposto por [de Carvalho Filho et al. 2014]; para a extração de características dos nódulos e classificação dos mesmos em malignos ou benignos, propomos uma Rede Neural Convolutiva 3D (CNN 3D); por fim, para a avaliação dos resultados utilizamos as métricas acurácia, índice kappa, sensibilidade, especificidade e área sob a curva ROC (AUC).

Vale destacar que um dos principais desafios encontrado nesse trabalho foi a necessidade do balanceamento do conjunto de imagens utilizado para o treinamento da CNN 3D, pois a quantidade de nódulos benignos segmentados era bem maior que a quantidade de nódulos malignos. Com isso, técnicas de aumento de dados foram utilizadas com o objetivo de melhorar a eficiência do algoritmo proposto.

2. Trabalhos Relacionados

Nos últimos anos inúmeras técnicas vêm sendo aplicadas na detecção de anormalidades em imagens de TC na região torácica, com o objetivo de auxiliar o trabalho de especialistas na busca pelo diagnóstico precoce. Nesse cenário, o uso de descritores de textura, descritores de forma e redes neurais convolucionais têm obtido destaque. O último método tem ganho maior notoriedade nos últimos anos, especialmente, pelo fato de apresentar recentemente os melhores resultados. A seguir são apresentados alguns trabalhos relacionados a essas três técnicas.

Os trabalhos de [Safta and Frigui 2018] e [Deep et al. 2016] são exemplos de métodos que utilizam descritores de textura. Em [Safta and Frigui 2018] os autores combinaram as características obtidas por meio dos descritores Matriz de Co-ocorrência de Níveis de Cinza (GLCM) e Padrões Binários Locais (LBP) e os utilizaram na classificação realizada através de uma Máquinas de Vetores de Suporte para Aprendizado de várias Instâncias (MI-SVM). Eles obtiveram resultados de AUC, especificidade, sensibilidade e acurácia de: 0,9696, 0,9855, 0,6979 e 0,9111, respectivamente.

Em [Deep et al. 2016] foi produzido um novo método denominado Padrão Extremo Quantizado Ternário Local Direcional (DLTerQEP), que extrai a relação ternária

entre os vizinhos para um determinado pixel central. O algoritmo proposto foi testado em três bases de dados de imagens distintas: LIDC-IDRI-CT, OASIS-MRI e VIA/I-ELCAP-CT. O método citado obteve melhorias significativas em relação a outros descritores clássicos na literatura, como Padrões Ternários Locais (LTP) e Padrões Extremos Quantizados Locais Nomeados (LQEP).

Se tratando de descritores de forma, temos como exemplos os trabalhos de [Shaffie 2019] e [de Carvalho Filho et al. 2017]. Em [Shaffie 2019], os autores apresentam uma nova estrutura para auxiliar no diagnóstico de nódulos malignos que combina descritores de aparência e forma para indicar a taxa de crescimento do nódulo. Neste trabalho as métricas precisão, sensibilidade e especificidade do diagnóstico do sistema proposto foram de 94,95%, 94,62%, 95,20%, respectivamente.

Em [de Carvalho Filho et al. 2017] foram utilizadas medidas funcionais de *Minkowski*, distâncias, representação do vetor de pontos, medidas de triangulação e diâmetros de *Feret* para diferenciar os padrões de formas malignas e benignas. Eles também utilizaram algoritmo genético para selecionar o melhor modelo e uma máquina de vetores de suporte para classificação. Na fase de teste, aplicaram a metodologia proposta a 1405 (394 malignos e 1011 benignos) nódulos do banco de dados LIDC-IDRI. Os resultados obtidos para a diferenciação de nódulos malignos e benignos alcançaram uma acurácia de 93,19%, sensibilidade de 92,75% e especificidade de 93,33%.

Finalmente, abaixo são comentados alguns trabalhos que empregaram CNNs na tarefa de detecção em imagens de CT de tórax. Em [Yan et al. 2016], os autores propõem uma CNN 3D na coleção de 1018 pacientes da base LIDC-IDRI. Neste trabalho, foram comparadas três estratégias diferentes de arquiteturas de redes neurais convolucionais: CNN 2D em nível de fatia, CNN 2D em nível de nódulo e CNN 3D em nível de nódulo. Os resultados de precisão de classificação de risco de malignidade em nódulos obtidos foram respectivamente de 86,7%, 87,3% e 87,4% para cada uma das três técnicas avaliadas. No trabalho de [da Nobrega et al. 2018] é proposto um método que utiliza diferentes arquiteturas de Redes Neurais Convolucionais (CNN) para a extração de características das imagens da coleção de imagens do LIDC-IDRI. Em seguida, as imagens foram classificadas utilizando os classificadores: *Naive bayes*, Perceptron Multicamada (MLP), Máquinas de Vetores de Suporte (SVM), Algoritmo K-vizinhos mais Próximos (KNN) e Floresta Aleatória (RF). Como resultado foi observado que as melhores combinações de extração profunda e classificação foram obtidos com CNN-ResNet50 e com SVM com *kernel* função de Base Radial, alcançando Acurácia de 88,41% e AUC de 93,19%.

Em [Sharma et al. 2018] foi proposto um algoritmo de duas etapas para a detecção precoce do câncer de pulmão. Na 1ª etapa foi utilizado o método Otsu seguido de operações morfológicas para a segmentação. Na 2ª etapa, foi empregada uma CNN para a classificação de nódulos em malignos ou benignos. Nesse experimento foram utilizados 6.306 imagens de TC do banco de dados do LIDC-IDRI sendo alcançado uma precisão de 84,13%, com a sensibilidade e especificidade de 91,69% e 73,16%.

No trabalho de [Zhao et al. 2018] foi proposta uma CNN híbrida (*LeNet* e *AlexNet*) constituída através da combinação das configurações de camada do *LeNet* e dos parâmetros do *AlexNet*. Eles utilizaram 743 amostras da LIDC para treinar e avaliar o modelo proposto. Os resultados obtidos após a otimização das configurações da CNN

foram acurácia e AUC de 0,822 e 0,877, respectivamente. Ainda nesse trabalho, os autores concluíram que a precisão da CNN depende significativamente do tamanho do *kernel*, taxa de aprendizado, tamanho do lote de treinamento e inicializações de peso.

Em [Dey et al. 2018] são propostas quatro arquiteturas de CNN 3D: *Basic CNN*, *Multi-Output*, *DenseNet* e *MoDenseNet*. Essas quatro redes são avaliadas no conjunto de dados público LIDC-IDRI e superam a maioria dos métodos existentes. Além disso, as redes pré-treinadas no conjunto de dados LIDC-IDRI podem ser estendidas para lidar com conjuntos de dados com poucas amostras usando o transferência de aprendizado. Os resultados obtidos em outro conjunto de dados foram 0,8634 (AUC) e 85,45 (Acurácia) sem transferência de aprendizado e 0,9010 (AUC) e 86,84 (Acurácia) com transferência de aprendizado. Utilizando o conjunto LIDC-IDRI obteve uma acurácia de 0.9040, sensibilidade de 0.9047, especificidade de 0.9033 e AUC de 0.9548.

Como pode ser observado na análise da literatura, as técnicas de CNNs 3D estão sendo amplamente exploradas e obtendo resultados promissores em relação as CNNs 2D. Portanto, além da utilização de uma arquitetura de CNN 3D iremos investigar a aplicação de aumento de dados (utilizando rotação e aplicação de ruído aleatório) para resolver o problema de desbalanceamento, com o objetivo de melhorar a classificação dos nódulos.

3. Materiais e métodos

A metodologia proposta está dividida em quatro etapas: aquisição das imagens, segmentação, extração de características, classificação e avaliação dos resultados. Um resumo dos procedimentos que fazer parte dessa metodologia é apresentado na Figura 1.

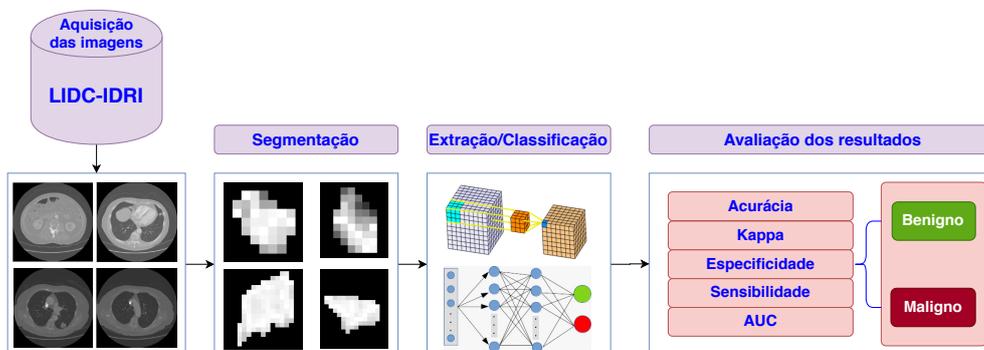


Figura 1. Fluxograma das quatro etapas desenvolvidas neste trabalho: aquisição das imagens, segmentação, extração/classificação e avaliação dos resultados.

3.1. Aquisição das imagens

A coleção de imagens LIDC-IDRI consiste no diagnóstico e rastreamento do câncer pulmonar em TCs com lesões marcadas. O conjunto contém 1018 exames, cada um inclui imagens de uma TC torácica clínica e um arquivo XML associado que registra os resultados de um processo de anotação de imagens de duas fases realizado por quatro radiologistas [Armato III et al. 2011]. A Figura 2 mostra cinco fatias do exame do paciente LIDC-IDRI-0001, onde o número médio de fatias por exame nesse banco de dados é 248.



Figura 2. Fatias 30, 34, 38, 42 e 46 do exame do paciente “LIDC-IDRI-0001”.

3.2. Segmentação

No intuito de encontrar a região de interesse nas imagens, no trabalho de [de Carvalho Filho et al. 2014] os autores realizaram uma segmentação das regiões dos nódulos. Neste procedimento foi utilizado um algoritmo baseado em *quality threshold clustering*. Foram selecionados 800 exames no banco de dados LIDC-IDRI e foi usado o limiar de 90 para toda a metodologia. Após a segmentação das imagens, foram obtidos 1.011 nódulos benignos e 394 malignos. Destacamos que na etapa de extração de características e classificação dos nódulos, foram utilizadas as regiões segmentadas proposta por [de Carvalho Filho et al. 2014] e as mesmas estão disponíveis em¹.

Observamos que o número de exemplos de cada uma das classes é bem desbalanceado, ou seja, temos mais exemplos da classe benigno em relação a classe maligna. Esse efeito pode produzir resultados ruins, principalmente para a detecção de elementos da classe com menor número de exemplos, e que é a mais importante no nosso problema. Portanto, utilizamos técnicas de aumento de dados (utilizando rotação e aplicação de ruído aleatório) durante o treinamento da CNN 3D utilizada na etapa seguinte.

3.3. Extração/Classificação

Para a extração de características e classificação de nódulos, utilizamos uma CNN 3D com dois blocos de camadas convolucionais. Cada bloco contendo duas camadas convolucionais e sucedidas por uma camada de *pooling*. Em seguida, existem duas camadas totalmente conectadas e uma camada de *softmax* que calcula a saída da rede. A Figura 3 apresenta um diagrama completo da arquitetura de CNN 3D utilizada neste trabalho.

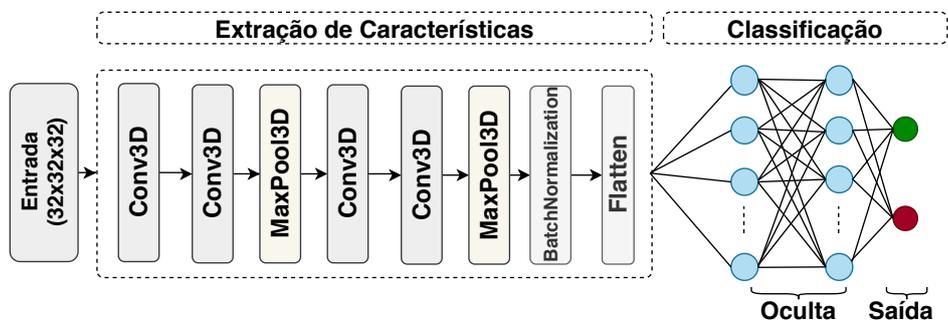


Figura 3. Diagrama completo da arquitetura de CNN 3D utilizada na classificação dos nódulos pulmonares em benigno ou maligno.

¹<https://github.com/pavic-ufpi/Bases/tree/master/Lung/Diagnosis>

A Tabela 1 mostra os principais parâmetros da arquitetura utilizada. O treinamento foi realizado por 2000 épocas, com taxa de aprendizado de 0,0001, *dropout* de 0,4 e tamanho de lote de 128.

Tabela 1. Principais parâmetros da arquitetura utilizada.

Parâmetros	Valor correspondente
Tamanho da camada de entrada	(32, 32, 32)
Número de camadas convolucionais	4
Número de filtros por camadas	(8, 16, 32 e 64)
Função de ativação	relu
Tamanho dos <i>kernels</i> convolucionais	(3, 3, 3)
Tamanho dos <i>kernels</i> de <i>pooling</i>	(2, 2, 2)
Valor do <i>kernel regularizer(L2)</i>	0.0001
Camadas totalmente conectada	2 (256 e 128 neurônios)
<i>Softmax</i>	2 neurônios

3.4. Avaliação dos resultados

Para avaliação dos resultados foram utilizadas as métricas: Acurácia, coeficiente *Kappa*, Sensibilidade, Especificidade e Área sob curva (AUC). Antes de entender cada uma dessas, é necessário a introdução a alguns conceitos elementares utilizados em suas composições, estes são demonstrados na Tabela 2.

Tabela 2. Termos utilizados para cálculo das métricas, sendo benigna a classe positiva e maligna a classe negativa.

Termo	Valor Correspondente
Verdadeiros Positivos (VP)	Classificação correta da classe Positivo.
Verdadeiros Negativos (VN)	Classificação correta da classe Negativo.
Falsos Positivos (FP)	Erro que acontece em razão do algoritmo ter previsto resultado Positivo quando o esperado seria Negativo.
Falsos Negativos (FN)	Erro que acontece em razão do algoritmo ter previsto resultado Negativo quando o esperado seria Positivo.

A Acurácia (Equação 1), considerada por diversos pesquisadores, uma das métricas mais simples para avaliar a qualidade de uma classificação, foi utilizada nesse trabalho para medir o número de predições corretas sem levar em conta positivos e negativos [Mishra 2018].

$$Acurácia (AC) = \frac{VP + VN}{VP + VN + FP + FN} \quad (1)$$

A Sensibilidade (Equação 2), ou taxa de verdadeiros positivos, é a métrica computada pela relação entre verdadeiros positivos e todos os casos positivos[Mishra 2018].

$$Sensibilidade (SE) = \frac{VP}{VP + FN} \quad (2)$$

Já a Especificidade (Equação 3), ou taxa de falsos positivos, corresponde a proporção de de falsos positivos com relação a todos os outros dados negativos [Mishra 2018].

$$Especificidade (ES) = \frac{VN}{VN + FP} \quad (3)$$

O índice *Kappa* (Equação 4), é uma métrica muito útil quando se trabalha com problemas de várias classes ou quando as classes do problema são desbalanceadas. Apesar de não haver um padrão de interpretação de seus dados, [Landis and Koch 1977] apresentam uma maneira para realizar o entendimento, que possibilita verificar o grau de concordância entre avaliadores, onde: um valor menor que 0 indica não concordância, entre 0 e 0,20 concordância leve, entre 0,21 e 0,40 como concordância razoável, entre 0,41 e 0,60 como concordância moderada, entre 0,61 e 0,80 como concordância substancial e entre 0,81 e 1 como concordância quase perfeita.

$$Kappa (K) = \frac{P_o - P_e}{1 - P_e} \quad (4)$$

onde,

$$P_o = \frac{VP + VN}{VP + VN + FP + FN} \quad (5)$$

e

$$P_e = \frac{[(VP + FN)(VP + FP)] + [(VN + FN)(VN + FP)]}{(VP + VN + FP + FN)^2} \quad (6)$$

Por fim, Área sob curva ROC é utilizada quando a classificação é uma problema de solução binário, como é o caso deste trabalho. Na prática ela corresponde a probabilidade de um exemplo positivo escolhido aleatoriamente ser mais alta que a de um exemplo negativo ser escolhido também aleatoriamente, e demonstra o contraste na predição [Avelar 2019, Mishra 2018].

4. Resultados e discussão

Os nódulos segmentados foram divididos em três conjuntos: treino, validação e teste, contendo 70%, 10% e 20% respectivamente de cada classe (benigno e maligno). Realizamos uma normalização dos dados e as intensidades das imagem formam escaladas entre 0 e 1. Os nódulos que possuíam apenas uma fatia foram descartados, ao final obtivemos 956 benignos e 389 malignos. Destes, 939 foram utilizados para treino (668 benigno e 271 maligno), 135 para validação (96 benigno e 39 maligno) e 271 para teste (192 benigno e 79 maligno).

Como objetivo foi avaliar o efeito do aumento de dados usando os diferentes cenários da Tabela 3, utilizamos técnicas essas técnica para formar quatro conjuntos de treino diferentes: O primeiro continha todos os nódulos segmentados (sem aumento de dados) que foram selecionados para o conjunto de treino, 668 benignos e 271 malignos; No segundo replicamos os nódulos malignos (gerando uma cópia), obtendo um total 542 nódulos malignos e permanecendo com os 668 nódulos benigno; Para o terceiro conjunto, replicamos os nódulos malignos usando rotação (45, 90, 135 e 180 graus) e aplicando ruídos aleatórios, em seguida selecionamos aleatoriamente 668 nódulos malignos

para que esse conjunto ficasse com a mesma quantidade de exemplos das duas classes. Por fim, para a criação do quarto conjunto, dobramos a quantidade de exemplos benigno utilizando rotação (45, 90, 135 e 180 graus) e aplicando ruídos aleatórios. Em seguida, aplicamos aumento de dados para a classe maligna, usando as mesmas operações, gerando 1336 desta classe também. Vale ressaltar que o aumento de dados foi utilizado apenas no conjunto de treino. Além das técnicas abordadas, outras possibilidades podem ser utilizadas como: variações no espaço de cores, transformações geométricas e outras [Shorten and Khoshgoftaar 2019].

A Tabela 3 mostra os resultados obtidos com a aplicação do método proposto e na Figura 4 são apresentados os resultados da matriz de confusão. As métricas utilizadas: Acurácia (AC), Kappa (K), Sensibilidade (SE), Especificidade (ES) e AUC, foram apresentadas na seção 3.4. Os resultados apresentados são referentes ao conjunto de teste, contendo 192 nódulos benigno e 79 maligno.

Tabela 3. Resultados obtidos após a aplicação do método proposto.

Experimento	Imagens treino	AC	K	SE	ES	AUC
T1 - Imagens segmentadas	939	0,9003	0,7465	0,7468	0,9635	0,8551
T2 - Imagens replicadas (maligno)	1210	0,9003	0,7578	0,8227	0,9322	0,8775
T3 - Rotação e ruído (maligno)	1336	0,9077	0,7774	0,8481	0,9322	0,8901
T4 - Rotação e ruído (ambos)	2672	0,8966	0,7441	0,7848	0,9427	0,8637

Experimento T1		Predito	
		Maligno	Benigno
Real	Maligno	59 VP	7 FP
	Benigno	20 FN	185 VN

Experimento T2		Predito	
		Maligno	Benigno
Real	Maligno	65 VP	13 FP
	Benigno	14 FN	179 VN

Experimento T3		Predito	
		Maligno	Benigno
Real	Maligno	67 VP	13 FP
	Benigno	12 FN	179 VN

Experimento T4		Predito	
		Maligno	Benigno
Real	Maligno	62 VP	11 FP
	Benigno	17 FN	181 VN

Figura 4. Matriz de confusão referente aos experimentos **T1**, **T2**, **T3** e **T4**, onde apresentamos os valores reais e valores preditos para cada classe.

Analisando o experimento **T1** podemos observar que o valor de sensibilidade foi menor que nos demais experimentos, em contrapartida o valor da especificidade foi maior. Esses resultados indicam que mais exemplos da classe benigna estão sendo corretamente classificados, no entanto, mais falsos negativos estão sendo obtidos. Isso ocorre porque o conjunto de treino **T1** possui bem mais exemplos da classe benigna, e isso influencia no treinamento da rede, que tende a identificar menos nódulos malignos.

Uma das principais medidas a ser analisada para dados desbalanceados é o índice *Kappa*, pois ele tende a produzir melhores resultados quando a predição das duas classes

são balanceadas. No experimento **T2** podemos observar que o valor de *Kappa* aumentou em relação ao **T1** e **T4**. Isso aconteceu devido ao aumento do número de exemplos da classe maligna no treino, que diminuiu o número de falsos negativos produzidos pela rede.

No experimento **T3**, realizamos operações de rotação e adição de ruídos nos exemplos da classe maligna. Com isso, houve uma melhora na capacidade de generalização da rede devido ao aumento da diversidade dos nódulos presentes no conjunto de treino. Dessa forma, esse experimento produziu o menor número de falsos negativos, o que é refletido no índice *Kappa* e Sensibilidade, que obtiveram os maiores valores dentre os testes realizados. Além disso, obtivemos também com o experimento **T3** a melhor acurácia e AUC, e perdendo apenas no quesito de Especificidade para os experimentos **T1** e **T4**.

Já no experimento **T4**, mesmo aumentando o número de dados de ambas as classes o resultado não melhorou. Como os exemplos da classe maligna foram aumentados em quase 5 vezes (de 271 para 1336), muitas imagens desse conjunto ficaram com características similares, isso influenciou no resultado obtido nesse experimento, pois a rede não obteve uma maior capacidade de generalização. Comparando o experimento **T4** com o experimento **T1**, obtivemos um melhor resultado nos quesitos de Sensibilidade e AUC, no entanto as métricas Acurácia, *Kappa* e Especificidade, foram inferiores aos demais experimentos.

Diante disso, o melhor resultado foi produzido pelo conjunto **T3**, no qual o número de nódulos malignos foram aumentados por meio de operações de rotação e adição de ruído aleatório até que o conjunto de treino ficasse com a mesma quantidade de exemplos de ambas as classes. A Tabela 4 fornece uma comparação do melhor resultado obtido pela metodologia proposta com trabalhos presentes na literatura e que realizam a classificação de nódulos pulmonares em benignos e malignos. É importante mencionar que os resultados apresentados em cada trabalho foram obtidos com diferentes critérios de avaliação e para número de imagens distintas da LIDC-IDRI.

Tabela 4. Comparação dos nossos resultados com outros trabalhos presentes na literatura e que realizam a classificação de nódulos pulmonares em benignos e malignos.

Trabalho	AC	SE	ES	AUC
[Safta and Frigui 2018]	0,9111	0,6979	0,9855	0,9696
[Shaffie 2019]	0,9495	0,9462	0,9520	-
[de Carvalho Filho et al. 2017]	0,9319	0,9275	0,9333	-
[Yan et al. 2016]	0,874	0,894	0,852	0,947
[da Nobrega et al. 2018]	0,8841	-	-	0,9319
[Sharma et al. 2018]	0,8413	0,9169	0,7316	-
[Zhao et al. 2018]	0,822	-	-	0,877
[Dey et al. 2018]	0,9040	0,9047	0,9033	0,9548
Nosso trabalho	0,9077	0,8481	0,9322	0,8901

5. Conclusão

Neste trabalho, propomos um algoritmo automatizado, baseado em uma CNN 3D, para classificar nódulos pulmonares em benignos ou malignos. Este algoritmo pode ser utili-

zado para a construção de sistemas CAD que auxiliem especialistas na interpretação do exame, fornecendo uma segunda opinião durante a análise do mesmo.

A metodologia proposta apresenta resultados promissores, principalmente com o uso de técnicas de aumento de dados. Observamos que ao balancear as classes no conjunto de treino, houve um aumento significativo na predição de exemplos da classe maligna, que contém a menor quantidade de exemplos. Os melhores resultados foram alcançados com o balanceamento do conjunto de treino por meio do aumento do número de nódulos malignos com operações de rotação e adição de ruído aleatório.

Neste trabalho, realizamos operações de rotação e aplicação de ruído em nível de fatia (2D). Em trabalhos futuros, pretende-se aplicar técnicas de aumento de dados em nível de nódulo, ou seja, aplicadas ao nódulo 3D. Para isso, investigaremos a utilização de técnicas de *Autoencoders* e *Generative Adversarial Network* (GANs), afim de gerar novos nódulos a partir dos exemplos disponíveis no conjunto de treino.

Agradecimentos

O presente trabalho foi realizado com apoio da Fundação de Amparo à Pesquisa do Estado do Piauí – FAPEPI e da Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES) - Código de Financiamento 001.

Referências

- Armato III, S. 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*, 38:915–931.
- Avelar, A. (2019). O que é auc e roc nos modelos de machine learning. Disponível em: <https://medium.com/@eam.avelar/o-que-%C3%A9-auc-e-roc-nos-modelos-de-machine-learning-2e2c4112033d>. Accessed: 2020-02-15.
- da Nobrega, R. V. M. et al. (2018). Lung nodule classification via deep transfer learning in ct lung images. In *2018 IEEE 31st International Symposium on Computer-Based Medical Systems (CBMS)*, pages 244–249.
- de Carvalho Filho, A. O., de Sampaio, W. B., Silva, A. C., de Paiva, A. C., Nunes, R. A., and Gattass, M. (2014). Automatic detection of solitary lung nodules using quality threshold clustering, genetic algorithm and diversity index. *Artificial intelligence in medicine*, 60(3):165–177.
- de Carvalho Filho, A. O., , et al. (2017). Computer-aided diagnosis system for lung nodules based on computed tomography using shape analysis, a genetic algorithm, and svm. *Medical & Biological Engineering & Computing*, 55(8):1129–1146.
- Deep, G., Kaur, L., and Gupta, S. (2016). Directional local ternary quantized extrema pattern: A new descriptor for biomedical image indexing and retrieval. *Engineering Science and Technology, an International Journal*, 19(4):1895 – 1909.
- Dey, R., Lu, Z., and Hong, Y. (2018). Diagnostic classification of lung nodules using 3d neural networks. In *2018 IEEE 15th International Symposium on Biomedical Imaging (ISBI 2018)*, pages 774–778.

- Inca (2019). Instituto Nacional do Câncer - ministério da saúde, câncer de pulmão. disponível em: <https://www.inca.gov.br/tipos-de-cancer/cancer-de-pulmao>. Acessado em: 02/08/2019.
- Landis, J. R. and Koch, G. G. (1977). The measurement of observer agreement for categorical data. *biometrics*, pages 159–174.
- Masood, A. et al. (2018). Computer-assisted decision support system in pulmonary cancer detection and stage classification on ct images. *Journal of Biomedical Informatics*, 79:117 – 128.
- Mishra, A. (2018). Metrics to evaluate your machine learning algorithm. Disponível em: <https://towardsdatascience.com/metrics-to-evaluate-your-machine-learning-algorithm-f10ba6e38234>. Acessado em: 15/02/2020.
- Safta, W. and Frigui, H. (2018). Multiple instance learning for benign vs. malignant classification of lung nodules in ct scans. In *2018 IEEE International Symposium on Signal Processing and Information Technology (ISSPIT)*, pages 490–494. IEEE.
- Shaffie, Ahmed e Soliman, A. e. K. H. A. e. T. F. e. G. M. e. D. N. e. E. A. e. K. R. e. E.-B. A. (2019). A novel ct-based descriptors for precise diagnosis of pulmonary nodules. In *2019 Conferência Internacional da IEEE sobre Processamento de Imagens (ICIP)*, pages 1400–1404.
- Sharma, M., Bhatt, J. S., and Joshi, M. V. (2018). Early detection of lung cancer from ct images: nodule segmentation and classification using deep learning. In *Tenth International Conference on Machine Vision (ICMV 2017)*, volume 10696, page 106960W. International Society for Optics and Photonics.
- Shorten, C. and Khoshgoftaar, T. M. (2019). A survey on image data augmentation for deep learning. *Journal of Big Data*, 6(1):60. <https://doi.org/10.1186/s40537-019-0197-0>.
- Yan, X., Pang, J., Qi, H., Zhu, Y., Bai, C., Geng, X., Liu, M., Terzopoulos, D., and Ding, X. (2016). Classification of lung nodule malignancy risk on computed tomography images using convolutional neural network: A comparison between 2d and 3d strategies. In *Asian Conference on Computer Vision*, pages 91–101. Springer.
- Zhao, X., Liu, L., Qi, S., Teng, Y., Li, J., and Qian, W. (2018). Agile convolutional neural network for pulmonary nodule classification using ct images. *International journal of computer assisted radiology and surgery*, 13(4):585–595.