

Redes Neurais Convolucionais Aplicadas no Diagnóstico do Câncer de Pulmão

Willian Ramon Barbosa Bessa¹, Rhaylson Silva do Nascimento¹,
Sebastião Emídio Alves Filho¹, Marcelino Pereira Dos Santos Silva¹,
Francisco Milton Mendes Neto¹

¹Programa de Pós-Graduação em Ciência da Computação
(UFERSA-UERN) Mossoró - RN - Brazil

{willian.bessa, rhaylson.nascimento}@alunos.ufersa.edu.br,

{sebastiaoalves, marcelinopereira}@uern.br, miltonmendes@ufersa.edu.br

Abstract. *This paper presents a comparative analysis of the application of different convolutional neural network in the diagnosis of lung cancer. Three network models are described for the diagnosis of the disease, based on architectures ResNet50v2, Xception and sequential. In order to improve the classification result, we also proposed the use of a voting. The training and validation of the networks was based on the IQ-OTHNCCD dataset from Wasit University. The voting-based model had an overall accuracy of 92%, being able to detect 99% of malignant tumors, 95% of normal CT scans and 71% of benign tumors.*

Resumo. *Este trabalho apresenta uma análise comparativa da aplicação de diferentes arquitetura de redes neurais convolucionais no diagnóstico do câncer de pulmão. São detalhados três modelos de redes implementados para o diagnóstico da enfermidade, baseados nas arquiteturas ResNet50v2, Xception e rede sequencial. A fim de melhorar a classificação final, propõe-se também o uso de um método baseado em votação. O treinamento e validação das redes se deu com base no conjunto de dados IQ-OTHNCCD da Universidade de Wasit. O modelo baseado em votação apresentou acurácia geral de 92%, sendo capaz de detectar 99% dos tumores malignos, 95% das tomografias normais e 71% dos tumores benignos.*

1. Introdução

Segundo o Instituto Nacional do Câncer Brasileiro [INCA 2019], o câncer é o principal problema de saúde pública no mundo e já está entre as quatro principais causas de morte prematura, antes dos 70 anos de idade, na maioria dos países. No Brasil, estimasse que a cada ano durante o triênio 2020-2022, 625 mil novos casos sejam diagnosticados, sendo 30 mil deles cancros da traqueia, brônquio e pulmão. Para [Araujo et al. 2018], a quantidade elevada de óbitos por câncer de pulmão se dá pela ineficiência no diagnóstico precoce e pela falta do acesso da população à exames especializados. Em seu trabalho [Franceschini e Santoro 2020], citam que os exames por imagem, como a tomografia por emissão de pósitrons e o raio-X são fortes aliados para identificação precoce de câncer, e aumentam as chances do paciente se recuperar, quando submetidos a um tratamento adequado [Araujo et al. 2018].

Tradicionalmente, o diagnóstico por imagem se dá por meio da análise feita por profissionais de saúde com base na sua expertise, todavia cabe destacar o notório crescimento do uso de inteligência computacional como mecanismo de suporte ao diagnóstico médico, principalmente daquelas baseadas em *Deep Learning* [Rahimzadeh et al. 2021]. Muitos trabalhos foram publicados durante todo o ano de 2020 descrevendo o uso de redes convolucionais aplicadas na extração de características de imagens da região do pulmão em exames de raio-X, tomografias e ultrassonografias, visando o diagnóstico da COVID-19 [Wang et al. 2020] [Gunraj et al. 2020] [Rahimzadeh et al. 2021]. Observa-se nesse tipo de abordagem a ideia de treinar um modelo inteligente, através de exemplos reais, para que assim o mesmo consiga generalizar soluções gerais para amostras que ainda lhe foram apresentadas.

Segundo [Rahimzadeh et al. 2021], as redes convolucionais trouxeram um novo leque de possibilidades para o campo da visão computacional e hoje permitem identificar, com precisão, padrões, objetos de maneira automática que podem não estar totalmente visíveis ao olho humano. Dada a eficiência das redes convolucionais em diagnosticar a COVID-19, pressupõe-se que as mesmas também funcionariam bem para o diagnóstico de câncer de pulmão. Levando isso em consideração, esse artigo descreve a implementação de três arquiteturas de redes neurais convolucionais aplicadas ao diagnóstico do câncer de pulmão.

Os modelos de redes propostos seguem diferentes arquiteturas. O primeiro deles baseia-se na arquitetura tradicional de uma rede neural convolucional, com a intercalação de camadas. Já o segundo e o terceiro baseiam-se em transferência de conhecimento e utilizam redes convolucionais já descritas na literatura (*ResNet50V2* e *Xception*). O treinamento e validação das redes foram realizadas utilizando a base de dados *IQ-OTHNCCD* da Universidade de Wasit, Iraque [Alyasriy e Muayed 2021]. Nesse conjunto de dados, há um total de 1097 tomografias computadorizadas em escala de cinza, previamente classificadas em três classes (normal, maligno e benigno), tendo cada imagem 512x512 píxeis. A fim de melhorar a classificação, foi implementado um método baseado em votação para agregar os resultados das três redes e assim fornecer uma classificação mais eficiente.

2. Trabalhos Relacionados

Redes neurais convolucionais vem sendo aplicadas como mecanismos de apoio ao diagnóstico médico nas mais diversas especialidades. Por conta da pandemia do novo coronavírus, se tornou bastante comum o uso desses modelos visando o diagnóstico precoce, a partir de imagens da região do pulmão. Durante todo o ano de 2020, novas arquiteturas foram apresentadas, aumentando assim o leque de possibilidades para aplicação de modelos baseados em *deep learning*. Anteriormente, diversas redes convolucionais já haviam sido propostas e validadas para o apoio aos mais diversos tipos de diagnósticos, como câncer de mama [Baffa 2018], tumores cerebrais [Balasooriya e Nawarathna 2017], lesões da pele [Dorj et al. 2018], entre outros.

No que diz respeito a análise de tomografias e raios-X da área dos pulmões, cabe destacar os trabalhos desenvolvidos por [Rahimzadeh et al. 2021] e [Narin et al. 2021] nos quais são descritas arquiteturas de rede cuja acurácia no diagnóstico da COVID-19 chega a 99,5% para imagens de raio-X e 98,49% em tomografias. Em ambos os artigos, observa-se a aplicação de transferência de conhecimento, utilizando as arquiteturas

ResNet50V2 e *Xception* e no último a presença de redes piramidais. Tais arquiteturas caracterizam-se pela menor quantidade de parâmetros e boa eficiência, tornando mais rápido o treinamento da rede mesmo com uma grande quantidade de imagens.

Redes neurais convolucionais também podem trabalhar em conjunto com outros algoritmos de aprendizagem. Em seu trabalho [Demir et al. 2020] descreve o uso dessas redes em conjunto com as máquinas de vetores de suporte no diagnóstico de doenças pulmonares a partir da conversão de sons em imagens. O autor propõe o uso transformadas de *Fourier* de curto tempo (STFT) para conversão de espectrogramas em imagem, que são então usadas no o treinamento de uma rede VGG-16. A saída da rede nesse caso fornece a entrada da SVM. Segundo o autor, essa arquitetura trouxe um acréscimo de 2% na acurácia quando comparada com a utilização de uma rede VGG-16 tradicional, obtendo-se 65,5% e 63,09% respectivamente.

Levando em consideração os resultados dos dois primeiros estudos, cuja acurácia ficou próxima dos 100%, bem como a semelhança do problema proposto nos dois trabalhos, optou-se pela implementação de arquiteturas baseadas em *ResNet50V2* e a *Xception*, além de uma rede sequencial a fim de identificar o quão superior seria a classificação diante das duas primeiras.

3. Descrição da Base de Dados

Para treinamento e teste das três redes convolucionais foi utilizado o *dataset* do Hospital Nacional de Ensino Oncológico do Iraque, disponibilizado pela Universidade de Wasit [Alyasriy e Muayed 2021]. O conjunto de dados possui 1097 tomografias computadorizadas em escala de cinza, cada imagem tem resolução de 512x512 píxeis, coletadas durante três meses no ano de 2019. As tomografias foram previamente classificadas em três classes (maligno, normal e benigno) por radiologistas e não contém referências ao paciente a quem pertencem. A Figura 1 contém amostras aleatórias retiradas do conjunto de dados original.

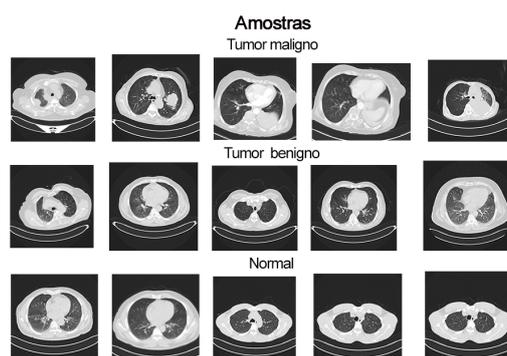


Figura 1. Amostras distribuídas por classes

Segundo os criadores do *dataset*, nas tomografias cuja classificação é do tipo maligno há uma variação quanto a extensão das lesões, isto é, há pacientes em estágio mais avançados do que outros. Na distribuição geral, em relação a quantidade de casos, têm-se 40 casos diagnosticados como tumor maligno, 55 casos diagnosticados como normais e 15 casos diagnosticados como benignos.

Como o conjunto de dados encontra-se previamente classificado, é possível aplicar inteligência computacional para que um determinado modelo consiga generalizar um diagnóstico dentro das três classes conhecidas. Para isso, serão utilizadas diferentes arquiteturas de redes convolucionais, bem como uma abordagem baseada em votação. A próxima seção descreve a arquitetura de cada tipo de rede implementada, bem como a tecnologia e hardware utilizados no projeto. A tabela 1 descreve a quantidade de imagens por classe.

Tabela 1. Distribuição da base IQ-OTHNCCD por classe

Classes	Quantidade de casos	Quantidade de imagens
Maligno	40	561
Normal	55	416
Benigno	15	120
Total	110	1097

4. Metodologia

Toda a implementação do projeto foi feita utilizando a linguagem python, a API Keras e Tensorflow. Também utilizou-se as bibliotecas *scikit learning* para geração dos relatórios de desempenho. O *framework* Keras fornece uma série de abstrações que facilitam a implementação de redes convolucionais, tornando simples e intuitiva a definição de novas arquiteturas. Além disso, a API oferece suporte à transferência de conhecimento, provendo uma série de modelos já descritos na literatura como *ResNet*, *Xception*, *Inception*, *AlexNet*. Treinamento e testes foram executados num computador de mesa com processador AMD Ryder 2700X, com oito núcleos e velocidade de clock de 3.7Ghz, memória RAM de 16GB e placa de vídeo Nvidia Gforce 1060 com 6GB dedicados.

A escolha dos modelos se deu por meio da análise da literatura, dando-se preferência àqueles que apresentaram bons resultados na detecção de pacientes infectados pela COVID-19. A título de comparação optou-se por desenvolver, além dos dois modelos baseados em transferência de conhecimento (ResNet50v2 [He et al. 2015] e Xception [Chollet 2017]), uma rede convolucional simples, cuja arquitetura está pautada na descrição básica dada por uns dos primeiros artigos sobre o tema [LeCun et al. 1998], a qual denominamos TwoRConv. Buscando melhorar a classificação, propõe-se também o uso de um método baseado em votação a fim de agregar resultados dos três modelos desenvolvidos. Nesse caso, o diagnóstico invés de ficar restrito a um modelo específico, se dará pela verificação da saída que mais se repete dentre todas as redes treinadas.

A base original foi dividida de maneira aleatória em duas novas pastas, sendo a primeira usada para treinamento e a segunda para teste. Utilizou-se a proporção de 80% das amostras para teste e o restante para validação. Como o banco de imagens é relativamente pequeno, durante o treinamento utilizou-se a técnica denominada em inglês *Image Augmentation*, a qual consiste na aplicação diferentes tipos de transformação numa mesma imagem para que ganhe diferentes representações e assim evite o *overfitting*.

5. Redes Implementadas

Os três modelos foram treinados e validados com base na nova subdivisão do *dataset*. Para não criar nenhum tipo de viés, os parâmetros e quantidade de épocas de treinamento

foram iguais para todos os modelos. As imagens do banco tiveram que passar por um redimensionamento antes de serem repassadas como entrada para as redes. Foram utilizadas imagens em escala de cinza com resolução 300x300 píxeis.

5.1. TwoRConv

Essa rede convolucional utiliza a arquitetura definida por [LeCun et al. 1998] na LeNet, ou seja, ela intercala camadas de convolução, *pooling* e realiza o processo denominado *flattening*. Por se tratar de uma rede do tipo sequencial, a TwoRConv também possui uma segunda parte na forma de uma rede neural multicamadas do tipo MLP responsável por receber o resultado do *flattening* e realizar a classificação.

A TwoRConv possui apenas duas camadas de convolução e duas camadas de *Max-Pooling*. Na camada escondida, existem 512 neurônios, que utilizam a função de ativação RELU. Já na camada de saída existem 3 neurônios que utilizam a função sigmóide, permitindo a classificação multiclases.

5.2. ResNet50V2 e Xception

Como ambos os modelos já possuem implementação na biblioteca *Keras*, bem como pesos pré-estabelecidos, utilizou-se o conceito de transferência de conhecimento. Nesse caso, os pesos já conhecidos são utilizados para se chegar mais rápido ao resultado desejado. Ocorre que os pesos pré-estabelecidos de ambos os modelos foram definidos para imagens com três canais de cores, porém as tomografias estão em escala de cinza. Por conta disso, houve a necessidade de adicionar uma camada de entrada para intermediar a comunicação com os modelos base. Algo semelhante ocorre na camada de saída, visto que tanto a *ResNet50V2*, quanto a *Xception* foram projetadas para ter 1000 classes na saída. Para contornar essa situação, a última camada de 1000 neurônios foi retirada e substituída por uma camada de *pooling* global, responsável por converter as saídas da camada anterior em um vetor e repassar esses valores para uma rede neural densa. No caso do modelo *Xception*, a rede densa é composta por apenas três neurônios, que utilizam função de ativação sigmóide. No caso do *ResNet50V2*, a rede MLP é composta por uma camada escondida de 1024 neurônios e uma camada de saída de três. Na camada escondida do *ResNet50V2*, utilizou-se a função de ativação RELU, já em sua saída a função é sigmóide foi usada.

O processo de transferência de conhecimento se deu de forma diferente para os dois modelos. As camadas de convolução da *Xception* foram retreinadas para que se adaptassem ao conjunto de dados, já no caso da *ResNet50V2* verificou-se que os pesos pré-estabelecidos retornavam resultados melhores ao não sofrer alteração.

6. Experimentos

A base de treinamento continha 890 imagens, destas 96 tomografias contém tumores benignos, 334 são de pacientes normais e 449 apresentam tumores malignos. Já a pasta de validação possui 24 imagens da classe benigno, 83 da classe normal e 112 da classe maligno. Os modelos foram treinados na ordem apresentada na seção anterior e, ao final da execução do treinamento de cada um deles, foi gerado um relatório de desempenho a fim de medir a precisão, acurácia geral e o *recall*. Para os três treinamentos, foram definidos mecanismos de *callback* visando a recuperação da melhor configuração baseada na

acurácia geral no diretório de teste. Para geração dos relatórios, utilizou-se o melhor modelo retornado após execução de 50 épocas. A tabela 2 apresenta os parâmetros utilizados no treinamento de cada um dos modelos.

Tabela 2. Parâmetros de treinamento

Parâmetros	TwoR2conv	Xception	ResNet50V2
Taxa de aprendizagem	1e-4	1e-4	1e-4
Tamanho do batch	14	14	14
Otimizador	adam	adam	adam
Função de perda	Categorical Crossentropy	Categorical Crossentropy	Categorical Crossentropy
Épocas	50	50	50
Horizontal/vertical flipping	Yes	Yes	Yes
Zoom range	5%	5%	5%
Width/height shifting	5%	5%	5%

Para facilitar a visualização dos erros e acertos sobre as imagens de validação, foram geradas matrizes de confusão para a *TwoRConv*, *ResNet50V2*, *Xception*, conforme a Figura 2. Observa-se que todas as redes possuem índices de acerto bastante altos para as classes normal e maligno, todavia, o mesmo não ocorre para a classe benigno. Esse comportamento pode ser explicado, em partes, pelo menor número de exemplos dessa classe dentro do conjunto de treinamento, fazendo com que a classificação tenda mais para as classes que possuem mais amostras.



Figura 2. Matrizes de confusão

O relatório de desempenho apresentado na Tabela 3 mostra que a precisão e *recall* para as classes normal e maligno seguiram mais altas do que para a classe benigno. De modo geral, a rede *Xception* foi a que apresentou melhor desempenho, destacando-se o fato de ter obtido a maior acurácia geral. A rede também obteve os três melhores resultados de precisão para as três classes. Por fim, o *Xception* apresentou melhores valores de *recall*. Um ponto a ser destacado é o fato das precisões da classe normal

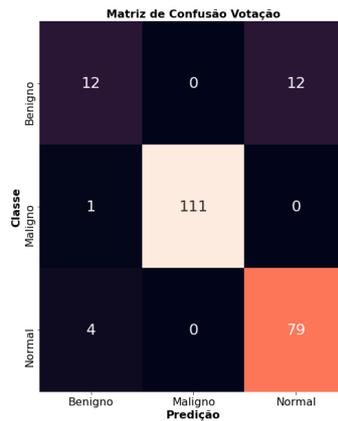


Figura 3. Matriz de confusão da Votação

ficarem próximas dos 85%, o que indica que algumas imagens dessa classe estão sendo classificadas como maligno ou benigno.

Tabela 3. Relatório de desempenho

	Acurácia Geral	Precisão			Recall		
		Benigno	Maligno	Normal	Benigno	Maligno	Normal
TwoRConv	0.89	0.53	1	0.83	0.33	0.97	0.95
Resnet50V2	0.90	0.74	1	0.83	0.71	0.91	0.94
Xception	0.91	0.64	1	0.84	0.38	0.99	0.95
Votação	0.92	0.71	1	0.87	0.50	0.99	0.95

Para melhorar a classificação, agregou-se o resultado da predição dos três modelos em uma só. Para isso, uma estratégia baseada em votação foi definida. Nela se apresenta a mesma imagem para as três redes, realiza-se a predição e ao fim há uma contagem de votos. Em caso de empate, a rede *Xception* possui valor mandatário e decide a saída. Levando em consideração o resultado de todas as métricas, constata-se que votação conseguiu balancear o desempenho da rede, melhorando a classificação.

7. Conclusão

No decorrer do trabalho foram descritos três modelos de redes convolucionais, bem como um mecanismo baseado em votação visando o retorno do diagnóstico mais adequado. As três redes treinadas apresentaram desempenho semelhante ao classificar imagens com tumores malignos, mas tiveram dificuldade para identificar lesões benignas. Cabe destacar o melhor desempenho da *Resnet50V2* na identificação de imagens consideradas benignas, enquanto a *Xception* apresentou maior precisão para identificar pacientes normais.

Diante dos resultados, é possível concluir que a aplicação de redes neurais convolucionais pode ser uma alternativa viável para o diagnóstico do câncer de pulmão. A classificação baseada em votação, a qual agrega resultado de três modelos implementados, atingiu uma acurácia geral de 92%, acertando 99% dos casos malignos e 87% dos casos normais e 50% dos benignos. A votação apresentou uma limitação quanto a classificação de casos benignos, isto é atribuído ao fato que o *dataset* apresentar poucos exemplos para

essa classe, impossibilitando assim uma classificação mais satisfatória. Porém, no geral, o modelo se mostrou bastante eficiente ao diferenciar casos malignos dos normais.

Referências

- Alyasriy, H. e Muayed, A.-H. (2021). The iq-othnccd lung cancer dataset.
- Araujo, L. H., Baldotto, C., Castro Jr, G. d., Katz, A., Ferreira, C. G., Mathias, C., Mascarenhas, E., Lopes, G. d. L., Carvalho, H., Tabacof, J., et al. (2018). Câncer de pulmão no brasil. *Jornal Brasileiro de Pneumologia*, 44(1):55–64.
- Baffa, Matheus de Freitas Oliveira e Lattari, L. G. (2018). Convolutional neural networks for static and dynamic breast infrared imaging classification. In *2018 31st SIBGRAPI Conference on Graphics, Patterns and Images (SIBGRAPI)*, pages 174–181. IEEE.
- Balasoorya, N. M. e Nawarathna, R. D. (2017). A sophisticated convolutional neural network model for brain tumor classification. In *2017 IEEE International Conference on Industrial and Information Systems (ICIIS)*, pages 1–5. IEEE.
- Chollet, F. (2017). Xception: Deep learning with depthwise separable convolutions.
- Demir, F., Sengur, A., e Bajaj, V. (2020). Convolutional neural networks based efficient approach for classification of lung diseases. *Health information science and systems*, 8(1):1–8.
- Dorj, U.-O., Lee, K.-K., Choi, J.-Y., e Lee, M. (2018). The skin cancer classification using deep convolutional neural network. *Multimedia Tools and Applications*, 77(8):9909–9924.
- Franceschini, J. P. e Santoro, I. L. (2020). Estadiamento do câncer de pulmão: uma visão epidemiológica brasileira. *Jornal Brasileiro de Pneumologia*, 46(1).
- Gunraj, H., Wang, L., e Wong, A. (2020). Covidnet-ct: A tailored deep convolutional neural network design for detection of covid-19 cases from chest ct images. *Frontiers in medicine*, 7.
- He, K., Zhang, X., Ren, S., e Sun, J. (2015). Deep residual learning for image recognition. 7.
- INCA (2019). *Estimativa 2020 : incidência de câncer no Brasil*. Instituto Nacional de Câncer José Alencar Gomes da Silva, Rio de Janeiro.
- LeCun, Y., Bottou, L., Bengio, Y., e Haffner, P. (1998). Gradient-based learning applied to document recognition. *Proceedings of the IEEE*, 86(11):2278–2324.
- Narin, A., Kaya, C., e Pamuk, Z. (2021). Automatic detection of coronavirus disease (covid-19) using x-ray images and deep convolutional neural networks. *Pattern Analysis and Applications*, pages 1–14.
- Rahimzadeh, M., Attar, A., e Sakhaei, S. M. (2021). A fully automated deep learning-based network for detecting covid-19 from a new and large lung ct scan dataset. *Bio-medical Signal Processing and Control*, 68:102588.
- Wang, L., Lin, Z. Q., e Wong, A. (2020). Covid-net: A tailored deep convolutional neural network design for detection of covid-19 cases from chest x-ray images. *Scientific Reports*, 10(1):1–12.