

# Classificação de doenças respiratórias a partir de imagens Raios-X de tórax usando redes neurais

Aldenis França  
 Campus Jaboatão dos Guararapes  
 Federal Institute of Pernambuco  
 Jaboatão dos Guararapes-PE, Brazil  
[aeagf@discente.ifpe.edu.br](mailto:aeagf@discente.ifpe.edu.br)

Carolina Torres  
 Campus Jaboatão dos Guararapes  
 Federal Institute of Pernambuco  
 Jaboatão dos Guararapes-PE, Brazil  
[carolina.torres@jaboatao.ifpe.edu.br](mailto:carolina.torres@jaboatao.ifpe.edu.br)

Luciano de Souza Cabral  
 Campus Jaboatão dos Guararapes  
 Federal Institute of Pernambuco  
 Jaboatão dos Guararapes-PE, Brazil  
[luciano.cabral@jaboatao.ifpe.edu.br](mailto:luciano.cabral@jaboatao.ifpe.edu.br)  
<https://orcid.org/0000-0002-4235-5753>

**Abstract**—A detecção precoce e precisa de doenças respiratórias, como COVID-19 e pneumonia, é de extrema importância para o tratamento eficaz e controle da disseminação dessas doenças. Com avanços recentes em técnicas de aprendizado de máquina, particularmente redes neurais convolucionais (CNNs), tornou-se possível utilizar imagens de raios-X para auxiliar na detecção dessas doenças. Neste artigo, exploramos a aplicação de redes neurais na detecção de COVID-19 e pneumonia usando imagens de raios-X, discutindo os desafios, abordagens e resultados alcançados até agora.

**Keywords**—COVID-19, pneumonia, redes neurais, raio-X, inteligência artificial.

## I. INTRODUÇÃO

A pandemia de COVID-19 representa um desafio significativo para a saúde global, exigindo esforços contínuos para aprimorar os métodos de detecção e diagnóstico de doenças. A pneumonia, por outro lado, continua sendo uma das principais causas de morbidade e mortalidade em todo o mundo. O uso de imagens de raios-X provou ser uma ferramenta valiosa na detecção e monitoramento dessas doenças respiratórias.

O objetivo deste estudo é explorar o uso de redes neurais convolucionais na detecção de COVID-19 e pneumonia por meio de imagens de raios-X. Discutiremos os principais desafios enfrentados nesta tarefa, as abordagens adotadas e os resultados obtidos até agora usando três redes neurais convolucionais: AlexNet, VGG16 e Xception.

## II. EMBASAMENTO TEÓRICO

### A. Covid-19

A COVID-19 é uma doença causada pelo coronavírus SARS-CoV-2, afetando principalmente o sistema respiratório. Seus sintomas variam de leves a graves, podendo levar à insuficiência respiratória e à morte. A detecção precoce do COVID-19 é crucial para limitar sua disseminação e fornecer tratamento adequado aos pacientes.

### B. Pneumonia

A pneumonia é uma infecção que afeta os pulmões e pode ser causada por diversos agentes infecciosos, como bactérias, vírus e fungos. Esta condição leva a sintomas característicos, incluindo febre, tosse, falta de ar e dor no peito. É de extrema importância identificar rapidamente a pneumonia, a fim de fornecer o tratamento adequado com antibióticos e outros cuidados médicos necessários.

### C. Imagens Raio-X

As radiografias de tórax, que utilizam imagens de raios X, são amplamente empregadas como uma ferramenta valiosa na detecção e diagnóstico de doenças pulmonares. Essas imagens fornecem uma visão detalhada da estrutura pulmonar, permitindo a identificação de anormalidades

associadas a problemas respiratórios. Com sua capacidade de fornecer informações precisas, as radiografias de tórax desempenham um papel crucial no auxílio aos profissionais de saúde na identificação e compreensão das doenças pulmonares.

### D. Convolutional Neural Network (CNN)

Uma rede neural convolucional (CNN) é uma arquitetura de rede neural profunda projetada especificamente para processamento de imagens. Consiste em camadas convolucionais que aplicam filtros às imagens de entrada, seguidas por camadas de agrupamento que reduzem a dimensionalidade dos dados.

As CNNs são conhecidas por sua capacidade de aprender recursos relevantes em imagens, tornando-as particularmente eficazes em tarefas de visão computacional, como classificação de objetos, detecção de objetos e segmentação de imagens.

Várias referências bibliográficas importantes contribuíram para o desenvolvimento e aplicação de redes neurais convolucionais. Alguns deles incluem:

- AlexNet [10]
- VGGNet [4]
- InceptionNet [11]
- ResNet [12]

Essas referências desempenharam um papel significativo no avanço do campo das CNNs e foram fundamentais para moldar sua arquitetura e desempenho.

## III. METODOLOGIA

Nesta seção, será fornecida uma descrição detalhada do conjunto de dados utilizado, juntamente com uma explicação dos métodos aplicados nas análises realizadas.

### A. Dataset

Para treinar e avaliar a rede neural, é necessário o acesso a um banco de dados contendo imagens de raios X de pacientes com COVID-19, pneumonia e casos saudáveis. Existem várias fontes de dados disponíveis, como repositórios públicos e colaborações entre instituições médicas.

O conjunto de dados utilizado possui uma estrutura de diretórios de imagens médicas ramificadas em três subpastas, contendo imagens de radiografia de tórax [1, 2] conforme Figura 1, distribuídas da seguinte forma:

- COVID: 1626 imagens (31.1%)
- NORMAL: 1802 imagens (34.5%)
- PNEUMONIA: 1800 imagens (34.4%)

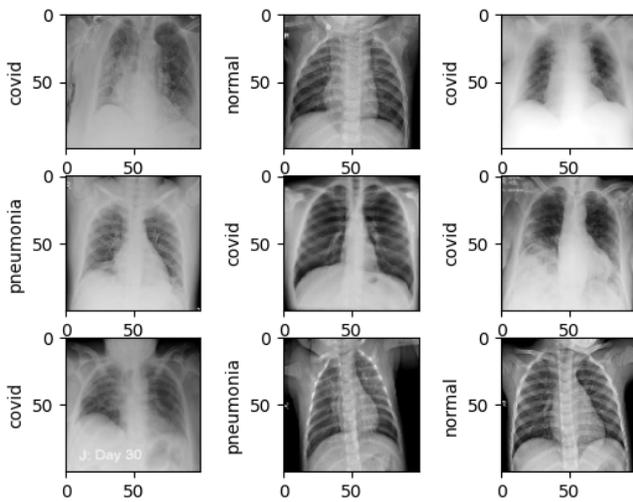


Fig. 1. Amostra de imagens de tórax do dataset em [1].

Essas imagens foram cuidadosamente selecionadas e rotuladas para garantir a precisão e relevância do conjunto de dados para os objetivos da pesquisa. Ao aproveitar esse conjunto de dados, espera-se obter insights significativos e desenvolver modelos eficazes para a detecção e classificação de casos de COVID-19, pneumonia e normais usando técnicas de aprendizado de máquina.

Prosseguindo, a próxima seção se concentrará em discutir a metodologia empregada para o pré-processamento dos dados e treinamento da rede neural.

### B. Pré Processamento das imagens

Antes de alimentar as imagens na rede neural, é crucial realizar um pré-processamento adequado. Isso envolve redimensionar as imagens para um tamanho padrão, normalizar os valores dos pixels, remover artefatos e aplicar realces de contraste, se necessário.

A etapa de redimensionamento garante consistência no tamanho de entrada, permitindo um processamento eficiente e a utilização de recursos computacionais. A normalização dos valores dos pixels traz os dados para uma escala comum, o que auxilia na redução de viés durante o treinamento e melhora a convergência da rede neural.

Ao incorporar técnicas de aumento de dados, a variabilidade do conjunto de dados é expandida, reduzindo o risco de overfitting e aprimorando a capacidade da rede de generalizar para exemplos não vistos.

No conjunto de dados utilizado, todas as imagens passaram por pré-processamento e foram redimensionadas para 100x100 no formato PNG. Além disso, técnicas de aumento de dados foram aplicadas, incluindo rotações e espelhamentos em uma faixa de 10 graus, para aumentar a diversidade do conjunto de treinamento.

No geral, essas etapas de pré-processamento desempenham um papel vital na preparação das imagens para entrada na rede neural, aprimorando seu desempenho e robustez em análises subsequentes e tarefas de classificação.

### C. Divisão em conjunto de treinamento e teste

Para avaliar o desempenho da rede neural, é importante dividir o conjunto de dados em conjuntos de treinamento e teste. O conjunto de treinamento é usado para ajustar os pesos da rede neural, enquanto o conjunto de teste é utilizado para avaliar sua capacidade de generalização [3]. É crucial garantir que esses conjuntos sejam representativos e não contenham imagens duplicadas.

Como o conjunto de dados fornecido não veio pré-estratificado em imagens de treinamento e teste, essa divisão foi realizada criando dois diretórios, chamados "treino" e "teste", e alocando as respectivas imagens em uma proporção de 80% e 20%, respectivamente. Além disso, uma amostra do conjunto de treinamento foi separada para fins de validação, permitindo a comparação de diferentes modelos.

Ao aderir a essas práticas, podemos avaliar e comparar com confiança o desempenho de diferentes modelos, garantindo a confiabilidade e validade dos resultados obtidos a partir do processo de treinamento da rede neural.

## IV. ARQUITETURA DA REDE NEURAL

A arquitetura da rede neural utilizada para a detecção de COVID-19 e pneumonia consiste em várias camadas convolucionais, seguidas por camadas de pooling e, posteriormente, camadas totalmente conectadas. Essa estrutura permite que a rede neural aprenda gradualmente características cada vez mais complexas nas imagens de raio-X.

### A. VGG16

A rede neural VGG16 é uma arquitetura de rede neural convolucional amplamente utilizada em tarefas de processamento de imagens, como classificação e detecção de objetos. Ela foi proposta pela Visual Geometry Group (VGG) na Universidade de Oxford [4].

A característica distintiva da VGG16 é a sua profundidade, com 16 camadas de convolução e pooling. A arquitetura consiste em várias camadas convolucionais empilhadas, seguidas por camadas de pooling e, posteriormente, camadas totalmente conectadas, como mostrado na Fig. 2 [4]. A VGG16 é conhecida por ter filtros de convolução pequenos (3x3) com profundidade constante em todas as camadas, o que permite uma representação mais rica de características.

A VGG16 alcançou resultados impressionantes em desafios de classificação de imagens, como o ImageNet Large Scale Visual Recognition Challenge (ILSVRC) em 2014 [5]. Ela demonstrou uma alta capacidade de aprendizado de recursos discriminativos, permitindo a extração de características complexas em imagens.

No entanto, a VGG16 também apresenta algumas desvantagens. Devido à sua profundidade, a arquitetura tem um grande número de parâmetros, o que pode levar a um alto consumo de memória e tempo de treinamento.

Além disso, a VGG16 não utiliza técnicas como a ativação de função ReLU - *Rectified Linear Unit* (Unidade Linear Retificada) [7], após as camadas de convolução, o que pode levar a problemas de saturação dos gradientes durante o treinamento.

Apesar das suas limitações, a VGG16 continua sendo uma arquitetura popular e serve como referência para o desenvolvimento de outras redes neurais convolucionais.

### B. Xception

A rede neural Xception é uma arquitetura de rede neural convolucional que foi proposta por François Chollet, um engenheiro do Google, em 2016 [8]. O nome "Xception" é uma junção de "Extreme Inception" e indica que a arquitetura é uma extensão da popular rede neural Inception.

A característica principal da Xception é a utilização de convoluções separáveis em profundidade, também conhecidas como convoluções em duas etapas [9]. Ao contrário das convoluções tradicionais, que misturam os

canais de entrada, as convoluções separáveis aplicam uma convolução espacial para cada canal individualmente e, em seguida, uma convolução em  $1 \times 1$  é aplicada para combinar as informações dos canais [8].

A Xception foi projetada com base na arquitetura Inception, que utiliza módulos Inception para extrair recursos em várias escalas. No entanto, a Xception leva a ideia um passo adiante, ela realiza uma convolução espacial  $1 \times 1$  seguida por uma convolução em profundidade  $3 \times 3$ , o que reduz o número de parâmetros e torna a arquitetura mais

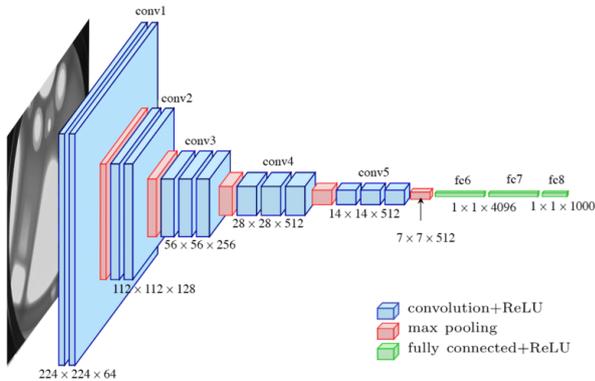


Fig. 2. A arquitetura de rede padrão VGG-16 proposta em [4].

eficiente em termos computacionais [9], como mostrado na Fig. 3. Isso permite que a rede aprenda características mais discriminativas em diferentes escalas e leve em consideração as relações espaciais entre os canais.

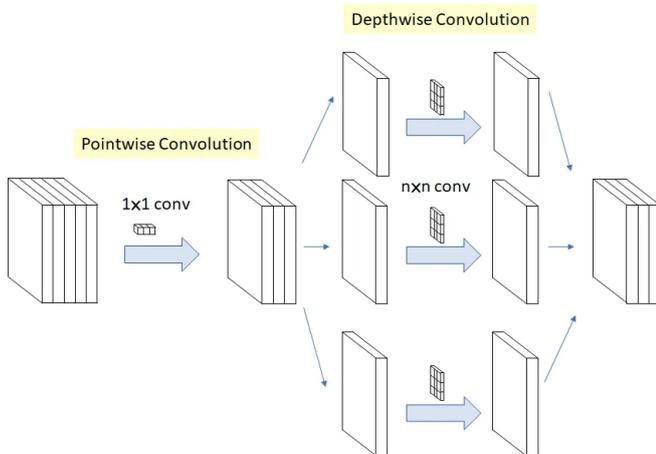


Fig. 3. Modelo de convoluções pontuais seguidas de convoluções em profundidade [9].

A arquitetura Xception demonstrou resultados impressionantes em desafios de classificação de imagens, como o ImageNet Large Scale Visual Recognition Challenge (ILSVRC) em 2016 [5]. Ela atingiu uma alta precisão de classificação e superou muitas arquiteturas anteriores em termos de desempenho e eficiência.

### C. Treinamento e otimização

O treinamento da rede neural é realizado por meio do ajuste dos pesos das suas camadas para minimizar uma função de perda. A otimização é realizada utilizando algoritmos como o gradiente descendente estocástico (SGD) ou variações mais avançadas, como o Adam [14].

Durante o treinamento, é importante monitorar a acurácia e a perda nos conjuntos de treinamento e teste para evitar problemas de overfitting.

### D. Transferência de aprendizado

A transferência de aprendizado é uma técnica comum em que uma rede neural pré-treinada em um conjunto de dados grande, como o ImageNet [5], é utilizada como ponto de partida. Os pesos das camadas iniciais são mantidos fixos, enquanto as camadas finais são adaptadas para a tarefa específica de detecção de COVID-19 e pneumonia. Isso permite que a rede neural aproveite o conhecimento prévio aprendido em tarefas relacionadas.

### V. RESULTADOS E DISCUSSÃO

Para a realização do trabalho foram escolhidos alguns modelos pré-treinados disponíveis na biblioteca Keras [13], utilizando o Python como linguagem de programação em notebooks no Google Colaboratory.

Com a finalidade de emitir diagnóstico automatizado de Covid-19 e Pneumonia, a serem confirmados pelo profissional de medicina, foram utilizadas as redes neurais convolucionais profundas na arquitetura AlexNet, VGG16 e Xception. Foram avaliados os desempenhos de cada uma nos conjuntos de treinamento e teste, retornando valores de acurácia e perda, além do histórico desses valores a cada época de análise.

A fim de padronizar as análises, utilizou-se imagens na proporção de  $100 \times 100$ , as análises foram feitas em 5 épocas, com batch size de 32 e random state de 42.

#### A. Desempenho da rede neural AlexNet

Após realizar a carga das imagens de tórax e convertê-las em arrays numpy, dividiu-se a base em treinamento e teste em uma proporção 80/20, estratificados pelos targets Normal, Covid e Pneumonia. Foram aplicadas camadas de convolução e Max Pooling bidimensionais e uma camada densa na saída, com funções de ativação ReLU e softmax.

No treinamento do modelo utilizou-se o método estocástico Adam [14], que é uma extensão do gradiente estocástico descendente para atualizar os pesos da rede utilizando *Momentum and Adaptive Learning Rates* para convergir mais rápido. Para esta rede neural foram definidos os hiperparâmetros presentes na Tabela I.

Tabela I. Hiperparâmetros do modelo AlexNet

Batch size	Input shape	Random state	Alpha	Epochs
32	(150,150,3)	42	0.00001	5

Após as cinco épocas de treinamento, chegou-se aos valores de acurácia e perdas mostrados na Tabela II.

Tabela II. Acurácias e perdas em treino e teste do modelo AlexNet

Conjunto de Dados	Acurácia Top-1 (%)	Acurácia Top-5 (%)	Perda Top-1 (%)	Perda Top-5 (%)
Treino	72.91	92.09	70.34	21.24
Teste	88.62	92.45	33.13	18.74

Ainda, nas Fig. 4 e Fig. 5 pode-se constatar a evolução desses elementos graficamente.

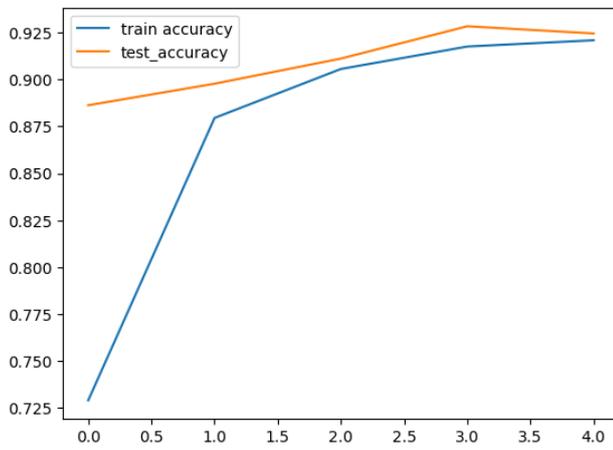


Fig. 4. Evolução da acurácia nos datasets de treino e teste (AlexNet).

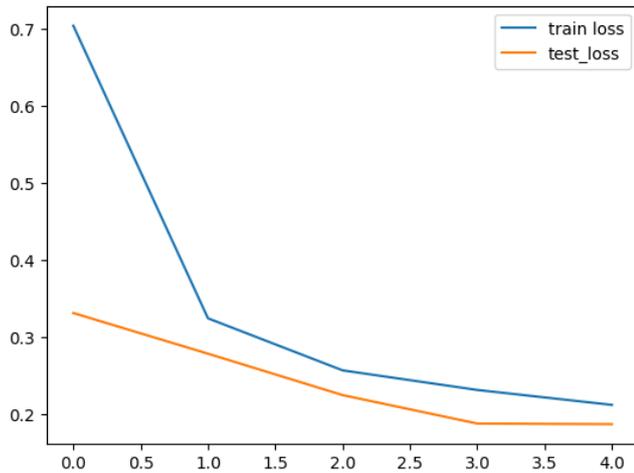


Fig. 5. Evolução da perda nos datasets de treino e teste (AlexNet).

Note que a cada época a acurácia foi se elevando, aproximando-se de um, enquanto as perdas foram sendo reduzidas, aproximando-se de zero.

### B. Desempenho da rede neural VGG16

Feita a carga das imagens de tórax de cada diretório (Normal, Covid e Pneumonia), elas foram redimensionadas para o tamanho 100x100 e convertidas em arrays numpy. Como padronização, dividiu-se a base em treinamento (80%) e teste (20%). Foram aplicadas camadas densa e plana, com funções de ativação ReLU e sigmoid.

Também foi utilizado o método estocástico Adam [14], como otimizador dos pesos da rede. Para esta rede neural foram definidos os hiperparâmetros presentes na Tabela III.

Tabela III. Hiperparâmetros do modelo VGG16.

Batch size	Learning rate	Decay steps	Decay rate	Epochs
32	0.0001	10	1	5

Após as cinco épocas de treinamento, chegou-se aos valores de acurácia e perdas mostrados na Tabela IV.

Tabela IV. Acurácias e perdas em treino e teste do modelo AlexNet

Conjunto de Dados	Acurácia Top-1 (%)	Acurácia Top-5 (%)	Perda Top-1 (%)	Perda Top-5 (%)
Treino	56.45	60.82	0.00	0.00
Teste	58.10	60.48	0.00	0.00

Ainda, nas Fig. 6 e Fig. 7 pode-se constatar a evolução desses elementos graficamente.

A cada época a acurácia aumenta, aproximando-se de 61%, enquanto as perdas foram muito baixas, em valores próximos a zero.

### C. Desempenho da rede neural Xception

De forma semelhante aos casos anteriores, concluída a carga das imagens de tórax, os seus caminhos (paths) foram adicionados a um *dataframe*, em que se aplicou a divisão de treino e teste em 80% e 20%, respectivamente, estratificados pelos targets Normal, Covid e Pneumonia.

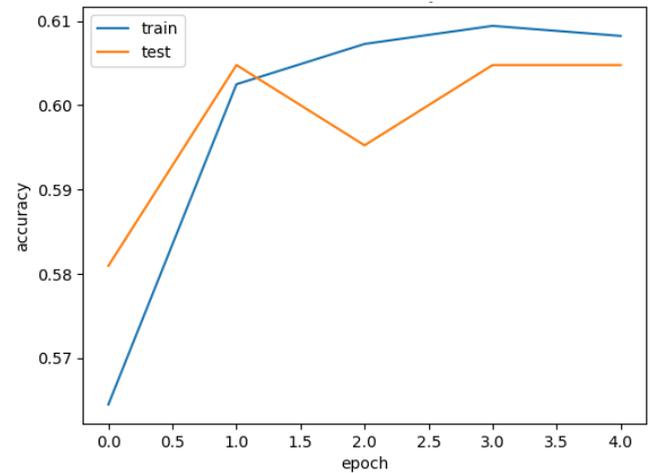


Fig. 6. Evolução da acurácia nos datasets de treino e teste (VGG16).

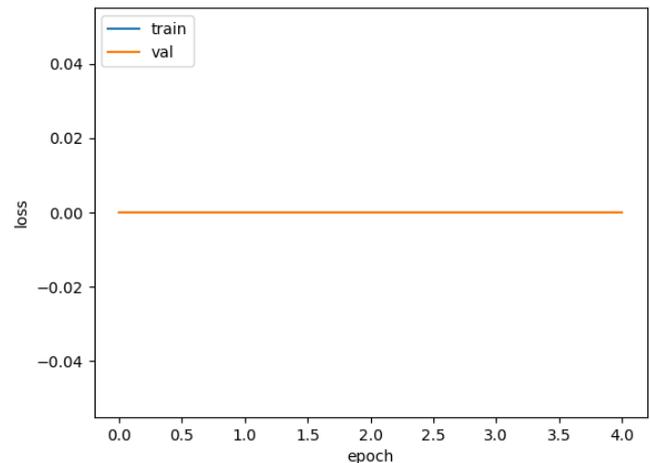


Fig. 7. Evolução da perda nos datasets de treino e teste (VGG16).

Desta forma, foram criados dois diretórios separados com os arquivos de imagem de treino e os arquivos de imagem de teste. Todos com as subpastas Normal, Covid e Pneumonia. Após isso, foram aplicados redimensionamento, rotação e mudança de zoom nas imagens

Além disso, foram aplicadas camadas densa e plana, com funções de ativação ReLU e softmax, e utilizou-se o método estocástico Adam [14], para os pesos da rede convergirem mais rápido. Para esta rede neural foram definidos os hiperparâmetros presentes na Tabela V.

Tabela V. Hiperparâmetros do modelo Xception.

Batch size	Target size	Random state	Rotation range	Epochs
32	(100, 100)	42	10	5

Após as cinco épocas de treinamento, chegou-se aos valores de acurácia e perdas mostrados na Tabela VI.

Tabela VI. Acurácias e perdas em treino e teste do modelo Xception.

Conjunto de Dados	Acurácia Top-1 (%)	Acurácia Top-5 (%)	Perda Top-1 (%)	Perda Top-5 (%)
Treino	70.39	87.67	247.63	33.02
Teste	90.24	90.24	25.04	26.87

Ainda, nas Fig. 8 e Fig. 9 pode-se constatar a evolução desses elementos graficamente.

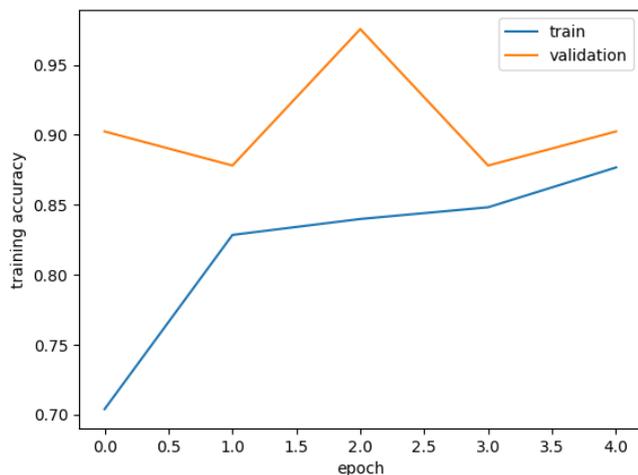


Fig. 8. Evolução da acurácia nos datasets de treino e teste (Xception).

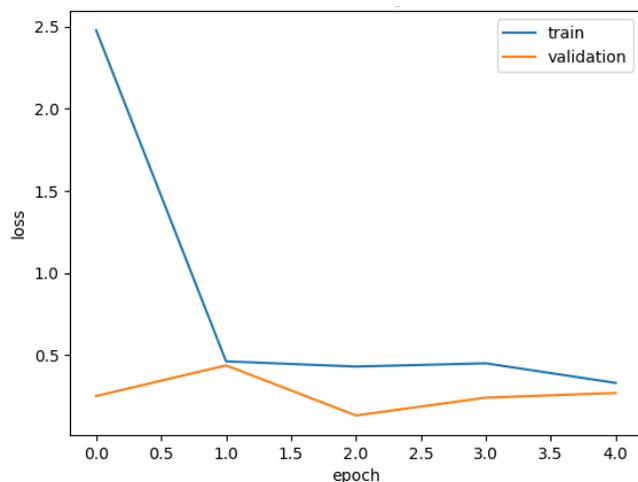


Fig. 7. Evolução da perda nos datasets de treino e teste (Xception).

Percebe-se que a cada época a acurácia foi se elevando, aproximando-se de 90%, enquanto as perdas foram sendo reduzidas, aproximando-se dos 20%.

#### D. Discussão geral sobre os modelos

Comparando a detecção de COVID-19 e pneumonia utilizando redes neurais com abordagens tradicionais, como classificadores lineares e métodos baseados em regras, observa-se um desempenho significativamente superior das CNNs. A capacidade das redes neurais de aprender automaticamente características relevantes das imagens de raio-X permite uma detecção mais precisa e robusta.

De forma semelhante, comparando-se os modelos de CNN utilizados e baseando-se nas acurácias e perdas apresentadas, vê-se que o modelo AlexNet performou melhor, com valores altos de acurácia (acima de 92%) e baixos níveis de perda (próximos a 20%).

Entretanto, mesmo o modelo VGG16 apresentando perdas em zero, ele teve o pior desempenho em acurácia (~61%). Ainda, o Xception teve uma boa performance com acurácia próximo a 90% e perdas em 20%, então é um modelo que deve ser levado em consideração também na detecção de doenças respiratórias por radiografias de tórax. Na Tabela VII, pode-se ver um resumo desses valores.

Tabela VII. Resumo das acurácias e perdas em treino e teste.

Dataset	Treino		Teste	
	Acurácia Top-5 (%)	Perda Top-5 (%)	Acurácia Top-5 (%)	Perda Top-5 (%)
AlexNet	92.09	21.24	92.45	18.74
VGG16	60.82	0.00	60.48	0.00
Xception	87.67	33.02	90.24	26.87

#### E. Limitações e desafios

Apesar dos avanços alcançados, existem algumas limitações e desafios a serem considerados. A disponibilidade de bases de dados equilibradas e representativas, a interpretabilidade das decisões tomadas pela rede neural, o poder computacional disponível para executar os modelos e a generalização para diferentes populações e condições clínicas são aspectos que requerem atenção adicional.

#### VI. CONCLUSÃO

A utilização de redes neurais convolucionais para a detecção de COVID-19 e pneumonia por meio de imagens de raio-X mostra-se promissora, apresentando resultados encorajadores em termos de sensibilidade e especificidade.

No futuro, é importante continuar aprimorando as técnicas de detecção utilizando redes neurais, explorando abordagens como a segmentação de regiões de interesse e o uso de técnicas de interpretabilidade para entender melhor as decisões tomadas pela rede. Além disso, é fundamental colaborar com profissionais de saúde para validar e incorporar essas abordagens em ambientes clínicos reais, contribuindo para aprimorar o diagnóstico e o tratamento dessas doenças respiratórias.

#### AGRADECIMENTOS

Gostaríamos de agradecer ao Instituto Federal de Pernambuco (IFPE) - Campus Jaboatão dos Guararapes pelo suporte provendo infraestrutura necessária para transcorrer com a pesquisa e pelo suporte financeiro parcial para o evento.

#### REFERÊNCIAS

- [1] Shastri, S., Kansal, I., Kumar, S. et al. CheXImageNet: a novel architecture for accurate classification of Covid-19 with chest x-ray digital images using deep convolutional neural networks. Health Technol. 12, 193–204 (2022). <https://doi.org/10.1007/s12553-021-00630-x>.
- [2] Kumar S, Shastri S, Mahajan S, et al. LiteCovidNet: A lightweight deep neural network model for detection of COVID-19 using X-ray images. Int J Imaging Syst Technol. 2022;1-17. DOI: <https://doi.org/10.1002/ima.22770>.
- [3] Thomas, R. How (and why) to create a good validation set. Fast.ai, 2017. <https://www.fast.ai/posts/2017-11-13-validation-sets.html>
- [4] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," arXiv preprint arXiv:1409.1556, 2014.
- [5] O. Russakovsky et al., "Imagenet large scale visual recognition challenge," International Journal of Computer Vision, vol. 115, no. 3, pp. 211–252, 2015.

- [6] Deng, J., Dong, W., Socher, R., Li, L. J., Li, K., & Fei-Fei, L. (2009). ImageNet: A large-scale hierarchical image database. IEEE Conference on Computer Vision and Pattern Recognition (CVPR).
- [7] Ceccon, D. Funções de ativação: definição, características, e quando usar cada uma. Expert Academy, 2020. <https://jaexpert.academy/2020/05/25/funcoes-de-ativacao-definicao-caracteristicas-e-quando-usar-cada-uma/>
- [8] François Chollet. "Xception: Deep Learning with Depthwise Separable Convolutions." Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR). 2017.
- [9] Tsang Sik-Ho. "Review: Xception with depthwise separable convolution". Towards Data Science, 2018. <https://towardsdatascience.com/review-xception-with-depthwise-separable-convolution/-better-than-inception-v3-image-dc967dd42568>.
- [10] Krizhevsky, A., Sutskever, I., & Hinton, G. E. (2012). ImageNet classification with deep convolutional neural networks. In: Advances in neural information processing systems (pp. 1097-1105).
- [11] Szegedy, C., Liu, W., Jia, Y., Sermanet, P., Reed, S., Anguelov, D., ... & Rabinovich, A. (2015). Going deeper with convolutions. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 1-9).
- [12] He, K., Zhang, X., Ren, S., & Sun, J. (2016). Deep residual learning for image recognition. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 770-778).
- [13] Keras. Keras applications. <https://keras.io/applications/>. [Online; acessado em: 18 de Maio de 2023].
- [14] Kingma, D and Ba, J . (2015) Adam: A method for Stochastic Optimization. Available at: <https://arxiv.org/pdf/1412.6980.pdf>