

CLASSIFICAÇÃO DE DOENÇAS RESPIRATÓRIAS A PARTIR DE IMAGENS RAIO-X DE TÓRAX USANDO REDES NEURAIAS

CLASSIFICATION OF RESPIRATORY DISEASES FROM CHEST X-RAY IMAGES USING NEURAL NETWORKS

Aldenis Everton Alves Guilherme de França

aeagf@discente.ifpe.edu.br

Maria Carolina Torres da Silva

carolina.torres@jaboatao.ifpe.edu.br

Luciano de Souza Cabral

luciano.cabral@jaboatao.ifpe.edu.br

RESUMO

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 raio-X para auxiliar na classificação dessas doenças. Neste artigo, exploramos a aplicação de redes neurais na classificação de COVID-19 e pneumonia usando imagens de raio-X, discutindo os desafios, abordagens e resultados alcançados até agora.

Palavras-chave: COVID-19. Pneumonia. Redes neurais. Raio-X. Inteligência artificial.

ABSTRACT

Early and accurate detection of respiratory diseases, such as COVID-19 and pneumonia, is extremely important for effective treatment and control of the spread of these diseases. With recent advances in machine learning techniques, particularly convolutional neural networks (CNNs), it has become possible to use X-ray images to aid in the classification of these diseases. In this article, we explore the application of neural networks in classification COVID-19 and pneumonia using X-ray images, discussing the challenges, approaches and results achieved so far.

Keywords: COVID-19. Pneumonia. Neural networks. X-ray. Artificial intelligence.

1 INTRODUÇÃO

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 (Biblioteca Virtual em Saúde, 2021). Já a pneumonia é uma infecção que afeta os pulmões e pode ser causada por diversos agentes infecciosos, como bactérias, vírus e fungos (Varella, 2011). Esta condição leva a certos sintomas, incluindo febre, tosse, falta de ar e dor no peito, os quais são comuns às doenças supracitadas.

Desta forma, a pandemia de COVID-19 representou 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. Com isso, o uso de imagens de raio-X provou ser uma ferramenta valiosa na detecção e monitoramento dessas doenças respiratórias. Pois, essas imagens fornecem uma visão detalhada da estrutura pulmonar, permitindo a identificação de anormalidades associadas a problemas respiratórios.

Logo, 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. Pois, há casos de pessoas que vieram a óbito ou estão com sequelas por terem recebido um diagnóstico errado de doença respiratória e receberam um tratamento inadequado, como relatado em Marcos (2022).

Com avanços recentes em técnicas de aprendizado de máquina, tornou-se possível utilizar imagens raio-X para auxiliar na classificação dessas doenças. Uma dessas técnicas é a rede neural convolucional (CNN), um tipo especial de algoritmo de inteligência artificial inspirado na forma como o cérebro humano processa informações visuais. Pode ser comparável a um “cérebro artificial”, o qual pode aprender a reconhecer padrões em imagens (Krizhevsky *et al.*, 2012).

Estas redes são conhecidas por sua capacidade de aprender recursos relevantes em imagens, tornando-as particularmente eficazes em segmentação de imagens, classificação e detecção de objetos (Krizhevsky *et al.*, 2012). Deste modo, entende-se que ela pode ser uma ferramenta auxiliar no diagnóstico de doenças respiratórias, servindo como mais um ponto de confirmação do diagnóstico prescrito pelos profissionais da saúde.

Sendo assim, a pergunta que queremos responder aqui é: “Qual a rede neural convolucional com maior taxa de acerto (acurácia) ao classificar doenças respiratórias a partir de imagens de raio-X de tórax?”

Portanto, o objetivo deste artigo é explorar o uso de redes neurais convolucionais na identificação precoce e precisa de doenças respiratórias, como a COVID-19 e a pneumonia, por meio de imagens raio-X de tórax humano,

culminando em um tratamento eficaz e no controle da propagação dessas doenças. Para isso, pretende-se realizar o tratamento e divisão da base de dados de imagens raio-X, a fim de padronizá-la e executar treinos e validações necessários.

Além disso, após a aplicação dos modelos de redes neurais convolucionais, busca-se apresentar uma evolução das métricas estatísticas, para garantir a eficácia do modelo na classificação de doenças respiratórias. Com isso, entende-se a relevância deste trabalho no meio científico e acadêmico: garantir melhor qualidade de vida para a população e ser mais um aliado no combate de doenças que dizimam a vida de tantos ao redor do mundo.

2 CONCEITOS BÁSICOS

Nesta seção, serão apresentados alguns conceitos básicos em relação ao que são imagens de raio-X, o que são redes neurais convolucionais e alguns exemplos de redes neurais que existem, as quais foram utilizadas na construção dos modelos de *machine learning* neste artigo. Estes dois tópicos são primordiais para o entendimento do assunto tratado aqui.

2.1 Imagens Raio-X

O raio-X do tórax é um exame de imagem que utiliza radiações para visualizar o interior do tórax. Dessa forma, consegue obter informações sobre o coração, pulmões, ossos, vasos sanguíneos e outras estruturas localizadas na região (Cerquetani, 2023).

Este exame produz imagens do tórax por meio de pequenas doses de radiação ionizante (Cerquetani, 2023). As imagens raio-X geradas fornecem uma visão detalhada da estrutura pulmonar, permitindo a identificação de anormalidades associadas a problemas respiratórios.

Com essa 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, como a COVID-19 e a pneumonia.

2.2 Redes Neurais Convolucionais

Inicialmente, redes neurais são modelos computacionais inspirados no cérebro humano. Elas consistem em várias "camadas" de nós (neurônios) conectados, que processam e transmitem informações, aprendendo a identificar padrões a partir dos dados de entrada (HAYKIN, 2001).

Já especificamente, uma rede neural convolucional (*Convolutional Neural Network* - CNN) é como um scanner inteligente que analisa uma imagem em partes menores, extraíndo implicitamente características relevantes nessas partes e, então,

combina essas informações para entender o que está na imagem completa (Goodfellow, *et al.*, 2016).

Ela possui uma arquitetura de rede neural profunda projetada para processar dados que possuem uma grade de estrutura, como imagens (Goodfellow, *et al.*, 2016). 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 implícitos em imagens, tornando-as particularmente eficazes em tarefas de visão computacional, como classificação e detecção de objetos e segmentação de imagens.

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

- AlexNet (Krizhevsky *et al.*, 2012);
- VGGNet (Simonyan *et al.*, 2015);
- InceptionNet (Szegedy *et al.*, 2015);
- ResNet (He *et al.*, 2016).

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

3 ARQUITETURA DA REDE NEURAL

A arquitetura da rede neural utilizada para a classificação 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.

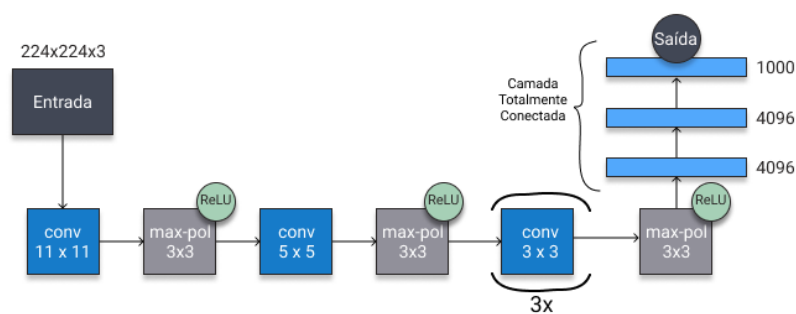
Desta forma, nesta seção será apresentado um resumo sobre cada uma das redes neurais utilizadas na modelagem deste artigo, a saber, a AlexNet, a VGG-16 e a *Xception*. Além de uma explicação sobre treinamento, otimização e transferência de aprendizado em redes neurais convolucionais.

3.1 AlexNet

A AlexNet é uma arquitetura de rede neural convolucional (CNN) revolucionária, introduzida por Alex Krizhevsky, Ilya Sutskever e Geoffrey Hinton em 2012. Sua principal contribuição foi demonstrar o poder das redes profundas na classificação de imagens, vencendo o *ImageNet Large Scale Visual Recognition Challenge (ILSVRC)* de 2012 com uma margem significativa de desempenho em relação a abordagens anteriores (Krizhevsky *et al.*, 2012).

A arquitetura da AlexNet é composta por oito camadas treináveis: cinco camadas convolucionais e três totalmente conectadas, como pode ser visto na Figura 1. Cada camada convolucional usa a função de ativação *ReLU* (*Rectified Linear Unit*), que acelera o treinamento em redes profundas, permitindo a extração eficaz de características complexas das imagens (Krizhevsky *et al.*, 2012). A AlexNet também incorpora camadas de *pooling*, que reduzem a dimensionalidade dos mapas de características, diminuindo a carga computacional e *dropout*, que desativa aleatoriamente neurônios durante o treinamento para evitar o *overfitting* (Krizhevsky *et al.*, 2012).

Figura 1 - Representação da arquitetura da rede neural AlexNet



Fonte: Krizhevsky *et al.* (2012)

Além disso, a AlexNet introduziu o uso intensivo de GPUs para acelerar o treinamento, dividindo o processo entre múltiplas unidades gráficas (Krizhevsky *et al.*, 2012). Essa implementação foi crucial para lidar com o grande volume de dados e a complexidade da arquitetura, tornando possível o uso eficiente de redes mais profundas em aplicações práticas.

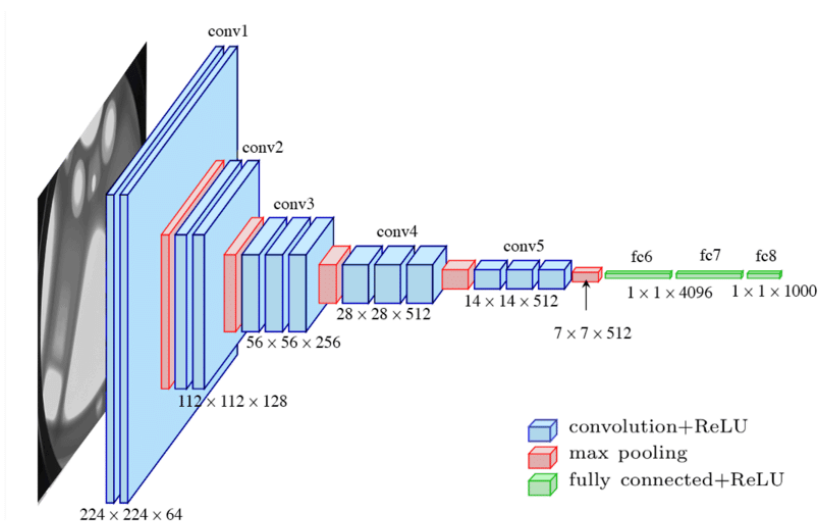
No geral, a AlexNet abriu caminho para arquiteturas mais profundas e complexas, como VGG (Simonyan *et al.*, 2015) e ResNet, tornando-se um marco na evolução do *deep learning*.

3.2 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 (Simonyan *et al.*, 2015).

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 Figura 2 (Simonyan *et al.*, 2015). 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.

Figura 2 - A arquitetura de rede padrão VGG-16 proposta por Simonyan e Zisserman



Fonte: Simonyan *et al.* (2015)

A VGG16 alcançou resultados impressionantes em desafios de classificação de imagens, como o *ImageNet Large Scale Visual Recognition Challenge (ILSVRC)* em 2014 (Russakovsky *et al.*, 2015). 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) (Ceccon, 2020), 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.

3.3 Xception

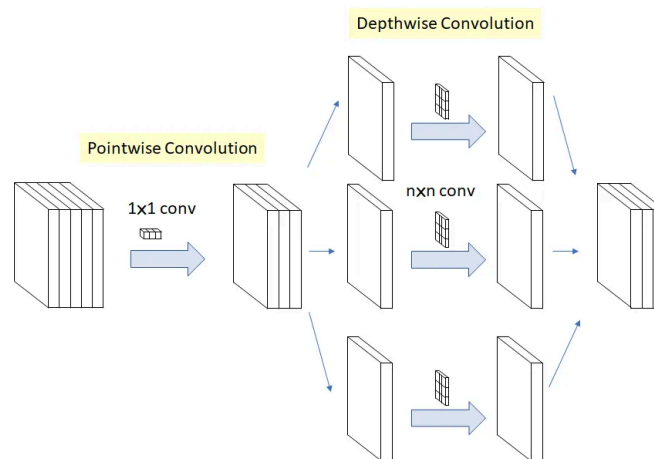
A rede neural *Xception* é uma arquitetura de rede neural convolucional que foi proposta por *François Chollet*, um engenheiro do *Google*, em 2016 (Chollet, 2017). O nome "*Xception*" é uma junção de "*Extreme Inception*" e indica que a arquitetura é uma extensão da popular rede neural *Inception*.

A característica principal da *Xception* é a utilização de convoluções separáveis em profundidade, também conhecidas como convoluções em duas etapas (Tsang, 2018). 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 1x1 é aplicada para combinar as informações dos canais (Chollet, 2017).

A rede neural *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 1x1 seguida por uma convolução em profundidade 3x3, o que reduz o número de parâmetros e torna a arquitetura mais eficiente em termos computacionais (Tsang, 2018), como mostrado na Figura 3. Isso permite que a rede aprenda características mais discriminativas em diferentes escalas e leve em consideração as relações espaciais entre os canais.

Figura 3 - Modelo de convoluções pontuais seguidas de convoluções em profundidade



Fonte: Tsang (2018)

A arquitetura *Xception* demonstrou resultados impressionantes em desafios de classificação de imagens, como o *ImageNet Large Scale Visual Recognition Challenge (ILSVRC)* em 2016 (Russakovsky *et al.*, 2015). Ela atingiu uma alta precisão de classificação e superou muitas arquiteturas anteriores em termos de desempenho e eficiência.

3.4 Treinamento, otimização e transferência de aprendizado

O treinamento de uma rede neural pode ser realizado por meio do ajuste dos pesos das suas camadas para minimizar uma função de perda. Já a otimização pode ser realizada utilizando algoritmos, como o gradiente descendente estocástico (SGD) ou variações mais avançadas, como o Adam (Kingma *et al.*, 2015).

Na fase de treinamento, é importante monitorar a acurácia e a perda nos conjuntos de treinamento e teste para evitar problemas de *overfitting* (sobreajuste).

Outra forma de treinar a rede neural é pela transferência de aprendizado. Trata-se de uma técnica comum em que uma rede neural pré-treinada em um conjunto de dados grande, como o ImageNet (Russakovsky *et al.*, 2015), é utilizada como ponto de partida.

No caso deste artigo, os pesos das camadas iniciais são mantidos fixos, enquanto as camadas finais são adaptadas para a tarefa específica de classificação em COVID-19 e pneumonia. Isso permite que a rede neural aproveite o conhecimento prévio aprendido em tarefas relacionadas.

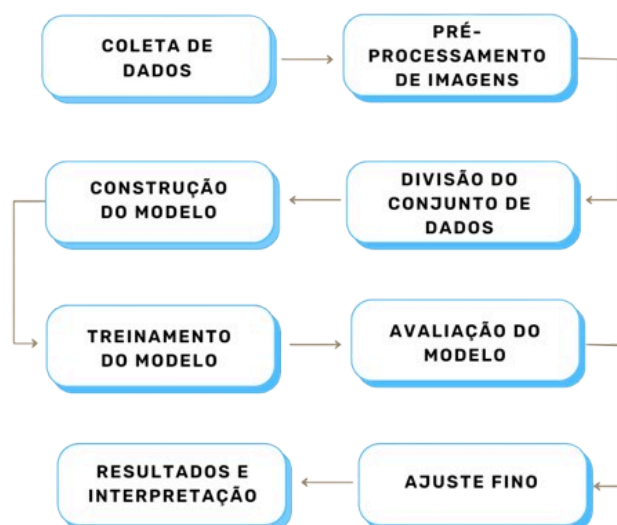
4 METODOLOGIA

Nesta seção, será fornecida uma descrição detalhada do conjunto de dados utilizado, juntamente com uma explicação dos métodos aplicados às imagens desse conjunto de dados.

Este artigo científico, que está baseado na metodologia de Lakatos e Marconi (LAKATOS; MARCONI, 2017), utiliza dois tipos principais de pesquisa: a pesquisa aplicada e a pesquisa quantitativa. A pesquisa aplicada se manifesta na busca por uma solução prática para a classificação automática de doenças respiratórias a partir de imagens de raio-X, com potencial impacto direto na área da saúde. Já a pesquisa quantitativa é evidenciada pelo uso de dados numéricos e estatísticos, especialmente na fase de treinamento e teste das redes neurais, onde se avalia a acurácia, precisão e outras métricas relevantes para a performance do modelo.

Além disso, o estudo se apoia em uma metodologia experimental, utilizando técnicas de *machine learning* e *deep learning* para testar hipóteses sobre a eficácia de diferentes arquiteturas de redes neurais na tarefa proposta.

Figura 4 - Desenho metodológico da pesquisa



Fonte: elaborado pelo autor.

A Figura 4 apresenta o Desenho metodológico da pesquisa, que perpassa pelas seguintes etapas:

- **Coleta de Dados:** momento em que ocorre a pesquisa das imagens de raio-X de tórax, reunindo as mais relevantes e confiáveis para a análise.
- **Pré-processamento das Imagens:**
 - **Redimensionamento:** ajustar o tamanho das imagens para um formato uniforme.
 - **Normalização:** alterar valores de *pixel* para melhorar a performance da rede neural.
 - **Aumento de Dados:** aplicar técnicas de *data augmentation* (rotação, *flip*, etc.) para melhorar a generalização do modelo.
- **Divisão do Conjunto de Dados:** separação das imagens em treinamento (80%) e teste (20%).
- **Construção do Modelo:**
 - **Definição da Arquitetura da Rede Neural Convolutiva:** seleção de camadas convolucionais, *pooling*, completamente conectadas, etc.
 - **Hiperparâmetros:** definição de parâmetros como número de camadas, taxa de aprendizado, número de épocas, tamanho do *batch*.
- **Treinamento do Modelo:** ajusta-se os pesos da rede neural com base nos dados de treinamento e monitora-se a precisão e erro nas fases de treinamento e teste.
- **Avaliação do Modelo:** testa-se o modelo usando o conjunto de teste e obtém-se as métricas de avaliação, como acurácia, sensibilidade, especificidade, matriz de confusão.
- **Ajuste Fino:** faz-se o ajuste de hiperparâmetros ou arquitetura para melhorar o desempenho do modelo.
- **Resultados e Interpretação:** realiza-se a análise dos resultados obtidos e compara-os com abordagens anteriores.

Tais etapas auxiliaram na condução da pesquisa e dos experimentos, servindo como uma *timeline* para organização da pesquisa.

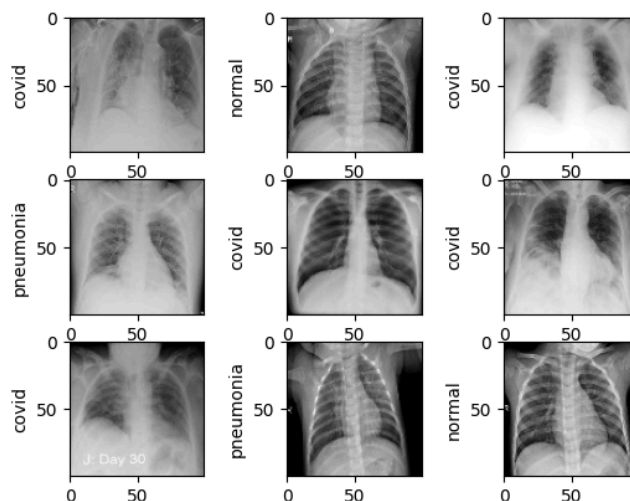
4.1 Dataset das imagens de raio-X

Para treinar e avaliar a rede neural, é necessário o acesso a um banco de dados contendo imagens de raio-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 (Shastri, 2022) (Kumar, 2022) conforme Figura 5, distribuídas da seguinte forma:

- COVID: 1.626 imagens (31.1%)
- NORMAL: 1.802 imagens (34.5%)
- PNEUMONIA: 1.800 imagens (34.4%)

Figura 5 - Amostra de imagens de tórax do dataset



Fonte: Shastri (2022)

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 classificação de casos de COVID-19, pneumonia e normais usando técnicas de aprendizado de máquina.

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

4.2 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.

4.3 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, baseado na técnica *train-test split* (divisão treino-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 (Thomas, 2017). 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 aleatoriamente as respectivas imagens em uma proporção de 80% e 20%, respectivamente, para reduzir o *overfitting*. 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.

4.4 Aplicação da modelagem

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

Conforme a recomendação da UNESCO sobre Ciência Aberta (Unesco, 2022), ao disponibilizar o código de um trabalho publicamente, o pesquisador amplia o acesso ao conhecimento científico e alinha-se à transparência de dados, metodologias e código, permitindo que outros validem, repliquem ou ampliem suas descobertas. Desta forma, o link do *dataset* das imagens de raio-X e *notebooks* aqui construídos estão no Apêndice 1 - Link para o *dataset* das imagens de raio-X e Apêndice 2 - Link para os *notebooks* dos modelos.

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 100x100, as análises foram feitas em 5 épocas, com *batch size* de 32 e *random state* de 42.

5 RESULTADOS E DISCUSSÕES

Nesta seção, serão apresentados os resultados das modelagens de cada tipo de arquitetura anteriormente mencionada, aplicadas às imagens de raio-X de tórax humano. Além de revelar medidas estatísticas e gráficos gerados a partir da execução desses modelos de redes neurais convolucionais.

5.1 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 *MaxPooling* bidimensionais e uma camada densa na saída, com funções de ativação *ReLU* e *softmax*.

No treinamento do modelo utilizou-se o método estocástico Adam (Kingma *et al.*, 2015), 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 definiu-se os hiperparâmetros presentes no Quadro 1.

Quadro 1 - Hiperparâmetros do modelo AlexNet.

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

Fonte: elaborado pelo autor.

Após as cinco épocas de treinamento, chegou-se aos valores de acurácia e perdas mostrados no Quadro 2.

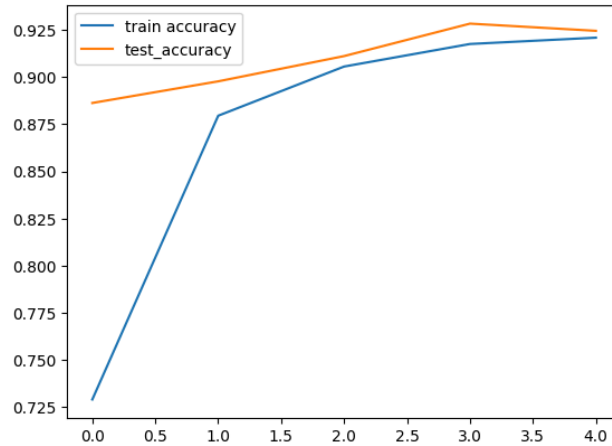
Quadro 2 - Acurácias e perdas em treino e teste do modelo AlexNet.

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

Fonte: elaborado pelo autor.

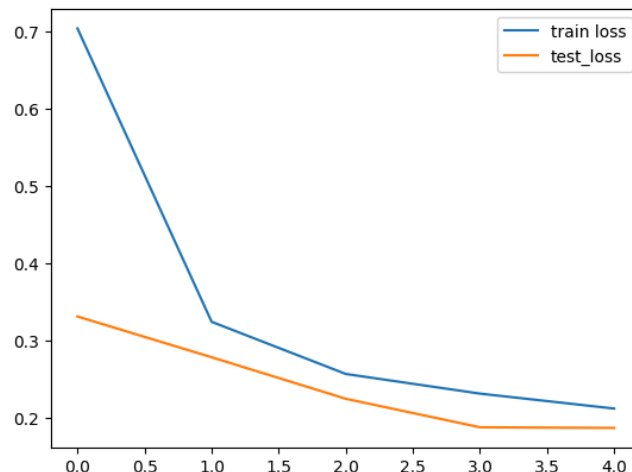
Ainda, na Figura 6 e na Figura 7 pode-se constatar a evolução desses elementos graficamente.

Figura 6 - Evolução da acurácia nos datasets de treino e teste (AlexNet)



Fonte: elaborado pelo autor.

Figura 7 - Evolução da perda nos datasets de treino e teste (AlexNet)

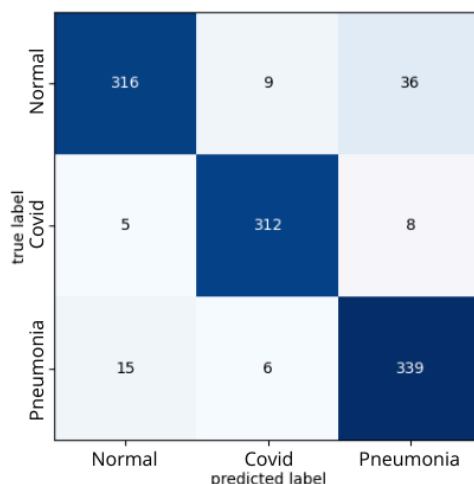


Fonte: elaborado pelo autor.

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

Na Figura 8 está a Matriz de Confusão desta modelagem em que, as linhas representam as classes verdadeiras e as colunas, as previsões do modelo. Nela, a classe Normal foi corretamente classificada 316 vezes, com 9 previsões incorretas como Covid e 36 como Pneumonia, o que indica uma leve confusão entre as classes. A classe Covid teve um desempenho sólido, com 312 acertos e apenas 13 erros. Para a classe Pneumonia, o modelo foi altamente preciso, com 339 previsões corretas e poucos erros (15 como Normal e 6 como Covid). Logo, percebe-se uma baixa taxa de erro, sugerindo que o modelo está bem treinado e ajustado para essa tarefa de classificação.

Figura 8 - Matriz de Confusão do modelo da rede neural AlexNet



Fonte: elaborado pelo autor.

5.2 Desempenho da rede neural VGG16

Concluída 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*. Utilizou-se a técnica de divisão treino-teste na proporção 80/20 e foram aplicadas camadas densa e plana, com funções de ativação *ReLU* e *sigmoide*.

Também foi utilizado o método estocástico Adam (Kingma *et al.*, 2015), como otimizador dos pesos da rede. Para esta rede neural foram definidos os hiperparâmetros presentes no Quadro 3.

Quadro 3 - Hiperparâmetros do modelo VGG16.

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

Fonte: elaborado pelo autor.

Após as cinco épocas de treinamento, chegou-se aos valores de acurácia e perdas mostrados no Quadro 4.

Quadro 4 - Acurácias e perdas em treino e teste do modelo VGG16.

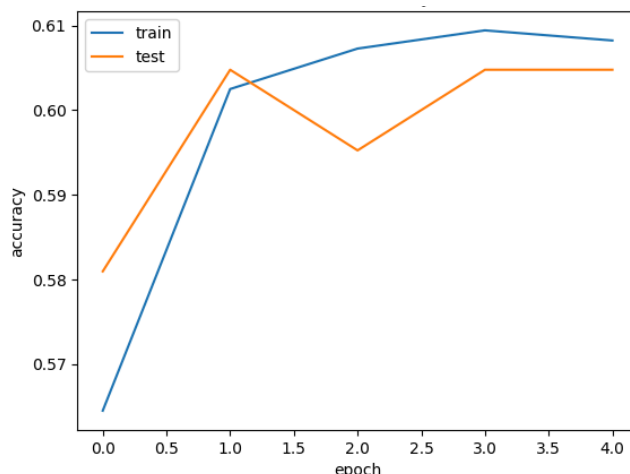
Conjunto de Dados	Acurácia Top-1 (%)	Acurácia Top-5 (%)	Perda Top-1 (%)	Perda Top-5 (%)
Treino	56.45	60.82	0.00	0.00
Teste	58.10	60.48	0.00	0.00

Fonte: elaborado pelo autor.

Ainda, na Figura 9 e na Figura 10 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.

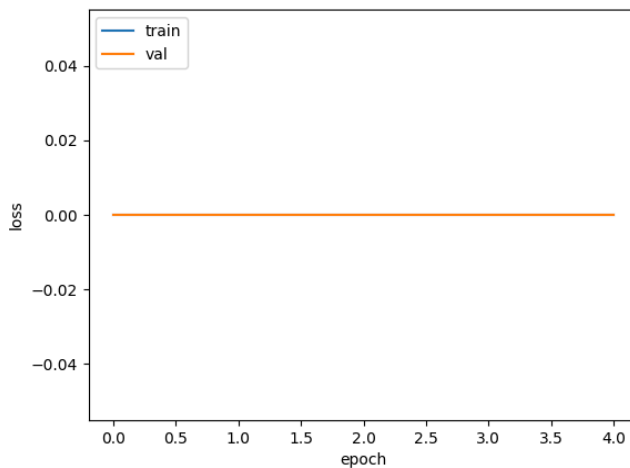
Um dos possíveis motivos dos valores das perdas terem dado zero são os valores dos hiperparâmetros, que precisam ser ajustados para evitar o *overfitting*.

Figura 9 - Evolução da acurácia nos datasets de treino e teste (VGG16)



Fonte: elaborado pelo autor.

Figura 10 - Evolução da perda nos datasets de treino e teste (VGG16)



Fonte: elaborado pelo autor.

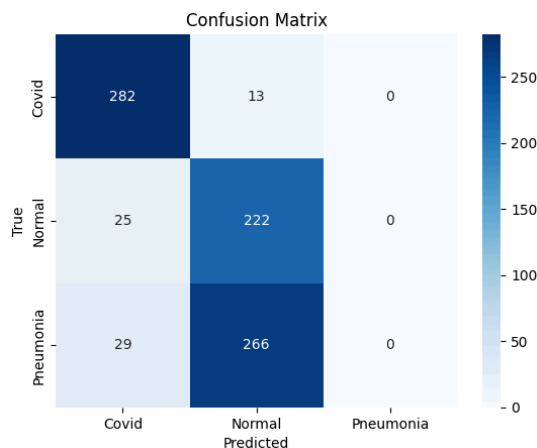
Ainda, para esta rede neural, percebe-se uma estabilização no valor da acurácia, pois uma variação de 56% na primeira época para 60% na quinta época representa pouca variação e baixo crescimento desta métrica.

Baseando-se na matriz de confusão da Figura 11, entende-se que o modelo mostrou um bom desempenho ao prever a classe Covid com 282 acertos e apenas 13 erros, confundindo-os com a classe Normal. No entanto, há uma confusão significativa entre Pneumonia e Normal, onde 266 imagens de Pneumonia foram

incorretamente classificadas como Normal. Isso indica que o modelo tem dificuldade em distinguir entre esses dois casos. Além disso, a classe Pneumonia não foi prevista corretamente nenhuma vez.

Melhorias podem ser feitas no balanceamento dos dados ou ajustes no modelo para melhorar a diferenciação entre Normal e Pneumonia.

Figura 11 - Matriz de Confusão do modelo da rede neural VGG16



Fonte: elaborado pelo autor.

5.3 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.

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* (Kingma *et al.*, 2015), para os pesos da rede convergirem mais rápido. Para esta rede neural foram definidos os hiperparâmetros presentes no Quadro 5.

Quadro 5 - Hiperparâmetros do modelo Xception.

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

Fonte: elaborado pelo autor.

Após as cinco épocas de treinamento, chegou-se aos valores de acurácia e perdas mostrados no Quadro 6.

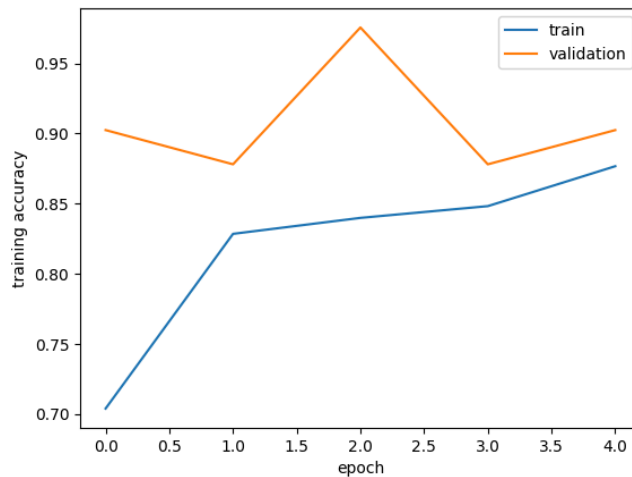
Quadro 6 - Acurácias e perdas em treino e teste do modelo *Xception*.

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

Fonte: elaborado pelo autor.

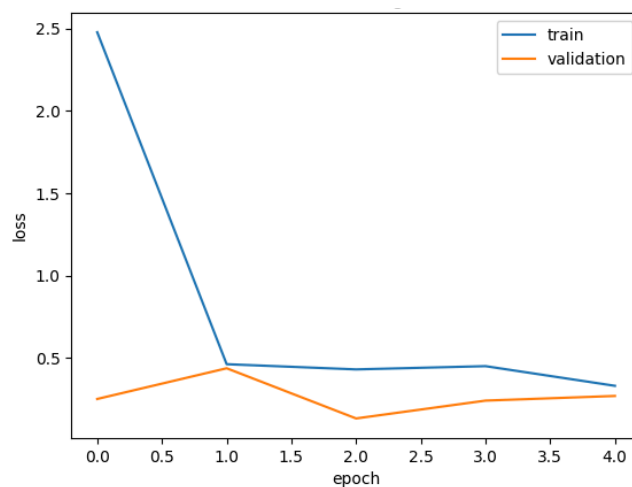
Ainda, na Figura 12 e na Figura 13 pode-se constatar a evolução desses elementos graficamente.

Figura 12 - Evolução da acurácia nos datasets de treino e teste (*Xception*)



Fonte: elaborado pelo autor.

Figura 13 - Evolução da perda nos datasets de treino e teste (*Xception*)



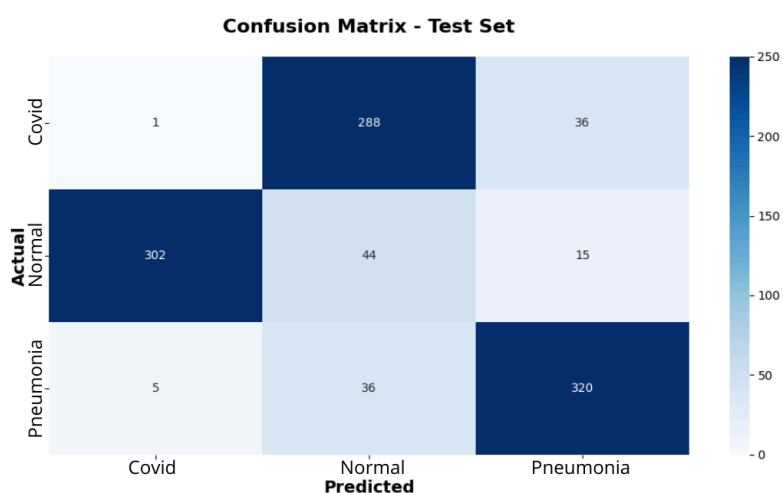
Fonte: elaborado pelo autor.

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%.

Para mais detalhes, podemos ver a Matriz de Confusão na Figura 14. Nela, a classe Covid foi mal classificada, com apenas uma predição correta e a maioria das amostras (288) sendo classificadas incorretamente como Normal. Para a classe Normal, o modelo teve 302 acertos, mas com 44 previsões erradas como Covid e 15 como Pneumonia, o que indica uma confusão significativa entre as classes. A classe Pneumonia teve um bom desempenho, com 320 acertos, mas ainda assim apresentou 36 erros, sendo esses classificados como Normal.

No geral, o modelo Xception mostra dificuldades especialmente na diferenciação entre Covid e Normal, sugerindo a necessidade de ajustes nos dados ou no treinamento para melhorar a precisão entre essas classes.

Figura 14 - Matriz de Confusão do modelo da rede neural Xception



Fonte: elaborado pelo autor.

5.4 Discussões gerais sobre os modelos

Comparando a classificação de COVID-19 e pneumonia utilizando redes neurais com abordagens tradicionais da área médica, como classificadores lineares e métodos baseados em regras, observa-se que as redes neurais convolucionais são mais um aliado na classificação dessas doenças. A capacidade dessas redes de aprender automaticamente características relevantes das imagens de raio-X permite uma classificação mais precisa e robusta.

De forma semelhante, comparando-se os modelos de redes neurais convolucionais 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%).

Contudo, mesmo o modelo VGG16 apresentando perdas em zero, ele teve o pior desempenho em acurácia e elevados erros nas classificações entre as classes.

Ainda, o *Xception* teve um bom valor de acurácia próximo a 90% e perdas em 20%, mas ele ainda apresenta dificuldade na diferenciação entre Covid e Normal. Então, é um modelo que precisa de ajuste para melhorar a classificação de doenças respiratórias por radiografias de tórax. No Quadro 7, pode-se ver um resumo das acurácias e perdas.

Quadro 7 - Resumo das acurácias e perdas em treino e teste.

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

Fonte: elaborado pelo autor.

Sendo assim, entende-se que, baseado nos modelos das redes neurais convolucionais aqui estudadas e aplicadas, o modelo da rede *AlexNet* é o mais adequado para classificar doenças respiratórias através de imagens raio-X de tórax, garantindo cerca de 92% de certeza no diagnóstico da doença.

Vale frisar que foi realizado mais de um experimento utilizando esta mesma base de dados de imagens, em divisão treino-teste de forma randômica e aqui foi mostrado o melhor desempenho dentre todos os experimentos. Além disso, foram executadas simulações com outras redes neurais convolucionais, como a ResNet, porém, devido à limitação do poder computacional, não se conseguiu finalizá-las e optou-se atuar com redes neurais mais leves computacionalmente.

6 CONSIDERAÇÕES FINAIS

Neste estudo, foi desenvolvida e avaliada uma abordagem baseada em redes neurais convolucionais (CNN) para a classificação de doenças respiratórias utilizando imagens de raio-X de tórax. A metodologia proposta demonstrou ser eficaz na classificação de condições respiratórias, como pneumonia e COVID-19, com destaque para a rede neural AlexNet, que obteve mais valor de acurácia e menor valor de erro.

O modelo treinado atingiu uma acurácia considerável no conjunto de teste, o que valida o uso de técnicas de aprendizado profundo (*deep learning*) para a análise de imagens médicas. Além disso, o uso de estratégias de aumento de dados e ajuste fino dos hiperparâmetros contribuiu para a robustez e a capacidade de generalização do modelo.

A sensibilidade e a especificidade obtidas indicam que o sistema se mostra uma ferramenta poderosa capaz de classificar corretamente tanto a presença quanto a ausência de doenças respiratórias. Assim sendo, esta ferramenta revela-se promissora para auxiliar médicos no diagnóstico clínico, com potencial de impacto real na prática médica.

6.1 LIMITAÇÕES

Apesar dos avanços alcançados nesta pesquisa, existem algumas limitações e desafios a serem considerados.

Como limitação tem-se a ausência de bases de dados mais representativas, as quais pudessem incluir outras doenças respiratórias como tuberculose e bronquite, a fim de expandir o uso das redes neurais com um diagnóstico assertivo.

Ainda, a falta do poder computacional disponível para executar os modelos acarretou na limitação da quantidade de épocas e conseqüente avanço na qualidade das métricas e a ausência da generalização para diferentes populações e condições clínicas implica em uma abordagem focada mais em determinado grupo social em detrimento de outros. Estes são aspectos que requerem atenção adicional.

6.2 TRABALHOS FUTUROS

Como recomendação de trabalhos futuros, é importante continuar aprimorando as técnicas de classificação e 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. Também, sugere-se a otimização de modelos mais leves, que possam ser implementados em ambientes com recursos computacionais limitados, como hospitais com infraestrutura tecnológica restrita.

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.

Ainda, uma maior diversidade nos dados de entrada, incluindo pacientes de diferentes faixas etárias, etnias e condições geográficas, pode melhorar a generalização do modelo em contextos clínicos reais. No geral, os resultados obtidos apontam para o potencial significativo da inteligência artificial na área da saúde, especialmente na automatização de diagnósticos médicos baseados em imagens.

AGRADECIMENTOS

Agradeço ao Instituto Federal de Pernambuco (IFPE) - Campus Jaboatão dos Guararapes pelo suporte provendo infraestrutura necessária para transcorrer com a pesquisa e pela assistência técnica.

REFERÊNCIAS BIBLIOGRÁFICAS

Biblioteca Virtual em Saúde do Ministério da Saúde. **Covid 19**. Brasil: out. 2021. Disponível em: <https://bvsmms.saude.gov.br/covid-19-2/>. Acesso em: 25 maio 2023.

CECCON, Denny. **Funções de ativação: definição, características, e quando usar cada uma**. Expert Academy, 2020. Disponível em: <https://iaexpert.academy/2020/05/25/funcoes-de-ativacao-definicao-caracteristicas-e-quando-usar-cada-uma/>. Acesso em: 25 jul 2024.

CERQUETANI, Samantha. **Raio-X do tórax: o que é, para que serve e como é feito**. Nav DASA, 19 jul. 2023. Disponível em: <https://nav.dasa.com.br/blog/raio-x-do-torax>. Acesso em: 12 jul. 2024.

CHOLLET, François. Xception: Deep Learning with Depthwise Separable Convolutions. *In: IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2017, Honolulu. Anais [...]. Honolulu: IEEE Conference on Computer Vision and Pattern Recognition, p. 1800-1807. Online.* Disponível em: <https://arxiv.org/pdf/1610.02357>. Acesso em: 25 jul. 2024.

GOODFELLOW, Ian; BENGIO, Yoshua; COURVILLE, Aaron. **Deep Learning**. Cambridge: MIT Press, 2016. Disponível em: <https://www.deeplearningbook.org>. Acesso em: 12 jul. 2024.

HAYKIN, Simon. **Redes Neurais: Princípios e Prática**. 2. ed. Porto Alegre: Bookman, 2001.

HE, Kaiming; ZHANG, Xiangyu; REN, Shaoqing; SUN, Jian. **Deep Residual Learning for Image Recognition**. *In: IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2016, Las Vegas. Anais [...]. Las Vegas: IEEE Conference on Computer Vision and Pattern Recognition, p. 770-778, DOI: 10.1109/CVPR.2016.90. Online.* Disponível em: https://www.cv-foundation.org/openaccess/content_cvpr_2016/papers/He_Deep_Residual_Learning_CVPR_2016_paper.pdf. Acesso em: 10 jul. 2024.

Keras. **Keras applications**. <https://keras.io/applications/>. Acesso em: 28 ago. 2024.

KINGMA, Diederik P.; BA, Jimmy Lei. Adam: A method for Stochastic Optimization. *In: 3rd International Conference on Learning Representations (ICLR 2015), 2015, San Diego. Anais [...]. San Diego: International Conference on Learning Representations. Online.* Disponível em: <https://arxiv.org/pdf/1412.6980.pdf>. Acesso em: 31 jul 2024.

KRIZHEVSKY, Alex; SUTSKEVER, Ilya; HINTON, Geoffrey. ImageNet classification with deep convolutional neural networks. *In: 26th Annual Conference on Neural Information Processing Systems, 2012, Lake Tahoe, Nevada. Anais [...]. Lake Tahoe, Nevada: Advances in Neural Information Processing Systems 25, p. 1097-1105. Online.* Disponível em:

https://proceedings.neurips.cc/paper_files/paper/2012/file/c399862d3b9d6b76c8436e924a68c45b-Paper.pdf. Acesso em: 31 ago. 2024.

KUMAR, Sachin; SHASTRI, Sourabh; MAHAJAN, Shilpa *et al.* LiteCovidNet: A lightweight deep neural network model for detection of COVID-19 using X-ray images. **International Journal of Imaging Systems and Technology**, Boston, v. 32, n. 5, p. 1464-1480, set. 2022. Disponível em: <https://doi.org/10.1002/ima.22770>. Acesso em: 28 ago. 2024.

LAKATOS, Eva Maria; MARCONI, Marina de Andrade. **Fundamentos de metodologia científica**. 8. ed. São Paulo: Atlas, 2017.

MARCOS, Marcelo. Família aponta que possível erro de diagnóstico na UPA pode ter levado mulher à morte em Francisco Beltrão. **Grupo RBJ de Comunicação**, Francisco Beltrão, 30 dez. 2022. Disponível em: <https://rbj.com.br/familia-aponta-que-possivel-erro-de-diagnostico-na-upa-pode-ter-levado-mulher-a-morte-em-francisco-beltrao/>. Acesso em: 27 maio 2023.

RUSSAKOVSKY, Olga *et al.* Imagenet large scale visual recognition challenge. **International Journal of Computer Vision**, v. 115, n. 3, p. 211-252, dez. 2015. Disponível em: <https://arxiv.org/pdf/1409.0575>. Acesso em: 24 jul. 2024.

SHASTRI, Sourabh; KANSAL, Isha; KUMAR, Sachin *et al.* CheXImageNet: a novel architecture for accurate classification of Covid-19 with chest x-ray digital images using deep convolutional neural networks. **Health and Technology**, v. 12, n. 1, p. 193-204, jan. 2022. Disponível em: <https://doi.org/10.1007/s12553-021-00630-x>. Acesso em: 28 ago. 2024.

SIMONYAN, Karen; ZISSERMAN, Andrew. Very Deep Convolutional Networks for Large-Scale Image Recognition. *In*: 3rd International Conference on Learning Representations (ICLR 2015), 2015, San Diego. **Anais [...]**. San Diego: International Conference on Learning Representations. *Online*. Disponível em: <https://arxiv.org/pdf/1409.1556>. Acesso em: 10 jul. 2024.

SZEGEDY, Christian *et al.* Going deeper with convolutions. *In*: IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2015, Boston. **Anais [...]**. Boston: IEEE Conference on Computer Vision and Pattern Recognition, p. 1-9, doi: 10.1109/CVPR.2015.7298594. Disponível em: <https://arxiv.org/abs/1409.4842>. Acesso em: 09 jul. 2024.

THOMAS, R. **How (and why) to create a good validation set**. Fast.ai, 2017. Disponível em: <https://www.fast.ai/posts/2017-11-13-validation-sets.html>. Acesso em: 28 ago. 2024.

TSANG, S.-H. **Review: Xception - with depthwise separable convolution, better than Inception-v3 (Image Classification)**. Towards Data Science, 2018. Disponível em: <https://towardsdatascience.com/review-xception-with-depthwise-separable-convolution-better-than-inception-v3-image-dc967dd42568>. Acesso em: 28 ago. 2024.

UNESCO. **Recomendação da UNESCO sobre Ciência Aberta**. Tradução para o português. Paris: UNESCO, 2022. Disponível em: https://unesdoc.unesco.org/ark:/48223/pf0000379949_por. Acesso em: 10 set. 2024.

VARELLA, Drauzio. **Pneumonia**. Biblioteca Virtual em Saúde do Ministério da Saúde. Brasil: maio 2011. Disponível em: <https://bvsmms.saude.gov.br/pneumonia-5/>. Acesso em: 25 maio 2023.

APÊNDICE

1. Link para o *dataset* das imagens de raio-X:
Google Drive: <https://lnkd.in/dZpMfSBa>
2. Link para os *notebooks* dos modelos:
Github: https://github.com/AldenisFranca/cnn_artigo_tcc