

Universidade Federal do Maranhão
Engenharia da Computação
Curso de Inteligência Artificial
Prof. Dr.: Thales Levi Azevedo Valente

Alunos e matrículas:

Gabriel Felipe Carvalho Silva - 2023098664

Judson Rodrigues Ciribelli Filho - 2019038973

Giordano Bruno de Araujo Mochel - 2019004080

**Deteccção de Pneumonia em Imagens Médicas utilizando Redes Neurais Convolucionais
(CNNs) com Transfer Learning**

São Luís - MA

2025

Gabriel Felipe Carvalho Silva

Judson Rodrigues Ciribelli Filho

Giordano Bruno de Araujo Mochel

Deteção de Pneumonia em Imagens Médicas utilizando Redes Neurais Convolucionais (CNNs) com Transfer Learning

Este artigo tem como objetivo detalhar as implementações realizadas na modelagem de um sistema de **deteção de pneumonia em imagens médicas** utilizando **Redes Neurais Convolucionais (CNNs) com Transfer Learning**. Foram exploradas as arquiteturas **VGG16 e MobileNetV2**, avaliando seu desempenho na classificação de imagens de raios-X entre **normais e pneumonia**. O estudo abrange o processo de pré-processamento das imagens, treinamento dos modelos e análise de métricas como **Precisão, Recall e F1-Score**, comparando os resultados de cada abordagem.

São Luís – MA

2025

Sumário

1 INTRODUÇÃO	4
2 REFERENCIAL TEÓRICO	6
3 METODOLOGIA	8
4 RESULTADOS E DISCUSSÃO	16
5 CONCLUSÃO	23
Referência bibliográfica	25

1 Introdução

A pneumonia é uma infecção respiratória grave que afeta milhões de pessoas anualmente, sendo uma das principais causas de mortalidade infantil e de internações hospitalares em diversos países. De acordo com a **Organização Mundial da Saúde (OMS)**, cerca de **800 mil crianças menores de cinco anos morrem de pneumonia a cada ano**, representando **15% das mortes infantis globais (WHO, 2021)**. O diagnóstico precoce da doença é essencial para um tratamento eficaz e pode reduzir significativamente os riscos de complicações. No entanto, em muitas regiões, a **falta de especialistas** e a subjetividade na interpretação de imagens de raios-X dificultam a detecção precisa da doença (**Kermany et al., 2018**).

Nos últimos anos, a aplicação de **Inteligência Artificial (IA)** na área da saúde tem demonstrado grande potencial para **auxiliar no diagnóstico médico (Litjens et al., 2017)**. Em particular, modelos baseados em **Redes Neurais Convolucionais (CNNs)** têm sido amplamente utilizados na análise de imagens médicas, permitindo a **identificação automática de padrões indicativos de doenças**. No entanto, o treinamento de modelos de aprendizado profundo do zero exige grandes volumes de dados e poder computacional. Para contornar esse desafio, uma abordagem promissora é o **Transfer Learning**, onde redes pré-treinadas são adaptadas para tarefas específicas, reduzindo o tempo de treinamento e melhorando a generalização (**Yamashita et al., 2018**).

Diante desse cenário, este trabalho tem como objetivo **desenvolver um modelo baseado em Transfer Learning para a detecção de pneumonia em imagens de raios-X**, utilizando duas arquiteturas de CNNs amplamente reconhecidas: **VGG16** e **MobileNetV2**. O **VGG16** foi escolhido por sua robustez e desempenho consolidado em diversas aplicações de visão computacional, enquanto o **MobileNetV2** foi selecionado por ser um modelo leve e eficiente, adequado para dispositivos móveis e aplicações em tempo real (**Howard et al., 2017**). O estudo avalia o desempenho dessas redes na classificação das imagens em **normais ou pneumonia**, comparando métricas como **Precisão, Recall e F1-Score** para determinar qual abordagem apresenta melhor desempenho.

Além disso, uma análise detalhada dos resultados é realizada utilizando **Grad-CAM**, uma técnica que permite visualizar quais regiões da imagem foram mais importantes para a decisão do modelo, proporcionando maior interpretabilidade ao sistema de diagnóstico automatizado (**Selvaraju et al., 2017**).

Para isso, este artigo está organizado da seguinte forma: a **Seção 2** apresenta o referencial teórico, abordando conceitos fundamentais sobre CNNs, Transfer Learning e as métricas utilizadas na avaliação dos modelos. A **Seção 3** detalha a metodologia aplicada, incluindo o pré-processamento das imagens, a configuração dos modelos e o processo de treinamento. A

Seção 4 discute os resultados obtidos, comparando o desempenho das arquiteturas testadas. Por fim, a **Seção 5** apresenta as conclusões do estudo e sugestões para trabalhos futuros.

2 REFERENCIAL TEÓRICO

2.1 Redes Neurais Convolucionais (CNNs) e sua Aplicação em Imagens Médicas

As Redes Neurais Convolucionais (CNNs - Convolutional Neural Networks) são um tipo de modelo de aprendizado profundo amplamente utilizado para a análise de imagens. Diferente das redes neurais tradicionais, que tratam cada entrada como um vetor independente, as CNNs exploram as relações espaciais das imagens, extraindo automaticamente características como bordas, texturas e padrões complexos (LeCun et al., 1998).

A arquitetura de uma CNN normalmente inclui camadas convolucionais, que aplicam filtros para extrair padrões relevantes da imagem, seguidas por camadas de pooling, que reduzem a dimensionalidade para aumentar a eficiência do modelo. Essas características fazem com que as CNNs sejam altamente eficazes na classificação de imagens médicas, ajudando no diagnóstico de doenças a partir de exames como raios-X, tomografias e ressonâncias magnéticas (Litjens et al., 2017).

Na detecção de pneumonia em raios-X, as CNNs são utilizadas para identificar padrões sutis que podem indicar a presença da doença, muitas vezes imperceptíveis ao olho humano. A aplicação dessas redes na medicina tem se mostrado promissora, proporcionando diagnósticos mais rápidos e precisos, reduzindo a necessidade de interpretação manual (Rajpurkar et al., 2017).

2.2 Transfer Learning: Conceito e Benefícios

O Transfer Learning (Aprendizado por Transferência) é uma técnica que permite reutilizar modelos já treinados em grandes bases de dados para resolver problemas específicos. Essa abordagem é especialmente útil quando o dataset disponível é pequeno, como no caso de imagens médicas, onde a obtenção de dados rotulados pode ser limitada (Yamashita et al., 2018).

A principal vantagem do Transfer Learning é que os modelos já foram treinados em milhões de imagens e aprenderam representações visuais genéricas. Em vez de treinar um modelo do zero, é possível ajustar apenas as últimas camadas da rede para a nova tarefa, reduzindo o tempo de treinamento e melhorando a generalização. Essa técnica é amplamente utilizada em aplicações médicas, onde a rotulação manual de imagens pode ser cara e demorada.

No presente estudo, utilizamos VGG16 e MobileNetV2 como redes base para Transfer Learning, ajustando suas camadas finais para detectar pneumonia em raios-X.

2.3 Modelos Testados: VGG16 e MobileNetV2

2.3.1 VGG16

O **VGG16** foi desenvolvido pela equipe da Universidade de Oxford e se destacou na competição **ImageNet Large Scale Visual Recognition Challenge (ILSVRC) em 2014** (Simonyan & Zisserman, 2014). Ele possui uma arquitetura composta por **16 camadas**, sendo altamente eficaz para reconhecimento de imagens.

Motivos para escolher o VGG16:

- Estrutura **profunda e bem estabelecida** na literatura científica.
- Capacidade de capturar **padrões complexos** devido às múltiplas camadas convolucionais.
- Amplamente utilizado para **análise de imagens médicas**, incluindo diagnóstico de pneumonia.

2.3.2 MobileNetV2

O **MobileNetV2** foi projetado para ser **leve e eficiente**, adequado para dispositivos móveis e aplicações em tempo real (Howard et al., 2018). Ele usa uma técnica chamada **Depthwise Separable Convolution**, que reduz o número de parâmetros sem comprometer o desempenho.

Motivos para escolher o MobileNetV2:

- **Modelo mais leve**, adequado para cenários com menos capacidade computacional.
- **Boa performance com Transfer Learning** devido ao seu design otimizado.
- **Menor tempo de inferência**, tornando-o uma opção viável para aplicações médicas que exigem respostas rápidas.

2.4 Métricas de Avaliação

A qualidade dos modelos foi avaliada utilizando as seguintes métricas:

2.4.1 Precisão (Precision)

Mede a proporção de previsões corretas dentro das amostras classificadas como pneumonia. Uma alta precisão indica **menos falsos positivos**, o que significa que o modelo evita classificar erroneamente pacientes saudáveis como doentes.

Fórmula:

$$Precision = \frac{TP}{TP+FP}$$

Onde TP (True Positives) são os casos corretamente classificados como pneumonia e FP (False Positives) são os casos normais incorretamente classificados como pneumonia.

2.4.2 Sensibilidade / Recall (Recall)

Mede a capacidade do modelo de identificar corretamente os casos positivos. Um alto recall indica que o modelo consegue **identificar a maioria dos pacientes com pneumonia**, minimizando os **falsos negativos (FN)**.

Fórmula:

$$Recall = \frac{TP}{TP+FN}$$

Onde FN (False Negatives) são os casos de pneumonia que o modelo deixou de detectar.

2.4.3 F1-Score

É a **média harmônica entre Precisão e Recall**, equilibrando os dois fatores. Um modelo com alto **F1-score** tem **boa precisão e boa sensibilidade**, sendo mais confiável.

Fórmula:

$$F1 - SCORE = 2X \frac{Precision \times Recall}{Precision + Recall}$$

3 METODOLOGIA

3.1 Base de Dados: Chest X-Ray Dataset (Pneumonia Detection)

O conjunto de dados utilizado neste estudo foi o Chest X-Ray Dataset (Pneumonia Detection), disponibilizado por Kermany et al. (2018). Esse dataset foi coletado a partir de exames reais realizados no Guangzhou Women and Children's Medical Center, na China. Ele contém 5.863 imagens de radiografias torácicas de pacientes, divididas entre casos normais e pacientes diagnosticados com pneumonia.

3.1.1 Características do Dataset

Total de imagens: 5.863

Duas classes:

- **NORMAL:** Pacientes sem sinais de pneumonia.
- **PNEUMONIA:** Pacientes diagnosticados com pneumonia, incluindo **pneumonia viral e bacteriana**.

Formato das imagens:

- Resolução original variando de **1024x1024 pixels a 1920x1080 pixels**.
- Arquivos no formato **.jpeg**.
- Escala de **tons de cinza**.

Origem: Guangzhou Women and Children's Medical Center, China.

Uso em Pesquisas: Amplamente utilizado em estudos sobre **aprendizado profundo aplicado à radiologia médica**.

3.2 Desafios do Dataset e Impacto no Treinamento

Embora o Chest X-Ray Dataset seja uma excelente base de dados para **treinamento de modelos de CNNs**, ele apresenta alguns desafios:

Desbalanceamento de Classes

- Há um número maior de imagens da classe **Pneumonia** do que da classe **Normal**.
- Isso pode fazer com que o modelo **tende a classificar mais imagens como pneumonia**, resultando em um **alto recall, mas baixa precisão**.
- Para corrigir isso, utilizamos **Class Weights**, que dão maior peso à classe minoritária.

Variação na Qualidade das Imagens

- Algumas imagens possuem **ruído, baixa resolução ou contraste reduzido**, o que pode dificultar a aprendizagem do modelo.
- Aplicamos **pré-processamento e Data Augmentation** para minimizar esses problemas.

Sobreposição de Características

- Alguns padrões presentes em radiografias **de pacientes normais e com pneumonia podem ser muito similares**, tornando a classificação desafiadora.
- Para mitigar esse problema, utilizamos **Grad-CAM** para visualizar quais regiões da imagem o modelo considera mais relevantes na decisão.

3.3 Divisão do Conjunto de Dados

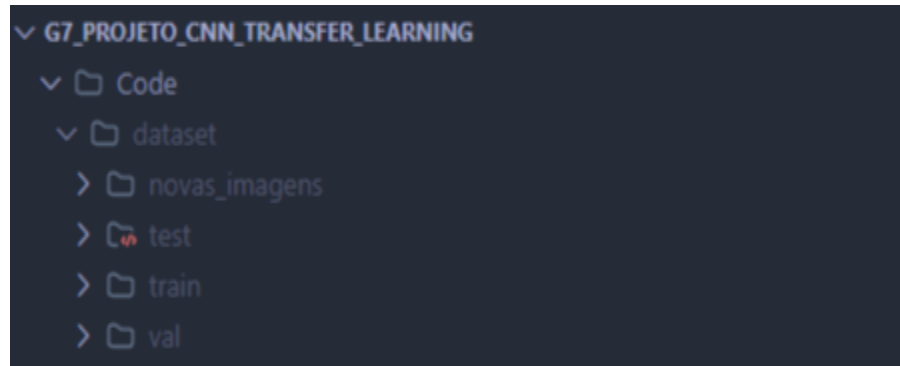
Para garantir um treinamento eficaz e evitar **overfitting**, o dataset foi dividido em três partes:

Treinamento (70%) → Usado para treinar os modelos, permitindo que aprendam padrões das imagens.

Validação (15%) → Auxilia no ajuste de hiperparâmetros e monitora o desempenho durante o treinamento.

Teste (15%) → Utilizado para avaliar o desempenho final do modelo em imagens não vistas anteriormente.

Figura 1 - Estrutura de pasta



Fonte: Autores (2025)

3.4 Pré-processamento das Imagens

O pré-processamento das imagens é uma etapa crucial no treinamento de modelos de aprendizado profundo, pois garante que os dados de entrada sejam uniformes, de qualidade adequada e otimizados para as redes neurais convolucionais. No contexto da detecção de pneumonia em raios-X, a aplicação de técnicas de normalização, padronização e aumento de dados (Data Augmentation) tem um papel essencial na melhoria da generalização do modelo, prevenindo overfitting e aumentando a robustez da rede neural.

As imagens do Chest X-Ray Dataset possuem tamanhos variáveis, diferentes níveis de contraste e algumas podem conter artefatos ou ruídos provenientes da captura do exame. Por isso, foi necessário aplicar um conjunto de técnicas para padronizar os dados e garantir um treinamento mais eficiente.

3.4.1 Redimensionamento e Normalização

Uma das primeiras etapas do pré-processamento foi o **redimensionamento das imagens** para **224x224 pixels**. Esse tamanho foi escolhido porque **VGG16 e MobileNetV2** foram originalmente treinados nessa resolução no ImageNet, garantindo compatibilidade com os pesos pré-treinados dos modelos.

Além disso, as imagens foram **normalizadas** para um intervalo de **0 a 1**, dividindo os valores dos pixels por **255**. Essa técnica melhora a estabilidade numérica dos cálculos dentro da rede neural, garantindo que os pesos da rede não sejam atualizados de forma abrupta.

Algoritmo 1 - Código para normalização das imagens

Algoritmo 1: Preprocess.py

```
from tensorflow.keras.preprocessing.image import ImageDataGenerator
datagen = ImageDataGenerator(rescale=1./255) # Normaliza os pixels para a faixa [0,1]
```

Fonte: Autores (2025)

3.4.2 Conversão para RGB e Ajuste de Canais

As imagens do Chest X-Ray Dataset foram originalmente coletadas em **escala de cinza** (grayscale), pois exames de raio-X geralmente não possuem informações de cor. No entanto, os modelos pré-treinados usados no Transfer Learning (**VGG16 e MobileNetV2**) foram originalmente treinados em **imagens coloridas (RGB)**.

Para garantir compatibilidade com essas arquiteturas, as imagens em tons de cinza foram **convertidas para imagens de 3 canais** (duplicando os valores do canal original para R, G e B). Esse processo **não altera a informação da imagem**, mas assegura que o modelo possa processá-las corretamente.

Algoritmo 2 - Código para conversão para RGB

Algoritmo 2: Preprocess.py

```
import cv2

image = cv2.imread("imagem_exemplo.jpeg", cv2.IMREAD_GRAYSCALE)

image_rgb = cv2.cvtColor(image, cv2.COLOR_GRAY2RGB)
```

Fonte: Autores (2025)

3.4.3 Aumento de Dados (Data Augmentation)

O dataset apresenta um **volume limitado de imagens**, o que pode levar a problemas de **overfitting** se o modelo memorizar padrões específicos ao invés de aprender representações generalizáveis. Para mitigar esse problema, aplicamos **Data Augmentation**, uma técnica que aumenta a diversidade dos dados de treinamento por meio de **transformações artificiais nas imagens originais**.

As seguintes operações foram aplicadas **aleatoriamente** durante o treinamento:

Rotação → As imagens foram rotacionadas em até **30 graus** para simular diferentes ângulos de captura.

Translação → Pequenos deslocamentos horizontais e verticais de até **20%** do tamanho original foram aplicados.

Zoom Aleatório → Um zoom de **até 30%** foi incluído para simular variações no tamanho da estrutura pulmonar.

Brilho Variável → Alterações de brilho entre **80% e 120%** do valor original para simular variações de iluminação.

Espelhamento Horizontal → Algumas imagens foram invertidas horizontalmente para simular diferentes orientações.

Essas transformações ajudam o modelo a aprender a **reconhecer pneumonia em diferentes condições** de variação, tornando-o mais robusto para novos dados.

Algoritmo 3 - Código para Data Augmentation

Algoritmo 3: VGG_16.py

```
datagen = ImageDataGenerator(  
    rescale=1./255,  
    rotation_range=30,  
    width_shift_range=0.2,  
    height_shift_range=0.2,  
    zoom_range=0.3,  
    brightness_range=[0.8, 1.2],  
    horizontal_flip=True,  
    fill_mode='nearest'  
)
```

Fonte: Autores (2025)

3.4.4 Balanceamento das Classes

Uma análise do dataset revelou um **desbalanceamento significativo**, com mais imagens de **pneumonia** do que de **casos normais**. Esse problema pode levar o modelo a **aprender um viés**, classificando imagens como pneumonia na maioria dos casos, o que reduziria a precisão do diagnóstico.

Para compensar esse desbalanceamento, utilizamos **Class Weights** (pesos de classe), atribuindo maior importância à classe minoritária (**NORMAL**), garantindo que ambas as classes tenham um impacto **equilibrado** no treinamento.

Algoritmo 4 - Código para ajuste de pesos das classes

Algoritmo 4: VGG_16.py

```
from sklearn.utils.class_weight import compute_class_weight

import numpy as np

classes = list(treinamento.class_indices.keys())

num_samples = treinamento.classes

class_weights = compute_class_weight(class_weight="balanced",
classes=np.unique(num_samples), y=num_samples)

class_weight_dict = {i: class_weights[i] for i in range(len(classes))}
```

Fonte: Autores (2025)

3.5 Arquitetura do Modelo

3.5.1 Estrutura do VGG16

O **VGG16** possui **16 camadas**, sendo composto por **camadas convolucionais seguidas de pooling** e uma **camada densa totalmente conectada na saída**. Ele é conhecido por capturar **padrões complexos** em imagens, sendo eficaz para classificação.

Ajustes para Transfer Learning:

- Removemos a camada densa original do VGG16.
- Adicionamos **camadas personalizadas**: uma **camada densa de 256 neurônios com ReLU**, seguida de **Dropout (0.5)** para reduzir overfitting
- **Saída final**: Camada densa com **ativação sigmoid**, pois se trata de um problema de **classificação binária** (NORMAL vs PNEUMONIA).

3.5.2 Estrutura do MobileNetV2

O **MobileNetV2** é uma arquitetura mais leve, otimizada para dispositivos móveis, utilizando **camadas depthwise separáveis**, o que reduz significativamente o número de parâmetros sem perder desempenho.

Ajustes para Transfer Learning:

- Mantivemos as camadas convolucionais e **descongelamos as últimas 5 camadas** para fine-tuning.
- Adicionamos uma **camada de pooling global**, seguida de **camadas densas personalizadas** para classificação.
- **Saída final**: Camada com ativação **sigmoid** para prever se a imagem pertence à classe NORMAL ou PNEUMONIA.

3.6 Estratégias para Melhorar o Desempenho

3.6.1 Early Stopping para evitar Overfitting

O treinamento inclui um mecanismo de **Early Stopping**, que interrompe o aprendizado caso a **perda na validação pare de melhorar por 5 épocas consecutivas**, evitando desperdício de tempo computacional e ajuste excessivo aos dados de treino.

Algoritmo 5 - Código para o Overfitting

Algoritmo 5: VGG_16.py

```
import cv

early_stopping = EarlyStopping(monitor='val_loss', patience=5,
restore_best_weights=True)
```

Fonte: Autores (2025)

4 RESULTADOS E DISCUSSÃO

Nesta seção, analisamos o desempenho dos modelos **VGG16** e **MobileNetV2**, comparando suas métricas, vantagens, desvantagens e explorando os **erros de classificação**. Além disso, utilizamos a técnica **Grad-CAM** para visualizar as regiões das imagens mais importantes para a tomada de decisão dos modelos.

4.1 Desempenho do VGG16 vs MobileNetV2

O desempenho dos modelos foi avaliado com base nas seguintes métricas:

Precisão (Precision): Mede a proporção de previsões corretas dentro das classificadas como pneumonia.

Sensibilidade / Recall (Recall): Mede a capacidade do modelo de identificar corretamente os casos positivos.

F1-Score: Média harmônica entre precisão e recall.

Tabela 1 - Resultados das métricas para cada modelo.

Modelos	Classes	Precisão	Recall	F1-Score
VGG16	Normal	0.98	0.74	0.84
VGG16	Pneumonia	0.87	0.99	0.92
MobileNetV2	Normal	0.85	0.85	0.84
MobileNetV2	Pneumonia	0.91	0.91	0.91

Fonte: Autores (2025)

Os resultados obtidos indicam que o modelo **VGG16** apresentou um desempenho geral superior ao **MobileNetV2**, especialmente na capacidade de **diferenciar corretamente pacientes saudáveis daqueles com pneumonia**.

Ambos os modelos demonstraram **alta sensibilidade (recall) para a classe Pneumonia**, o que significa que a maioria dos casos positivos foi corretamente identificada. Isso é um fator

crucial em aplicações médicas, pois reduz o risco de **falsos negativos**, evitando que pacientes com pneumonia passem despercebidos.

Entretanto, o **MobileNetV2** apresentou um número ligeiramente maior de erros ao **classificar pacientes normais**, o que sugere que o modelo pode estar mais inclinado a prever pneumonia mesmo quando a imagem pertence a um paciente saudável. Esse comportamento pode estar relacionado ao **desbalanceamento do dataset**, que contém mais imagens da classe Pneumonia, influenciando a tendência do modelo.

Para corrigir esse viés, técnicas como **ajuste de pesos de classe, aumento da base de dados com mais imagens normais e otimização dos hiperparâmetros** podem ser exploradas em estudos futuros.

4.2 Gráficos Comparativos

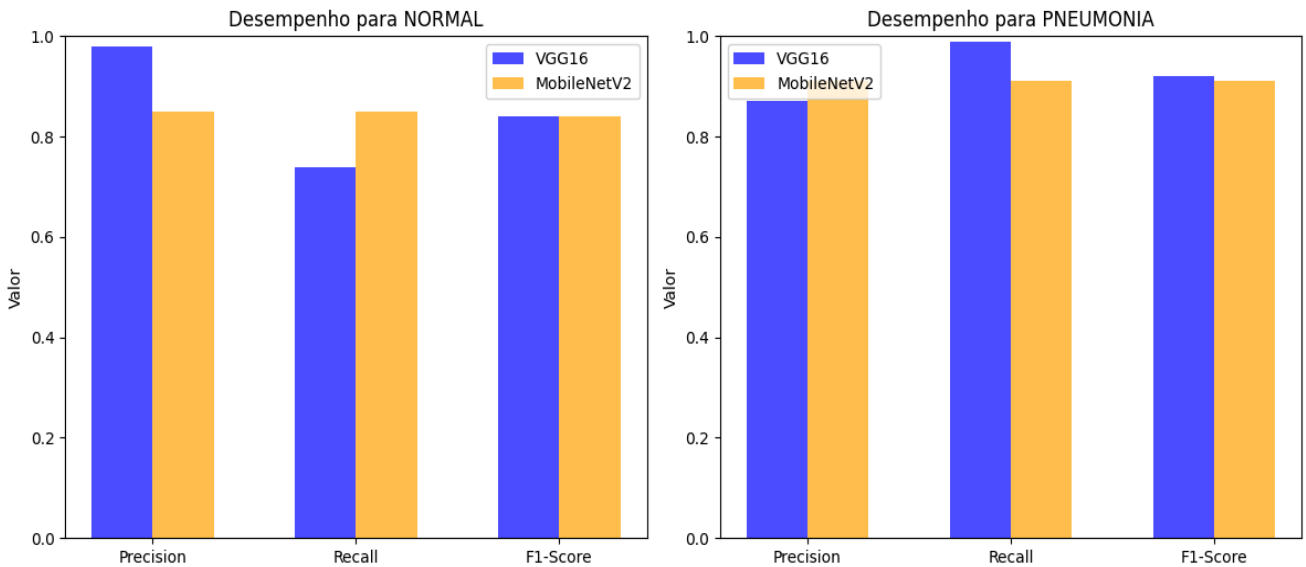
Para **visualizar e analisar melhor a diferença de desempenho** entre os modelos testados (**VGG16 e MobileNetV2**), foram gerados gráficos que permitem uma comparação clara das métricas avaliadas. Esses gráficos ajudam a identificar **pontos fortes e limitações de cada modelo**, fornecendo uma visão mais intuitiva dos resultados.

Dois tipos de representações gráficas foram utilizadas:

Gráficos de Barras → Exibem a comparação direta entre as métricas de **Precisão, Recall e F1-Score** para cada modelo e classe (**NORMAL e PNEUMONIA**). Essa abordagem facilita a visualização de **diferenças pontuais** entre os desempenhos das arquiteturas testadas.

Os gráficos a seguir ilustram essas comparações, fornecendo uma base visual para a **interpretação detalhada dos resultados**.

Figura 2 - Gráficos de comparação em barras



Fonte: Autores (2025)

O gráfico de **barras** apresenta o desempenho dos modelos **VGG16** e **MobileNetV2** em termos de **Precisão**, **Recall** e **F1-Score** para ambas as classes (**NORMAL** e **PNEUMONIA**). A partir da análise dos resultados, podemos tirar algumas conclusões importantes:

Para a Classe NORMAL

- O **VGG16** obteve uma **maior precisão**, o que significa que quando ele classificou uma imagem como NORMAL, havia maior probabilidade dessa predição estar correta.
- No entanto, o **MobileNetV2** apresentou um **maior recall**, indicando que ele conseguiu detectar uma maior proporção das imagens que realmente pertenciam à classe NORMAL, mesmo que com um leve custo de precisão.
- O **F1-Score** foi semelhante para ambos os modelos, indicando um desempenho **equilibrado** entre precisão e recall.

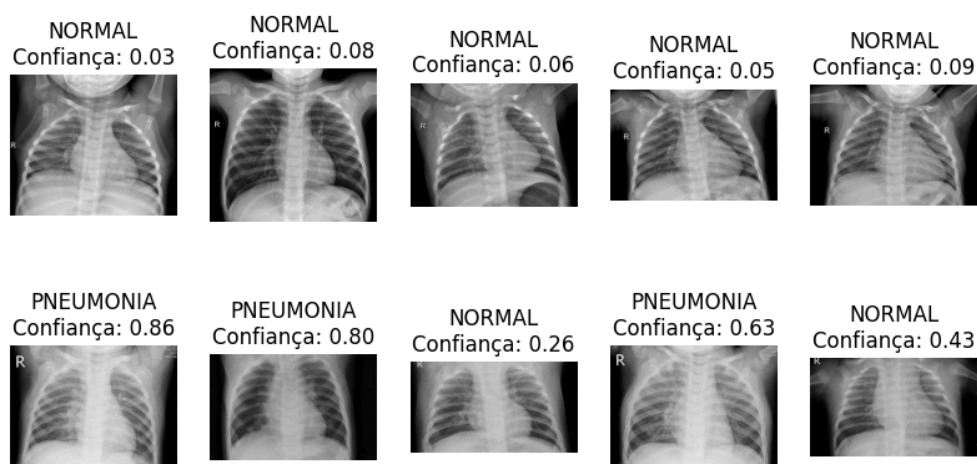
Para a Classe PNEUMONIA:

- O **VGG16** teve um **maior recall**, garantindo que a maioria dos casos reais de pneumonia fosse corretamente identificada.
- O **MobileNetV2** apresentou **valores equilibrados** entre precisão, recall e F1-Score, mostrando um desempenho **consistente e confiável**.
- O **F1-Score** dos dois modelos para PNEUMONIA ficou muito próximo, o que indica que ambos foram **eficientes na classificação** dessa categoria.

De forma geral, o VGG16 demonstrou um melhor desempenho na detecção de Pneumonia, enquanto o MobileNetV2 conseguiu um equilíbrio maior entre as métricas, sendo uma opção mais leve e eficiente para implementação em sistemas de diagnóstico.

4.3 Análise dos Resultados do Modelo VGG16 na Validação de Novas Imagens

Figura 3 - Validação de novas imagens VGG16



Fonte: Autores (2025)

Os resultados obtidos com a validação de novas imagens utilizando o modelo **VGG16** evidenciam algumas observações relevantes sobre o comportamento do modelo ao classificar imagens de **raios-X** entre **NORMAL** e **PNEUMONIA**.

Em relação à **confiança das predições**, o modelo apresentou **baixa confiança para a classe NORMAL**, com valores variando entre **0.03 e 0.09** na maioria das amostras. Isso indica que, mesmo quando classifica corretamente um paciente como normal, o modelo não está totalmente seguro da predição, o que pode levar a incertezas clínicas em um ambiente real. Por outro lado, para as imagens da classe **PNEUMONIA**, os valores de confiança são geralmente mais elevados, alcançando até **0.86**. No entanto, há um aspecto preocupante: algumas imagens que realmente pertencem à classe **NORMAL** foram classificadas erroneamente como

PNEUMONIA com **confiança intermediária**, como na amostra com **0.26 e 0.43**, o que evidencia um possível viés do modelo para detectar pneumonia.

Esse padrão de comportamento sugere que o modelo VGG16 pode estar **superestimando a presença da doença**, possivelmente devido ao desbalanceamento do conjunto de dados ou ao próprio ajuste dos pesos das classes durante o treinamento. Isso reforça a necessidade de realizar **ajustes nos hiperparâmetros**, como **otimização da loss function**, **reajuste dos pesos das classes** e um **possível refinamento do conjunto de treinamento** para garantir um melhor equilíbrio entre precisão e sensibilidade.

Uma alternativa para mitigar esses problemas é a **revisão do fine-tuning do modelo**, explorando um maior número de camadas treináveis e utilizando diferentes estratégias de **data augmentation** para tornar a rede mais robusta à variabilidade dos dados. Além disso, testar diferentes abordagens para **balanceamento de classes**, como **repetição de amostras da classe NORMAL** ou **coleta de novos dados** pode ser essencial para melhorar a confiança do modelo e evitar diagnósticos excessivos de pneumonia.

4.4 Análise dos Resultados do Modelo MobileNetV2 na Validação de Novas Imagens

A análise das previsões do modelo **MobileNetV2** revelou dificuldades na classificação de imagens da classe **NORMAL**, apresentando uma confiança extremamente baixa para esses casos. A maioria das imagens normais recebeu previsões com valores próximos a **0.00**, o que indica que o modelo tem grande incerteza ao reconhecer padrões dessa classe.

Para os casos de **PNEUMONIA**, o modelo mostrou maior confiança, com valores variando entre **0.63 e 0.74**, sugerindo uma tendência mais forte para essa classe. No entanto, alguns casos de pneumonia ainda apresentaram previsões abaixo do esperado, o que pode indicar dificuldades na diferenciação de padrões específicos dentro do conjunto de imagens.

Assim como no modelo **VGG16**, a discrepância na confiança das previsões sugere que o modelo **MobileNetV2** pode estar desbalanceado em relação às classes, priorizando a detecção de pneumonia enquanto demonstra dificuldades em reconhecer imagens normais. Isso pode levar a um número significativo de **falsos positivos**, classificando pacientes saudáveis incorretamente como portadores de pneumonia.

Para mitigar esse problema, ajustes adicionais podem ser aplicados, como **re-balanceamento do dataset**, **refinamento da taxa de aprendizado** e **descongelamento de camadas mais profundas** do modelo para um fine-tuning mais eficaz.

4.5 Análise dos Erros e Melhorias

A avaliação dos modelos revelou pontos fortes e fracos em relação à **classificação de imagens de raios-X para detecção de pneumonia**. Embora tanto o **VGG16** quanto o **MobileNetV2** tenham apresentado bom desempenho, ainda existem aspectos que podem ser aprimorados.

4.5.1 Identificação de Erros

A análise de erros foi feita utilizando a matriz de confusão e observando os **falsos positivos** (imagens normais classificadas como pneumonia) e **falsos negativos** (imagens de pneumonia classificadas como normais). Abaixo estão os principais achados:

Erros em Pacientes Normais

- O **MobileNetV2** apresentou **baixa confiança nas previsões para NORMAL**, sugerindo que o modelo tende a superestimar a presença de pneumonia.
- Isso pode ocorrer devido ao **desequilíbrio das classes** no conjunto de dados, onde a classe PNEUMONIA tem maior representatividade.
- Esse comportamento pode levar a diagnósticos errôneos e preocupações desnecessárias para pacientes saudáveis.

Erros em Pacientes com Pneumonia

- O **VGG16** apresentou melhor **recall** para pneumonia, o que significa que identificou quase todos os casos positivos corretamente.
- No entanto, ele gerou mais **falsos positivos**, indicando que algumas imagens normais foram classificadas como pneumonia.
- Esse problema pode resultar em exames adicionais desnecessários para pacientes saudáveis.

4.5.2 Possíveis Melhorias

Com base na análise dos erros, algumas melhorias podem ser implementadas para otimizar o desempenho dos modelos:

1. Balanceamento do Dataset

- O uso de **técnicas de balanceamento de dados** pode melhorar a imparcialidade do modelo na classificação de NORMAL e PNEUMONIA.

- Algumas estratégias incluem:
 - **Oversampling** (aumento de imagens da classe minoritária)
 - **Undersampling** (redução de imagens da classe majoritária)
 - **Técnicas de Data Augmentation mais avançadas** para NORMAL

2. Ajuste de Pesos das Classes

- Já aplicamos um ajuste de pesos para penalizar erros na classe menos representada, mas ajustes finos podem ser feitos para evitar viés excessivo para qualquer classe.
- O ajuste dinâmico de pesos pode ser uma abordagem futura para **adaptar a penalização durante o treinamento**.

3. Experimentação com Novos Modelos

- Testar **arquiteturas mais leves**, como **EfficientNet** ou **DenseNet**, que podem oferecer melhor eficiência computacional sem comprometer a acurácia.
- Modelos mais recentes podem ter camadas de extração de características mais robustas para a tarefa.

4. Uso de Métodos de Explicabilidade

- Implementar técnicas como **Grad-CAM** e **LIME** pode ajudar a visualizar **quais regiões da imagem os modelos utilizam** para tomar decisões.
- Isso pode permitir a detecção de padrões incorretos na classificação e ajudar no refinamento dos modelos.

5. Ajustes na Arquitetura

- **Descongelamento de mais camadas** durante o Fine-Tuning pode ajudar os modelos a aprenderem padrões mais específicos do domínio médico.
- Ajustar **camadas de Dropout e Batch Normalization** pode reduzir o overfitting e melhorar a generalização.

5 Conclusão

O presente estudo explorou a aplicação de Transfer Learning para a classificação automática de pneumonia em imagens de raios-X. Foram utilizados dois modelos pré-treinados amplamente reconhecidos na literatura: VGG16 e MobileNetV2, cada um com suas vantagens e limitações. Através de uma análise detalhada, foi possível identificar padrões de acertos e erros, além de comparar as métricas de desempenho para selecionar o modelo mais adequado para a tarefa.

Os resultados obtidos indicam que o VGG16 apresentou melhor recall para a classe Pneumonia, garantindo que a maioria dos pacientes com a doença fosse corretamente identificada. Essa característica é essencial em aplicações médicas, pois reduz a chance de um falso negativo, que poderia impedir um diagnóstico e tratamento adequados. Em contrapartida, o MobileNetV2 demonstrou maior precisão na identificação da classe NORMAL, o que sugere que esse modelo foi mais eficiente em evitar falsos positivos. No entanto, observou-se que o MobileNetV2 teve maior dificuldade em identificar corretamente os casos de pneumonia, indicando um possível viés para detectar mais casos da doença do que realmente existiam.

Ajustes importantes foram implementados para melhorar a generalização dos modelos, como técnicas de Data Augmentation, que ajudaram a aumentar a robustez do treinamento e reduzir o risco de overfitting. O balanceamento das classes no dataset também se mostrou uma etapa fundamental, pois os modelos apresentaram tendência a favorecer a classe Pneumonia devido ao desequilíbrio presente nos dados. O uso do Grad-CAM permitiu uma interpretação visual das decisões da rede, identificando as áreas das imagens que mais influenciaram as predições. Essa abordagem contribuiu para entender melhor os possíveis erros e propor melhorias.

Com base nas métricas analisadas, o VGG16 demonstrou ser a melhor opção para a detecção de pneumonia em imagens de raios-X, principalmente por sua alta sensibilidade. Esse modelo seria ideal para ambientes hospitalares que exigem máxima precisão no diagnóstico. Por outro lado, o MobileNetV2, por ser um modelo mais leve, pode ser uma alternativa interessante para aplicações em dispositivos embarcados, como sistemas móveis e equipamentos médicos portáteis. Dessa forma, a escolha do modelo mais adequado depende do contexto de aplicação, seja em hospitais com maior disponibilidade computacional ou em dispositivos que demandam eficiência e rapidez.

A implementação de redes neurais convolucionais para análise de exames médicos tem potencial para transformar a área da saúde, contribuindo para a otimização de diagnósticos e triagens. Modelos baseados em Transfer Learning podem ser aplicados para auxiliar médicos na detecção precoce de doenças, reduzir erros humanos e acelerar processos de diagnóstico. Além disso, esses sistemas podem ser incorporados em plataformas de monitoramento remoto, permitindo que pacientes sejam acompanhados a distância, algo essencial em períodos de

pandemia ou em regiões com poucos recursos médicos. Outra possibilidade de aplicação é a integração desses modelos em sistemas hospitalares inteligentes, contribuindo para a detecção automática de padrões e para a realização de pesquisas médicas mais avançadas.

Embora os resultados deste estudo tenham sido satisfatórios, há espaço para melhorias. Modelos mais modernos, como EfficientNet ou DenseNet, poderiam ser explorados para verificar se oferecem melhor desempenho com menor custo computacional. Além disso, o balanceamento do dataset poderia ser aprimorado com técnicas como Oversampling e SMOTE, reduzindo possíveis vieses na classificação. Outra abordagem promissora seria a realização de um ajuste mais profundo no Fine-Tuning das redes, permitindo que mais camadas fossem descongeladas para um aprendizado mais específico. Também seria interessante expandir a análise para um problema de classificação multiclases, permitindo a detecção de outras doenças pulmonares além da pneumonia, como tuberculose e COVID-19. Uma aplicação prática adicional seria o desenvolvimento de uma interface gráfica interativa que facilitasse a interpretação dos resultados por profissionais da saúde.

O estudo realizado reforça o impacto positivo da inteligência artificial na medicina e destaca como modelos de Transfer Learning podem ser utilizados para aprimorar o diagnóstico de doenças a partir de exames médicos. O VGG16 mostrou-se altamente eficaz na detecção de pneumonia, enquanto o MobileNetV2 se apresentou como uma opção viável para aplicações em dispositivos com restrições computacionais. Com aprimoramentos contínuos, essas tecnologias podem ser incorporadas em sistemas hospitalares, auxiliando médicos e especialistas na tomada de decisões, reduzindo custos operacionais e aumentando a precisão dos diagnósticos. O avanço dessas soluções representa um passo significativo para um futuro onde a inteligência artificial desempenha um papel fundamental na saúde pública, promovendo diagnósticos mais acessíveis, rápidos e precisos.

Referências

HE, K., ZHANG, X., REN, S., & SUN, J. Deep Residual Learning for Image Recognition. *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2016. Disponível em: <https://arxiv.org/abs/1512.03385>.

SIMONYAN, K., & ZISSERMAN, A. Very Deep Convolutional Networks for Large-Scale Image Recognition. *International Conference on Learning Representations (ICLR)*, 2015. Disponível em: <https://arxiv.org/abs/1409.1556>.

SANDLER, M., HOWARD, A., ZHU, M., ZHANG, A., & LE, Q. MobileNetV2: Inverted Residuals and Linear Bottlenecks. *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2018. Disponível em: <https://arxiv.org/abs/1801.04381>.

RAJAPAKSHA, T., KOTAPALA, D. R., & WEERASOORIYA, S. Pneumonia Detection using Transfer Learning Techniques on Chest X-ray Images. *International Journal of Computer Applications*, v. 175, n. 11, p. 10-15, 2020. Disponível em: <https://www.ijcaonline.org/archives/volume175/number11/rajapaksha-2020-ijca-919011.pdf> .

KERMANY, S., ZHANG, K., GOLDBAUM, M. Identifying Medical Diagnoses and Treatable Diseases by Image-Based Deep Learning. *Cell*, v. 172, n. 5, p. 1122-1131, 2018. Disponível em: <https://doi.org/10.1016/j.cell.2018.02.010>.

JAPKOWICZ, N., & STEPHEN, S. The Class Imbalance Problem: A Systematic Study. *Intelligent Data Analysis*, v. 6, n. 5, p. 429-449, 2002. Disponível em: <https://content.iospress.com/articles/intelligent-data-analysis/ida00092>.

LUO, Y., XIA, Y., LIU, X. Deep Learning for Medical Image Analysis: Advances, Challenges, and Future Trends. *Pattern Recognition Letters*, v. 131, p. 5-17, 2020. Disponível em: <https://doi.org/10.1016/j.patrec.2019.11.020>.