



INSTITUTO FEDERAL DE EDUCAÇÃO, CIÊNCIA E TECNOLOGIA DO PIAUÍ
CAMPUS CORRENTE
TECNOLOGIA EM ANÁLISE E DESENVOLVIMENTO DE SISTEMAS

GABRIEL ROCHA PAES DA COSTA

**SISTEMA DE APOIO À DECISÃO BASEADO EM DEEP LEARNING PARA
DIAGNÓSTICO DE PNEUMONIA EM RADIOGRAFIAS TORÁCICAS**

CORRENTE, PIAUÍ
2025

GABRIEL ROCHA PAES DA COSTA

SISTEMA DE APOIO À DECISÃO BASEADO EM DEEP LEARNING PARA
DIAGNÓSTICO DE PNEUMONIA EM RADIOGRAFIAS TORÁCICAS

Trabalho de Conclusão de Curso
apresentado como exigência parcial para
obtenção do diploma do Curso de Análise e
Desenvolvimento de Sistemas do Instituto
Federal de Educação, Ciência e Tecnologia
do Piauí, Campus Corrente.

Orientador: Prof. M^e. Carlos Estevão
Bastos Sousa

CORRENTE, PIAUÍ
2025

Dados Internacionais de Catalogação na Publicação (CIP) de acordo com ISBD

Costa, Gabriel Paes da

S837q Sistema de apoio à decisão baseado em Deep Learning para diagnóstico de pneumonia em radiografias torácicas / Gabriel Paes da Costa. - 2025.
44 p.: il. color.

Trabalho de Conclusão de Curso (Tecnologia em Análise e Desenvolvimento de Sistemas) - Instituto Federal de Educação, Ciência e Tecnologia do Piauí, Campus Corrente, 2025.

Orientador : Prof Me. Carlos Estevão Bastos Sousa

1. Análise e Desenvolvimento de Sistemas - estudo e ensino. 2. Detecção de Pneumonia. 3. Redes Neurais Convolucionais. I. Título.

CDD - 004

Elaborado por Tefischer Huanderson Soares e Sousa CRB 3/1251

RESUMO

A pneumonia é uma das principais causas de mortalidade global, especialmente em crianças e idosos, demandando diagnósticos rápidos e precisos. Este trabalho desenvolve um sistema computacional de apoio ao diagnóstico de pneumonia em radiografias de tórax utilizando *Deep Learning*. Foram avaliadas oito arquiteturas de redes neurais convolucionais pré-treinadas: *Xception*, *VGG16*, *VGG19*, *ResNet50*, *InceptionV3*, *InceptionResNetV2*, *DenseNet121* e *DenseNet169*, aplicando *Transfer Learning* e *Fine-Tuning* com diferentes combinações de otimizadores, taxas de aprendizado e tamanhos de *batch* no conjunto *Chest X-Ray Images*. O *F1-score* foi adotado como métrica principal devido ao desbalanceamento das classes, proporcionando avaliação mais equilibrada. O *InceptionResNetV2* treinado com SGD obteve melhor desempenho, alcançando *F1-score* de 96%. Foi implementada interface web em *Streamlit*. Os resultados demonstram que a solução pode auxiliar profissionais na detecção precoce de pneumonia.

Palavras-chaves: Detecção de Pneumonia; Redes Neurais Convolucionais; *Deep Learning*; Radiografia de Tórax; Transferência de Aprendizado; *Fine-tuning*; Análise de Imagens Médicas.

ABSTRACT

Pneumonia is one of the leading causes of global mortality, especially among children and the elderly, requiring rapid and accurate diagnoses. This work develops a computational system to support the diagnosis of pneumonia in chest radiographs using *Deep Learning*. Eight pre-trained convolutional neural network architectures were evaluated: *Xception*, *VGG16*, *VGG19*, *ResNet50*, *InceptionV3*, *InceptionResNetV2*, *DenseNet121*, and *DenseNet169*, applying *Transfer Learning* and *Fine-Tuning* with different combinations of optimizers, learning rates, and *batch* sizes on the *Chest X-Ray Images* dataset. The *F1-score* was adopted as the main metric due to class imbalance, providing a more balanced evaluation. *InceptionResNetV2* trained with SGD achieved the best performance, reaching an *F1-score* of 96%. A web interface was implemented using *Streamlit*. The results demonstrate that the solution can assist professionals in the early detection of pneumonia.

Keywords: Pneumonia Detection; Convolutional Neural Networks (CNN); Deep Learning; Chest X-Ray; Transfer Learning; Fine-tuning; Medical Image Analysis.

LISTA DE ILUSTRAÇÕES

Figura 1 – Imagem de Pneumonia vs. pulmão normal.....	14
Figura 2 – Modelo de TL com FT baseado em uma CNN pré-treinada no <i>ImageNet</i>	17
Figura 3 – Arquitetura de uma CNN com camadas de convolução, <i>pooling</i> e densas.....	18
Figura 4 – Etapas metodológicas adotadas no estudo	23
Figura 5 – CNN genérica com a camada adicional de classificação.....	25
Figura 6 – Representação da arquitetura <i>Xception</i> com suas principais camadas.	26
Figura 7 – Representação da arquitetura <i>VGG16</i> com suas principais camadas.	26
Figura 8 – Representação da arquitetura <i>InceptionV3</i> com suas principais camadas.....	27
Figura 9 – Representação da arquitetura <i>InceptionResNetV2</i> com suas principais camadas.....	28
Figura 10 – Representação da arquitetura <i>DenseNet121</i> com suas principais camadas.....	28
Figura 11 – Representação da arquitetura <i>ResNet50</i> com suas principais camadas.	29
Figura 12 – Fluxo dos experimentos realizados.....	30
Figura 13 – Matrizes de confusão — <i>InceptionResNetV2</i> (conjunto de teste)	37
Figura 14 – Matrizes de confusão — <i>VGG16</i> (conjunto de teste)	38
Figura 15 – Tela inicial — aguardando upload de exame individual.....	40
Figura 16 – Exemplo de upload de imagem individual, com pré-visualização e metadados do arquivo	40
Figura 17 – Resultado de análise indicando <i>Normal</i>	41
Figura 18 – Resultado de análise indicando <i>Pneumonia</i>	42
Figura 19 – Tela de upload em lote	42
Figura 20 – Resultados do processamento em lote	43

LISTA DE TABELAS

Tabela 1 – Síntese dos trabalhos relacionados	22
Tabela 2 – Avaliação do Xception em diferentes combinações de hiperparâmetros	31
Tabela 3 – Avaliação do VGG16 em diferentes combinações de hiperparâmetros	32
Tabela 4 – Avaliação do VGG19 em diferentes combinações de hiperparâmetros	32
Tabela 5 – Avaliação do ResNet50 em diferentes combinações de hiperparâmetros	33
Tabela 6 – Avaliação do InceptionV3 em diferentes combinações de hiperparâmetros	34
Tabela 7 – Avaliação do InceptionResNetV2 em diferentes combinações de hiperparâmetros	34
Tabela 8 – Avaliação do DenseNet121 em diferentes combinações de hiperparâmetros	35
Tabela 9 – Avaliação do DenseNet169 em diferentes combinações de hiperparâmetros	36
Tabela 10 – Comparação dos dois melhores modelos com otimizador SGD ...	37
Tabela 11 – Desempenho por classe do InceptionResNetV2 com Adam e SGD.....	39

LISTA DE ABREVIATURAS E SIGLAS

IFPI	Instituto Federal de Educação, Ciência e Tecnologia do Piauí
OMS	Organização Mundial da Saúde
UNICEF	Fundo das Nações Unidas para a Infância
CNN	Rede Neural Convolucional (<i>Convolutional Neural Network</i>)
DL	Aprendizado Profundo (<i>Deep Learning</i>)
ML	Aprendizado de Máquina (<i>Machine Learning</i>)
AI	Inteligência Artificial (<i>Artificial Intelligence</i>)
TL	Aprendizado por Transferência (<i>Transfer Learning</i>)
FT	Ajuste Fino (<i>Fine-Tuning</i>)
LR	Taxa de Aprendizado (<i>Learning Rate</i>)
BS	Tamanho de Lote (<i>Batch Size</i>)
SGD	Gradiente Estocástico Descendente (<i>Stochastic Gradient Descent</i>)
Adam	Otimizador adaptativo para redes neurais (<i>Adaptive Moment Estimation</i>)
F1-score	Métrica combinada de precisão e revocação
Precision	Precisão: proporção de predições corretas positivas
Recall	Revocação: proporção de positivos corretamente identificados
AUC	Área sob a Curva (<i>Area Under the Curve</i>)
ReLU	Unidade Linear Retificada (<i>Rectified Linear Unit</i>)
Dropout	Técnica de regularização que desativa neurônios durante o treinamento
ImageNet	Base de imagens para pré-treinamento
Kaggle	Plataforma de datasets e competições de aprendizado de máquina
X-ray	Radiografia (<i>Radiograph</i>)
SDRA	Síndrome do Desconforto Respiratório Agudo

PACS	Sistema de Comunicação e Arquivamento de Imagens (<i>Picture Ar-chiving and Communication System</i>)
TCC	Trabalho de Conclusão de Curso
DATASUS	Departamento de Informática do
SUS SUS	Sistema Único de Saúde
VGG	<i>Visual Geometry Group</i> (arquitetura de CNN) ResNet
	Rede Residual (<i>Residual Network</i>) InceptionV3
	Arquitetura <i>Inception</i> versão 3
	InceptionResNetV2 <i>Inception-ResNet</i> versão 2
DenseNet	Rede Densamente Conectada (<i>Densely Connected Network</i>) Xception <i>Extreme Inception</i> (arquitetura de CNN)
MobileNet	Arquitetura CNN otimizada para dispositivos móveis NASNet <i>Neural Architecture Search Network</i>
U-Net	Arquitetura CNN para segmentação de imagens
Streamlit	Framework Python para criação de aplicações web Keras Biblioteca de alto nível para redes neurais em Python NumPy Biblioteca Python para computação numérica
Pillow	Biblioteca Python para processamento de imagens (PIL) Python Linguagem de programação
SVM	Máquina de Vetores de Suporte (<i>Support Vector Machine</i>)
GoogleNet	Arquitetura CNN desenvolvida pelo Google

SUMÁRIO

1	INTRODUÇÃO	10
1.1	Objetivos	11
1.2	Organização do Trabalho	12
2	CONCEITOS BÁSICOS	13
2.1	Pneumonia	13
2.2	Inteligência Artificial na Medicina	15
2.2.1	Inteligência Artificial	15
2.2.2	Aprendizado de Máquina	16
2.2.3	Deep Learning	16
2.2.4	Redes Neurais Convolucionais (CNNs)	17
3	TRABALHOS RELACIONADOS	20
4	METODOLOGIA	23
4.1	Revisão da Literatura	23
4.2	Coleta e Preparação dos Dados	24
4.3	Seleção e Implementação	24
4.3.1	Modelos Utilizados	25
4.3.1.1	Xception	25
4.3.1.2	VGG16 e VGG19	26
4.3.1.3	InceptionV3	27
4.3.1.4	InceptionResNetV2	27
4.3.1.5	DenseNet121 e DenseNet169	28
4.3.1.6	ResNet50	29
4.4	Parâmetros e Treinamento	29
4.5	Ordem dos Experimentos	30
5	RESULTADOS	31
5.1	Avaliação dos Modelos	31
5.2	Comparação de Desempenho	37
5.2.1	Implementação do Sistema de Análise de Radiografias	39
6	CONCLUSÃO	44
6.1	Trabalhos Futuros	44
	REFERÊNCIAS	45

1 INTRODUÇÃO

A pneumonia é uma infecção respiratória aguda que afeta o parênquima pul-monar, comprometendo a função respiratória e, em casos graves, podendo evoluir para fibrose pulmonar. É uma síndrome infecciosa que é causada por agentes diversos, como bactérias, vírus, fungos e aspiração. De acordo com a Organização Mundial da Saúde (OMS) (2023) e o Fundo das Nações Unidas para a Infância (UNICEF) (2024), trata-se de uma das doenças respiratórias mais comuns no mundo, com centenas de milhões de casos e milhões de mortes anuais, afetando principalmente crianças menores de cinco anos e idosos acima de 75 anos, sendo que, globalmente, mais de 700 mil crianças nessa faixa etária morrem anualmente devido à pneumonia. Segundo Chumbita *et al.* (2020), continua sendo a principal causa de morte em crianças menores de cinco anos e em adultos acima de 70 anos, com redução de 36% na mortalidade infantil entre 2007 e 2017, mas aumento de 34% entre os idosos.

No Brasil, os dados do Sistema de Informação da Atenção Básica (DATA-SUS/SUS) confirmam que entre janeiro e agosto de 2022 ocorreram 44.523 óbitos por pneumonia, enquanto no mesmo período de 2021 foram registradas 31.027 mortes, evidenciando um aumento significativo na mortalidade (Departamento de Informática do SUS – DATASUS, 2022). Esses dados reforçam a relevância clínica e social da doença e a necessidade de estratégias que melhorem a precisão e a rapidez no diagnóstico.

O diagnóstico da pneumonia pode ser realizado por diferentes métodos, como o exame clínico, radiografia de tórax e a tomografia computadorizada (Ibrahim *et al.*, 2021). Embora esses recursos sejam valiosos, eles dependem fortemente da experiência e do julgamento do profissional de saúde, o que pode ocasionar atrasos ou reduzir a precisão do diagnóstico. Em muitos casos clínicos, identificar a causa exata da doença ainda representa um desafio: estima-se que, em mais de 50% dos casos, o agente causador não seja claramente determinado, dificultando a escolha do tratamento mais adequado (Mahesse, 2023). Além disso, divergências entre médicos na interpretação de imagens radiográficas são comuns, reforçando o caráter subjetivo do processo.

Essa dificuldade diagnóstica se agrava em casos avançados da doença, quando se faz necessário o início imediato de tratamento com antibióticos e medidas de precaução adequadas (Silva *et al.*, 2020). A complexidade em identificar corretamente a pneumonia, somada ao crescente número de exames médicos e à subjetividade do diagnóstico radiográfico, evidencia a necessidade de métodos auxiliares que sejam capazes de fornecer resultados mais precisos, rápidos e consistentes.

Com o avanço da computação, a Inteligência Artificial (*Artificial Intelligence* - AI) tem desempenhado um papel cada vez mais relevante na medicina, especialmente no diagnóstico de diferentes doenças (Vasile *et al.*, 2021). O uso de técnicas de Aprendizado Profundo (*Deep Learning* - DL) tem sido amplamente explorado na análise de imagens médicas e no apoio à tomada de decisão clínica, possibilitando reduzir a necessidade de exames caros ou invasivos, além de oferecer elevada confiabilidade nos resultados (Braga *et al.*, 2019). Essa tecnologia se mostra promissora ao lidar com grandes volumes de dados e extrair padrões complexos que muitas vezes não são perceptíveis a olho nu.

Entre as arquiteturas de DL mais utilizadas para processamento de imagens destacam-se as Redes Neurais Convolucionais (*Convolutional Neural Network* - CNNs) (Alzubaidi *et al.*, 2021). Essas redes são compostas por camadas de convolução, responsáveis pela extração de características como bordas e texturas, seguidas por camadas de subamostragem (*pooling*), que reduzem a dimensionalidade preservando informações relevantes, e por camadas densas totalmente conectadas, responsáveis pela classificação final (Khan *et al.*, 2020). A aplicação de CNNs tem obtido grande sucesso em tarefas de visão computacional e se consolidado como uma das principais abordagens em diagnóstico por imagem.

Neste contexto, este trabalho tem como objetivo desenvolver um sistema computacional de apoio ao diagnóstico de pneumonia em radiografias de tórax, utilizando técnicas de DL. O sistema proposto emprega arquiteturas de CNNs pré-treinadas com técnicas de Aprendizado por Transferência (*Transfer Learning* - TL) e *Fine-Tuning*, além da avaliação de combinações de Otimizadores, Taxas de Aprendizado (*Learning Rate*

- LR) e Tamanhos de Lote (*Batch Size* - BS), utilizando o conjunto de dados *Chest X-Ray Images* (Pneumonia). Para tornar a solução acessível e aplicável em ambientes clínicos, foi desenvolvida uma interface web interativa utilizando *Streamlit*, permitindo que profissionais de saúde realizem análises de radiografias de forma rápida e intuitiva. Assim, espera-se contribuir para o desenvolvimento de uma ferramenta prática de apoio ao diagnóstico, com impacto potencial na redução da mortalidade e na melhoria do cuidado clínico de pacientes com pneumonia.

1.1 OBJETIVOS

O objetivo deste Trabalho de Conclusão de Curso (TCC) é desenvolver um sistema computacional de apoio ao diagnóstico de pneumonia em radiografias de tórax, utilizando técnicas de DL. As principais contribuições deste trabalho são:

- Revisar a literatura e implementar diferentes modelos de CNN com TL e *Fine-Tuning* para detecção de pneumonia em radiografias de tórax, utilizando o conjunto de dados *Chest X-Ray Images* (Pneumonia) (Kermany *et al.*, 2018).
- Comparar o desempenho de oito arquiteturas pré-treinadas: *Xception*, *VGG16*, *VGG19*, *ResNet50*, *InceptionV3*, *InceptionResNetV2*, *DenseNet121* e *DenseNet169*.
- Avaliar o impacto de diferentes configurações de treinamento, incluindo os otimizadores *Adam* e *SGD*, LR e BS, além de utilizar métricas adequadas para dados desbalanceados, priorizando o *F1-score*.
- Identificar a melhor combinação entre modelo e hiperparâmetros, buscando otimizar o desempenho preditivo ao mesmo tempo em que minimiza o custo computacional.
- Implementar uma interface interativa com *Streamlit*, um *Framework Python* que permite criar aplicações web dinâmicas, permitindo o envio de imagens de raio-X para análise pelo modelo.

1.2 ORGANIZAÇÃO DO TRABALHO

Este trabalho está organizado em seis seções. Na Seção 2, são apresentados os conceitos básicos sobre pneumonia, inteligência artificial na medicina, aprendizado de máquina e redes neurais convolucionais. Na Seção 3, são revisados os trabalhos relacionados, destacando abordagens recentes de classificação automatizada. Na Seção 4, é detalhada a metodologia adotada, abrangendo a revisão da literatura, a coleta e preparação dos dados, a implementação dos modelos de CNN e os parâmetros de treinamento. Na Seção 5, são analisados e discutidos os resultados obtidos. Por fim, na Seção 6, são apresentadas as conclusões, considerações finais e perspectivas para trabalhos futuros.

2 CONCEITOS BÁSICOS

Nesta seção, são apresentados os principais conceitos que sustentam a aplicação de CNNs no diagnóstico automatizado de pneumonia por meio de radiografias de tórax. Primeiramente, são descritos o quadro clínico e epidemiológico da pneumonia, seguido de uma introdução à AI e ao ML, com ênfase nas abordagens de DL. Por fim, são detalhados os princípios e a estrutura das CNNs, tecnologia central desta pesquisa.

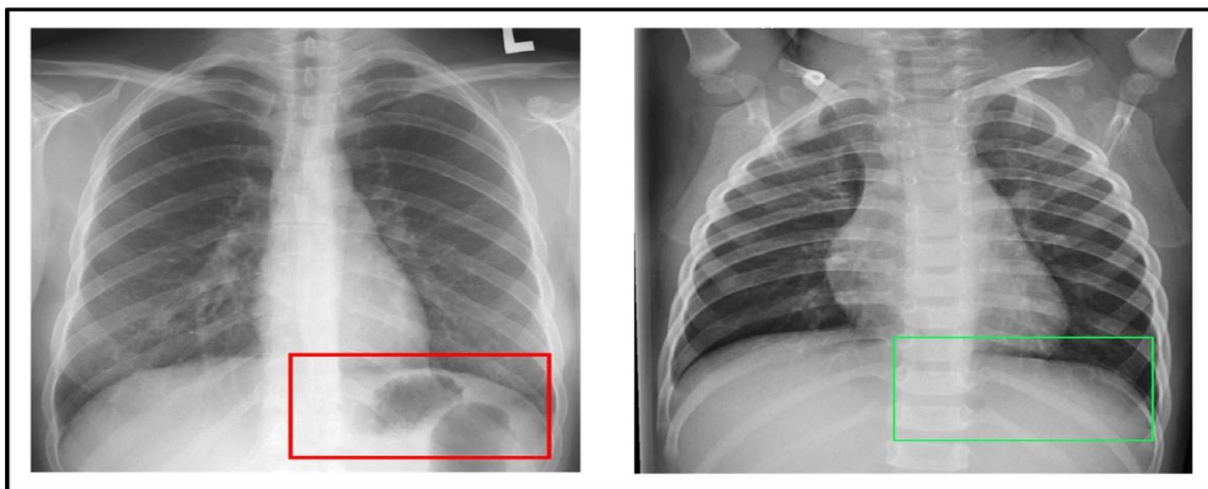
2.1 PNEUMONIA

A pneumonia é uma infecção aguda que compromete o parênquima pulmonar, resultando em inflamação dos alvéolos e acúmulo de exsudato, o que prejudica a oxigenação e pode desencadear insuficiência respiratória (Mayo Clinic Staff, 2020). Sua etiologia é variada, incluindo bactérias, vírus, fungos e microrganismos atípicos, e a resposta inflamatória desencadeada no tecido pulmonar leva à consolidação alveolar e, em casos mais graves, pode evoluir para complicações como sepse, síndrome do desconforto respiratório agudo (SDRA) e até fibrose pulmonar residual (Ali *et al.*, 2021; Mayo Clinic Staff, 2024).

Do ponto de vista clínico, os principais sintomas incluem tosse produtiva ou seca, febre, calafrios, dor torácica pleurítica e dispneia, embora em populações vulneráveis, como crianças, idosos e imunossuprimidos, a apresentação possa ser atípica (Mayo Clinic Staff, 2020). O diagnóstico combina avaliação clínica e exames de imagem, sobretudo radiografia de tórax, o método mais utilizado por sua disponibilidade e baixo custo. Entretanto, limitações diagnósticas são comuns: divergências entre radiologistas, baixa sensibilidade em fases iniciais e dificuldade em distinguir pneumonia de outras doenças pulmonares contribuem para diagnósticos tardios ou equivocados (Rios *et al.*, 2024).

Na Figura 1, são apresentadas duas subfiguras, uma de um pulmão saudável e outra com pneumonia, destacando a região afetada.

Figura 1 – Imagem de Pneumonia vs. pulmão normal.



Fonte: *Kaggle*, (Kermany *et al.*, 2018).

Como mostrado na Figura 1, a imagem ilustra de forma clara a diferença entre um pulmão saudável e um afetado por pneumonia. Enquanto o pulmão normal apresenta campos aéreos uniformes, o pulmão com pneumonia mostra opacidades nas regiões inferiores, indicativas de acúmulo de líquido ou inflamação.

No contexto epidemiológico, a pneumonia representa um dos maiores desafios de saúde pública. Segundo estimativas globais, é responsável por milhões de óbitos anuais, especialmente em crianças menores de cinco anos e idosos (Rios *et al.*, 2024). Esse impacto mostra a necessidade de desenvolver estratégias para minimizando atrasos e consequências adversas.

Essa distinção visual, como apresentado na Figura 1, embora perceptível nesta imagem, nem sempre é facilmente identificável: estudos demonstram que divergências entre radiologistas são comuns, especialmente em casos iniciais (Mahesse, 2023).

Dessa forma, torna-se evidente que a avaliação clínica isolada e a interpretação manual de radiografias podem ser limitadas. Métodos auxiliares baseados em técnicas computacionais, como DL, oferecem uma alternativa promissora. Eles podem analisar imagens de maneira sistemática, detectando padrões sutis que podem passar despercebidos ao olho humano, aumentando a acurácia do diagnóstico e reduzindo o tempo necessário para a tomada de decisão.

Portanto, o uso de técnicas computacionais aplicadas a radiografias de tórax não apenas complementa a análise clínica tradicional, mas também tem potencial para melhorar a detecção precoce da pneumonia, reduzir erros diagnósticos e contribuir significativamente para área da saúde. Diante da complexidade das imagens, surge a necessidade de ferramentas capazes de identificar padrões sutis, abrindo espaço para a aplicação da AI na medicina.

2.2 INTELIGÊNCIA ARTIFICIAL NA MEDICINA

As tecnologias computacionais avançadas no ambiente clínico tem revolucionado a forma como diagnósticos são realizados. A AI é uma das principais ferramentas dessa transformação, abrangendo desde sistemas de apoio à decisão até métodos de análise automatizada de imagens médicas. Os fundamentos de AI, ML e DL permitem compreender como essas abordagens evoluíram até as CNNs, amplamente aplicadas na detecção de pneumonia em radiografias de tórax.

2.1.1 Inteligência Artificial

A AI é um campo da ciência da computação dedicado ao desenvolvimento de sistemas capazes de simular capacidades cognitivas humanas, como aprendizado, raciocínio, percepção e tomada de decisão (Kourounis *et al.*, 2023). Nas últimas décadas, este campo tem experimentado avanços significativos impulsionados pelo aumento da capacidade computacional, pela disponibilidade massiva de dados e pelo desenvolvimento de algoritmos sofisticados de ML e DL (Connor *et al.*, 2021).

Na área da saúde, estas tecnologias têm se mostrado promissoras para transformar a prática médica. Baseando-se em algoritmos capazes de aprender padrões complexos a partir de grandes volumes de informações, suas aplicações abrangem desde o diagnóstico até o planejamento terapêutico (Kumar *et al.*, 2023). A principal vantagem reside na capacidade de processar e analisar dados de forma mais rápida e precisa do que métodos convencionais, auxiliando profissionais na detecção precoce de doenças e na otimização de fluxos clínicos (Chong *et al.*, 2025).

No contexto do diagnóstico por imagem, especialmente em radiografias de tórax, algoritmos baseados em AI têm demonstrado desempenho comparável ao de radiologistas experientes na detecção de pneumonia, inclusive em ambientes de alta demanda ou com escassez de especialistas (Becker *et al.*, 2022). Sua integração com Sistema de Comunicação e Arquivamento de Imagens (Picture Archiving and Communication System - PACS) permite análises em larga escala e em tempo quase real, viabilizando triagens automatizadas e priorização de casos críticos em unidades de pronto-atendimento (Hwang *et al.*, 2023; Nguyen *et al.*, 2022).

Além do diagnóstico, perspectivas futuras apontam para modelos mais interpretáveis, que ofereçam não apenas resultados quantitativos, mas também justificativas visuais, como mapas de calor ou regiões de interesse, aumentando a confiança dos profissionais nas decisões assistidas por AI (Radočaj *et al.*, 2025). Essa tendência reforça que a tecnologia não substitui o julgamento médico, mas atua como ferramenta complementar, potencializando diagnósticos mais precisos, reduzindo erros e otimizando recursos, especialmente em áreas com limitações de infraestrutura. Para viabilizar essas aplicações clínicas, a AI fundamenta-se em paradigmas específicos de aprendizado, sendo o ML a base computacional que sustenta o desenvolvimento de sistemas diagnósticos inteligentes.

2.1.2 Aprendizado de Máquina

O ML é um subcampo da AI voltado para o desenvolvimento de algoritmos capazes de identificar padrões em dados e melhorar seu desempenho preditivo sem programação explícita. Segundo Esteva Esteva *et al.* (2021), esses algoritmos podem ser aplicados à estratificação de risco, previsão de desfechos e apoio ao diagnóstico, tornando os processos mais ágeis e assertivos.

Métodos clássicos de ML, como Máquina de Vetores de Suporte (*Support Vector Machine* - SVMs) ou *Random Forests*, têm sido utilizados com sucesso em diversas aplicações médicas. No entanto, esses métodos dependem de uma etapa manual de extração de características, na qual especialistas definem quais atributos dos dados são relevantes para a tarefa (Zhang; Qie, 2023).

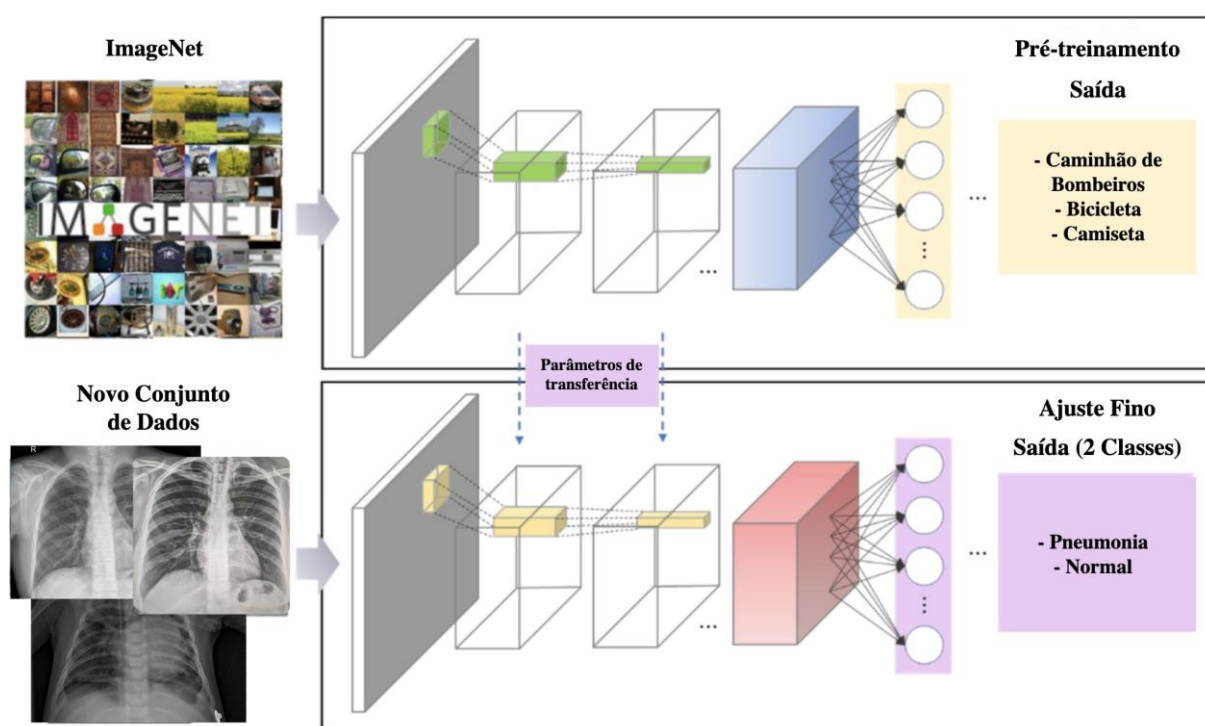
Essa limitação motivou o desenvolvimento de técnicas mais avançadas, capazes de realizar essa extração de forma automática, especialmente para dados complexos como imagens médicas. Nesse contexto, surge o DL como uma evolução natural do ML, trazendo capacidades superiores para o processamento de dados médicos.

2.1.3 Deep Learning

Dentro do ML, o DL representa um avanço significativo. Trata-se de um conjunto de técnicas baseadas em redes neurais profundas, compostas por múltiplas camadas de processamento que aprendem representações hierárquicas dos dados (Liu *et al.*, 2021). Enquanto métodos clássicos de ML dependem de extração manual de características, as redes profundas realizam essa extração de forma automática, o que se mostra particularmente vantajoso para dados complexos, como imagens médicas de raios-X ou tomografias. Essa diferença explica por que o DL tem superado os métodos tradicionais em tarefas de reconhecimento de padrões visuais, alcançando resultados comparáveis ou superiores aos de especialistas humanos em alguns contextos (Fu; Man, 2022; Zhang; Qie, 2023).

Um dos fatores que impulsionaram o uso do DL na área médica é a técnica de TL. Nessa abordagem, modelos previamente treinados em grandes bases de dados genéricas, como o *ImageNet*, são adaptados para uma tarefa específica com menor quantidade de dados rotulados (Khan *et al.*, 2020). Esse processo reduz o custo computacional e o tempo de treinamento, além de melhorar a generalização em domínios médicos, onde a obtenção de grandes volumes de imagens anotadas é desafiadora. A ilustração desse processo é apresentada na Figura 2.

Figura 2 – Modelo de TL com FT baseado em uma CNN pré-treinada no *ImageNet*.



Fonte: Elaborado com base em Park *et al.* (2021).

Complementarmente, o *Fine-Tuning* consiste em descongelar parte das camadas de um modelo pré-treinado e atualizar seus pesos de acordo com o novo conjunto de dados. Essa estratégia permite que as camadas superiores aprendam padrões específicos do domínio médico, preservando ao mesmo tempo o conhecimento prévio das camadas inferiores (Karimi *et al.*, 2021).

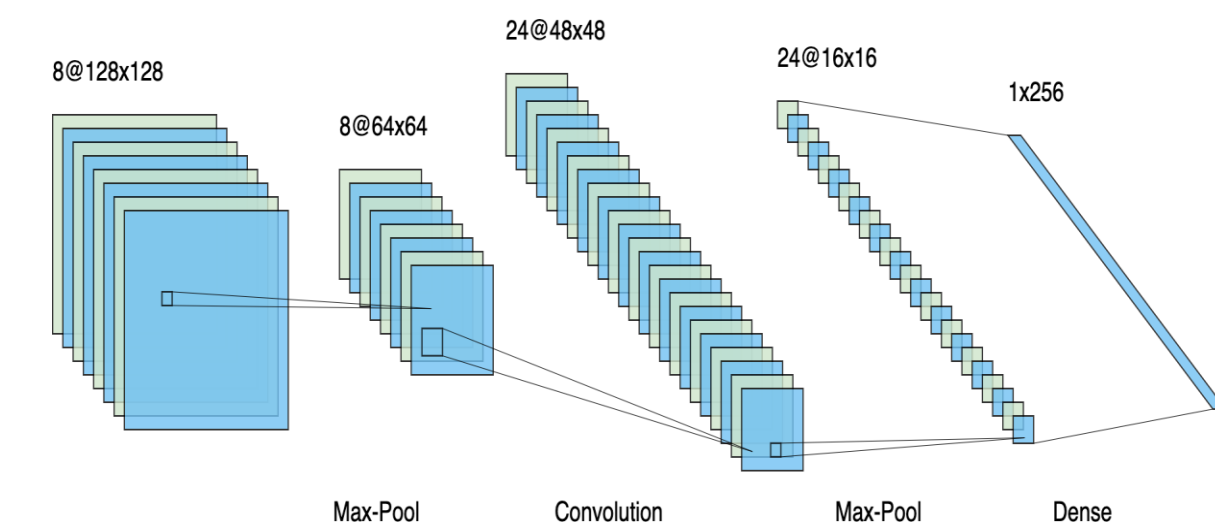
A combinação de TL e *Fine-Tuning* tem sido amplamente empregada em estudos recentes para a detecção de pneumonia e outras doenças em radiografias torácicas, alcançando desempenhos notáveis com conjuntos de dados relativamente pequenos (Esteve *et al.*, 2021). Entre as diversas arquiteturas de DL existentes, as CNNs se destacam como a abordagem mais adequada para análise de imagens médicas, sendo a técnica central para a detecção de pneumonia em radiografias de tórax.

2.1.4 Redes neurais convolucionais (cnns)

As CNNs são uma classe de redes profundas que se dedica ao processamento de dados estruturados em matrizes, como imagens. A principal característica das CNNs é a capacidade de extrair automaticamente características espaciais hierárquicas a partir dos dados brutos, o que as torna altamente eficazes em tarefas de visão computacional, incluindo diagnóstico médico (Kourounis *et al.*, 2023).

A estrutura básica de uma CNN compreende três tipos principais de camadas: convolução, *pooling* e densas (*fully connected*). As camadas de convolução aplicam filtros que detectam padrões locais, como bordas, texturas e formas. Cada filtro gera um mapa de características (*feature map*) que preserva informações espaciais relevantes para a tarefa de classificação (Krichen, 2023). As camadas de *pooling*, por sua vez, reduzem a dimensionalidade dos mapas de características, mantendo as informações mais importantes e conferindo invariância a pequenas translações e distorções nas imagens. Finalmente, as camadas densas consolidam as informações extraídas pelas camadas convolucionais e realizam a classificação final, produzindo probabilidades associadas a cada classe (Xie *et al.*, 2025; Hossain *et al.*, 2024), conforme representado na Figura 3.

Figura 3 – Arquitetura de uma CNN com camadas de convolução, *pooling* e densas.



Fonte: Autor.

A Figura 3 representa o fluxo de processamento em uma CNN. A imagem de entrada passa por camadas de convolução e *pooling*, que reduzem progressivamente o tamanho das representações e aumentam a quantidade de mapas de características. Essas transformações extraem padrões cada vez mais complexos, até que os dados sejam achatados e processados por camadas densas, responsáveis pela etapa final de classificação.

Na área médica, as CNNs têm se mostrado particularmente eficazes na análise de radiografias, incluindo a detecção de pneumonia. Esses modelos são capazes de identificar padrões sutis de consolidação alveolar, infiltração e opacidades que podem passar despercebidos em análises humanas, oferecendo suporte à decisão clínica e reduzindo a variabilidade entre profissionais (Sharma; Guleria, 2023). A aplicação de CNNs em radiografias de tórax permite não apenas a classificação binária, Pneumonia e Normal, mas também a geração de mapas de ativação que destacam as regiões relevantes da imagem, facilitando a interpretação médica e o rastreamento de anomalias (Rahman *et al.*, 2024; Tsuji *et al.*, 2023).

O *fine-tuning* desempenha papel estratégico na utilização de CNNs em imagens médicas. Ao empregar modelos pré-treinados em grandes conjuntos de imagens genéricas, como o *ImageNet*, e ajustar suas camadas finais ao conjunto específico de radiografias, é possível aproveitar o conhecimento previamente aprendido, ao mesmo tempo em que adapta a rede aos padrões clínicos particulares da pneumonia (Oliveira *et al.*, 2021; Karimi *et al.*, 2021).

Essa abordagem é especialmente valiosa em contextos de baixa disponibilidade de dados rotulados, comuns em hospitais e centros de pesquisa, pois permite reduzir o tempo de treinamento e melhorar a generalização do modelo. Camadas mais profundas podem ser descongeladas gradualmente para capturar características mais específicas do domínio médico, enquanto as camadas iniciais, responsáveis por detectar padrões visuais genéricos, permanecem congeladas. Dessa forma, o aprendizado de máquina, especialmente o *deep learning* e suas

CNNs, oferecem uma base sólida para o desenvolvimento de sistemas de diagnóstico assistido por computador. A utilização de CNNs combinadas com estratégias de *transfer learning* e *fine-tuning* potencializa a detecção precoce de pneumonia e a padronização de análises em radiografias de tórax, contribuindo para um atendimento mais rápido, preciso e confiável (Vardhan *et al.*, 2023; Salehi *et al.*, 2021). Com esses fundamentos estabelecidos, o Capítulo 3 são revisados os estudos relevantes, com o objetivo de identificar lacunas e fundamentar as escolhas experimentais deste trabalho.

3 TRABALHOS RELACIONADOS

Diversos trabalhos recentes têm explorado o uso de técnicas de ML na detecção de doenças por meio de exames de imagem, como Radiografias de Tórax. Nesse contexto, modelos baseados em CNNs se destacam pela capacidade de extrair automaticamente características complexas e relevantes das imagens, superando métodos tradicionais de processamento e classificação. Essas abordagens têm obtido resultados promissores na identificação de doenças como pneumonia e COVID-19.

Os autores Apostolopoulos *et al.* (2020) utilizaram cinco modelos diferentes de CNN como *VGG19*, *MobileNetV2*, *Inception*, *Xception* e *InceptionResNetV2*, para a classificação de imagens de raios X em três classes: COVID-19, pneumonia bacteriana comum e condições normais. As imagens foram obtidas a partir de bases públicas, incluindo o repositório *COVID-19 Radiography Database* e o conjunto *Chest X-Ray Images (Pneumonia)*, totalizando 1.427 imagens. Dentre os modelos avaliados, o *VGG19* apresentou o melhor desempenho, alcançando 98,75% de acurácia e 92,85% de sensibilidade na tarefa de classificação.

No artigo de Bhadra & Kar (2020), foi proposta uma arquitetura CNN de múltiplas camadas, menos complexa e com um treinamento significativamente mais rápido em comparação com redes neurais mais avançadas, como *MobileNet* e *ResNet*. O estudo utilizou 1.330 casos de COVID-19, 1.330 casos de pneumonia e 1.330 imagens de casos normais, divididos em 80% para treinamento, 10% para validação e 10% para teste cego, onde os dados são completamente novos para a avaliação do modelo, assegura uma avaliação imparcial.

O modelo passou por validação cruzada em 10 partições (*k-fold*) e obteve uma precisão de 99,1%. Em comparação, as redes pré-treinadas *VGG16* alcançaram 93,9% em 549 segundos, a *VGG19* obteve 92,2% em 668 segundos, o *MobileNet* apresentou 98,8% em 224 segundos, o *InceptionV3* alcançou 99,5% em 457 segundos e o *ResNet50* registrou 98,3% em 537 segundos.

No estudo de Gupta *et al.* (2021) foi proposto o modelo *InstaCovNet-19*, uma CNN baseada em aprendizado profundo. Esse modelo integra várias arquiteturas pré-treinadas de CNN, incluindo *ResNet-101*, *Inception v3*, *Xception*, *MobileNetv2* e *NASNe*. Essas arquiteturas são combinadas por meio da técnica de empilhamento integrado, que reúne os recursos para alcançar os resultados mais precisos. O modelo foi aplicado à detecção de três classes: normal, pneumonia e COVID-19, alcançando uma precisão de 99,08%. Para o treinamento, foi utilizado um conjunto de radiografias de tórax de bases públicas do *Kaggle*, balanceado por amostragem aleatória para conter cerca de 360 imagens por classe (normal, pneumonia e COVID-19).

O trabalho de Oliveira *et al.* (2021) classifica imagens de radiografia do tórax em duas categorias: COVID-19 e não COVID-19, utilizando um conjunto de imagens de alta qualidade para evitar vieses. Foram usadas 7.875 imagens de COVID-19 e 16.841 de não COVID-19. 80% foram usadas para treinamento (6.300 de COVID-19 e 13.472 de não COVID-19) e 20% para teste (1.575 de COVID-19 e 3.369 de não COVID-19). As imagens foram segmentadas com o modelo *U-Net*, que atingiu um *IoU* de 93,33%, antes da classificação por redes neurais convolucionais. Entre 25 modelos avaliados, 5 modelos obtiveram uma acurácia entre 91% e 92%, foram eles: *VGG-11*, *ResNet-152*, *Densenet-161*, *Densenet-201* e *GoogleNet*. Para cada um dos cinco modelos que se destacaram, é gerado um mapa de calor. A partir desses mapas, é calculada a média para obter o mapa de calor final.

Os autores Narin *et al.* (2021) aplicaram modelos baseados em redes neurais convolucionais (*ResNet50*, *ResNet101*, *ResNet152*, *InceptionV3* e *Inception-ResNetV2*) para a detecção de COVID-19 em imagens de raios-X torácicos. Foram utilizados três conjuntos de dados distintos: COVID-19 vs normal, COVID-19 vs pneumonia viral e COVID-19 vs pneumonia bacteriana, compostos por 341 imagens de COVID-19 e aproximadamente 2 800 imagens das demais classes, obtidas de repositórios públicos (*GitHub de Cohen et al.*, *Kaggle* e *Radiopaedia*). Entre os modelos testados, o *ResNet50* apresentou o melhor desempenho, com acurácias de 96,1%, 99,5% e 99,7% nos respectivos conjuntos de dados.

O estudo conduzido pelos autores Diniz *et al.* (2023) apresentou um método para detecção automática da COVID-19 em radiografias de tórax, utilizando uma abordagem baseada em seleção automática de realce de imagem combinada com redes neurais convolucionais. Os resultados obtidos foram promissores, com a arquitetura *Xception* alcançando acurácia de até 99,39% e *F1-score* de 98,71%. Além da *Xception*, outras arquiteturas avaliadas incluíram *ResNet*, *DenseNet*, *VGG* e *Inception*, todas apresentando *F1-score* superior a 90%. Os experimentos foram realizados utilizando o otimizador Adam, taxa de aprendizado de 0,001 e *batch size* de 16.

Por fim, Hasan *et al.* (2024) analisaram diversos modelos baseados em CNN aplicados à detecção de pneumonia, incluindo arquiteturas amplamente utilizadas como *VGG16*, *ResNet50*, *InceptionV3*, *DenseNet121* e *Xception*. O estudo destacou o uso recorrente do aprendizado por transferência como estratégia para aprimorar o desempenho em conjuntos de dados limitados e avaliou a complexidade computacional, a acurácia e a aplicabilidade clínica desses modelos.

Na Tabela 1 é apresentada uma comparação entre os trabalhos relacionados, destacando as abordagens utilizadas e os modelos de aprendizado profundo empregados em cada estudo.

Tabela 1 – Síntese dos trabalhos relacionados.

Referência	Modelos / Dados	Resultados / Diferencial
Apostolopoulos <i>et al.</i> (2020)	VGG19, MobileNetV2, Inception, Xception, InceptionResNetV2	Comparação entre arquiteturas; acurácia reportada alta (ex.: até 98,8% em determinadas
Bhadra & Kar (2020)	CNN customizada, VGG16/19, MobileNet, InceptionV3, ResNet50; 1.330 casos/cls	CNN customizada rápida; precisão até 99,1% em teste cego.
Gupta <i>et al.</i> (2021)	InstaCovNet-19 (ResNet-101, Inception v3,	Empilhamento de pré-treinados; precisão reportada 99,08%.
NASNe) Oliveira <i>et al.</i> (2021)	Xception, MobileNetV2, U-Net + VGG-11, ResNet-152,	Uso de segmentação (U-Net) e mapas de calor; acurácias dos top-5 entre 91–92%.
	DenseNet-161/201, GoogleNet; 7.875 COVID / 16.841 não-COVID	ResNet50 destacado como melhor entre avaliados (autor relata desempenho superior).
Narin <i>et al.</i> (2021)	ResNet50/101/152, InceptionV3, Inception-ResNetV2; múltiplas bases	
Diniz <i>et al.</i> (2023)	Xception, ResNet, DenseNet, VGG, Inception	Ênfase em realce automático + CNN; Xception alcançando acurácia até 99,39%.
Hasan <i>et al.</i> (2024)	Revisão de VGG16, ResNet50, InceptionV3, DenseNet121, Xception	Revisão crítica sobre transferência de aprendizado e complexidade computacional.
Este trabalho	Xception, VGG16, VGG19, ResNet50, InceptionV3, InceptionResNetV2, DenseNet121, DenseNet169	Avaliação experimental sistemática com variações controladas de hiperparâmetros (LR e BS); priorização de <i>F1-score</i> ponderado; análise detalhada por classe; comparação de otimizadores (Adam vs SGD); desenvolvimento de protótipo em Streamlit para aplicação prática.

Fonte: Autor.

Na Tabela 1, são resumidas as arquiteturas de CNNs avaliadas e os principais resultados reportados. Os modelos escolhidos para comparação, *Xception*, *VGG16*, *VGG19*, *ResNet50*, *InceptionV3*, *InceptionResNetV2*, *DenseNet121* e *DenseNet169*, foram selecionados para representar distintas famílias arquiteturais, desde arquiteturas clássicas até redes modernas e densas, permitindo avaliar a influência da profundidade sobre o problema de classificação binária.

O diferencial do presente trabalho está na proposta de uma avaliação experi-mental sistemática e reproduzível: serão comparadas oito arquiteturas sob variações controladas de hiperparâmetros ($LR = \{1 \times 10^{-3} \text{ e } 1 \times 10^{-4}\}$, $BS = \{16 \text{ e } 32\}$), priorizando o *F1-score* ponderado para mitigar o desbalanceamento de classes, apresentando análise detalhada por classe (*recall* para Normal vs Pneumonia), e investigando o efeito de diferentes otimizadores (Adam vs SGD). Além disso, será desenvolvido um protótipo em *Streamlit* para demonstrar a aplicabilidade prática do modelo selecionado.

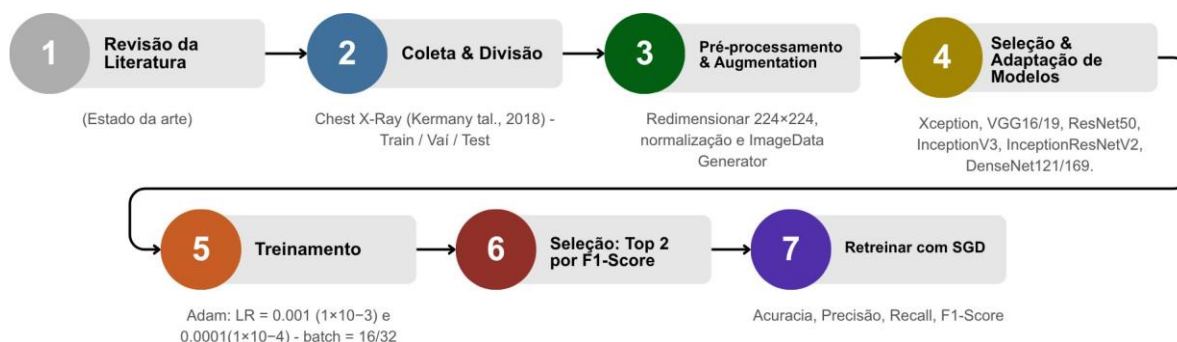
A seguir, no Capítulo 4, é detalhado o conjunto de dados, o pré-processamento, a configuração experimental, os procedimentos de treino e validação, bem como a implementação do protótipo.

4 METODOLOGIA

A pesquisa em questão tem como objetivo avaliar e comparar diferentes arquitetura de CNNs. Para alcançar esse objetivo, é adotada uma abordagem sistemática que inclui a implementação dos algoritmos e a realização de uma avaliação experimental detalhada.

Como mostrado na Figura 4, o fluxo metodológico seguido neste trabalho, destacando as principais etapas desde a revisão da literatura até avaliação final dos modelos.

Figura 4 – Etapas metodológicas adotadas no estudo.



Fonte: Autor.

A Figura 4 sintetiza sete etapas: 1) revisão da literatura; 2) coleta e divisão do *dataset*; 3) pré-processamento e aumento; 4) seleção e adaptação das arquiteturas; 5) treinamento inicial com Adam; 6) seleção dos dois melhores por *F1-score*; 7) retreinamento com SGD e avaliação final (acurácia, precisão, *recall* e *F1-score*). A seguir, detalha-se a revisão da literatura que fundamentou a escolha das arquiteturas e parâmetros.

4.1 REVISÃO DA LITERATURA

Esta seção inicia com uma revisão de artigos e estudos recentes sobre técnicas de classificação de imagens médicas, com foco especial na detecção de pneumonia em radiografias de tórax. A revisão contempla métodos modernos de DL, com destaque para arquiteturas: *Xception*, *VGG16*, *VGG19*, *ResNet50*, *InceptionV3*, *InceptionRes-NetV2*, *DenseNet121* e *DenseNet169*. As CNNs têm sido amplamente utilizadas em diagnósticos automatizados devido à sua capacidade de extrair de forma eficiente padrões visuais relevantes em dados médicos (Mall *et al.*, 2023). Como já foi abordado nas seções anteriores, vários estudos aplicam essas arquiteturas especificamente à detecção de pneumonia. A revisão foi realizada em bases de dados acadêmicas utilizando as palavras-chave: "Classificação de Imagens Médicas", "Radiografias do Tórax", "Diagnóstico de Pneumonia", "Redes Neurais Convolucionais", "Deep Learning em Imagens Médicas", "Transferência de Aprendizado", "Fine-Tuning" e "Avaliação de Modelos em Diagnóstico Médico".

4.2 COLETA E PREPARAÇÃO DOS DADOS

Neste estudo, utilizou-se o conjunto de dados *Chest X-Ray Images* (Pneumonia) disponibilizado por Kermany *et al.* (2018) e disponível no *Kaggle*. A base original contém 5.856 radiografias, distribuídas entre as classes *normal* (1.583; 27,03%) e *pneumonia* (4.273; 72,97%).

Após o pré-processamento utilizado neste trabalho, o conjunto final usado foi de 5.840 imagens, particionado em 4.173 imagens para treino (71,46%), 1.043 para validação (17,86%) e 624 para teste (10,68%). Essa proporção segue uma prática amplamente utilizada em estudos de DL e também é adotada em trabalhos recentes de detecção de pneumonia por radiografias de tórax (Chollet, 2021; Guo *et al.*, 2023).

Todas as imagens foram padronizadas para 224×224 *pixels* e normalizadas no intervalo [0,1]. No treinamento, aplicamos aumento de dados via *ImageDataGenerator* (*Keras*): rotações até 20°, deslocamentos horizontal/vertical de até 15%, zoom de até 15%, inversão horizontal, variação de brilho entre 0,8 e 1,2, cisalhamento até 10° e preenchimento por vizinhança mais próxima.

Essas transformações foram selecionadas com base em práticas consolidadas na literatura e em testes preliminares, os quais mostraram uma melhora na capacidade de generalização do modelo. Os conjuntos de validação e teste foram tratados apenas com normalização, garantindo que a avaliação reflita o desempenho em imagens não aumentadas.

4.3 SELEÇÃO E IMPLEMENTAÇÃO

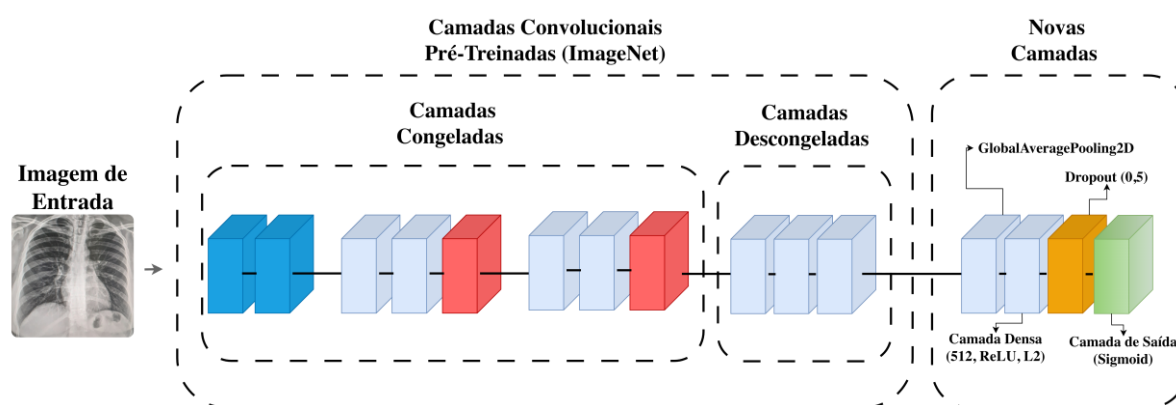
A partir da revisão da literatura, oito modelos pré-treinados foram selecionados, todos com pesos treinados no *ImageNet* e amplamente utilizados na área médica. Os modelos escolhidos foram: *Xception*, *VGG16*, *VGG19*, *ResNet50*, *InceptionV3*, *InceptionResNetV2*, *DenseNet121* e *DenseNet169*.

Para cada modelo, empregou-se a técnica de TL com ajuste fino inicial (*early fine-tuning*). Essa abordagem consiste em descongelar as últimas camadas convolucionais da arquitetura, permitindo que seus pesos sejam ajustados ao novo domínio das imagens médicas, enquanto as camadas restantes permanecem congeladas, preservando o conhecimento previamente adquirido.

Além disso, foi incorporada uma nova estrutura de classificação ao final de cada modelo, composta por uma camada de *Global Average Pooling*, seguida de uma camada densa com 512 neurônios, função de ativação *ReLU* e regularização L2 para reduzir *overfitting*, uma camada de *Dropout* com taxa de 0,5 e uma camada de saída com ativação sigmóide, adequada para classificação binária. Essa configuração permite que os modelos se adaptem especificamente à tarefa de detecção de padrões em imagens médicas, mantendo a robustez proporcionada pelo treinamento prévio em *ImageNet*.

Na Figura 5, é ilustrada uma arquitetura genérica de CNN, destacando-se a adição da nova estrutura de classificação proposta neste trabalho.

Figura 5 – CNN genérica com a camada adicional de classificação.



Fonte: Autor.

Como mostrado na Figura 5, após as Camadas Convolucionais Pré-Treinadas, foi inserida uma camada de *Global Average Pooling*, seguida das camadas densas e da saída *sigmoidal*. Essa modificação permite que os modelos pré-treinados sejam ajustados de forma eficiente à tarefa específica de classificação binária proposta neste estudo.

4.1.1 Modelos Utilizados

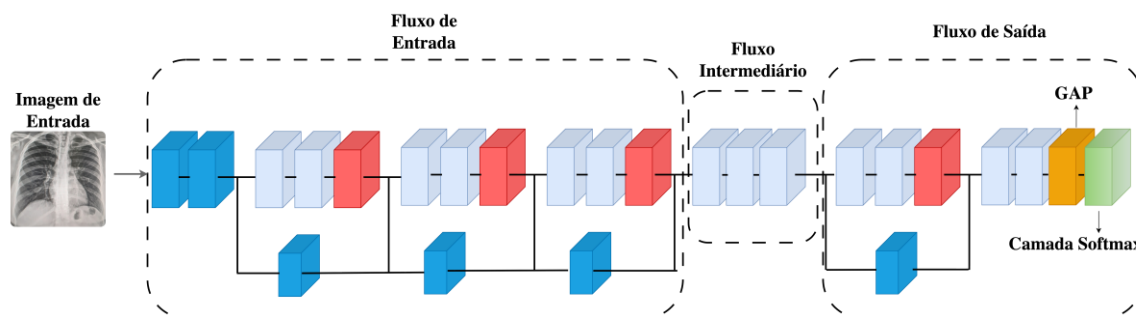
Oito modelos pré-treinados disponíveis na biblioteca *Keras* foram analisados, todos com pesos previamente ajustados no conjunto *ImageNet*. Estes modelos foram escolhidos devido à sua eficiência comprovada em tarefas de classificação de imagens médicas, como demonstrado em estudos recentes.

4.1.1.1 Xception

O modelo *Xception*, originalmente proposto por Chollet (2017), é baseado inteiramente em convoluções separáveis por profundidade, o que resulta em uma arquitetura eficiente e de alto desempenho. Com 36 camadas, destaca-se por sua capacidade de extrair características complexas com menor custo computacional. No contexto da classificação de imagens radiológicas, o modelo foi empregado por (Diniz *et al.*, 2023), demonstrando resultados promissores.

Na Figura 6, é ilustrada de forma simplificada, a arquitetura *Xception* e sua organização geral em etapas de entrada, processamento intermediário e saída.

Figura 6 – Representação da arquitetura *Xception* com suas principais camadas.



Fonte: Autor.

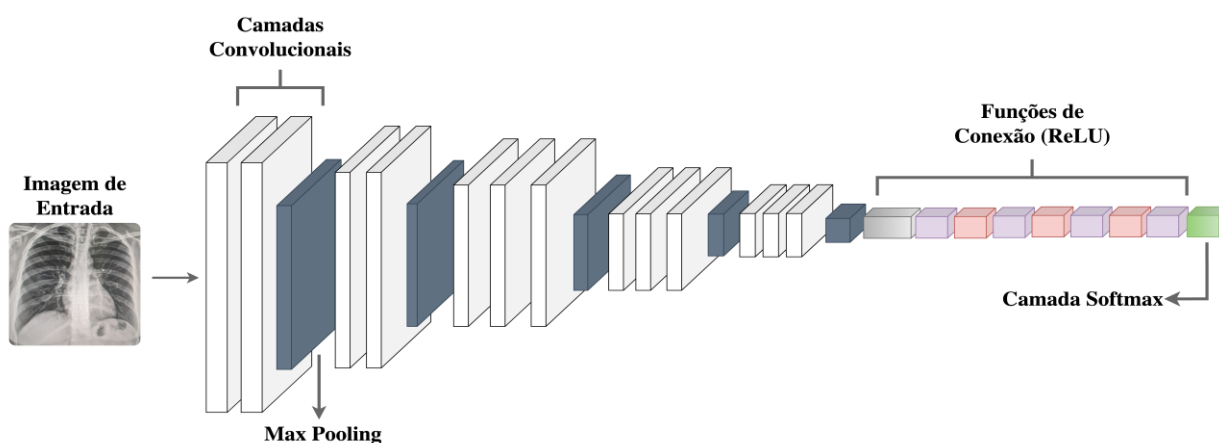
Como mostrado na Figura 6, a arquitetura utiliza blocos de convoluções separáveis por profundidade, responsáveis por aprimorar a eficiência e o desempenho do modelo.

4.1.1.2 VGG16 e VGG19

As arquiteturas *VGG16* e *VGG19*, desenvolvidas por Simonyan & Zisserman (2014). A *VGG16* possui 16 camadas ponderáveis e cerca de 138 milhões de parâmetros, enquanto a *VGG19* apresenta 19 camadas e aproximadamente 143 milhões de parâmetros. Ambas utilizam convoluções com filtros 3×3 empilhados, o que possibilita uma extração detalhada de características. Estudos anteriores (Hasan *et al.*, 2024) demonstram sua eficácia em tarefas como a detecção de pneumonia em radiografias de tórax, evidenciando seu potencial em sistemas de diagnóstico médico assistido por AI.

Na Figura 7, é apresentada uma representação resumida das arquiteturas *VGG16* e *VGG19*, destacando as pilhas de convoluções intercaladas com camadas de *pooling*.

Figura 7 – Representação da arquitetura *VGG16* com suas principais camadas.



Fonte: Autor.

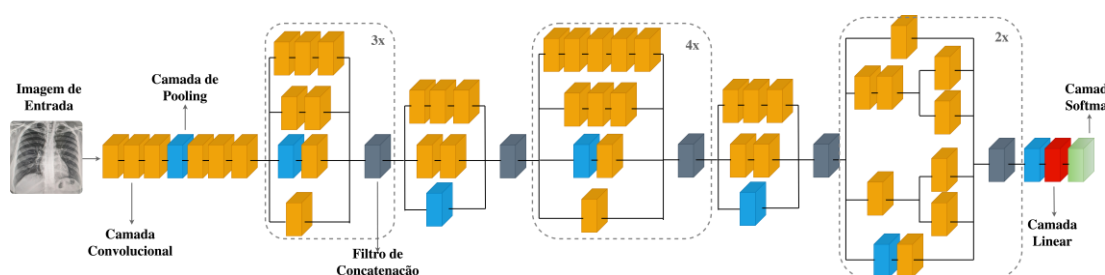
Como ilustrado na Figura 7, ambas as arquiteturas utilizam pilhas de convoluções intercaladas com camadas de *pooling*, responsáveis por reduzir a dimensionalidade e preservar as características mais relevantes das imagens.

4.1.1.3 InceptionV3

Desenvolvido por Szegedy *et al.* (2016), o modelo *InceptionV3* possui 48 camadas e aproximadamente 23 milhões de parâmetros. Sua principal inovação está na introdução dos módulos *Inception*, que permitem a extração de características em múltiplas escalas simultaneamente. Essa abordagem modular favorece uma representação mais eficiente e expressiva de informações visuais complexas, mostrando-se altamente eficaz em aplicações de diagnóstico por imagem (Bhadra; KAR, 2020).

Na Figura 8, é apresentada uma visão simplificada do *InceptionV3*, enfatizando os módulos *Inception* que extraem características em múltiplas escalas.

Figura 8 – Representação da arquitetura *InceptionV3* com suas principais camadas.



Fonte: Autor.

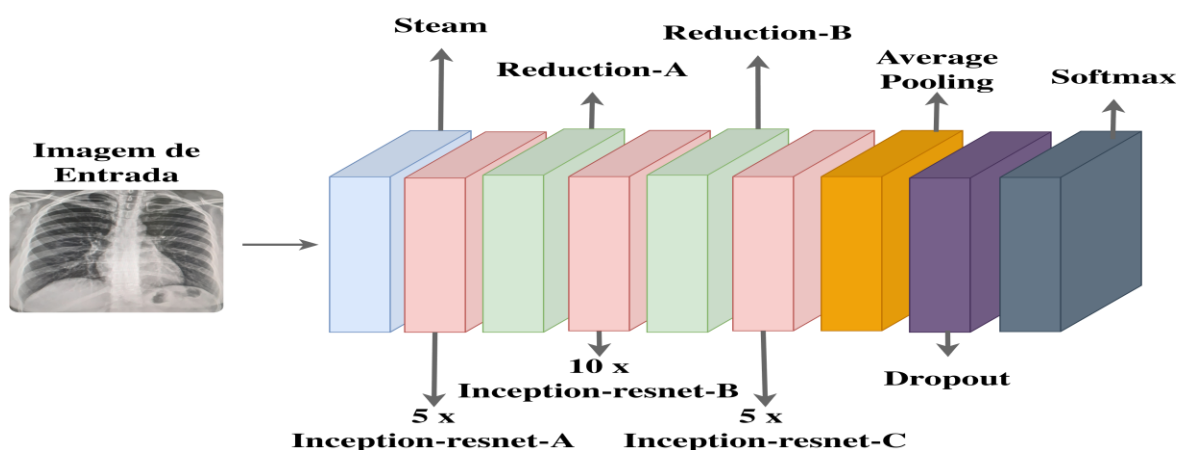
Como evidenciado na Figura 8, os módulos *Inception* realizam operações paralelas de convolução em diferentes tamanhos de filtro, permitindo a captura eficiente de informações em diversas escalas.

4.1.1.4 InceptionResNetV2

O modelo *InceptionResNetV2*, introduzido por Szegedy *et al.* (2017), combina a profundidade e a capacidade multiescala dos módulos *Inception* com as conexões residuais das redes *ResNet*. Com 164 camadas e aproximadamente 55 milhões de parâmetros, destaca-se pelo alto desempenho preditivo em tarefas que exigem elevada precisão, como o reconhecimento de padrões em exames radiológicos (Apostolopoulos *et al.*, 2020).

Na Figura 9, é ilustrada de forma resumida a arquitetura *InceptionResNetV2*, evidenciando a combinação de módulos *Inception* com conexões.

Figura 9 – Representação da arquitetura *InceptionResNetV2* com suas principais camadas.



Fonte: Autor.

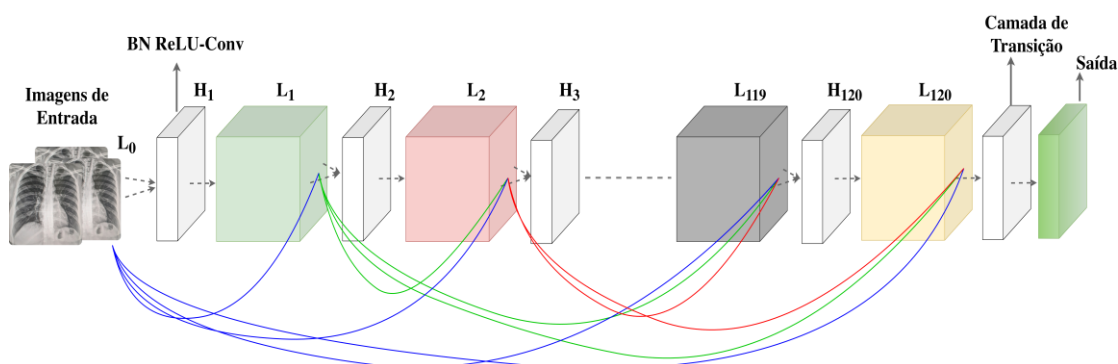
Como ilustrado na Figura 9, essa arquitetura combina a extração multiescala dos módulos *Inception* com as conexões residuais, favorecendo o aprendizado mais profundo e estável.

4.1.1.5 DenseNet121 e DenseNet169

As arquiteturas *DenseNet121* e *DenseNet169*, propostas por Huang *et al.* (2017), destacam-se pelo uso de conexões densas entre as camadas, o que favorece a reutilização de características e a redução no número de parâmetros. A *DenseNet121* possui 121 camadas e aproximadamente 8 milhões de parâmetros, enquanto a *DenseNet169* apresenta 169 camadas e cerca de 14 milhões de parâmetros. Ambas têm demonstrado alto desempenho em tarefas de diagnóstico por imagem, mesmo em cenários com desbalanceamento de dados, como evidenciado em (Hasan *et al.*, 2024; Diniz *et al.*, 2023).

Na Figura 10, é apresentada uma visão resumida das arquiteturas *DenseNet121* e *DenseNet169*, destacando o uso de conexões densas entre camadas.

Figura 10 – Representação da arquitetura *DenseNet121* com suas principais camadas.



Fonte: Autor.

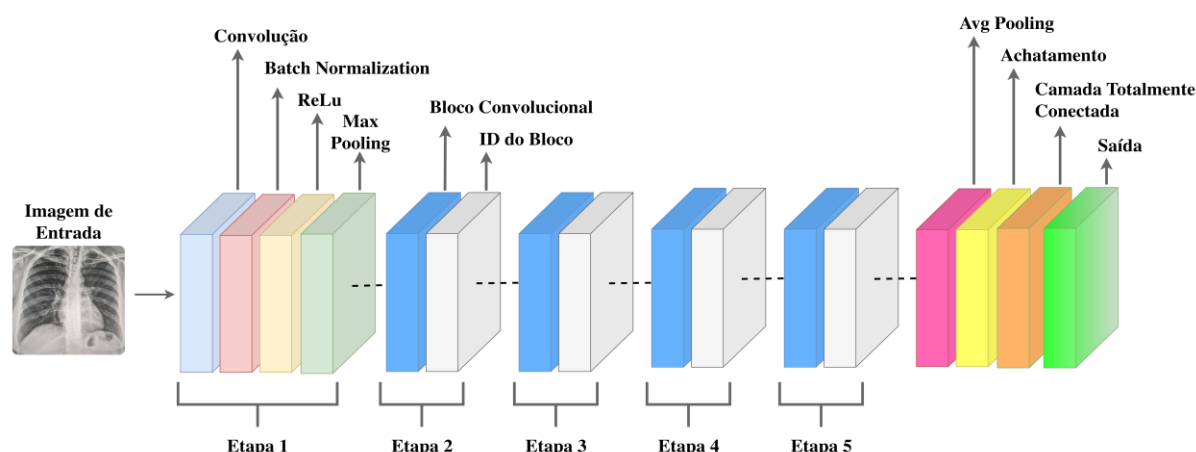
Como ilustrado na Figura 10, as *DenseNets* promovem conexões diretas entre todas as camadas, permitindo o reuso de *features* e aprimorando o fluxo de gradiente durante o treinamento.

4.1.1.6 ResNet50

As redes *ResNet50*, propostas por He *et al.* (2016), introduzem conexões residuais que possibilitam o treinamento de redes muito profundas sem perda de desempenho devido ao problema do gradiente evanescente. A *ResNet50* conta com 50 camadas e cerca de 23 milhões de parâmetros. Sua capacidade de aprendizado profundo as torna eficazes na detecção de padrões complexos em imagens médicas (Bhadra; KAR, 2020).

Na Figura 11, é apresentada uma visão simplificada da arquitetura *ResNet50*, destacando a presença dos blocos residuais.

Figura 11 – Representação da arquitetura *ResNet50* com suas principais camadas.



Fonte: Autor.

Como ilustrado na Figura 11, os blocos residuais permitem o reaproveitamento de informações entre camadas, facilitando o treinamento de redes mais profundas.

4.2 PARÂMETROS E TREINAMENTO

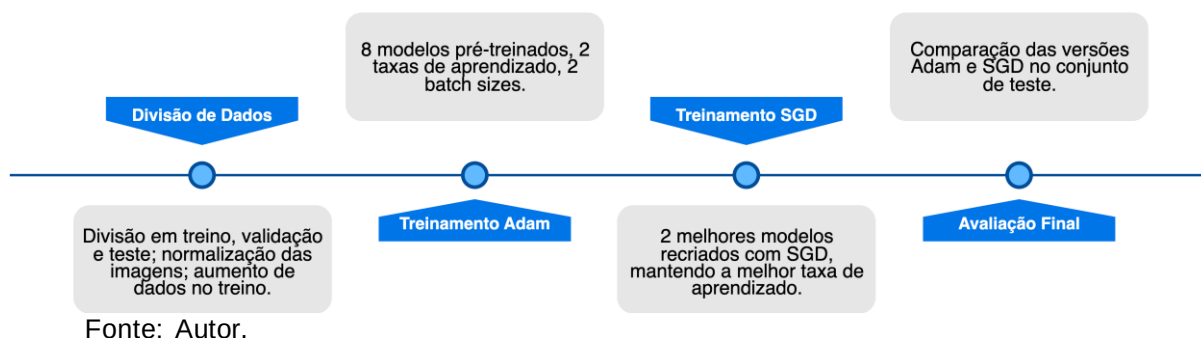
Os modelos foram treinados utilizando o otimizador Adam, testando duas taxas de aprendizado iniciais, 1×10^{-3} e 1×10^{-4} , valores frequentemente reportados na literatura por proporcionarem bom equilíbrio entre velocidade de convergência e estabilidade do treinamento (Sufian *et al.*, 2023). Também foram avaliados BS de 16 e 32, escolhidos por representarem configurações comumente utilizadas (Bressem *et al.*, 2020). Além disso, aplicou-se a estratégia de ajuste dinâmico da LR por meio do *callback ReduceLROnPlateau*, que reduz a taxa em 10% após duas épocas sem melhoria na *val_loss*, com limite mínimo de 1×10^{-6} .

Após os experimentos com o otimizador Adam, os dois modelos com melhor *F1-score* foram selecionados para testes adicionais com o otimizador SGD (com *momentum* de 0.9), utilizando a melhor LR identificada anteriormente. Essa etapa teve como objetivo avaliar se o uso do SGD poderia aprimorar o desempenho dos modelos, conforme indicado no estudo (Manickam *et al.*, 2021), que destacou sua eficácia na detecção de pneumonia.

4.3 ORDEM DOS EXPERIMENTOS

Os experimentos foram conduzidos em três etapas principais. Inicialmente, o conjunto de dados foi dividido em treino, validação e teste, sendo que técnicas de aumento de dados foram aplicadas apenas ao conjunto de treinamento. Em seguida, os oito modelos pré-treinados foram treinados utilizando o otimizador Adam, considerando duas LR e dois BS, totalizando 32 experimentos. Com base nos melhores resultados obtidos com Adam, os dois modelos mais promissores foram recriados utilizando o otimizador SGD, mantendo a melhor LR, resultando em dois experimentos adicionais. Por fim, todos os modelos foram avaliados no conjunto de teste, permitindo a comparação das métricas de desempenho entre as versões treinadas com Adam e SGD.

Na Figura 12, é apresentado um fluxograma que sintetiza as etapas desde a preparação dos dados até a avaliação final e a comparação entre versões treinadas com diferentes otimizadores.

Figura 12 – Fluxo dos experimentos realizados.

O pipeline descrito na Figura 12 abrange as etapas de pré-processamento e divisão dos dados, o treinamento inicial dos modelos com Adam, a seleção dos dois melhores modelos com base no *F1-score* e o retreinamento com SGD, e, por fim, a avaliação e comparação final.

Com o procedimento experimental consolidado e o pipeline visualizado, pas-samos à apresentação dos resultados. A seção seguinte detalha as métricas de desempenho (acurácia, precisão, recall e *F1-score*) para cada arquitetura testada, o impacto das variações de LR e BS, e a comparação entre os melhores modelos treinados com Adam e SGD.

5 RESULTADOS

Nesta seção, é apresentado o desempenho das arquiteturas de CNNs na detecção de pneumonia em radiografias de tórax. Os resultados são organizados em três etapas: i) avaliação individual de cada modelo usando o otimizador Adam, consi-derando diferentes LT e BS. ii) comparação entre os dois melhores modelos, usando otimizador SGD e considerando o *F1-Score* na escolha do melhor. iii) desenvolver um sistema interativo.

5.1 AVALIAÇÃO DOS MODELOS

A primeira etapa de avaliação envolveu a análise individual de cada modelo pré-treinado, considerando todas as combinações de LR e BS. Para cada configuração, foram calculadas métricas essenciais: acurácia, precisão, *Recall*, *F1-score*. Essa abordagem permite observar o impacto dos parâmetros de treinamento no desempenho de cada arquitetura.

A Tabela 2 sintetiza o desempenho do *Xception* nas combinações de BS e LR testadas, destacando a configuração que melhor equilibrou precisão e sensibilidade.

Tabela 2 – Avaliação do *Xception* em diferentes combinações de hiperparâmetros

Batch	LR	Acurácia (%)	Precisão (%)	Recall (%)	F1-score (%)
16	1×10^{-3}	90.0	91.0	90.0	90.0
16	1×10^{-4}	92.0	91.0	92.0	92.0
32	1×10^{-3}	91.0	91.0	91.0	91.0
32	1×10^{-4}	90.0	91.0	90.0	90.0

Fonte: Autor.

Conforme mostra a Tabela 2, o modelo *Xception* apresentou desempenho consistente em todas as combinações de hiperparâmetros, com acurácia variando entre 90,0% e 92,0% e *F1-score* variando entre 90,0% e 92,0%. A melhor configuração, com BS de 16 e LR de 1×10^{-4} , alcançou 92,0% em acurácia e *F1-score*, enquanto outras combinações mantiveram resultados próximos, mostrando estabilidade do modelo. Observa-se que, de modo geral, todas as variações mantiveram métricas equilibradas de precisão e *recall*, indicando que o modelo consegue distinguir de forma estável as classes Normal e Pneumonia.

Em termos de desempenho por classe, observa-se que as configurações com LR maior (1×10^{-3}) tenderam a apresentar maior *recall* para Pneumonia, chegando a 96,0%, enquanto a classe Normal apresentou *recall* entre 80,0% e 89,0%. Já as combinações com LR menor (1×10^{-4}), especialmente com BS de 16, mostraram resultados mais equilibrados, com *recall* de 92,0% para Normal e 91,0% para Pneumonia. Esses resultados sugerem que o *Xception* mantém boa generalização mesmo com variações moderadas de BS e LR, apresentando um desempenho consistente entre as classes.

Na Tabela 3, são resumidos os resultados do *VGG16*, permitindo observar como variações em BS e LR influenciaram as métricas principais.

Tabela 3 – Avaliação do *VGG16* em diferentes combinações de hiperparâmetros

Batch	LR	Acurácia (%)	Precisão (%)	Recall (%)	F1-score (%)
16	1×10^{-3}	88.0	89.0	88.0	88.0
16	1×10^{-4}	87.0	88.0	87.0	87.0
32	1×10^{-3}	93.0	93.0	93.0	93.0
32	1×10^{-4}	95.0	95.0	95.0	94.0

Fonte: Autor.

Conforme apresentado na Tabela 3, o *VGG16* apresentou um desempenho variado de 87,0% a 95,0% em acurácia e de 87,0% a 95,0% em *F1-score*. A melhor configuração, com BS de 32 e LR de 1×10^{-4} , alcançou 95,0% em acurácia e 94,0% em *F1-score*, enquanto outras combinações apresentaram resultados próximos, evidenciando estabilidade do modelo. Observa-se que o aumento do BS favoreceu melhorias consistentes nas métricas, especialmente em precisão e *F1-score*, demonstrando que o modelo se beneficia de mais amostras por atualização de gradiente.

Analisando o desempenho por classe, nota-se que o *recall* para Pneumonia variou entre 83,0% e 98,0%, enquanto para a classe Normal os valores ficaram entre 73,0% e 88,0%. Configurações com LR maior mantiveram um bom equilíbrio entre as classes, enquanto LR menor apresentou ligeiro aumento de *recall* para Pneumonia, mas redução para Normal. Esses resultados indicam que o *VGG16* consegue distinguir ambas as classes de maneira consistente, embora com desempenho global ligeiramente inferior ao *Xception*, principalmente em cenários com menor BS.

Na Tabela 4, são apresentados os resultados do *VGG19*, evidenciando a estabilidade do modelo frente às combinações avaliadas.

Tabela 4 – Avaliação do *VGG19* em diferentes combinações de hiperparâmetros

Batch	LR	Acurácia (%)	Precisão (%)	Recall (%)	F1-score (%)
16	1×10^{-3}	92.0	92.0	92.0	92.0
16	1×10^{-4}	93.0	93.0	93.0	93.0
32	1×10^{-3}	93.0	93.0	93.0	93.0
32	1×10^{-4}	92.0	92.0	92.0	92.0

Fonte: Autor.

Como pode ser observado na Tabela 4, todas as combinações resultaram em métricas próximas, indicando a estabilidade do modelo. O *VGG19* apresentou desempenho consistente em todas as combinações de hiperparâmetros, com acurácia variando entre 92,0% e 93,0% e *F1-score* ponderado entre 92,0% e 93,0%. A melhor configuração, com BS de 16 e LR de 1×10^{-4} , alcançou 93,0% em acurácia e *F1-score*, enquanto outras combinações mantiveram resultados próximos, evidenciando estabilidade do modelo. Observa-se que o aumento do BS trouxe ganhos modestos, especialmente nas métricas de precisão e *recall*, sugerindo que o modelo mais profundo consegue capturar melhor os exemplos positivos.

Considerando o desempenho por classe, o *recall* para Pneumonia variou entre 92,0% e 98,0%, enquanto a classe Normal apresentou *recall* entre 85,0% e 91,0%. Configurações com LR menor mostraram resultados ligeiramente mais equilibrados entre as classes, enquanto combinações com LR maior tenderam a favorecer a classe Pneumonia. Esses resultados indicam que o *VGG19* mantém boa capacidade de generalização, oferecendo desempenho consistente mesmo com variações moderadas de BS e LR, com ganhos discretos em relação ao *VGG16*.

Na Tabela 5, é mostrado o desempenho do *ResNet50* nas diferentes configurações, indicando maior variabilidade entre execuções em algumas combinações.

Tabela 5 – Avaliação do *ResNet50* em diferentes combinações de hiperparâmetros

Batch	LR	Acurácia (%)	Precisão (%)	Recall (%)	F1-score (%)
16	1×10^{-4}	70.0	72.0	70.0	66.0
32	1×10^{-3}	83.0	84.0	83.0	83.0
32	1×10^{-4}	84.0	84.0	84.0	83.0

Fonte: Autor.

Na Tabela 5 são apresentados os resultados obtidos com a arquitetura *Res-Net50*, que demonstrou desempenho relativamente estável, com acurácia variando entre 70,0% e 84,0% e *F1-score* ponderado entre 66,0% e 83,0%. A configuração com BS de 32 e LR de 1×10^{-4} obteve o melhor resultado, alcançando 84,0% em acurácia e 83,0% em *F1-score*, enquanto a configuração com menor BS (16) e LR 1×10^{-4} apresentou desempenho inferior, com acurácia de 70,0% e *F1-score* de 66,0%. Esses resultados indicam que o modelo se beneficia de um maior número de amostras por atualização de gradiente, refletindo sua capacidade de extrair características profundas mesmo em conjuntos de dados relativamente desbalanceados.

Analisando o desempenho por classe, observa-se que o *recall* para Pneumonia variou de 94,0% a 96,0%, enquanto a classe Normal apresentou *recall* mais baixo, entre 29,0% e 67,0%, evidenciando uma tendência do modelo a favorecer a detecção de casos positivos. Configurações com LR mais alta não mostraram vantagens significativas em relação a LR mais baixa, sugerindo que ajustes na LR tiveram impacto limitado nesse contexto. No geral, o *ResNet50* demonstrou desempenho aceitável, mas inferior às arquiteturas como *Xception* e *VGG*, principalmente na detecção equilibrada das duas classes.

Na Tabela 6, são reunidas as métricas do *InceptionV3*, destacando as configurações que melhor equilibraram precisão e *recall*.

Tabela 6 – Avaliação do *InceptionV3* em diferentes combinações de hiperparâmetros

Batch	LR	Acurácia (%)	Precisão (%)	Recall (%)	F1-score (%)
16	1×10^{-3}	91.0	91.0	91.0	91.0
16	1×10^{-4}	92.0	92.0	92.0	91.0
32	1×10^{-3}	87.0	89.0	87.0	87.0
32	1×10^{-4}	91.0	91.0	91.0	91.0

Fonte: Autor.

A Tabela 6 indica desempenho sólido para o *InceptionV3*, com acurácia variando entre 87,0% e 92,0% e *F1-score* entre 87,0% e 91,0%. A melhor configuração foi obtida com BS de 16 e LR de 1×10^{-4} , alcançando 92,0% em acurácia e 91,0% em *F1-score*. De modo geral, as combinações com BS de 16 apresentaram resultados superiores, sugerindo que atualizações de gradiente mais frequentes contribuíram para um melhor ajuste do modelo.

Em relação ao desempenho por classe, observa-se que as configurações com LR menor apresentaram maior equilíbrio entre as classes, com *recall* de 85,0% para Normal e 95,0% para Pneumonia. Já com LR mais alta e BS de 32, houve leve queda de desempenho, com *recall* reduzido para a classe Normal (68,0%) e aumento na sensibilidade para Pneumonia (99,0%). Esses resultados indicam que o *InceptionV3* mantém boa capacidade de generalização e estabilidade, destacando-se pelo equilíbrio entre precisão e *recall* nas condições mais favoráveis de treinamento.

Na Tabela 7, é sintetizado o comportamento do *InceptionResNetV2*, evidenciando ganhos expressivos quando treinado com taxas de aprendizado menores.

Tabela 7 – Avaliação do *InceptionResNetV2* em diferentes combinações de hiperparâmetros

Batch	LR	Acurácia (%)	Precisão (%)	Recall (%)	F1-score (%)
16	1×10^{-3}	83.0	86.0	83.0	82.0
16	1×10^{-4}	95.0	95.0	95.0	95.0
32	1×10^{-3}	90.0	90.0	90.0	90.0
32	1×10^{-4}	94.0	94.0	94.0	94.0

Fonte: Autor.

A Tabela 7 revela um ótimo desempenho do *InceptionResNetV2*, com acurácia variando entre 83,0% e 95,0% e *F1-score* entre 82,0% e 95,0%. A melhor configuração foi obtida com BS de 16 e LR de 1×10^{-4} , alcançando 95,0% em acurácia e *F1-score*, demonstrando elevada capacidade de generalização e estabilidade. De modo geral, o modelo se beneficiou de taxas de aprendizado menores, que favoreceram um ajuste mais preciso dos pesos durante o treinamento.

Em termos de desempenho por classe, a configuração ideal apresentou *recall* de 92,0% para a classe Normal e 97,0% para Pneumonia, indicando sensibilidade equilibrada entre ambas. Já nas configurações com LR mais alta, observou-se redução no desempenho, com acurácia em torno de 83,0% a 90,0%, especialmente pela queda no *recall* da classe Normal. Esses resultados reforçam que o *InceptionResNetV2*, quando ajustado com *learning rate* mais baixo, consegue explorar sua profundidade e estrutura híbrida de forma eficiente, garantindo excelente desempenho na detecção de Pneumonia.

Na Tabela 8, são resumidos os resultados do *DenseNet121*, mostrando a sensibilidade do modelo a certas combinações de hiperparâmetros.

Tabela 8 – Avaliação do *DenseNet121* em diferentes combinações de hiperparâmetros

Batch	LR	Acurácia (%)	Precisão (%)	Recall (%)	F1-score (%)
16	1×10^{-3}	91.0	92.0	91.0	90.0
16	1×10^{-4}	93.0	94.0	93.0	93.0
32	1×10^{-3}	93.0	93.0	93.0	93.0
32	1×10^{-4}	81.0	84.0	81.0	79.0

Fonte: Autor.

A Tabela 8 mostra que o *DenseNet121* apresentou desempenho consistente, com acurácia variando entre 81,0% e 93,0% e *F1-score* entre 79,0% e 93,0%. A melhor configuração foi obtida com BS de 16 e LR de 1×10^{-4} , alcançando 93,0% em acurácia e *F1-score*, indicando equilíbrio entre sensibilidade e precisão. De modo geral, o modelo se mostrou sensível a variações na LR, com queda expressiva nas métricas quando combinada com BS maior e *learning rate* mais baixo.

Nas melhores execuções, a classe Normal atingiu *recall* de até 85,0%, enquanto Pneumonia chegou a 98,0%, evidenciando forte desempenho na detecção de casos positivos. Apesar da leve assimetria entre as classes, o *DenseNet121* demonstrou boa capacidade de generalização e estabilidade durante o treinamento, mantendo resultados robustos em diferentes combinações de hiperparâmetros.

Na Tabela 9, são apresentados os resultados do *DenseNet169*, ressaltando sua consistência e elevado desempenho nas combinações testadas.

Tabela 9 – Avaliação do *DenseNet169* em diferentes combinações de hiperparâmetros

Batch	LR	Acurácia (%)	Precisão (%)	Recall (%)	F1-score (%)
16	1×10^{-3}	93.0	93.0	93.0	93.0
16	1×10^{-4}	94.0	94.0	94.0	94.0
32	1×10^{-3}	94.0	94.0	94.0	94.0
32	1×10^{-4}	93.0	93.0	93.0	93.0

Fonte: Autor.

Por fim, como mostrado na Tabela 9, o modelo *DenseNet169* o mais profundo da família DenseNet, apresentou resultados consistentes e de alto desempenho em todas as combinações de hiperparâmetros, com acurácia variando entre 93,0% e 94,0%. As métricas de precisão, *recall* e *F1-score* mantiveram-se equilibradas, também entre 93,0% e 94,0%, indicando estabilidade no processo de generalização.

Observou-se que a melhor configuração foi obtida com BS 32 e LR de 1×10^{-3} , alcançando precisão de 94%, *recall* de 94% e *F1-score* de 94%. Essa configuração destacou-se principalmente pelo bom equilíbrio entre as classes, refletindo a capacidade da arquitetura em extrair características mais refinadas e fornecer previsões mais confiáveis, sobretudo na distinção entre casos normais e de pneumonia.

De forma geral, os resultados mostraram desempenho consistente entre as arquiteturas avaliadas. Os modelos *VGG16* (Tabela 3) e *InceptionResNetV2* (Tabela 7) alcançaram as maiores acurácias, próximas de 95%, seguidos por *DenseNet169* (Tabela 9, 94%) e *DenseNet121* (Tabela 8, 93%).

As redes *Xception* (Tabela 2) e *InceptionV3* (Tabela 6) mantiveram resultados estáveis na faixa de 90–92%, enquanto a *ResNet50* (Tabela 5) apresentou variação mais acentuada, com acurácia entre 70% e 84%, dependendo da combinação de hiperparâmetros. Esses resultados indicam que maior profundidade nem sempre implica melhor desempenho, sobretudo em bases de dados menores.

Observou-se também que o BS de 32 e a LR de 1×10^{-4} favoreceram a maioria dos modelos, resultando em métricas mais equilibradas entre precisão, *recall* e *F1-score*. Por exemplo, o *Xception* (Tabela 2) com essa configuração atingiu *F1-score* de 92% tanto para Normal quanto para Pneumonia, enquanto no *VGG16* (Tabela 3) a mesma combinação elevou a precisão para 95% e *F1-score* para 94%. Modelos mais profundos, como *DenseNet169* (Tabela 9) e *InceptionResNetV2* (Tabela 7), apresentaram valores de *recall* para Pneumonia próximos de 98–99%, mas com maior variação para a classe Normal (entre 85% e 92%), evidenciando sensibilidade desigual entre as classes. Esses resultados reforçam a importância de analisar individualmente o desempenho por classe, sobretudo em bases desbalanceadas.

Em resumo, modelos intermediários como *VGG16* (Tabela 3) e *InceptionRes-NetV2* (Tabela 7) mostraram o melhor equilíbrio entre desempenho e estabilidade, enquanto redes mais profundas, como *DenseNet169* (Tabela 9), exigem ajustes finos para evitar sobreajuste. As arquiteturas mais leves, por sua vez, mantiveram boa generalização e menor sensibilidade a hiperparâmetros. A seguir, os melhores modelos serão treinados com SGD, mantendo as melhores taxas de aprendizado.

5.2 COMPARAÇÃO DE DESEMPENHO

A segunda etapa consistiu na comparação entre os modelos, considerando o impacto de diferentes configurações de treinamento sobre as métricas de desempenho. Foram selecionadas as melhores combinações de BS e LR identificadas anteriormente, sendo *VGG16* com BS igual a 32 e LR de 1×10^{-4} e *InceptionResNetV2* com BS igual a 16 e LR de 1×10^{-4} . Cada um desses modelos foi avaliado na versão treinada com Adam e com SGD, mantendo as respectivas combinações de BS e LR.

Na Tabela 10, são comparados os *F1-scores* ponderados dos dois melhores modelos treinados com Adam e com SGD, facilitando a análise do impacto do otimizador.

Tabela 10 – Comparação dos dois melhores modelos com otimizador SGD.

Modelo	Otimizador	Precision (%)	Recall (%)	F1-score (%)
InceptionResNetV2	Adam	95.0	95.0	95.0
InceptionResNetV2	SGD	96.0	96.0	96.0
VGG16	Adam	95.0	95.0	94.0
VGG16	SGD	91.0	90.0	90.0

Fonte: Autor.

Observa-se comportamentos distintos ao comparar as versões treinadas com Adam e com SGD (Tabela 10). O *InceptionResNetV2* melhorou com SGD, passando de *F1-score* ponderado de 95% (Adam) para 96% (SGD), com precisão e *recall* igualmente elevados (96%). O detalhamento por classe é apresentado na Tabela 11, onde se nota avanço tanto na classe *Normal* quanto em *Pneumonia*, além de maior estabilidade de generalização durante o treinamento.

Na Figura 13, são apresentadas as matrizes de confusão do modelo *Inception-ResNetV2* treinado com os otimizadores Adam e SGD, respectivamente.

Figura 13 – Matrizes de confusão — *InceptionResNetV2* (conjunto de teste)

(a) Adam			(b) SGD		
	Previsto Normal	Previsto Pneumonia		Previsto Normal	Previsto Pneumonia
Verdadeiro Normal	215	19	Verdadeiro Normal	220	14
Verdadeiro Pneumonia	10	380	Verdadeiro Pneumonia	12	378

Fonte: Autor.

A partir da análise nas matrizes de confusão, Figura 13, é possível afirmar para o *InceptionResNetV2*, a versão com Adam registrou 380 verdadeiros positivos e 10 falsos negativos para *Pneumonia*, enquanto a versão com SGD apresentou 378 verdadeiros positivos e 12 falsos negativos, com o *recall* para *Pneumonia* sofreu variação de 97% para 96.9%), mas a precisão aumentou de 95.2% para 96.4%, explicando o ganho no *F1-score* ponderado observado ao trocar para SGD. Em outras palavras, SGD elevou a confiança nas previsões positivas (menos falsos positivos) ao custo de uma pequena queda no *recall* de *Pneumonia*, resultando em melhor equilíbrio global para essa arquitetura.

Na Figura 14, são apresentadas as matrizes de confusão do modelo *VGG16* treinado com os otimizadores Adam e SGD, respectivamente.

Figura 14 – Matrizes de confusão — VGG16 (conjunto de teste)

(a) Adam			(b) SGD		
	Previsto Normal	Previsto Pneumonia		Previsto Normal	Previsto Pneumonia
Verdadeiro Normal	207	27	Verdadeiro Normal	217	17
Verdadeiro Pneumonia	7	383	Verdadeiro Pneumonia	44	346

Fonte: Autor.

Já no modelo VGG16, ilustrado na Figura 14, a troca para SGD produziu um efeito adverso: com Adam houve 383 verdadeiros positivos e apenas 7 falsos negativos (*recall* Pneumonia 98%), enquanto com SGD os verdadeiros positivos caíram para 346 e os falsos negativos aumentaram para 44 (*recall* 88.7%), causando a expressiva redução do *F1-score* ponderado registrada na Tabela 10. Apesar de SGD ter aumentado o número de verdadeiros negativos para a classe Normal (maior *recall* para Normal), o aumento de falsos negativos para Pneumonia representa um *trade-off* clínico desfavorável, pois perder casos positivos tem custo diagnóstico significativo.

Em suma, as matrizes reforçam a conclusão anterior: o *InceptionResNetV2* apresenta benefício ao treinar com SGD evidenciando melhor equilíbrio entre preci-são, *recall* e *F1-score* global, enquanto o VGG16 mostrou-se sensível à escolha do otimizador, com perda de sensibilidade para Pneumonia quando treinado com SGD. Esses achados reforçam a necessidade de considerar custos por classe, especialmente falsos negativos em cenários clínicos, ao selecionar otimizadores. Caso a troca para SGD seja adotada em modelos sensíveis, recomenda-se ajustar taxas de aprendizado, critérios de parada e/ou estratégias de balanceamento para minimizar falsos negativos.

Na Tabela 11, é detalhado o desempenho por classe do *InceptionResNetV2* nas versões com Adam e com SGD, permitindo avaliar o equilíbrio entre Normal e Pneumonia.

Tabela 11 – Desempenho por classe do InceptionResNetV2 com Adam e SGD.

Classe	Otimizador	Precision (%)	Recall (%)	F1-score (%)
Normal	Adam	96	92	94
Normal	SGD	95	94	94
Pneumonia	Adam	95	97	96
Pneumonia	SGD	96	97	97

Fonte: Autor.

Conforme Tabela 11, o ganho com SGD no *InceptionResNetV2* expressa-se tanto na classe Normal quanto em Pneumonia, com melhoria no equilíbrio entre precisão e recall. Em contraste, o *VGG16* mostrou queda de desempenho ao ser treinado com SGD, reduzindo o *F1-score* ponderado de 94% (Adam) para 90% (SGD), evidenciando sensibilidade maior à escolha do otimizador.

Embora o *recall* para a classe Normal tenha aumentado de 88% para 93%, houve perda de sensibilidade para Pneumonia de 98% para 89%, o que explica a redução da métrica. A versão com SGD do *VGG16* obteve *weighted avg* próximo de 90–91%, com *precision* de 91%, *recall* de 90%, *F1-score* de 90%, indicando que, no caso dessa arquitetura, a troca para SGD introduziu um *trade-off* entre equilíbrio por classe e desempenho.

Observou-se que o uso de SGD exigiu mais etapas/épocas de treino para convergência, porém trouxe comportamento de perda mais suave e generalização mais consistente no *InceptionResNetV2*. Isso sugere que o impacto do otimizador é dependente da arquitetura.

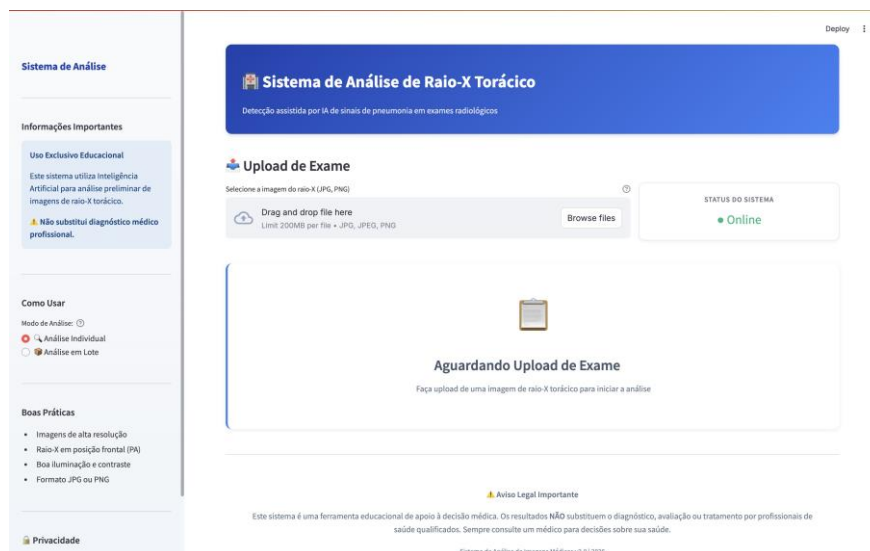
Em síntese, a comparação confirma que o ajuste do otimizador pode melhorar a generalização, como observado no *InceptionResNetV2*, embora o efeito não seja uniforme entre as arquiteturas. Esses resultados reforçam a importância de considerar o comportamento individual de cada modelo ao definir as estratégias de treinamento. Em seguida, é apresentado o sistema desenvolvido, que permite o envio e a análise automatizada de imagens de radiografias.

5.1.1 Implementação do Sistema de Análise de Radiografias

Como etapa final do estudo, foi desenvolvido um sistema interativo em *Python* com interface web em *Streamlit* para análise de radiografias torácicas. O modelo utilizado foi *InceptionResNetV2* com otimizador SGD (BS = 32, LR = 1×10^{-4}), apresentando *F1-score* médio de 96%. O carregamento do modelo é realizado a partir de *checkpoints* e a aplicação utiliza mecanismos de cache para melhorar a responsividade. O pré-processamento e I/O de imagens empregam *Pillow* e utilitários do *Keras*, enquanto as operações numéricas são realizadas com *NumPy*. O sistema permite o envio de imagens (JPG/JPEG/PNG) e realiza a classificação entre Normal e Pneumonia, exibindo a probabilidade de acerto e alertas visuais de confiança.

Na Figura 15, observa-se a tela inicial do sistema, onde o usuário é orientado a enviar uma imagem individual para análise.

Figura 15 – Tela inicial — aguardando upload de exame individual.

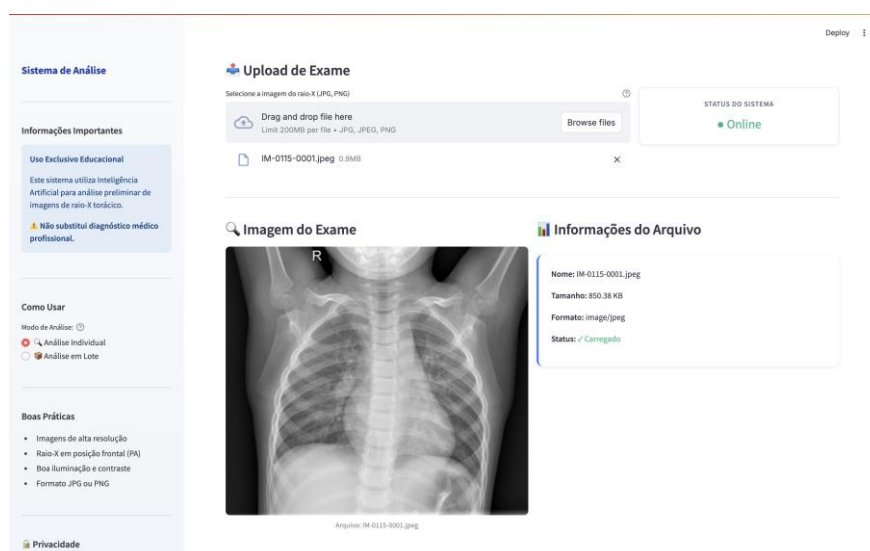


Fonte: Autor.

Como mostrado na Figura 15, a interface destaca o estado de espera, apresenta instruções essenciais e prepara o usuário para o fluxo de análise, garantindo clareza e previsibilidade na interação.

Na Figura 16, observa-se o fluxo de upload individual, no qual a imagem enviada é exibida juntamente com seus metadados antes do início da análise.

Figura 16 – Exemplo de upload de imagem individual, com pré-visualização e metadados do arquivo.

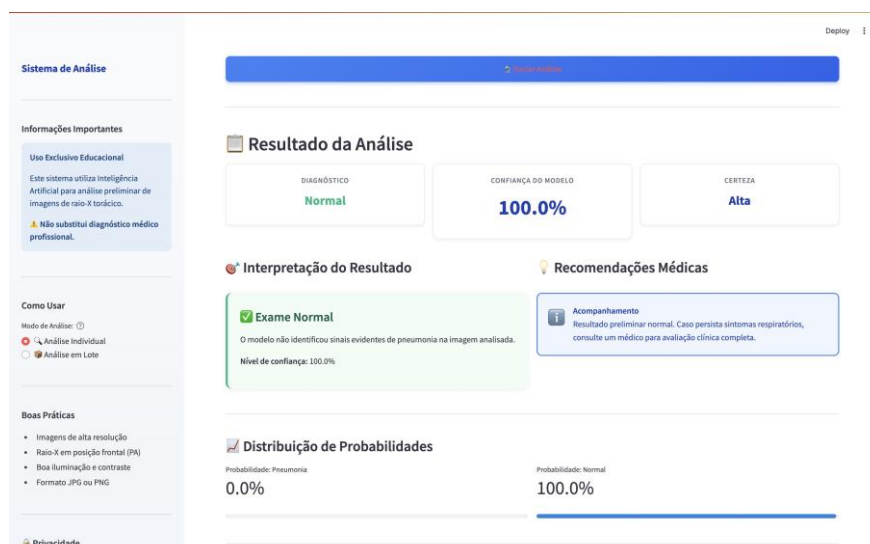


Fonte: Autor.

Conforme mostrado na Figura 16, o sistema permite ao usuário confirmar o exame carregado e verificar informações básicas do arquivo, tornando o processo de análise mais transparente e seguro.

Na Figura 17, observa-se o resultado de uma análise em que a imagem é classificada como *Normal*. O sistema exibe o rótulo previsto, a probabilidade correspondente e orientações relevantes para o usuário.

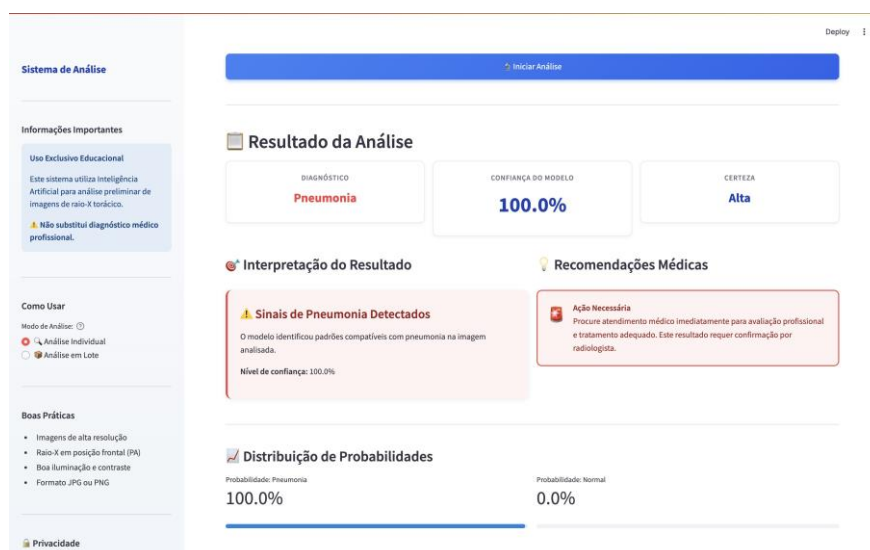
Figura 17 – Resultado de análise indicando *Normal*



Fonte: Autor.

Como ilustrado na Figura 17, a interface destaca o diagnóstico previsto e o nível de confiança, permitindo uma interpretação imediata e orientando o usuário sobre possíveis condutas, mesmo diante de um resultado normal.

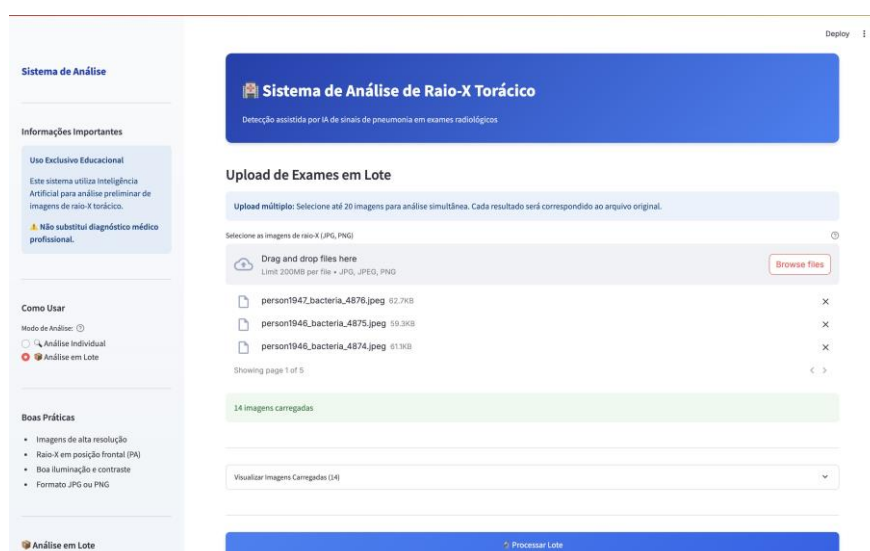
Na Figura 18, observa-se um resultado em que a imagem é classificada como *Pneumonia*, incluindo alerta visual, nível de confiança e orientações clínicas de enca-minhamento.

Figura 18 – Resultado de análise indicando *Pneumonia*

Fonte: Autor.

Como mostrado na Figura 18, o sistema evidencia o diagnóstico suspeito por meio de realces visuais e apresenta recomendações para avaliação presencial e confirmação especializada, reforçando a importância do acompanhamento clínico.

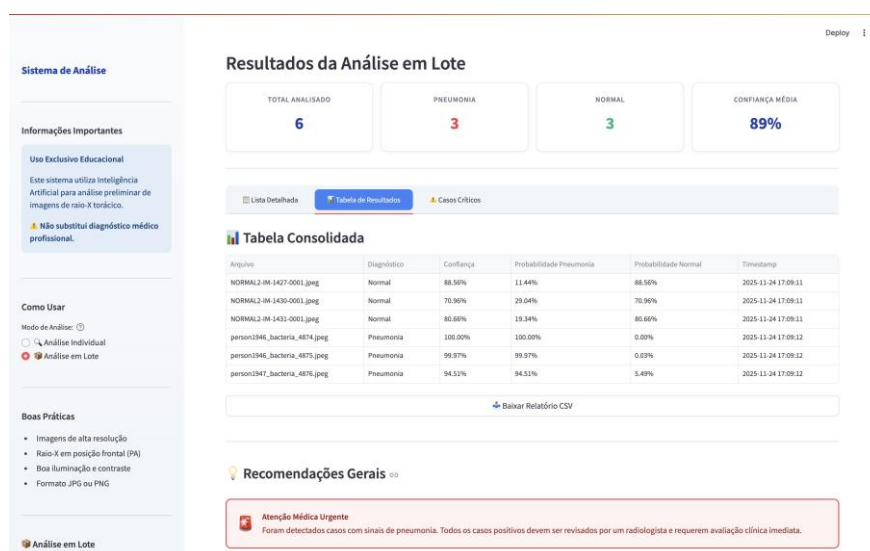
O sistema também disponibiliza um modo de processamento em lote. Na Figura 19, observa-se a tela de upload múltiplo, na qual o usuário pode selecionar até 20 imagens para análise conjunta.

Figura 19 – Tela de upload em lote.

Fonte: Autor.

Como ilustrado na Figura 19, a interface permite carregar vários exames simultaneamente, facilitando fluxos de trabalho que envolvem grandes volumes de imagens. Por fim, na Figura 20, observa-se a tela de resultados do processamento em lote, incluindo estatísticas gerais, visualização individual dos exames, tabela consolidada e opção de exportação em CSV.

Figura 20 – Resultados do processamento em lote.



Fonte: Autor.

Como ilustrado na Figura 20, o sistema sintetiza o desempenho do lote (total analisado, quantidades por classe e confiança média) ao mesmo tempo em que permite explorar cada exame individualmente e exportar os dados, tornando o processo adequado para avaliações em escala e registro documental.

Os resultados apresentados mostram que algumas arquiteturas pré-treinadas podem oferecer desempenho elevado e estável para análise de pneumonia em radiografias, com destaque o *InceptionResNetV2* treinado com o otimizador SGD e alcançando *F1-score* de 96%, além do *VGG16*, que também obteve resultados expressivos em suas melhores configurações. Complementado pelo protótipo em *Streamlit*, esse conjunto de resultados fornece base para avaliar aplicabilidades clínicas, mas também revela limitações relacionadas ao balanceamento de classes e à necessidade de validação externa. No capítulo de conclusão sintetizamos essas observações, destacamos as contribuições do trabalho e apontamos direções prioritárias para futuros estudos.

6 CONCLUSÃO

O objetivo deste trabalho foi desenvolver um sistema computacional interativo para auxílio no diagnóstico de pneumonia em radiografias de tórax, utilizando técnicas de DL com TL e *fine-tuning*. Os objetivos gerais e específicos foram atingidos por meio da implementação de uma solução completa que integra classificação automatizada baseada em CNNs pré-treinadas e interface web acessível.

A análise experimental revelou que arquiteturas como VGG16 e InceptionResNetV2 apresentaram desempenho superior, alcançando 95% em acurácia/*F1-score* nas melhores execuções. O InceptionResNetV2 treinado com o otimizador SGD, com BS de 32 e LR de 1×10^{-4} se destacou, alcançando um *F1-score* de 96%, sendo integrado ao sistema desenvolvido em *Streamlit*. A análise por classe revelou maior sensibilidade para a detecção de pneumonia, demonstrando a capacidade do sistema em auxiliar profissionais de saúde na triagem de casos suspeitos em ambientes clínicos.

6.1 TRABALHOS FUTUROS

Em trabalhos futuros, espera-se aprimorar e diversificar as bases de dados; aplicar segmentação para focar regiões pulmonares relevantes; e conduzir estudos com equipes clínicas. Essas medidas visam aumentar a robustez e a aplicabilidade clínica das soluções desenvolvidas.

REFERÊNCIAS

- Ali, R. M. M. *et al.* Post-covid-19 pneumonia lung fibrosis: a worrisome sequelae in surviving patients. **Egypt Journal of Radiology and Nuclear Medicine**, v. 52, n. 1, p. 101, 2021. PMID: PMC8047597; Acesso em: 20 ago. 2025. Disponível em: <<https://pmc.ncbi.nlm.nih.gov/articles/PMC8047597/>>. Citado na página 13.
- Alzubaidi, L. *et al.* Review of deep learning: concepts, cnn architectures, challenges, applications, future directions. **Journal of big Data**, Springer, v. 8, n. 1, p. 53, 2021. Citado na página 11.
- Apostolopoulos, I. D. *et al.* Covid-19: automatic detection from x-ray images utilizing transfer learning with convolutional neural networks. **Physical and engineering sciences in medicine**, Springer, v. 43, n. 2, p. 635–640, 2020. Citado 3 vezes nas páginas 20, 22 e 27.
- Becker, J. *et al.* Artificial intelligence-based detection of pneumonia in chest radiographs. **Diagnostics**, MDPI, v. 12, n. 6, p. 1465, 2022. Citado na página 15.
- Bhadra, R. *et al.* Covid detection from cxr scans using deep multi-layered cnn. In: **2020 IEEE Bombay Section Signature Conference (IBSSC)**. Mumbai, India: IEEE, 2020. p. 214–218. Citado 4 vezes nas páginas 20, 22, 27 e 29.
- Braga, A. V. *et al.* Machine learning: o uso da inteligência artificial na medicina. **Brazilian Journal of Development**, v. 5, n. 9, p. 16407–16413, 2019. Citado na página 11.

Bressem, K. K. *et al.* Comparing different deep learning architectures for classification of chest radiographs. **Scientific reports**, Nature Publishing Group UK London, v. 10, n. 1, p. 13590, 2020. Citado na página 29.

Chollet, F. Xception: Deep learning with depthwise separable convolutions. In: **Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)**. Honolulu, HI, USA: IEEE, 2017. p. 1251–1258. Citado na página 25.

Chollet, F. **Deep Learning with Python**. 2nd. ed. Shelter Island, NY: Manning Publications, 2021. ISBN 9781617296864. Citado na página 24.

Chong, P. L. *et al.* Integrating artificial intelligence in healthcare: applications, challenges, and future directions. **Future Science OA**, Taylor & Francis, v. 11, n. 1, p. 2527505, 2025. Citado na página 15.

Chumbita, M. *et al.* Can artificial intelligence improve the management of pneumonia.

Journal of clinical medicine, MDPI, v. 9, n. 1, p. 248, 2020. Citado na página 10.

Connor, K. L. *et al.* The future role of machine learning in clinical transplantation.

Transplantation, LWW, v. 105, n. 4, p. 723–735, 2021. Citado na página 15.

Departamento de Informática do SUS – DATASUS. **Mortalidade por pneumonia no Brasil - Dados de 2021 e 2022**. 2022. Acesso em: 6 jun. 2025. Disponível em: <<https://datasus.saude.gov.br/>>. Citado na página 10.

Diniz, J. O. B. *et al.* Detecção de covid-19 em imagens de raio-x de tórax através de seleção automática de pré-processamento e de rede neural convolucional. In: SBC. **Anais do 23º Simpósio Brasileiro de Computação Aplicada à Saúde (SBCAS)**. São Paulo, SP: Sociedade Brasileira de Computação, 2023. p. 162–173. ISSN 2763-8952. Citado 4 vezes nas páginas 21, 22, 25 e 28.

Esteva, A. *et al.* Deep learning-enabled medical computer vision. **Nature Medicine**, 2021. Citado 2 vezes nas páginas 16 e 17.

Fu, L. *et al.* Deep learning tomographic reconstruction through hierarchical decomposition of domain transforms. **Visual Computing for Industry, Biomedicine, and Art**, Springer, v. 5, n. 1, p. 30, 2022. Citado na página 16.

Fundo das Nações Unidas para a Infância (UNICEF). **Pneumonia em Crianças: Dados e Estatísticas**. Nova Iorque: Fundo das Nações Unidas para a Infância, 2024.<<https://data.unicef.org/topic/child-health/pneumonia/>>. Publicado em 2024. Acessado em 24 nov. 2025. Citado na página 10.

Guo, K. *et al.* Diagnosis and detection of pneumonia using weak-label based on x-ray images: a multi-center study. **BMC Medical Imaging**, Springer, v. 23, n. 1, p. 209, 2023. Citado na página 24.

Gupta, A. *et al.* Instacovnet-19: A deep learning classification model for the detection of covid-19 patients using chest x-ray. **Applied Soft Computing**, Elsevier, v. 99, p. 106859, 2021. Citado 2 vezes nas páginas 20 e 22.

Hasan, M. R. *et al.* Recent advancement of deep learning techniques for pneumonia prediction from chest x-ray image. **Medical Reports**, Elsevier, v. 6, p. 100106, 2024. ISSN 2949-9186. Citado 4 vezes nas páginas 21, 22, 26 e 28.

He, K. *et al.* Deep residual learning for image recognition. In: **Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)**. Las Vegas, NV, USA: IEEE, 2016. p. 770–778. Citado na página 29.

Hossain, M. *et al.* Artificial intelligence in healthcare: applications and future directions. **BMC Medical Informatics**, 2024. Citado na página 18.

Huang, G. *et al.* Densely connected convolutional networks. In: **Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)**. Honolulu, HI, USA: IEEE, 2017. p. 4700–4708. Citado na página 28.

Hwang, E. J. *et al.* Conventional versus artificial intelligence-assisted interpretation of chest radiographs in patients with acute respiratory symptoms in emergency department: a pragmatic randomized clinical trial. **Korean Journal of Radiology**, v. 24, n. 3, p. 259, 2023. Citado na página 15.

Ibrahim, A. U. *et al.* Pneumonia classification using deep learning from chest x-ray images during covid-19. **Cognitive computation**, Springer, p. 1–13, 2021. Citado na página 10.

Karimi, D. *et al.* Transfer learning in medical image analysis: A survey. **Medical Image Analysis**, 2021. Citado 2 vezes nas páginas 17 e 19.

Kermany, D. S. *et al.* Identifying medical diagnoses and treatable diseases by image-based deep learning. **cell**, Elsevier, v. 172, n. 5, p. 1122–1131, 2018. Citado 3 vezes nas páginas 11, 14 e 24.

Khan, A. *et al.* A survey of the recent architectures of deep convolutional neural networks. **Artificial intelligence review**, Springer, v. 53, p. 5455–5516, 2020. Citado 2 vezes nas páginas 11 e 16.

Kourounis, G. *et al.* Computer image analysis with artificial intelligence: a practical introduction to convolutional neural networks for medical professionals. **Postgraduate medical journal**, Oxford University Press, v. 99, n. 1178, p. 1287–1294, 2023. Citado 2 vezes nas páginas 15 e 17.

Krichen, M. Convolutional neural networks: A survey. **Computers**, MDPI, v. 12, n. 8, p. 151, 2023. Citado na página 18.

Kumar, Y. *et al.* Artificial intelligence in disease diagnosis: a systematic literature review, synthesizing framework and future research agenda. **Journal of ambient intelligence and humanized computing**, Springer, v. 14, n. 7, p. 8459–8486, 2023. Citado na página 15.

Liu, X. *et al.* A review of deep-learning-based medical image segmentation methods. **Sustainability**, MDPI, v. 13, n. 3, p. 1224, 2021. Citado na página 16.

Mahesse, G. J. C. Detecção e classificação de pneumonia com base na análise de imagens de radiografia torácica: caso de estudo-hospital central de maputo. Universidade Eduardo Mondlane, 2023. Citado 2 vezes nas páginas 10 e 14.

Mall, P. K. *et al.* A comprehensive review of deep neural networks for medical image processing: Recent developments and future opportunities. **Healthcare Analytics**, Elsevier, v. 4, p. 100216, 2023. Citado na página 23.

Manickam, A. *et al.* Automated pneumonia detection on chest x-ray images: A deep learning approach with different optimizers and transfer learning architectures. **Measurement**, Elsevier, v. 184, p. 109953, 2021. Citado na página 30.

Mayo Clinic Staff. **Pneumonia – Symptoms and causes**. 2020. <<https://www.mayoclinic.org/diseases-conditions/pneumonia/symptoms-causes/syc-20354204>>. Acesso em: 20 ago. 2025. Citado na página 13.

Mayo Clinic Staff. **Pulmonary fibrosis – Symptoms and causes**. 2024. <<https://www.mayoclinic.org/diseases-conditions/pulmonary-fibrosis/symptoms-causes/syc-20353690>>. Acesso em: 20 ago. 2025. Citado na página 13.

Narin, A. *et al.* Automatic detection of coronavirus disease (covid-19) using x-ray images and deep convolutional neural networks. **Pattern Analysis and Applications**, Springer, v. 24, p. 1207–1220, 2021. Citado 2 vezes nas páginas 21 e 22.

Nguyen, N. H. *et al.* Deployment and validation of an ai system for detecting abnormal chest radiographs in clinical settings. **Frontiers in Digital Health**, Frontiers Media SA, v. 4, p. 890759, 2022. Citado na página 15.

Oliveira, B. *et al.* Suporte para diagnóstico de covid-19 por meio de classificação automática de imagens de raio-x e modelos explicáveis. In: FILHO, C. J. A. B. *et al.* (Ed.). **Anais do 15º Congresso Brasileiro de Inteligência Computacional (CBIC 2021)**. Fortaleza, CE, Brasil: Sociedade Brasileira de Inteligência Computacional (SBIC), 2021. p. 1–8. Citado 3 vezes nas páginas 19, 20 e 22.

Organização Mundial da Saúde (OMS). **Pneumonia**. Genebra: Organização Mundial da Saúde, 2023. <<https://www.who.int/health-topics/pneumonia>>. Publicado em 2023. Acessado em 24 nov. 2025. Citado na página 10.

Park, M. *et al.* Multilabel image classification with deep transfer learning for decision support on wildfire response. **Remote Sensing**, MDPI, v. 13, n. 19, p. 3985, 2021. Citado na página 17.

Radočaj, P. *et al.* Interpretable deep learning for pediatric pneumonia classification using chest x-rays. **MDPI Electronics**, 2025. Citado na página 15.

Rahman, M. F. *et al.* Machine-learning-enabled diagnostics with improved visualization of disease lesions in chest x-ray images. **Diagnostics**, MDPI, v. 14, n. 16, p. 1699, 2024. Citado na página 18.

Rios, N. A. B. *et al.* Perfil epidemiológico das internações hospitalares por pneumonia no brasil. **LUMEN ET VIRTUS**, v. 15, n. 43, p. 8516–8525, 2024. Citado 2 vezes nas páginas 13 e 14.

Salehi, M. *et al.* Automated detection of pneumonia cases using deep transfer learning with paediatric chest x-ray images. **The British journal of radiology**, The British Institute of Radiology., v. 94, n. 1121, p. 20201263, 2021. Citado na página 19.

Sharma, S. *et al.* A deep learning based model for the detection of pneumonia from chest x-ray images using vgg-16 and neural networks. **Procedia Computer Science**, Elsevier, v. 218, p. 357–366, 2023. Citado na página 18.

Silva, L. *et al.* Redes neurais convolucionais aplicadas na detecção de pneumonia através de imagens de raio-x. In: **Anais do Congresso Brasileiro de Informática em Saúde (CBIS 2019)**. Vitória, ES, Brasil: Sociedade Brasileira de Informática em Saúde (SBIS), 2020. p. 1–8. Citado na página 10.

Simonyan, K. *et al.* Very deep convolutional networks for large-scale image recognition. **arXiv preprint arXiv:1409.1556**, 2014. Citado na página 26.

Sufian, M. M. *et al.* Covid-19 classification through deep learning models with three-channel grayscale ct images. **Big Data and Cognitive Computing**, MDPI, v. 7, n. 1, p. 36, 2023. Citado na página 29.

Szegedy, C. *et al.* Inception-v4, inception-resnet and the impact of residual connections on learning. In: **Proceedings of the Thirty-First AAAI Conference on Artificial Intelligence (AAAI 2017)**. San Francisco, CA, USA: AAAI Press, 2017. v. 31, n. 1. Citado na página 27.

Szegedy, C. *et al.* Rethinking the inception architecture for computer vision. In: **Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)**. Las Vegas, NV, USA: IEEE, 2016. p. 2818–2826. Citado na página 27.

Tsuji, T. *et al.* Classification of chest x-ray images by incorporation of medical domain knowledge into operation branch networks. **BMC Medical Imaging**, Springer, v. 23, n. 1, p. 62, 2023. Citado na página 18.

Vardhan, A. *et al.* A radiographic, deep transfer learning framework, adapted to estimate lung opacities from chest x-rays. **Bioelectronic Medicine**, Springer, v. 9, n. 1, p. 1, 2023. Citado na página 19.

Vasile, C. M. *et al.* Intelligent diagnosis of thyroid ultrasound imaging using an ensemble of deep learning methods. **Medicina**, MDPI, v. 57, n. 4, p. 395, 2021. Citado na página 11.

Xie, Y. *et al.* Diagnosis of pneumonia from chest x-ray images using yolo deep learning. **Frontiers in Neurorobotics**, 2025. Citado na página 18.

Zhang, H. *et al.* Applying deep learning to medical imaging: a review. **Applied Sciences**, MDPI, v. 13, n. 18, p. 10521, 2023. Citado na página 16.