

UNIVERSIDADE FEDERAL DE UBERLÂNDIA

Gustavo Pioli Resende

**Comparação de modelos de CNN na detecção
de câncer de mama utilizando imagens
infravermelhas multiângulo.**

Uberlândia, Brasil

2025

UNIVERSIDADE FEDERAL DE UBERLÂNDIA

Gustavo Pioli Resende

Comparação de modelos de CNN na detecção de câncer de mama utilizando imagens infravermelhas multiângulo.

Trabalho de conclusão de curso apresentado à Faculdade de Computação da Universidade Federal de Uberlândia, como parte dos requisitos exigidos para a obtenção título de Bacharel em Ciência da Computação.

Orientador: Renata dos Santos Melo

Coorientador: Henrique Coelho Fernandes

Universidade Federal de Uberlândia – UFU

Faculdade de Computação

Bacharelado em Ciência da Computação

Uberlândia, Brasil

2025

**UNIVERSIDADE FEDERAL DE UBERLÂNDIA****Faculdade de Computação**

Av. João Naves de Ávila, nº 2121, Bloco 1A - Bairro Santa Mônica, Uberlândia-MG, CEP 38400-902
Telefone: (34) 3239-4144 - <http://www.portal.facom.ufu.br/> facom@ufu.br

**ATA DE DEFESA - GRADUAÇÃO**

Curso de Graduação em:	Bacharel em Ciência da Computação				
Defesa de:	Projeto de Graduação 2 - GBC082				
Data:	09/05/2025	Hora de início:	08:04	Hora de encerramento:	09:00
Matrícula do Discente:	12111BCC010				
Nome do Discente:	Gustavo Pioli Resende				
Título do Trabalho:	Comparação de modelos de CNN na detecção de câncer de mama utilizando imagens infravermelhas multi-ângulo				
A carga horária curricular foi cumprida integralmente?		(X) Sim () Não			

Reuniu-se na sala [virtual pela plataforma MSTeams](#), a Banca Examinadora, designada pelo Colegiado do Curso de Graduação em [Ciência da Computação](#), assim composta: Professores: Bruno Augusto Nassif Travençolo - FACOM/UFU; Leandro Nogueira Couto - FACOM/UFU; Henrique Coelho Fernandes - FACOM/UFU; Renata dos Santos Melo - FACOM/UFU orientador(a) do(a) candidato(a).

Iniciando os trabalhos, o(a) presidente da mesa, Ma. [Renata dos Santos Melo](#), apresentou a Comissão Examinadora e o candidato(a), agradeceu a presença do público, e concedeu ao discente a palavra, para a exposição do seu trabalho. A duração da apresentação do discente e o tempo de arguição e resposta foram conforme as normas do curso.

A seguir o(a) senhor(a) presidente concedeu a palavra, pela ordem sucessivamente, aos(às) examinadores(as), que passaram a arguir o(a) candidato(a). Ultimada a arguição, que se desenvolveu dentro dos termos regimentais, a Banca, em sessão secreta, atribuiu o resultado final, considerando o(a) candidato(a):

(X) Aprovado(a) Nota [95] (Somente números inteiros)

OU

() Aprovado(a) sem nota.

Nada mais havendo a tratar foram encerrados os trabalhos. Foi lavrada a presente ata que após lida e achada conforme foi assinada pela Banca Examinadora.



Documento assinado eletronicamente por **Renata dos Santos Melo, Professor(a) Substituto(a) do Magistério Superior**, em 09/05/2025, às 09:31, conforme horário oficial de Brasília, com fundamento no art. 6º, § 1º, do [Decreto nº 8.539, de 8 de outubro de 2015](#).



Documento assinado eletronicamente por **Leandro Nogueira Couto, Professor(a) do Magistério Superior**, em 09/05/2025, às 14:37, conforme horário oficial de Brasília, com fundamento no art. 6º, § 1º, do [Decreto nº 8.539, de 8 de outubro de 2015](#).



Documento assinado eletronicamente por **Bruno Augusto Nassif Travençolo, Professor(a) do Magistério Superior**, em 09/05/2025, às 18:18, conforme horário oficial de Brasília, com fundamento no art. 6º, § 1º, do [Decreto nº 8.539, de 8 de outubro de 2015](#).



A autenticidade deste documento pode ser conferida no site https://www.sei.ufu.br/sei/controlador_externo.php?acao=documento_conferir&id_orgao_acesso_externo=0, informando o código verificador **6321951** e o código CRC **89D92C2F**.

Resumo

O câncer de mama é o segundo tipo mais comum entre as mulheres, representando um importante desafio para a área da saúde. Embora a mamografia seja amplamente utilizada como principal método de diagnóstico, sua confiabilidade é limitada em pacientes mais jovens. Neste contexto, este estudo investiga o uso de algoritmos de aprendizado de máquina aplicados a imagens termográficas como alternativa complementar para o diagnóstico precoce. Utilizando o banco de dados DMR-IR e redes neurais convolucionais (CNNs) pré-treinadas, foram adotadas modificações arquiteturais conhecidas na literatura com o intuito de avaliá-las em um cenário de múltiplas entradas com diferentes ângulos da mama. Dentre os modelos testados, a DenseNet apresentou o melhor desempenho, com acurácia de 94,44%, F1-score de 0,9474 e sensibilidade de 100%, evidenciando sua capacidade de identificar corretamente todos os casos positivos. Os resultados indicam que o aumento da densidade de neurônios pode impactar positivamente o desempenho, desde que a arquitetura consiga lidar com o maior volume de informações sem *overfitting* — como ocorre com a DenseNet, diferentemente da EfficientNet, que apresentou limitações nesse aspecto. Além disso, a abordagem de múltiplas entradas mostrou-se eficaz por permitir a análise de regiões anteriormente inacessíveis em arquiteturas de entrada única. Dessa forma, o uso de aprendizado de máquina com imagens termográficas revela-se promissor como ferramenta auxiliar na detecção precoce do câncer de mama.

Palavras-chave: Redes Neurais Convolucionais, Câncer de Mama, Imagens Termográficas, Classificação, Aprendizado de Máquina.

Lista de ilustrações

Figura 1 – Imagem Termográfica retirada do banco público DMR-IR (Departamento de Ciência da Computação (IC/UFF), 2014).	16
Figura 2 – Arquitetura básica de uma rede neural profunda (<i>Deep Neural Network</i>) extraída de (IBM Data and AI Team, 2023).	17
Figura 3 – Convolução com <i>kernel</i> de Sobel - extraída de (LIMA; RUBIK; MORAIS, 2020).	18
Figura 4 – Arquitetura dentro de um DenseBlock - extraída de (HUANG; LIU; WEINBERGER, 2016).	20
Figura 5 – Tabela comparativa dos modelos de DenseNet, “conv” na tabela representa o conjunto de processos BN-ReLU-Conv - extraído de (HUANG; LIU; WEINBERGER, 2016).	21
Figura 6 – Gráfico de erros de redes neurais profundas - extraído de (HE et al., 2015).	22
Figura 7 – Comparação do bloco tradicional (esquerda) com bloco residual da ResNet (direita) - extraído de (ZHANG; YU, 2021)	22
Figura 8 – Tabela comparativa de modelos ResNet - extraído de (HE et al., 2015).	23
Figura 9 – Arquitetura EfficientNet - extraído de (TAN; LE, 2019).	25
Figura 10 – Exemplo de visualizações de diferentes ângulos - Lateral Esquerda, Frontal e Lateral Direita.	30
Figura 11 – Curvas ROC com valores de AUC baseados em modelos hipotéticos - extraído de (Google Developers, 2023).	32
Figura 12 – À esquerda: Matriz de Confusão. À direita: Curva ROC da DenseNet E2.	36
Figura 13 – À esquerda: Matriz de Confusão. À direita: Curva ROC da ResNet50 E1.	37
Figura 14 – À esquerda: Matriz de Confusão. À direita: Curva ROC da EfficientNetB0 C1.	37
Figura 15 – Matriz de Confusão e Curva ROC para etapa 2 - ResNet E1.	39
Figura 16 – Matriz de Confusão e Curva ROC para etapa 2 - EfficientNet C1.	40
Figura 17 – Matriz de Confusão e Curva ROC para etapa 2 - DenseNet121 E2.	40

Lista de tabelas

Tabela 1	– Matriz de confusão genérica para classificação binária.	31
Tabela 2	– Métricas DenseNet121 - Experimento E2	36
Tabela 3	– Comparação das Métricas dos Experimentos A2 e E1	36
Tabela 4	– Métricas EfficientNetB0 - Experimento C1	37
Tabela 5	– Resultados Obtidos na Etapa 1 em Comparativo com TA	38
Tabela 6	– Métricas ResNet50 - E1 - Etapa 2	39
Tabela 7	– Métricas EfficientNetB0 - C1 - Etapa 2	39
Tabela 8	– Métricas DenseNet121 - E2 - Etapa 2	40
Tabela 9	– Resultados Obtidos na Etapa 2 em Comparativo com TA	41
Tabela 10	– Comparação entre modelos correlatos e desempenho do modelo proposto. Valores em %.	42

Lista de abreviaturas e siglas

CNN	<i>Convolutiona Neural Network</i>
DMR	<i>Database for Mastology Research</i>
DMR-IR	<i>Database for Mastology Research - Infrared</i>
IA	Inteligência Artificial
BN	<i>Batch Normalization</i>
GAP	<i>Global Average Pooling</i>
ROI	<i>Region of Interest</i>
AG	Algoritmo Genético
PSO	<i>Particle Swarm Optimization</i>
CG	Coeficiente Cognitivo
SVM	<i>Support Vector Machine</i>
k-NN	<i>K-Nearest Neighbor</i>
HUAP	Hospital Universitário Antônio Pedro
ACC	Acurácia (<i>Accuracy</i>)
SEN	Sensibilidade (<i>Sensitivity</i>)
SPE	Especificidade (<i>Specificity</i>)
PREC	Precisão (<i>Precision</i>)
F1-score	Métrica F1
TP	Verdadeiros Positivos (<i>True Positives</i>)
TN	Verdadeiros Negativos (<i>True Negatives</i>)
FP	Falsos Positivos (<i>False Positives</i>)
FN	Falsos Negativos (<i>False Negatives</i>)
TPR	Taxa de Verdadeiros Positivos

FPR	Taxa de Falsos Positivos
ROC	<i>Receiver Operation Characteristic</i>
AUC	<i>Area Under the Curve</i> (Área sob a Curva)
FC	Camada totalmente conectada (<i>Fully Connected</i>)

Sumário

1	INTRODUÇÃO	11
2	OBJETIVOS	13
2.1	Objetivos Gerais	13
2.2	Objetivos Específicos	13
3	FUNDAMENTAÇÃO TEÓRICA	14
3.1	Câncer de Mama	14
3.1.1	Diagnóstico	14
3.1.2	Imagens Infravermelhas ou Termográficas	15
3.2	Redes Neurais	15
3.2.1	Redes Neurais Convolucionais (CNN)	17
3.2.2	DenseNet	19
3.2.2.1	Dense Block	19
3.2.2.2	Transition Layer	20
3.2.2.3	Camada Final de Classificação	21
3.2.3	ResNet	21
3.2.3.1	Residual Block	22
3.2.3.2	Modelos ResNet	23
3.2.4	EfficientNet	23
3.2.4.1	Compound Scaling	24
3.2.4.2	Modelos EfficientNet	24
4	TRABALHOS RELACIONADOS	26
5	MATERIAIS E MÉTODOS	29
5.1	Banco de dados	29
5.2	Classificadores	30
5.3	Métricas	30
5.3.1	Matriz de Confusão	31
5.3.2	Curvas ROC	31
6	DESENVOLVIMENTO	33
6.1	Ajuste na base de dados	33
6.2	Classificação com CNNs selecionadas	34
6.3	Resultados e Discussões	35
6.3.1	Primeira Etapa	35

6.3.1.1	DenseNet121	35
6.3.1.2	ResNet50	35
6.3.1.3	EfficientNetB0	36
6.3.1.4	Comparativo	37
6.3.2	Segunda Etapa	38
6.3.2.1	ResNet	38
6.3.2.2	EfficientNet	39
6.3.2.3	DenseNet	39
6.3.2.4	Comparativo	41
6.3.3	Discussões	41
7	CONCLUSÃO	43
	REFERÊNCIAS	44

1 Introdução

O câncer de mama é o segundo tipo de câncer mais comum entre as mulheres americanas e também ocupa a segunda posição em número de mortes relacionadas ao câncer nesse grupo. De acordo com a *American Cancer Society*, estima-se que, em 2024, serão diagnosticados 310.720 novos casos de câncer de mama em mulheres, representando aproximadamente 30% dos cânceres femininos a cada ano (SOCIETY, 2020). No Brasil, a estimativa para o triênio de 2023 a 2025 é de 73.610 novos casos de câncer de mama, com uma taxa de incidência de 84,46 a cada 100 mil mulheres na região Sudeste (INCA, 2024). O diagnóstico precoce é essencial, pois possibilita o uso de tratamentos menos invasivos e aumenta as chances de sucesso, reduzindo a mortalidade relacionada à doença (INCA, 2024). As principais técnicas de diagnóstico incluem a mamografia, a ressonância magnética e o ultrassom, sendo a mamografia a mais amplamente utilizada, com sensibilidade de até 90% em mulheres mais velhas. No entanto, essa técnica apresenta uma redução significativa de confiabilidade em pacientes mais jovens, com sensibilidade em torno de 60% (GONÇALVES; SOUZA; FERNANDES, 2021). Diante desse desafio, surge a necessidade de explorar novas abordagens, como a análise de imagens por termografia infravermelha, que tem demonstrado grande potencial para superar as limitações da mamografia em pacientes mais jovens, ao detectar alterações relacionadas à atividade metabólica e circulação vascular em tecidos cancerosos (MILOSEVIC; JANKOVIC; PEULIC, 2014).

O aperfeiçoamento das câmeras de captura de imagem infravermelha, junto ao desenvolvimento de novas tecnologias de análise de imagem, como as Redes Neurais Convolucionais (CNNs), possibilita a exploração dos benefícios da termografia na detecção do câncer de mama. Essas inovações visam aumentar a confiabilidade do método, até então utilizado principalmente como exame complementar devido à sua baixa acurácia (ZULUAGA-GOMEZ et al., 2019). Nesse contexto, é relevante comparar os diversos modelos de CNNs desenvolvidos para análise de imagens, com o objetivo de aprimorar a acurácia e reduzir a ocorrência de falsos positivos e falsos negativos, buscando uma ferramenta mais eficaz para a detecção precoce do câncer de mama.

Diversos estudos utilizaram o banco de imagens DMR-IR em aplicações de Redes Neurais Convolucionais (CNNs) com o objetivo de detectar câncer de mama. Em geral, essas abordagens empregaram modelos pré-treinados e utilizaram imagens estáticas da visão frontal das pacientes como em (GONÇALVES; SOUZA; FERNANDES, 2021). Os resultados indicam que algumas arquiteturas de CNN se destacam no diagnóstico, alcançando métricas elevadas de desempenho. Enquanto alguns trabalhos exploraram imagens completas, outros como o trabalho de (DOMINGOS, 2021) investigaram a segmentação de regiões de interesse na área mamária, com variações nos resultados conforme a abordagem

adotada. Além disso, técnicas de otimização, como a busca por hiperparâmetros explorada por (ZULUAGA-GOMEZ et al., 2019), também foram aplicadas para aprimorar o desempenho das redes neurais.

Este trabalho busca a continuação do esforço para maior confiabilidade do diagnóstico via imagens termográficas por meio da análise comparativa de diferentes modelos de CNNs a fim de destacar cenários de aplicação que podem favorecer determinado modelo com foco em obter uma melhor acurácia e sensibilidade, além de outras métricas, no diagnóstico de câncer de mama.

Para atingir tais objetivos será utilizada a linguagem de programação *Python* em razão de seu amplo repertório de bibliotecas para aprendizado de máquina. Além disso, vale o destaque para a biblioteca *PyTorch*, por oferecer uma estrutura flexível e eficiente para a construção e treinamento de modelos preditivos, especialmente em tarefas que envolvem redes neurais e grandes volumes de dados.

2 Objetivos

Os objetivos deste trabalho estão divididos em objetivo geral e objetivos específicos, conforme descrito a seguir.

2.1 Objetivos Gerais

Este trabalho tem como objetivo principal treinar classificadores baseados em redes neurais convolucionais (CNNs), com a implementação de arquiteturas multiângulo, para distinguir pacientes saudáveis de pacientes com câncer, utilizando imagens termográficas do banco de dados DMR-IR.

2.2 Objetivos Específicos

Como objetivos específicos foram definidos:

- Explorar modificações em arquiteturas de CNNs consagradas no contexto da classificação de imagens;
- Utilizar redes modificadas selecionadas em um modelo para interpretar imagens frontais e laterais no processo de classificação;
- Comparar a performance dos modelos utilizando métricas como F1-score e acurácia com abordagens desenvolvidas em outros trabalhos da literatura, em especial o estudo desenvolvido por (GONÇALVES; SOUZA; FERNANDES, 2021).

3 Fundamentação Teórica

Para uma melhor compreensão do trabalho, este capítulo apresentará os conceitos fundamentais.

3.1 Câncer de Mama

Câncer engloba um conjunto de mais de 100 doenças caracterizadas pelo crescimento desordenado e rápido de células que formam aglomerados de células chamados tumores, invadindo tecidos adjacentes e podendo se espalhar para outras partes do corpo, fenômeno conhecido como metástase ([Ministério da Saúde, 2024](#)). A causa principal do câncer são mutações no DNA das células, que alteram suas instruções naturais de funcionamento. Cada célula sadia possui instruções sobre como deve crescer, dividir-se e morrer; quando ocorrem mutações nessas instruções, células defeituosas começam a se proliferar descontroladamente, levando ao surgimento do câncer ([INCA, 2024](#)).

O câncer de mama é um dos tipos mais incidentes no Brasil, com 73.610 novos casos estimados em 2022 e 18.361 mortes em 2021, incluindo 220 homens e 18.139 mulheres. Como a idade é um dos principais fatores de risco, com cerca de 80% dos casos ocorrendo após os 50 anos, a prevenção e o diagnóstico precoce são fundamentais para reduzir a mortalidade. Investir em conscientização, exames regulares e acesso ao tratamento pode salvar vidas e minimizar o impacto dessa doença na população ([INCA, 2024](#)).

3.1.1 Diagnóstico

Na maioria dos casos de câncer de mama a detecção é feita após o aparecimento dos sintomas, porém muitas mulheres que possuem câncer podem não apresentar sintomas, tornando a triagem regular um procedimento importante no combate a doença ([FIGUEIREDO et al., 2018](#)).

A triagem tem como objetivo detectar o câncer de mama precocemente, utilizando exames físicos e de imagem para rastrear a doença. Os exames físicos procuram identificar qualquer anomalia na região das mamas, seja pela experiência do profissional de saúde ou pelo autoconhecimento da mulher sobre seu corpo. Diferentes exames de imagem, como mamografia, ultrassonografia e ressonância magnética, podem ser utilizados para investigar e diagnosticar o câncer de mama. Com base nos resultados obtidos, avalia-se a necessidade de uma biópsia, que é o único método capaz de confirmar com precisão se o achado é um câncer ([FIGUEIREDO et al., 2018](#)).

Mamografias são técnicas que utilizam raios-X de baixa dose usados para examinar as mamas. Realizar mamografias regularmente pode ajudar a identificar o câncer de mama em um estágio inicial, quando o tratamento tem maior chance de sucesso. A mamografia é capaz de identificar alterações nas mamas potencialmente cancerígenas, frequentemente anos antes do surgimento de quaisquer sintomas clínicos. Resultados de décadas de pesquisas mostram claramente que mulheres que fazem mamografias regularmente têm mais probabilidade de ter o câncer de mama detectado precocemente, menos probabilidade de precisar de tratamentos agressivos, como cirurgia para remover toda a mama (mastectomia) e quimioterapia, e uma maior chance de cura (SOCIETY, 2020). No entanto, as mamografias apresentam limitações diagnósticas e podem falhar em detectar alguns casos de câncer de mama, especialmente em mulheres jovens, que possuem seios mais densos. A alta densidade mamária dificulta a identificação de tumores nas imagens e, além disso, é considerada um fator de risco adicional para o desenvolvimento da doença (BOYD et al., 2007). E, caso algo seja identificado em uma mamografia de rastreamento, é provável que a mulher precise de outros exames (como mais mamografias ou uma ultrassonografia mamária) para confirmar o diagnóstico (SOCIETY, 2020).

3.1.2 Imagens Infravermelhas ou Termográficas

A utilização de exames de imagem infravermelha apresenta-se como uma alternativa promissora para o diagnóstico precoce do câncer de mama. A Termografia Infravermelha da mama baseia-se no princípio de que células cancerígenas aumentam a taxa metabólica e, conseqüentemente, promovem a formação de novos vasos sanguíneos, processo conhecido como neovascularização e amplamente utilizado para determinar a agressividade tumoral, resultando em alterações de temperatura na área ao redor do tumor (PASTORE, 2003). Essa variação térmica gerada pelo tumor pode ser transmitida até a superfície da pele da mama, onde a termografia identifica essas mudanças por meio da radiação térmica. Esse fenômeno envolve a emissão de energia térmica em ondas eletromagnéticas, que são captadas por uma câmera infravermelha, permitindo a visualização das variações de temperatura na pele (FIGUEIREDO et al., 2018). A Figura 1 mostra uma imagem termográfica de uma paciente saudável publicada no banco de imagens DMR-IR¹.

3.2 Redes Neurais

As redes neurais são inspiradas no funcionamento do cérebro humano, onde existem diversos neurônios interconectados formando uma rede que, por pulsos elétricos, permitem que os seres humanos tomem suas decisões. As redes neurais, então, são um método

¹ Disponível em <<https://visual.ic.uff.br/dmi>>

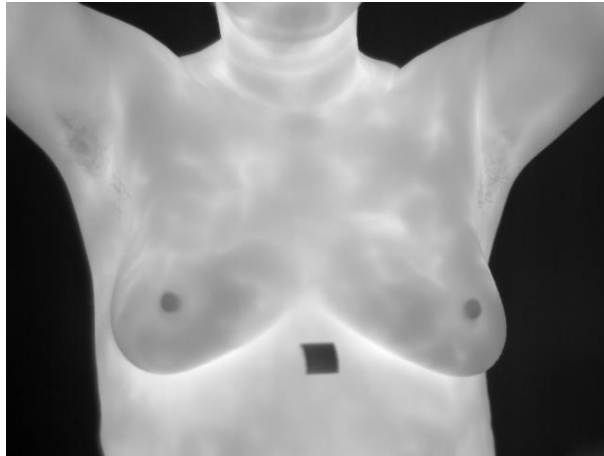


Figura 1 – Imagem Termográfica retirada do banco público DMR-IR ([Departamento de Ciência da Computação \(IC/UFF\), 2014](#)).

de inteligência artificial (IA) composto por diversos nós interconectados que permitem que computadores aprendam com erros e melhorem continuamente, resolvendo problemas complexos como classificação de imagens e sumarização de documentos ([IBM Data and AI Team, 2023](#)).

A arquitetura básica de uma rede neural se baseia em três tipos de camadas (Figura 2):

- Camadas de entrada: recebem as informações dos dados externos e os processa para enviar para a próxima camada;
- Camadas ocultas: são camadas intermediárias que recebem informações já processadas pelas camadas anteriores e novamente realiza novos procedimentos para repassá-los aos níveis subsequentes. Pode-se haver diversas camadas ocultas em um modelo de rede neural e são elas que dão a característica de profundas as redes neurais usadas em *deep learning*.
- Camadas de saída: são as responsáveis por fornecer o resultado final do processamento da entrada da rede, sua composição deve ser de acordo com o que se espera de saída para o modelo, ou seja, uma camada de saída deve possuir um nó de saída caso o problema se de classificação binária, dessa forma o resultado será dado como 1 ou 0. No entanto, para um problema de classificação múltipla a camada de saída deverá ser construída com mais nós de saída para abranger todos os tipos possíveis de classificação ([Amazon Web Services, 2024](#)).

As redes neurais profundas (*deep learning*) possuem múltiplas camadas ocultas interligadas, com milhões de nós conectados. Diferentemente de outros métodos de *machine learning*, onde a seleção de características é realizada manualmente, as redes profundas

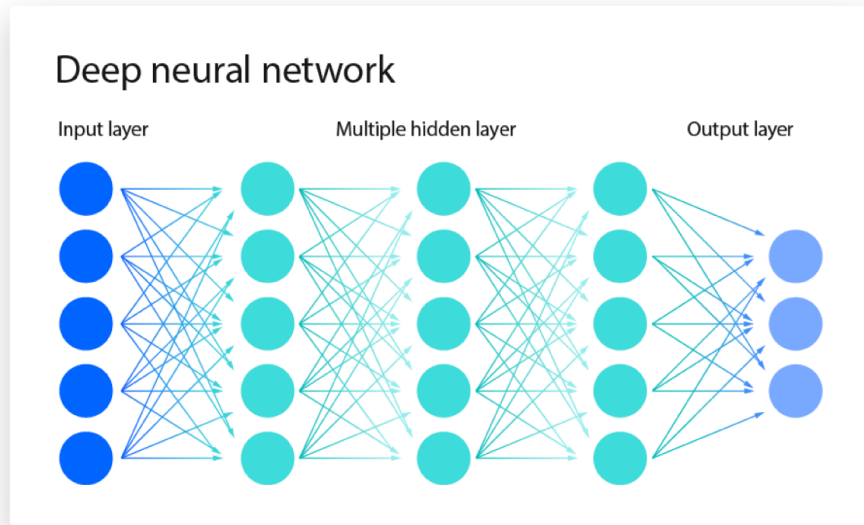


Figura 2 – Arquitetura básica de uma rede neural profunda (*Deep Neural Network*) extraída de (IBM Data and AI Team, 2023).

extraem automaticamente padrões dos dados de entrada. Além disso, atribuem pesos às conexões entre os nós, ajustando-se durante o treinamento para otimizar a precisão do modelo. Quanto mais camadas ocultas, maior a capacidade da rede de reconhecer padrões complexos, porém, essa sofisticação exige um volume significativo de dados para treinamento e maior poder computacional (Amazon Web Services, 2024).

3.2.1 Redes Neurais Convolucionais (CNN)

Em redes neurais tradicionais, a principal operação matemática realizada nas conexões entre os neurônios é a multiplicação matricial. Cada neurônio recebe uma combinação linear das ativações da camada anterior, ponderadas pelos pesos da rede, e aplica uma função de ativação para produzir a saída. Esse processo, expresso pela equação

$$y = f(Wx + b) \quad (3.1)$$

onde W é a matriz de pesos, x é a entrada, b é o viés e f é a função de ativação, permite que a rede aprenda padrões complexos nos dados (IBM, 2025). No entanto, ao lidar com dados estruturados em grade, como imagens, esse método pode ser ineficiente devido à grande quantidade de parâmetros necessários.

As CNNs, introduzidas por LeCun em 1989, são um tipo especializado de rede neural projetada para processar dados com uma estrutura em grade, como séries temporais e imagens. O uso da operação matemática chamada convolução, que substitui a multiplicação matricial, é o que distingue as CNNs e as torna mais eficientes na extração de características relevantes em dados mais complexos. A convolução envolve a aplicação de

um *kernel* (filtro), que é uma pequena matriz de pesos, sobre uma entrada (por exemplo, uma imagem). O *kernel* desliza sobre a imagem, multiplicando seus valores pelos valores correspondentes na região da imagem onde está posicionado e, em seguida, somando esses produtos para gerar um valor de saída (Figura 3). Essa abordagem é altamente eficaz em várias aplicações práticas (GOODFELLOW; BENGIO; COURVILLE, 2016b).

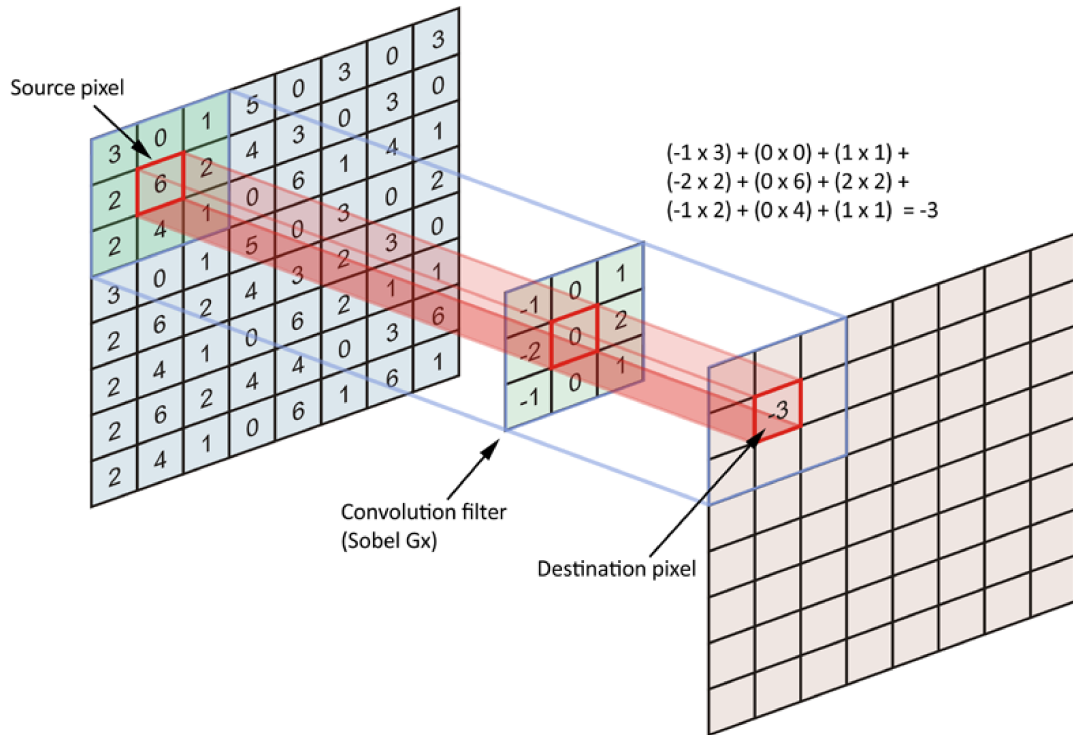


Figura 3 – Convolução com *kernel* de Sobel - extraída de (LIMA; RUBIK; MORAIS, 2020).

Neste contexto, a convolução permite a implementação de interações esparsas, onde as unidades de saída se conectam apenas a partes específicas da entrada. Por exemplo, em uma imagem, um pequeno *kernel*, pode detectar características locais significativas, como bordas, sem a necessidade de armazenar um grande número de parâmetros. Além disso, as CNNs utilizam o conceito de compartilhamento de parâmetros, onde um único conjunto de pesos é aplicado em diferentes posições da imagem, otimizando a eficiência do modelo (GOODFELLOW; BENGIO; COURVILLE, 2016b).

Contudo, existem grandes problemas na utilização de CNNs, quanto mais profunda é uma CNN, ou seja, mais camadas convolucionais, mais ela consegue extrair características da entrada, porém para isso as redes neurais convolucionais costumam ter milhares de parâmetros, mesmo as menores arquiteturas, e com isso seu treinamento demanda muito poder computacional e tempo de execução, geralmente feito em unidades de processamento gráfico (GONÇALVES; SOUZA; FERNANDES, 2021). Além disso, CNNs precisam de uma grande quantidade de dados para o seu treinamento e a disponibilidade

de informações para serem utilizadas como entrada durante esse processo é geralmente escassa, portanto é comum se aplicar métodos de aumento de dados para fornecer ao treinamento da rede. O aumento de dados foca em gerar novas entradas com as informações disponíveis, porém com o cuidado de manter a natureza dos dados (GONÇALVES; SOUZA; FERNANDES, 2021).

3.2.2 DenseNet

DenseNet (Rede Neural Convolutacional Densa) é uma arquitetura de CNNs apresentada por (HUANG; LIU; WEINBERGER, 2016). Ela foi desenvolvida para melhorar a propagação das informações entre as camadas, facilitando o aprendizado. Isso é alcançado por meio da reutilização das características aprendidas em camadas anteriores, o que também contribui para a redução do número de parâmetros necessários.

A principal inovação da *DenseNet* é a conectividade densa entre camadas convolucionais dentro de um bloco. Em vez de aplicar operações convolucionais de forma sequencial e independente, cada camada recebe como entrada as ativações de todas as camadas anteriores e passa suas próprias ativações para todas as camadas subsequentes. Se tivermos uma rede com L camadas, então, em vez de ter L conexões (como nas CNNs tradicionais), teremos $L(L+1)/2$ conexões, criando uma malha densa de comunicação entre camadas (GeeksforGeeks, 2024).

3.2.2.1 Dense Block

O núcleo da DenseNet é o bloco denso (*Dense Block*), onde ocorre a conectividade completa entre camadas. Dentro desse bloco, cada camada recebe como entrada todas as ativações das camadas anteriores e passa suas próprias ativações para todas as camadas subsequentes.

Cada camada dentro do bloco aplica a seguinte sequência de operações:

- *Batch Normalization* (BN): normalização da saída das camadas anteriores ajustando a escala e valores;
- *ReLU Activation*: uma das funções de ativação mais populares e amplamente utilizadas em redes neurais, pode ser definida como:

$$f(x) = \max(0, x) \quad (3.2)$$

onde \max representa o valor máximo entre 0 e a entrada x definida como o valor máximo entre zero e a entrada, ou seja, caso a entrada x seja positiva, a função retorna x ; caso x seja negativa, ela retorna 0. Em termos gráficos, isso significa que qualquer valor negativo é neutralizado, enquanto valores positivos permanecem iguais (GOODFELLOW; BENGIO; COURVILLE, 2016a);

- Convolução 3×3 : operação de convolução com um *kernel* de 3 linhas e 3 colunas.

A saída de cada camada é concatenada (e não somada) às saídas anteriores, resultando em um crescimento linear do número de canais dentro do bloco (Figura 4). Se a primeira camada de um bloco tem f mapas de características e adicionamos k mapas por camada, então, após m camadas, o número total de canais será:

$$total = f + m \cdot k \quad (3.3)$$

onde k é chamado de *growth rate* (taxa de crescimento) (HUANG; LIU; WEINBERGER, 2016).

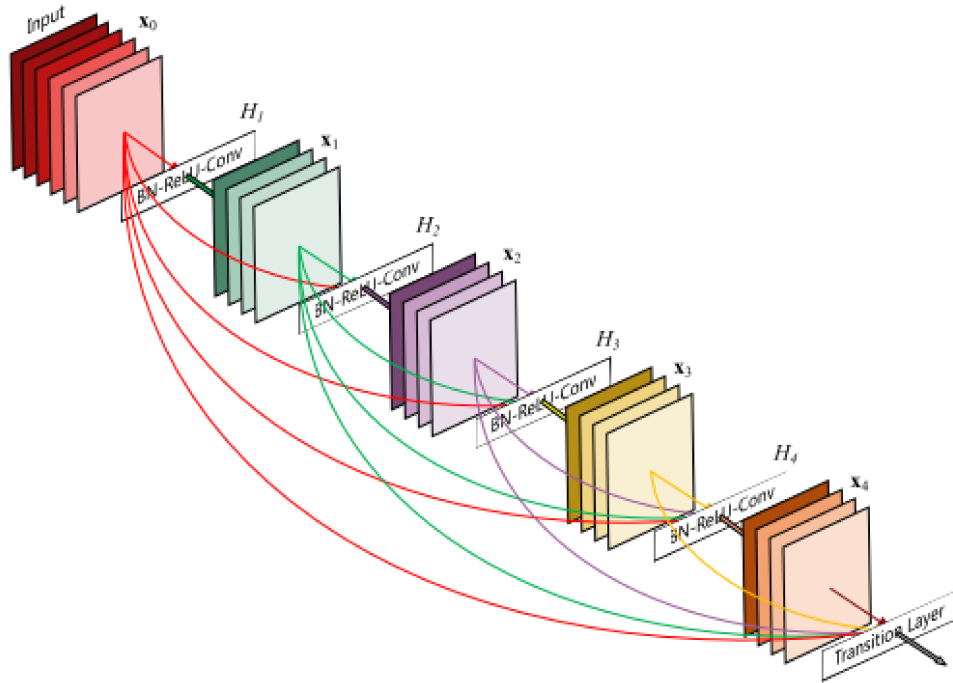


Figura 4 – Arquitetura dentro de um DenseBlock - extraída de (HUANG; LIU; WEINBERGER, 2016).

3.2.2.2 Transition Layer

Para evitar o crescimento excessivo do número de canais entre blocos densos, a DenseNet incorpora camadas de transição entre os blocos. Cada *Transition Layer* consiste em (HUANG; LIU; WEINBERGER, 2016):

- *Batch Normalization* (BN): normalização da saída das camadas anteriores ajustando a escala e valores;

- Convolução 1×1 : aplica um *kernel* 1×1 fornecendo para cada pixel um valor por meio da combinação linear dos valores de todos os canais daquele pixel fornecidos pelo *Dense Block*;
- Camada de *Pooling* 2×2 : processo realizado para reduzir o tamanho das representações espaciais das entradas, mantendo as informações mais importantes.

3.2.2.3 Camada Final de Classificação

Após passar pelos *Dense Blocks* e *Transition Layers*, os mapas de ativação são reduzidos utilizando *Global Average Pooling* (GAP), ou seja, é feito um procedimento em que cada canal fornece um único valor para a média de seus *pixels* concatenando em um único vetor de tamanho C, onde C é o número de canais. Tal vetor é, por fim, utilizado em uma camada *softmax* para classificação, ela transforma a sua entrada em um vetor de probabilidades onde cada valor representa a probabilidade da entrada pertencer a uma classe específica (HUANG; LIU; WEINBERGER, 2016).

As variações da arquitetura DenseNet, como DenseNet-121 e DenseNet-201, estão organizadas na Figura 5, que compara diferentes configurações de profundidade e número de parâmetros.

Layers	Output Size	DenseNet-121	DenseNet-169	DenseNet-201	DenseNet-264
Convolution	112×112	7×7 conv, stride 2			
Pooling	56×56	3×3 max pool, stride 2			
Dense Block (1)	56×56	$\begin{bmatrix} 1 \times 1 \text{ conv} \\ 3 \times 3 \text{ conv} \end{bmatrix} \times 6$	$\begin{bmatrix} 1 \times 1 \text{ conv} \\ 3 \times 3 \text{ conv} \end{bmatrix} \times 6$	$\begin{bmatrix} 1 \times 1 \text{ conv} \\ 3 \times 3 \text{ conv} \end{bmatrix} \times 6$	$\begin{bmatrix} 1 \times 1 \text{ conv} \\ 3 \times 3 \text{ conv} \end{bmatrix} \times 6$
Transition Layer (1)	56×56	1×1 conv			
	28×28	2×2 average pool, stride 2			
Dense Block (2)	28×28	$\begin{bmatrix} 1 \times 1 \text{ conv} \\ 3 \times 3 \text{ conv} \end{bmatrix} \times 12$	$\begin{bmatrix} 1 \times 1 \text{ conv} \\ 3 \times 3 \text{ conv} \end{bmatrix} \times 12$	$\begin{bmatrix} 1 \times 1 \text{ conv} \\ 3 \times 3 \text{ conv} \end{bmatrix} \times 12$	$\begin{bmatrix} 1 \times 1 \text{ conv} \\ 3 \times 3 \text{ conv} \end{bmatrix} \times 12$
Transition Layer (2)	28×28	1×1 conv			
	14×14	2×2 average pool, stride 2			
Dense Block (3)	14×14	$\begin{bmatrix} 1 \times 1 \text{ conv} \\ 3 \times 3 \text{ conv} \end{bmatrix} \times 24$	$\begin{bmatrix} 1 \times 1 \text{ conv} \\ 3 \times 3 \text{ conv} \end{bmatrix} \times 32$	$\begin{bmatrix} 1 \times 1 \text{ conv} \\ 3 \times 3 \text{ conv} \end{bmatrix} \times 48$	$\begin{bmatrix} 1 \times 1 \text{ conv} \\ 3 \times 3 \text{ conv} \end{bmatrix} \times 64$
Transition Layer (3)	14×14	1×1 conv			
	7×7	2×2 average pool, stride 2			
Dense Block (4)	7×7	$\begin{bmatrix} 1 \times 1 \text{ conv} \\ 3 \times 3 \text{ conv} \end{bmatrix} \times 16$	$\begin{bmatrix} 1 \times 1 \text{ conv} \\ 3 \times 3 \text{ conv} \end{bmatrix} \times 32$	$\begin{bmatrix} 1 \times 1 \text{ conv} \\ 3 \times 3 \text{ conv} \end{bmatrix} \times 32$	$\begin{bmatrix} 1 \times 1 \text{ conv} \\ 3 \times 3 \text{ conv} \end{bmatrix} \times 48$
Classification Layer	1×1	7×7 global average pool			
		1000D fully-connected, softmax			

Figura 5 – Tabela comparativa dos modelos de DenseNet, “conv” na tabela representa o conjunto de processos BN-ReLU-Conv - extraído de (HUANG; LIU; WEINBERGER, 2016).

3.2.3 ResNet

A partir de 2012, tornou-se tendência aumentar a profundidade das redes neurais para reduzir a taxa de erro. No entanto, esse aumento trouxe um problema comum: o

desaparecimento ou explosão do gradiente. Isso ocorre porque, à medida que os gradientes são propagados para camadas mais profundas, eles podem se tornar excessivamente pequenos, dificultando o ajuste dos pesos (desaparecimento do gradiente), ou crescer de forma descontrolada, tornando o treinamento instável (explosão do gradiente). Como consequência, algumas camadas podem prejudicar o aprendizado ao gerar atualizações que anulam os ajustes anteriores ou desviam o modelo do caminho ideal. Dessa forma, redes muito profundas, sem mecanismos adequados de compensação, tendem a apresentar um aumento nos erros de treinamento e teste (Figura 6). Com o objetivo de tentar resolver esse problema a arquitetura ResNet introduziu os chamados blocos residuais (Residual Blocks) (GFG, 2025).

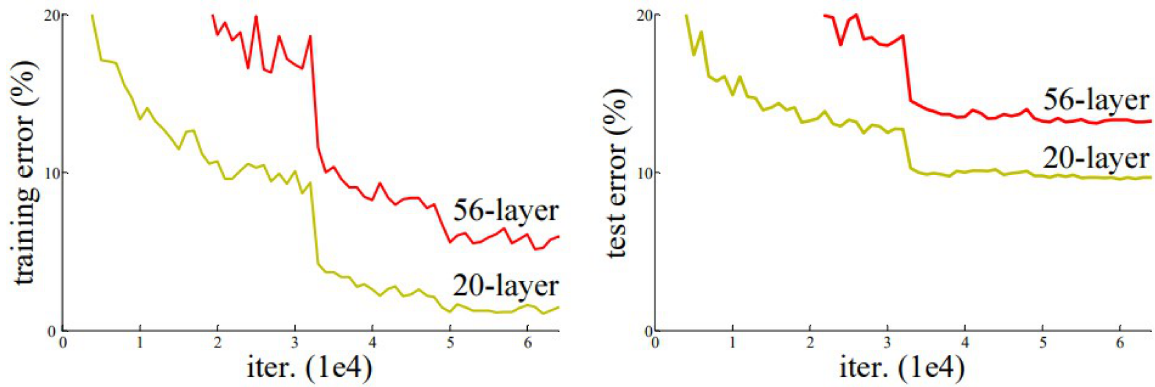


Figura 6 – Gráfico de erros de redes neurais profundas - extraído de (HE et al., 2015).

3.2.3.1 Residual Block

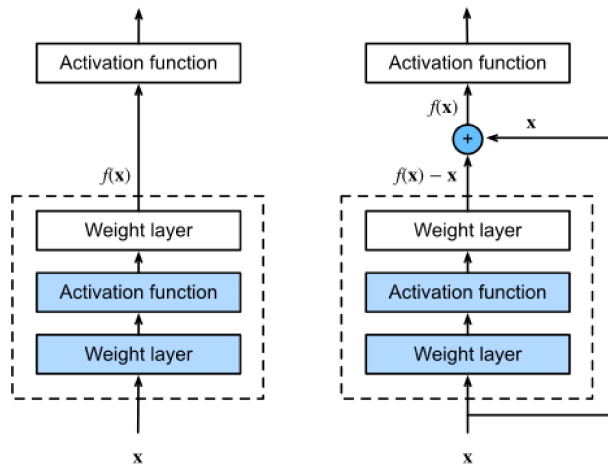


Figura 7 – Comparação do bloco tradicional (esquerda) com bloco residual da ResNet (direita) - extraído de (ZHANG; YU, 2021)

Como forma de evitar o desaparecimento de gradiente a entrada de um *residual block* seria repassada também à saída do bloco pelo que foi chamado de “*shortcut connec-*

tion”, dessa forma o bloco residual teria agora sua saída sendo a diferença entre a entrada e a saída tradicional ($f'(x) = f(x) - x$), essa sendo somada a entrada atuando de forma que a saída repassada seria dada por $y = f'(x) + x$, como demonstrado na Figura 7, ou seja, mesmo que $f'(x)$ seja pequena ou nula, a informação da entrada original x ainda será propagada, o que ajuda a mitigar o problema do desaparecimento do gradiente durante o treinamento de redes profundas (HE et al., 2015).

3.2.3.2 Modelos ResNet

Diversos modelos de ResNet foram desenvolvidos durante a competição Imagenet 2015, variando em termos de número de camadas e dos procedimentos aplicados dentro de cada bloco residual (Figura 8). Para modelos com menos de 18 ou 34 camadas, a arquitetura básica consistia em um bloco residual com duas convoluções 3×3 . No entanto, para arquiteturas mais profundas, como as de 50, 101 ou 152 camadas, convoluções 1×1 foram incorporadas, visando ajustar a dimensionalidade da saída. Além disso, para garantir que a entrada da *shortcut connection* fosse compatível com a saída do bloco residual, convoluções 1×1 também foram aplicadas, permitindo que a soma $f'(x) + x$ fosse realizada de forma adequada (HE et al., 2015).

layer name	output size	18-layer	34-layer	50-layer	101-layer	152-layer
conv1	112×112	7×7, 64, stride 2				
conv2_x	56×56	3×3 max pool, stride 2				
		$\begin{bmatrix} 3 \times 3, 64 \\ 3 \times 3, 64 \end{bmatrix} \times 2$	$\begin{bmatrix} 3 \times 3, 64 \\ 3 \times 3, 64 \end{bmatrix} \times 3$	$\begin{bmatrix} 1 \times 1, 64 \\ 3 \times 3, 64 \\ 1 \times 1, 256 \end{bmatrix} \times 3$	$\begin{bmatrix} 1 \times 1, 64 \\ 3 \times 3, 64 \\ 1 \times 1, 256 \end{bmatrix} \times 3$	$\begin{bmatrix} 1 \times 1, 64 \\ 3 \times 3, 64 \\ 1 \times 1, 256 \end{bmatrix} \times 3$
conv3_x	28×28	$\begin{bmatrix} 3 \times 3, 128 \\ 3 \times 3, 128 \end{bmatrix} \times 2$	$\begin{bmatrix} 3 \times 3, 128 \\ 3 \times 3, 128 \end{bmatrix} \times 4$	$\begin{bmatrix} 1 \times 1, 128 \\ 3 \times 3, 128 \\ 1 \times 1, 512 \end{bmatrix} \times 4$	$\begin{bmatrix} 1 \times 1, 128 \\ 3 \times 3, 128 \\ 1 \times 1, 512 \end{bmatrix} \times 4$	$\begin{bmatrix} 1 \times 1, 128 \\ 3 \times 3, 128 \\ 1 \times 1, 512 \end{bmatrix} \times 8$
conv4_x	14×14	$\begin{bmatrix} 3 \times 3, 256 \\ 3 \times 3, 256 \end{bmatrix} \times 2$	$\begin{bmatrix} 3 \times 3, 256 \\ 3 \times 3, 256 \end{bmatrix} \times 6$	$\begin{bmatrix} 1 \times 1, 256 \\ 3 \times 3, 256 \\ 1 \times 1, 1024 \end{bmatrix} \times 6$	$\begin{bmatrix} 1 \times 1, 256 \\ 3 \times 3, 256 \\ 1 \times 1, 1024 \end{bmatrix} \times 23$	$\begin{bmatrix} 1 \times 1, 256 \\ 3 \times 3, 256 \\ 1 \times 1, 1024 \end{bmatrix} \times 36$
conv5_x	7×7	$\begin{bmatrix} 3 \times 3, 512 \\ 3 \times 3, 512 \end{bmatrix} \times 2$	$\begin{bmatrix} 3 \times 3, 512 \\ 3 \times 3, 512 \end{bmatrix} \times 3$	$\begin{bmatrix} 1 \times 1, 512 \\ 3 \times 3, 512 \\ 1 \times 1, 2048 \end{bmatrix} \times 3$	$\begin{bmatrix} 1 \times 1, 512 \\ 3 \times 3, 512 \\ 1 \times 1, 2048 \end{bmatrix} \times 3$	$\begin{bmatrix} 1 \times 1, 512 \\ 3 \times 3, 512 \\ 1 \times 1, 2048 \end{bmatrix} \times 3$
	1×1	average pool, 1000-d fc, softmax				
FLOPs		1.8×10^9	3.6×10^9	3.8×10^9	7.6×10^9	11.3×10^9

Figura 8 – Tabela comparativa de modelos ResNet - extraído de (HE et al., 2015).

3.2.4 EfficientNet

Com o crescimento dos modelos de redes neurais e a crescente preocupação com os limites de hardware disponíveis, foi introduzido o modelo EfficientNet por Mingxing Tan e Quoc V. Le, do Google Research, no artigo de 2019 intitulado “EfficientNet: Rethinking Model Scaling for Convolutional Neural Networks”. O objetivo principal foi oferecer uma abordagem mais eficiente, mas igualmente precisa, para redes neurais convolucionais. Para adaptar grandes modelos aos recursos limitados, os autores focaram no escalonamento de redes, desenvolvendo técnicas inovadoras que, ao contrário de outros modelos também voltados para eficiência, ajustavam de maneira conjunta, utilizando o que denominaram

de *Compound Scaling*, as dimensões das entradas, possibilitando um equilíbrio entre desempenho e eficiência computacional, sem comprometer a precisão (TAN; LE, 2019).

3.2.4.1 Compound Scaling

A ideia central do *compound scaling* é encontrar uma forma eficiente de aumentar a largura, profundidade e resolução de uma rede ao mesmo tempo, em vez de escalar uma única dimensão isoladamente, como é comum em outros modelos.

- **Largura:** Refere-se ao número de filtros em cada camada da rede. Aumentar a largura pode melhorar a capacidade do modelo para aprender representações mais complexas.
- **Profundidade:** Refere-se ao número de camadas na rede. Aumentar a profundidade permite que o modelo capture informações mais complexas e de alto nível.
- **Resolução:** Refere-se ao tamanho das imagens de entrada para a rede. Imagens de maior resolução podem fornecer mais detalhes, mas também aumentam o custo computacional.

A ideia é usar um único fator de escala que aplica ajustes proporcionais a largura, profundidade e resolução, de modo que todas as dimensões sejam aumentadas de forma balanceada e eficiente.

Foram criados parâmetros α , β , γ atuando na largura, profundidade e resolução respectivamente, tais variáveis eram ajustadas de acordo com a necessidade de balanceamento entre custo computacional e precisão (TAN; LE, 2019).

3.2.4.2 Modelos EfficientNet

Os diferentes modelos do EfficientNet foram desenvolvidos a partir de variações de um único coeficiente de *compound scaling*, mantendo a base da arquitetura. Essa base é composta principalmente por camadas MBConv (*Mobile Inverted Bottleneck Convolution*), inspiradas na arquitetura MobileNet (Figura 9). Cada camada MBConv é estruturada a partir dos seguintes procedimentos:

- **Convolução 1×1 :** Reduzir o número de canais para menor custo computacional;
- **Convolução 3×3 (ou 5×5):** podendo aplicar uma diferenciação que foi denominada de *Depthwise Convolution* que se baseia na aplicação de um único filtro em cada um dos canais em vez do procedimento de convolução tradicional que pode ser mais custoso computacionalmente;

- *Squeeze and Excitation*: é uma técnica que recalibra dinamicamente os canais dos mapas de características (*feature maps*), destacando os mais relevantes e aprimorando a capacidade representativa da rede;
- Convolução 1x1: utilizada para ajustar a dimensionalidade dos mapas de características, adaptando-os aos requisitos da próxima camada;
- Operação de soma residual, inspirada na abordagem *skip connections* da ResNet, que adiciona os dados de entrada à saída da camada.

Cada camada MBConv possui variações como MBConv1 ou MBConv6 onde 6 representa a aplicação de 6 camadas de convoluções depthwise enquanto 1 representa a uma única aplicação (NIMBUS, 2025).

Stage i	Operator $\hat{\mathcal{F}}_i$	Resolution $\hat{H}_i \times \hat{W}_i$	#Channels \hat{C}_i	#Layers \hat{L}_i
1	Conv3x3	224×224	32	1
2	MBConv1, k3x3	112×112	16	1
3	MBConv6, k3x3	112×112	24	2
4	MBConv6, k5x5	56×56	40	2
5	MBConv6, k3x3	28×28	80	3
6	MBConv6, k5x5	14×14	112	3
7	MBConv6, k5x5	14×14	192	4
8	MBConv6, k3x3	7×7	320	1
9	Conv1x1 & Pooling & FC	7×7	1280	1

Figura 9 – Arquitetura EfficientNet - extraído de (TAN; LE, 2019).

4 Trabalhos Relacionados

O trabalho de (ZULUAGA-GOMEZ et al., 2019) utilizou imagens termográficas e CNNs para a detecção de câncer de mama, empregando a base de dados DMR-IR. Como parte do pré-processamento, as Regiões de Interesse (ROIs, do inglês *Region of Interest*) foram extraídas por meio de técnicas como corte, redimensionamento e normalização das imagens térmicas. Diversas CNNs de estado da arte foram avaliadas, além das CNNs propostas pelos próprios autores, denominadas *baseline*, utilizando também otimização de hiperparâmetros. O modelo *surrogate*, também sugerido, tem o objetivo de encontrar a melhor arquitetura de CNN para a base de dados em questão, utilizando uma otimização bayesiana com o estimador *Tree Parzen Estimator*. Os parâmetros considerados na otimização incluem: o número mínimo e máximo de blocos, o número de camadas convolucionais, o número de filtros por bloco, o tipo de otimizador, o tamanho do *kernel*, o tamanho da camada de *pooling*, o tipo de *batch normalization*, a taxa de *dropout* para melhorar a generalização dos modelos, o número de neurônios conectados nas últimas duas camadas e o tipo da camada final. Os autores afirmam que, para a base de dados utilizada, CNNs menores obtêm melhores resultados, fato que motivou a escolha por modelos mais leves.

No trabalho de (GONÇALVES; SOUZA; FERNANDES, 2022) também foi aproveitada a mesma base de dados para desenvolver CNNs para detecção de câncer de mama. Como forma de desenvolver o melhor os modelos de CNN utilizados em seu trabalho anterior (GONÇALVES; SOUZA; FERNANDES, 2021) foi proposto pelos autores a utilização de modelos bioinspirados. Foram exploradas a utilização de algoritmos genéticos (AG), da otimização por enxame de partículas (PSO, do inglês *Particle Swarm Optimization*) e de modelos *surrogate*. Os parâmetros que compõem as redes neurais propostas durante a experimentação com AG são: taxa de aprendizado, número de camadas densamente conectadas, número de neurônios em cada camada, presença ou não de *dropout* e seu valor quando presente. A aplicação de AGs com elitismo, *crossover* com posição de corte única e mutação realizada em dois locais do cromossomo, o binário que indica a presença ou não da camada representada pelo cromossomo e no valor correspondente, esse processo é escolhido aleatoriamente determinando em qual cromossomo será feita a mutação e em qual parte dele. Já os experimentos realizados com o PSO envolveram variações nos seguintes parâmetros: número de iterações, tamanho do enxame e o coeficiente cognitivo (CG, que indica a importância atribuída à melhor solução global encontrada até o momento). Como resultado os autores trouxeram melhorias como: utilizando o modelo *surrogate* com o AG foi possível melhorar o resultado da DenseNet de 0,92 para 1 de F1-score. Já a ResNet com o PSO e o modelo *surrogate* melhorou de 0,85 para 0,92 de F1-score e, mais significa-

tivamente com relação ao tempo, a DenseNet, rede que apenas com o AG demorou 7,99 horas e com o AG e o *surrogate* demorou 2,74 horas.

Já no trabalho de (MILOSEVIC; JANKOVIC; PEULIC, 2014) segue um caminho diferente ao não utilizar CNNs e aplicar funcionalidades baseadas em *Gray Level Co-occurrence Matrices* (GLCM) usadas para avaliar a eficácia das informações texturais contidas nas regiões de massa. Um total de 20 GLCMs foram extraídas de imagens termográficas. O processo de classificação foi realizado com base em características texturais e utilizando a metodologia de validação cruzada, empregando três classificadores: SVM (*Support Vector Machine*), k-NN (*K-Nearest Neighbor*) e *Naive Bayes*. O SVM, fundamentado na teoria do aprendizado estatístico, busca minimizar o erro em dados não vistos durante o treinamento e foca em exemplos difíceis de classificar, os chamados vetores de suporte, o que o torna eficaz para generalização. O k-NN classifica objetos com base na maioria dos votos de seus vizinhos mais próximos no espaço multidimensional, utilizando a distância entre os objetos como métrica. Já o Naive Bayes, um classificador probabilístico baseado no teorema de Bayes, opera em suposição de independência entre as características e se destaca por exigir poucos dados de treinamento. Como resultado os autores forneceram uma melhor performance na utilização de k-NN alcançando valores de precisão por volta de 92,5%, seguido de SVM com 85% e Naive Bayes com 80%.

Se dedicando a também auxiliar na detecção precoce do câncer de mama o trabalho de (SÁNCHEZ-CAUCE; PÉREZ-MARTÍN; LUQUE, 2021) opta por realizar uma abordagem de classificação utilizando múltiplas entradas, combinando imagens térmicas de diferentes ângulos (frontais e laterais) com dados pessoais e clínicos. Para cada entrada foram desenvolvidos modelos e treinados para classificar cada tipo de entrada, ao final as classificações são concatenadas e utilizadas em uma CNN final para fornecer o resultado final para aquele indivíduo. Utilizando o banco de imagens DMR-IR foi possível alcançar resultados como acurácia de 97%, área sob a curva ROC de 0,99, especificidade de 100% e sensibilidade de 83% demonstrando que adicionar imagens em diferentes ângulos ao estudo tem a capacidade de aprimorar a performance do modelo, alcançando melhores resultados quando associado também à dados pessoais e clínicos para reconhecer pacientes doentes.

Seguindo o mesmo caminho de classificação com múltiplas entradas o trabalho de (MAMMOOTTIL et al., 2022) aplica 5 ângulos de visualização diferentes utilizando, entre outros, o banco de imagens DMR-IR. Para cada um dos ângulos de visualização foram treinadas separadamente redes neurais convolucionais buscando obter a melhor performance individual de cada uma, além disso uma outra rede neural seria a responsável por treinar em cima dos dados clínicos de cada indivíduo, as classificações de cada CNN era representada em tuplas com informações de probabilidade de classificação em cada uma das classes, já para a sexta rede neural os resultados eram dados em 0 ou 1 (saúdável

ou doente, respectivamente). Os resultados das 6 redes foram usados para treinamento de uma rede neural final que em seu melhor modelo foi capaz de fornecer a CNN de múltiplas entradas acurácia de 93,8%, sensibilidade de 88,9%, especificidade de 90,6% e área sob a curva ROC de 0,9.

5 Materiais e Métodos

Para verificar o seguimento em direção ao objetivo proposto, foram obtidos dados do banco DMR-IR, que passaram por um devido pré-processamento: as imagens foram adaptadas ao formato de entrada esperado por cada um dos modelos de CNN estudados, incluindo o redimensionamento e a conversão para imagens com 3 canais, quando necessário. Além disso, foi aplicado aumento de dados (*data augmentation*) com o intuito de ampliar o repertório de treinamento das redes convolucionais.

Para a comparação da performance entre os modelos, foram definidos os parâmetros de operação de cada arquitetura, além de utilizadas métricas como acurácia e precisão, acompanhadas de matrizes de confusão para facilitar a visualização dos resultados.

Também foram realizados aperfeiçoamentos nos diversos modelos com o objetivo de otimizá-los e reduzir o número de falsos positivos e falsos negativos, contribuindo assim para a confiabilidade do uso de CNNs em imagens termográficas voltadas à detecção de câncer de mama.

5.1 Banco de dados

O banco de imagens *Database for Mastology Research with Infrared Image* (DMR-IR) oferece imagens térmicas mastológicas para detecção de câncer de mama com a devida separação entre pacientes saudáveis (*Healthy*) e doentes (*Sick*).

Ele contém termografias, matrizes térmicas e dados pessoais e clínicos de 287 pacientes do Hospital Universitário Antônio Pedro (HUAP), da Universidade Federal Fluminense, no Brasil. As imagens foram obtidas com uma câmera termográfica FLIR SC620, capaz de capturar variações de temperatura entre -40°C e 500°C . O processo de aquisição seguiu métodos padronizados, com controle de posicionamento dos pacientes e climatização adequada do ambiente.

As termografias são disponibilizadas de acordo com dois protocolos distintos: estático e dinâmico. No protocolo estático, são realizadas cinco capturas: uma frontal, duas laterais esquerdas (com ângulos de 45° e 90°) e duas laterais direitas (com os mesmos ângulos), representando o corpo em estado térmico estável (Figura 10).

No protocolo dinâmico, o corpo do paciente é submetido a resfriamento com o uso de um ventilador, e as imagens são capturadas ao longo de cinco minutos, durante o processo de recuperação térmica natural. Esse protocolo fornece 20 imagens frontais sequenciais e duas imagens laterais em ângulo de 90° .

Além das imagens termográficas, foram coletadas informações pessoais, como faixa etária, etnia e estado civil. Do ponto de vista clínico, foram considerados fatores como hábitos alimentares, uso de próteses, tratamento hormonal, realização de mamografias e histórico de radioterapia. Adicionalmente, foram registradas condições momentâneas no momento do exame, como a temperatura corporal e o consumo recente de substâncias como café ou álcool, que poderiam interferir nos resultados térmicos.

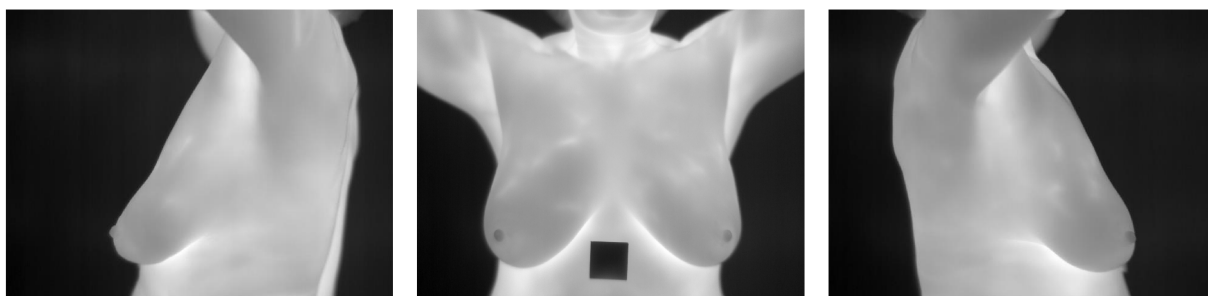


Figura 10 – Exemplo de visualizações de diferentes ângulos - Lateral Esquerda, Frontal e Lateral Direita.

5.2 Classificadores

Serão utilizadas arquiteturas de redes neurais convolucionais consolidadas para a classificação de imagens. Os experimentos feitos por (GONÇALVES; SOUZA; FERNANDES, 2021) exploram as arquiteturas DenseNet201, ResNet50 e VGG16, dessa forma, usando como base seus caminhos, inicialmente procura-se obter seus resultados utilizando uma versão mais leve do primeiro modelo, o mesmo segundo modelo e uma arquitetura diferente comparado com o terceiro, foram escolhidos respectivamente DenseNet121, ResNet50 e EfficientNetB0. Os treinamentos foram feitos com equipamento pessoal, uma placa de vídeo Nvidia 2060 com 6GB de VRAM e processador intel i5 10^a geração.

5.3 Métricas

Como meio de avaliar o desempenho de nossos classificadores foram utilizadas métricas amplamente aplicadas em experimentos semelhantes, as medidas de acurácia (ACC), sensibilidade (SEN), especificidade (SPE), precisão (PREC) e F1-score. Tais métricas são calculadas em termos de verdadeiros positivos (TP), que representam o número de pacientes classificados corretamente como doentes; verdadeiros negativos (TN), o número de pacientes classificados corretamente como saudáveis; falsos positivos (FP), pacientes saudáveis classificados incorretamente como doentes e falsos negativos (FN), pacientes doentes classificados incorretamente como saudáveis. Os valores das métricas

são obtidos utilizando as seguintes equações:

$$ACC = \frac{TP + TN}{TP + TN + FP + FN} \quad (5.1)$$

$$SEN = \frac{TP}{TP + FN} \quad (5.2)$$

$$SPE = \frac{TN}{TN + FP} \quad (5.3)$$

$$PREC = \frac{TP}{TP + FP} \quad (5.4)$$

$$\text{F1-score} = \frac{2 \times PREC \times SEN}{PREC + SEN} \quad (5.5)$$

Como forma de aprimorar a visualização dos resultados foram construídas também matrizes de confusão e curvas ROC (curva de característica de operação do receptor).

5.3.1 Matriz de Confusão

Uma matriz de confusão é um método de visualização de resultados de algoritmos classificadores, permitindo a análise detalhada do desempenho do modelo ao comparar as previsões feitas com os valores reais. Com ela, os dados são organizados em uma tabela que mostra a quantidade de acertos e erros em cada classe, facilitando a identificação de padrões e possíveis problemas no modelo, como vieses ou dificuldade em distinguir determinadas categorias.

A Tabela 1 a seguir apresenta o modelo genérico de uma matriz de confusão para problemas binários:

	Classe Predita	
	Positiva	Negativa
Classe Real Positiva	Verdadeiro Positivo	Falso Negativo
Classe Real Negativa	Falso Positivo	Verdadeiro Negativo

Tabela 1 – Matriz de confusão genérica para classificação binária.

5.3.2 Curvas ROC

A Curva ROC (Receiver Operating Characteristic) é uma ferramenta gráfica usada para avaliar o desempenho de modelos de classificação binária. Ela mostra a relação entre a taxa de verdadeiros positivos (TPR) e a taxa de falsos positivos (FPR) para diferentes limiares de decisão.

Ao analisar a curva ROC, se ela permanecer próxima da diagonal de 45° em relação ao eixo horizontal, isso indica que o modelo tem a mesma probabilidade de classificar um exemplo como positivo ou negativo, ou seja, está realizando uma classificação aleatória. No entanto, com o treinamento adequado, espera-se que a curva se desvie dessa diagonal, movendo-se para acima ou abaixo dela.

Se a curva ROC se mantiver acima da diagonal, significa que o modelo tem maior tendência a classificar corretamente os positivos, indicando um bom desempenho. Por outro lado, se a curva estiver abaixo da diagonal, o modelo pode estar invertendo as classificações, atribuindo maior probabilidade à classe errada.

Para quantificar esse desempenho, utilizamos a Área sob a Curva ROC (AUC - *Area Under the Curve*)(Figura 11). Quanto mais próximo de 1, melhor a capacidade do modelo em distinguir entre classes. Já valores próximos de 0,5 indicam um modelo com desempenho semelhante ao acaso ([Google Developers, 2023](#)).

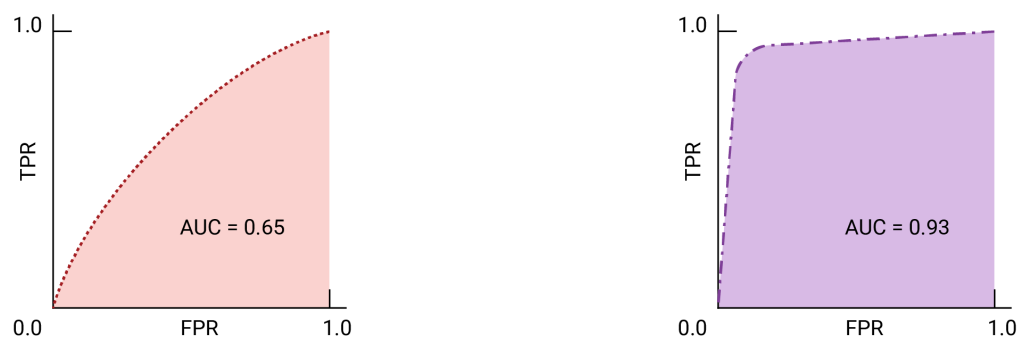


Figura 11 – Curvas ROC com valores de AUC baseados em modelos hipotéticos - extraído de ([Google Developers, 2023](#)).

6 Desenvolvimento

O estudo foi feito inicialmente buscando obter os melhores resultados utilizando imagens frontais procurando manter balanceamento entre as classes, foram feitas também modificações seguindo o estabelecido por (GONÇALVES; SOUZA; FERNANDES, 2021) com o intuito de registrar diferenças e semelhanças entre as arquiteturas recebendo as mesmas alterações. Em um segundo momento foram selecionadas as arquiteturas modificadas com as melhores performances para serem aplicadas em treinamentos utilizando imagens frontais e laterais para incrementar o campo de detecção para áreas não visualizadas nos primeiros testes, isso sob um treinamento mais longo para alcançar melhor generalização na classificação dos indivíduos.

6.1 Ajuste na base de dados

Para utilizar as imagens do banco DMR-IR no treinamento do modelo, foi necessário realizar alguns ajustes. Inicialmente, as imagens térmicas, originalmente em escala de cinza, foram convertidas para a coloração “hot” da biblioteca Matplotlib (Python), facilitando a interpretação visual. Além disso, as imagens foram armazenadas no banco tanto no formato .png quanto como matrizes de pixels em arquivos .txt. Para garantir maior fidelidade das informações, foi dada prioridade às imagens reconstruídas a partir dos arquivos .txt, sempre que disponíveis, a fim de evitar possíveis distorções decorrentes da compressão do formato .png.

O banco de dados apresentava um número significativamente maior de imagens de pacientes saudáveis em comparação aos diagnosticados com a doença. Para garantir um melhor balanceamento entre as classes, foi aplicada a técnica de *data augmentation*, buscando manter uma proporção de 50% para cada classe. As técnicas utilizadas incluíram o espelhamento horizontal das imagens de pacientes doentes e uma ampliação de 10%, mantendo a resolução original. Além disso, para evitar viés e garantir uma melhor generalização do modelo, assegurou-se que imagens de um mesmo paciente não estivessem presentes simultaneamente em mais de um conjunto (por exemplo, tanto no treinamento quanto no teste).

Considerando apenas imagens frontais, a base de dados original era composta por 386 pacientes saudáveis, com um total de 3473 imagens, e 67 pacientes doentes, com um total de 832 imagens. Com o uso de técnicas de *data augmentation*, foi possível, para o primeiro modelo de testes, compor um conjunto de dados com aproximadamente 3000 imagens frontais de pacientes saudáveis e 3000 imagens frontais de pacientes doentes. Esse conjunto foi dividido da seguinte forma: 70% para treinamento, 15% para validação e 15%

para teste.

Para o segundo modelo, foi considerada a seleção de apenas três imagens por paciente: uma frontal e duas laterais (uma de cada perspectiva). Novos pacientes foram gerados a partir de imagens modificadas de pacientes originais, também por meio de *data augmentation*. Dessa forma, a base de dados para este modelo totalizou 240 pacientes e 720 imagens, mantendo a distribuição de 50% de saudáveis e 50% de doentes, com a mesma divisão de 70% para treinamento, 15% para validação e 15% para teste.

6.2 Classificação com CNNs selecionadas

Cada treinamento foi realizado de forma independente em cinco execuções. Inicialmente, cada execução consistiu em 10 épocas, ou seja, todo o conjunto de dados foi percorrido dez vezes durante o treinamento do modelo. Ao final das 10 épocas, encerrava-se uma execução, e seus resultados eram armazenados. Após as cinco execuções, foi adotado um modelo de consenso por votação majoritária para definir as predições finais: para cada entrada, a classe prevista pela maioria dos modelos era considerada como a classificação final.

Essa estratégia visa reduzir os efeitos da aleatoriedade próprios do processo de inicialização dos modelos, permitindo uma avaliação mais robusta e representativa do desempenho médio esperado.

Na primeira etapa do estudo, buscou-se uma reprodução parcial do trabalho de (GONÇALVES; SOUZA; FERNANDES, 2021), que utilizou as arquiteturas DenseNet201, ResNet50 e VGG16, aplicando modificações nos modelos de CNNs e treinamentos com duração de 30 épocas. Para efeito de comparação direta, utilizou-se neste trabalho a ResNet50, além da DenseNet121 — como uma alternativa menos profunda à DenseNet201 — e a EfficientNetB0 — como uma opção de menor custo computacional. Todos os modelos propostos foram treinados por um número reduzido de épocas (10), em comparação às 30 utilizadas por (GONÇALVES; SOUZA; FERNANDES, 2021), devido ao escopo mais restrito deste estudo — voltado à seleção da melhor modificação a ser empregada em etapas posteriores — e também a limitações de hardware. Ainda assim, mantiveram-se modificações estruturais semelhantes às aplicadas no trabalho original, obedecendo a seguinte nomenclatura:

- (A) O modelo padrão da arquitetura sem modificações.
- (B) Duas camadas totalmente conectadas (2 FC) com 256 neurônios na primeira camada e 2 neurônios na saída da segunda camada ($256 \rightarrow 2$).
- (C) Duas camadas totalmente conectadas (2 FC) com 512 neurônios na primeira camada e 2 neurônios na segunda ($512 \rightarrow 2$).

- (D) Duas camadas totalmente conectadas (2 FC) com 1024 neurônios na primeira camada e 2 neurônios na segunda ($1024 \rightarrow 2$).
- (E) Três camadas totalmente conectadas (3 FC), com 4096 neurônios na primeira camada, 1024 na segunda e 2 na última camada de saída ($4096 \rightarrow 1024 \rightarrow 2$).
- Todos os casos realizados com taxas de aprendizado de (1) 0.01 e (2) 0.001 totalizando 10 testes diferentes para cada arquitetura.

Essas variações foram analisadas para avaliar o impacto do número de camadas e neurônios na performance do modelo. As modificações que obtiveram os melhores resultados foram utilizadas em novas experimentações.

Para a segunda etapa, foram criados modelos de CNN compostas por 3 sub-redes com as modificações da etapa anterior que obtiveram melhores resultados, agora sob uma execução mais longa utilizando 30 épocas, elas receberam como entrada a imagem frontal, lateral esquerda e lateral direita respectivamente, os resultados do processamento das sub-redes foram integrados e utilizados em camadas totalmente conectadas com o devido *bottleneck* para uma classificação binária fornecendo os resultados finais.

6.3 Resultados e Discussões

Todos os resultados obtidos representam a performance do modelo de consenso por votação para demonstrar o comportamento esperado do modelo específico sob tais parâmetros de treinamento.

6.3.1 Primeira Etapa

Foram realizadas 5 execuções de treinamentos de 10 épocas seguindo a nomenclatura de modificações citada anteriormente.

6.3.1.1 DenseNet121

Com a DenseNet121 houve destaque para o modelo E2 com resultados próximos de 80% para todas as métricas (Tabela 2) resultando em uma matriz de confusão com distribuição equilibrada entre as classes, com nível razoável de erros, além de uma área sob a curva ROC indicando boa capacidade de distinção entre as classes (Figura 12).

6.3.1.2 ResNet50

Para o caso dos experimentos feitos com a ResNet50 dois experimentos se mostraram bem próximos A2 e E1 (Tabela 3), apesar dos resultados melhores em acurácia, especificidade e precisão por parte de A2 a medida F1 foi menor que no experimento E,

Tabela 2 – Métricas DenseNet121 - Experimento E2

Métrica	Valor (%)
Acurácia	82,60
Sensibilidade	79,60
Especificidade	85,60
Precisão	84,68
F1-score	82,06

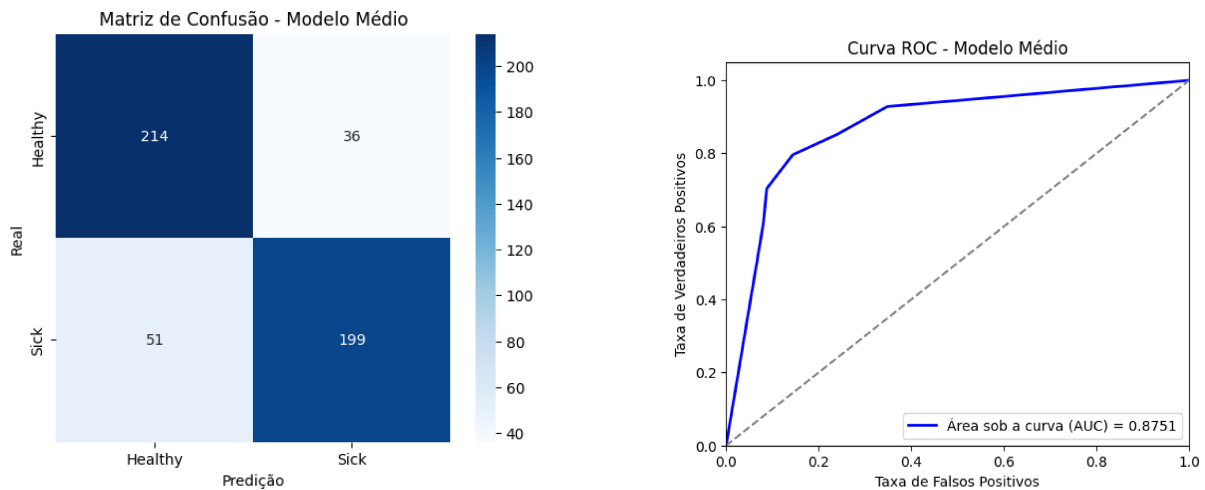


Figura 12 – À esquerda: Matriz de Confusão. À direita: Curva ROC da DenseNet E2.

dessa forma, apesar de E alcançar valores menores mas próximos em algumas das métricas ele se mostrou um modelo com resultados mais estáveis e foi selecionado para a segunda etapa. Nesse sentido, a ResNet50 E1 apresentou matriz de confusão equilibrada e área sob a curva ROC em níveis razoáveis acima de 0,80 (Figura 13).

Tabela 3 – Comparação das Métricas dos Experimentos A2 e E1

Métrica	Experimento A2	Experimento E1
Acurácia	82,60	82,00
Sensibilidade	75,20	78,40
Especificidade	90,00	85,60
Precisão	88,26	84,48
F1-score	81,21	81,33

6.3.1.3 EfficientNetB0

Para a EfficientNetB0 o experimento C1 demonstrou os melhores resultados com um equilíbrio entre todas as métricas com destaque para o valor de 90% na especificidade (Tabela 4). Também apresentou uma matriz de confusão equilibrada, porém com muitos falsos negativos reflexo da sensibilidade baixa. Além disso, obteve o melhor valor para área sob a curva ROC se aproximando de 0.9 (Figura 14).

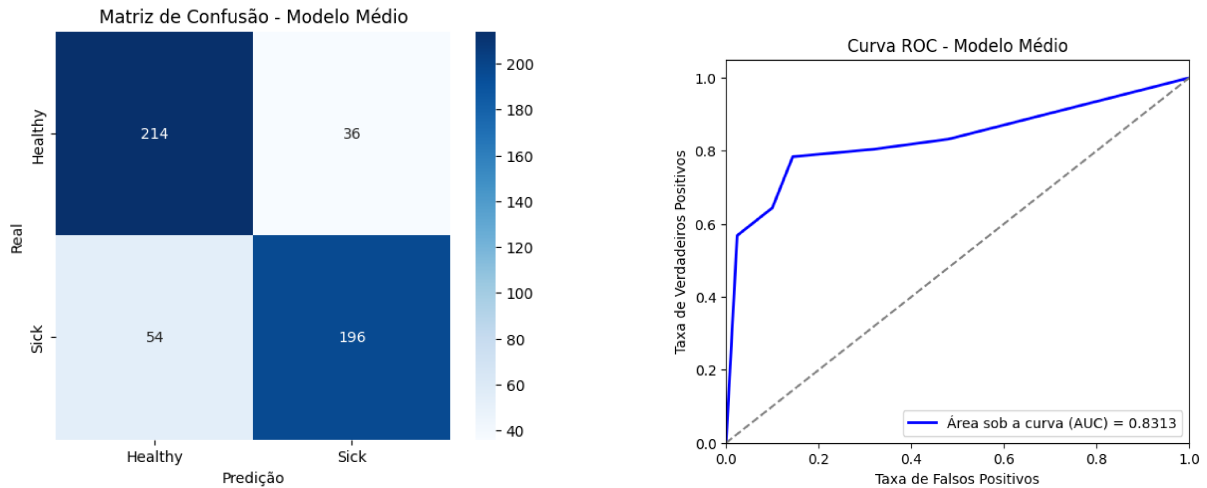


Figura 13 – À esquerda: Matriz de Confusão. À direita: Curva ROC da ResNet50 E1.

Tabela 4 – Métricas EfficientNetB0 - Experimento C1

Métrica	Valor (%)
Acurácia	81,40
Sensibilidade	72,40
Especificidade	90,40
Precisão	88,29
F1-score	79,56

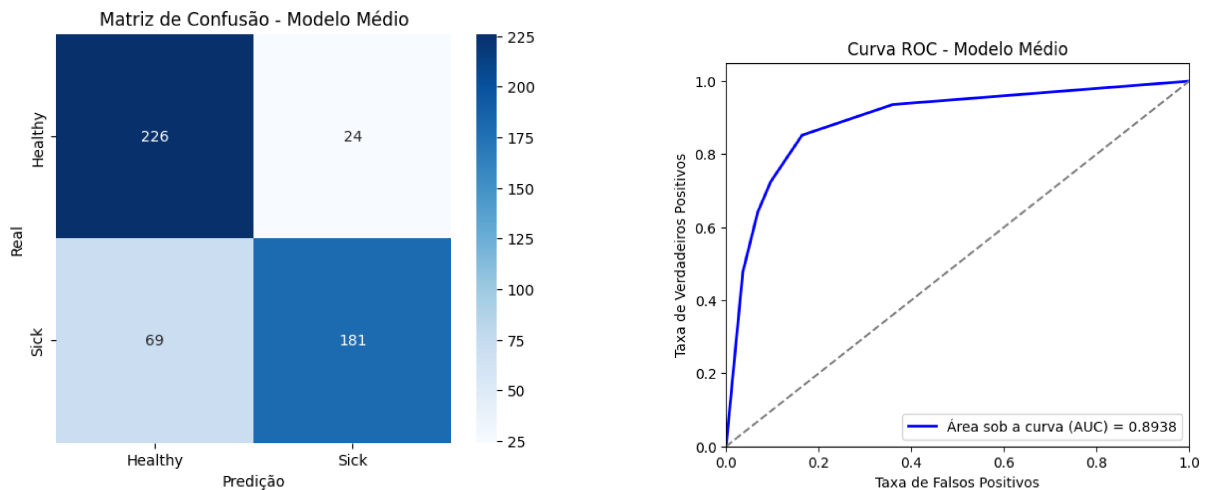


Figura 14 – À esquerda: Matriz de Confusão. À direita: Curva ROC da EfficientNetB0 C1.

6.3.1.4 Comparativo

A Tabela 5 apresenta uma comparação direta com o trabalho anterior de (GONÇALVES; SOUZA; FERNANDES, 2021), referido como TA (Trabalho Anterior). Os valores mostrados para as ResNets se aproximam na maioria dos casos, pois em ambos os trabalhos foram utilizados ResNet50 havendo diferenciações no modo como foi feita o *data*

augmentation e a duração do treinamento. Para a DenseNet, a diferença entre os modelos DenseNet121 e DenseNet201 tiveram impacto nos resultados, na maioria dos casos o primeiro se saiu pior nas métricas comparado ao segundo utilizado em TA.

Tabela 5 – Resultados Obtidos na Etapa 1 em Comparativo com TA

Experimento	CNN	Acurácia (%)	Acurácia TA (%)	F1-Score (%)	F1-Score TA (%)
A1	DenseNet	76,00	80,85	74,25	80,85
A2	DenseNet	81,20	80,85	78,64	80,85
A2	ResNet	82,60	80,85	81,21	81,63
B2	DenseNet	77,80	77,66	75,82	76,40
B2	ResNet	79,80	77,66	78,09	77,42
C2	ResNet	82,20	79,79	80,94	80,00
D1	DenseNet	77,20	80,85	77,56	80,85
D2	DenseNet	77,20	87,23	73,36	88,24
D2	ResNet	81,60	81,91	80,59	83,81
E1	DenseNet	78,60	84,04	76,48	84,85
E1	ResNet	82,00	81,91	81,33	83,17

6.3.2 Segunda Etapa

Com os resultados obtidos da primeira etapa foram selecionados os modelos DenseNet121 - E2, ResNet50 - E1 e EfficientNetB0 - C1 para serem aplicados no treinamento proposto para a segunda etapa. Percebe-se que em geral a aplicação de novas camadas totalmente conectadas com maior densidade de neurônios tende a afetar positivamente o desempenho dos modelos como mostrado pelo experimento E nas DenseNet e ResNet, contudo para a EfficientNet, um modelo desenvolvido para trabalhar com menos camadas em busca de um caminho que exige menos recursos computacionais, os resultados foram melhores obtidos em experimentos com menor densidade de neurônios como o experimento C1, camadas mais densas como nos experimentos subsequentes houveram sinais de *overfitting* ao mostrar algumas épocas de treinamento *train-loss* de 0,0846 e *validation-loss* de 0,9440, ou seja, um treinamento com resultados precisos porém uma validação com grandes perdas.

Como definido anteriormente cada modelo seria replicado em outros 3 submodelos que iriam agregar resultados para uma classificação final, utilizando de imagens frontais e duas laterais como entrada.

6.3.2.1 ResNet

Para o modelo E1 da ResNet50 foram obtidos os resultados menos satisfatórios (Tabela 6), porém ainda se mantendo em um nível ligeiramente superior à primeira etapa e com detecção aceitável dos casos negativos. Desmonstrou bons resultados para a matriz de confusão mantendo apenas alguns casos de erro e curva roc alcançando rapidamente taxa de 1,0 de verdadeiros positivos (Figura 15).

Tabela 6 – Métricas ResNet50 - E1 - Etapa 2

Métrica	Valor (%)
Acurácia	83,33
Sensibilidade	77,78
Especificidade	88,89
Precisão	87,50
F1-score	82,35

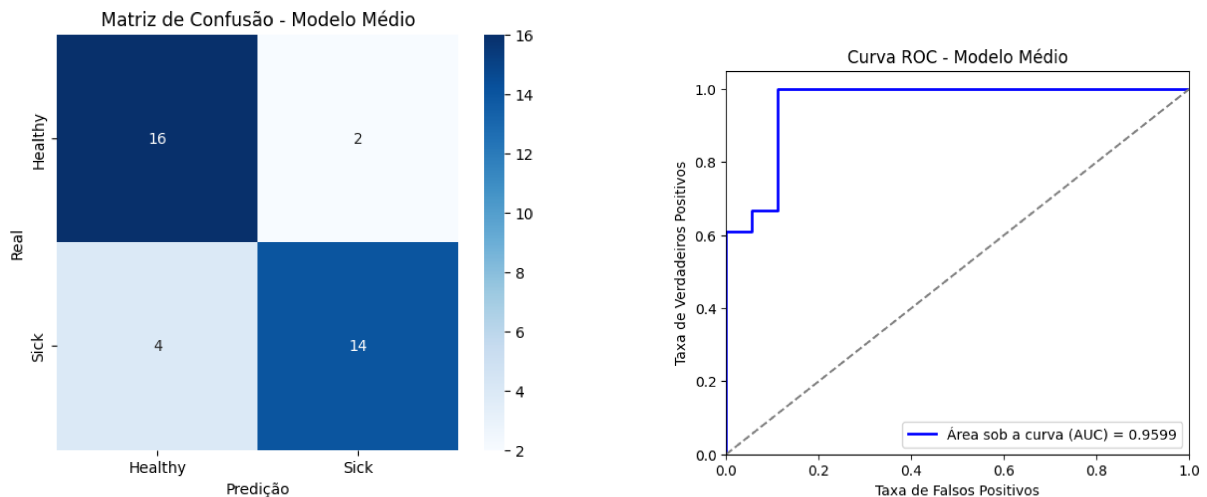


Figura 15 – Matriz de Confusão e Curva ROC para etapa 2 - ResNet E1.

6.3.2.2 EfficientNet

Para o modelo C da EfficientNetB0 seus resultados colocaram como o segundo melhor na segunda etapa, com a capacidade de detectar cerca de 94% dos casos positivos (Tabela 7). A matriz de confusão revelou apenas 3 erros de classificação entre os 36 indivíduos testados, enquanto a curva ROC apresentou os melhores resultados entre os modelos da etapa 2, com uma área sob a curva de 0,9969, indicando uma excelente capacidade de discriminação do modelo (Figura 16).

Tabela 7 – Métricas EfficientNetB0 - C1 - Etapa 2

Métrica	Valor (%)
Acurácia	91,67
Sensibilidade	94,44
Especificidade	88,89
Precisão	89,47
F1-score	91,89

6.3.2.3 DenseNet

Os melhores resultados foram então obtidos pelo modelo DenseNet121 - E2 (Tabela 8), sendo a DenseNet uma arquitetura treinada para trabalhar com grande volume

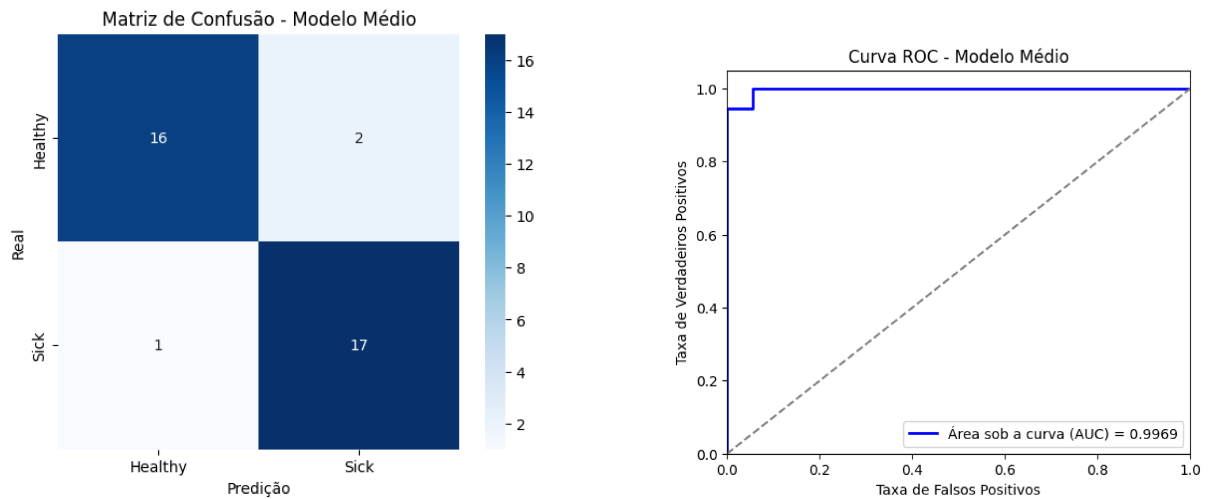


Figura 16 – Matriz de Confusão e Curva ROC para etapa 2 - EfficientNet C1.

de informações, o experimento E2 forneceu densas camadas FC para utilizar do grande número de parâmetros em cada sub-rede. Tais métricas resultam em uma matriz de confusão com apenas 2 erros de classificação e uma curva ROC com rápida ascensão alcançando área sob a curva de 0,9907 (Figura 17).

Tabela 8 – Métricas DenseNet121 - E2 - Etapa 2

Métrica	Valor (%)
Acurácia	94,44
Sensibilidade	100,00
Especificidade	88,89
Precisão	90,00
F1-score	94,74

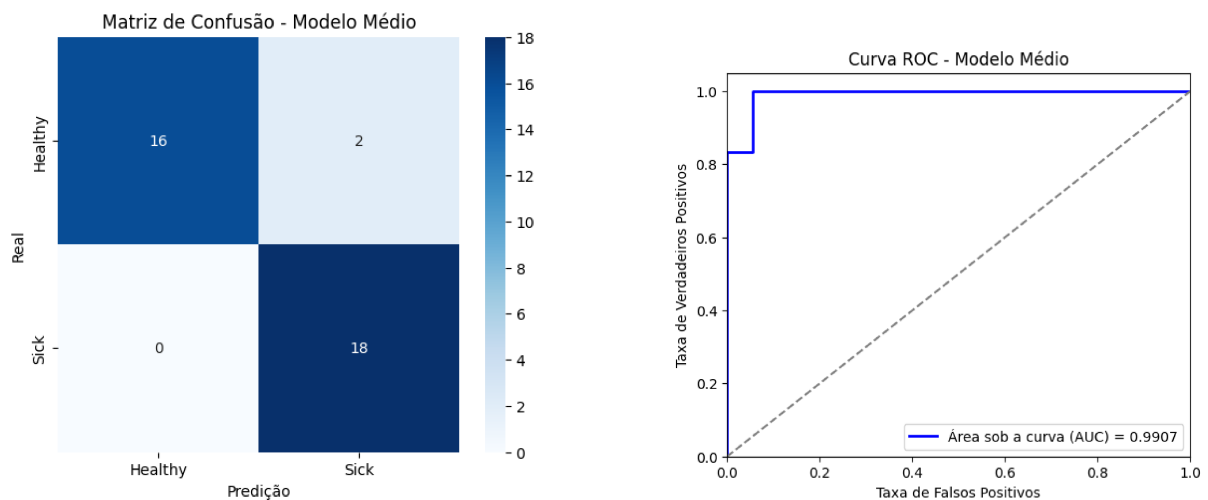


Figura 17 – Matriz de Confusão e Curva ROC para etapa 2 - DenseNet121 E2.

6.3.2.4 Comparativo

Ao visualizar na Tabela 9 os aprimoramentos feitos pela segunda etapa foram capazes de se equiparar com as melhores performances (experimento D2) no trabalho proposto por (GONÇALVES; SOUZA; FERNANDES, 2021).

Tabela 9 – Resultados Obtidos na Etapa 2 em Comparativo com TA

CNN	Acurácia (%)	F1-Score (%)
DenseNet201D2 (TA)	87,23	88,24
ResNet50D2 (TA)	81,91	83,81
DenseNet121E2	94,44	94,74
EfficientNetB0C1	91,67	91,89
ResNet50E1	83,33	82,35

6.3.3 Discussões

De modo geral, os experimentos da primeira etapa demonstraram que o aumento no número e na densidade de camadas totalmente conectadas tem um impacto positivo no desempenho dos modelos. No entanto, essas modificações exigem que a arquitetura seja capaz de lidar com a elevada quantidade de parâmetros gerados. Nesse contexto, a DenseNet, por contar com mecanismos estruturais que facilitam essa adaptação, apresentou desempenho superior à EfficientNet, cuja proposta é mais voltada à eficiência computacional e ao uso reduzido de recursos.

Na segunda etapa, a adaptação dos modelos para processar conjuntos compostos por três entradas permitiu uma interpretação mais abrangente da região de interesse, revelando áreas afetadas que, anteriormente, estavam ocultas em visualizações exclusivamente frontais. Em contraste com a etapa anterior — em que a análise era restrita à visão frontal —, essa abordagem expandida demonstrou-se vantajosa para a tarefa de classificação, conforme evidenciado pelos resultados mais promissores obtidos. Dos 36 pacientes (totalizando 108 imagens) utilizados na fase de testes, os modelos mais eficazes foram capazes de classificar corretamente mais de 90% dos casos.

A Tabela 10 apresenta a comparação do desempenho do modelo proposto com abordagens correlatas da literatura. Observa-se que as arquiteturas EfficientNetB0 e DenseNet121 mantiveram resultados competitivos, alcançando F1-scores semelhantes ou superiores a modelos mais complexos, como a DenseNet201 e a ResNet50 utilizadas por (GONÇALVES; SOUZA; FERNANDES, 2022).

Sobre estudos semelhantes utilizando a abordagem de múltiplas entradas, o modelo proposto foi capaz de superar razoavelmente os resultados do trabalho de (MAMMOOTIL et al., 2022) mantendo-se superior em acurácia e semelhante em sensibilidade e especificidade. Já o trabalho de (SÁNCHEZ-CAUCE; PÉREZ-MARTÍN; LUQUE,

2021) foi capaz de obter melhores resultados em acurácia demonstrando também foco na classificação dos casos negativos como pode ser observado com o valor de 100% para a especificidade.

Em comparação a métodos mais distintos, como o K-NN utilizado por (MILOSEVIC; JANKOVIC; PEULIC, 2014), o modelo baseado na DenseNet121 obteve melhor desempenho em termos de acurácia. No entanto, modelos mais recentes com estratégias de otimização de hiperparâmetros, como o proposto por (ZULUAGA-GOMEZ et al., 2019), apresentaram métricas superiores em precisão, semelhantes em acurácia e ligeiramente inferiores em F1-score. Ainda assim, os resultados obtidos neste trabalho demonstram a robustez e a eficácia da abordagem adotada, especialmente considerando o uso de arquiteturas mais leves.

Referência	Abordagem	Modelo	Acurácia	Precisão	F1-Score	SEN	SPE
Modelo Proposto	CNN	DenseNet121	94,44	90,00	94,74	100	88,89
		EfficientNetB0	91,67	89,47	91,89	94,44	88,89
		ResNet50	83,33	87,50	82,35	77,78	88,89
(GONÇALVES; SOUZA; FERNANDES, 2022)	AG	DenseNet201	91,66	-	92,30	100	83,33
		ResNet50	91,66	-	90,90	83,33	100
	PSO	DenseNet201	91,66	-	92,30	100	83,33
		ResNet50	83,33	-	85,71	100	66,66
(MILOSEVIC; JANKOVIC; PEULIC, 2014)	K-NN	-	92,5	75	-	78,6	100
(ZULUAGA-GOMEZ et al., 2019)	otimização de Hiperparâmetros	CNN-Hyp 1	94	92	91	92	-
(SÁNCHEZ-CAUCE; PÉREZ-MARTÍN; LUQUE, 2021)	CNN de Múltiplas Entradas	-	97	-	-	83	100
(MAMMOOTTIL et al., 2022)	CNN de Múltiplas Entradas	-	93,8	-	-	88,9	96,7

Tabela 10 – Comparação entre modelos correlatos e desempenho do modelo proposto. Valores em %.

7 Conclusão

A aplicação de CNNs na classificação de imagens termográficas tem se mostrado uma abordagem promissora no diagnóstico do câncer de mama. Com os avanços nas tecnologias de aquisição de imagem e no desenvolvimento de arquiteturas de redes neurais, foi possível aprimorar a identificação de alterações térmicas associadas a modificações metabólicas características da doença, permitindo análises mais sensíveis e precisas. Esses progressos contribuíram para classificações mais confiáveis, reforçando a utilidade da termografia como ferramenta diagnóstica. As modificações propostas por (GONÇALVES; SOUZA; FERNANDES, 2021) demonstraram o impacto da densidade de neurônios nas CNNs, evidenciando ganhos de desempenho em modelos projetados para lidar com alta dimensionalidade, enquanto redes mais simples apresentaram sinais de *overfitting*, comprometendo a capacidade de generalização.

Este estudo investigou arquiteturas consolidadas na tarefa de classificação de imagens, como DenseNet, ResNet e EfficientNet, com adaptações voltadas à melhoria do desempenho em imagens médicas. Entre os modelos avaliados, a DenseNet121, mesmo em sua versão mais compacta, obteve os melhores resultados, com um F1-score de 94,74%. Além disso, apresentou uma curva ROC consistente, com sensibilidade de 100%, o que é crucial para evitar a classificação errônea de pacientes com a doença como saudáveis. Em comparação com estudos anteriores, os resultados obtidos demonstraram-se competitivos, evidenciando o potencial da abordagem proposta para futuras aplicações e aprimoramentos no diagnóstico de câncer de mama.

Como perspectiva para trabalhos futuros, sugerem-se testes com modelos mais complexos e densos, que poderiam ser explorados com o uso de hardware mais robusto, buscando verificar se há ganhos adicionais de desempenho. Além disso, utilizar bases de dados maiores para melhorar a generalização dos modelos. Outra proposta é ampliar a arquitetura de múltiplas entradas, passando a considerar até cinco visualizações diferentes por paciente, o que pode enriquecer a análise das imagens. Por fim, destaca-se a possibilidade de delimitar as regiões de interesse, realizando testes com esses recortes, a fim de avaliar potenciais ganhos de desempenho na classificação.

Referências

Amazon Web Services. **What is a Neural Network?** 2024. <<https://aws.amazon.com/what-is/neural-network/>>. Citado 2 vezes nas páginas 16 e 17.

BOYD, N. F.; GUO, H.; MARTIN, L. J.; SUN, L.; STONE, J.; FISHELL, E.; JONG, R. A.; HISLOP, G.; CHIARELLI, A.; MINKIN, S. et al. Mammographic density and the risk and detection of breast cancer. **New England journal of medicine**, Mass Medical Soc, v. 356, n. 3, p. 227–236, 2007. Citado na página 15.

Departamento de Ciência da Computação (IC/UFF). **DMR-IR: Breast Thermography Image Dataset**. 2014. <<https://visual.ic.uff.br/dmi/>>. Acesso em: 11 março 2025. Citado 2 vezes nas páginas 5 e 16.

DOMINGOS, A. C. P. Redes neurais convolucionais para detecção de câncer de mama utilizando regiões de interesse de imagens infravermelhas. 2021. Disponível em: <<https://repositorio.ufu.br/handle/123456789/33864>>. Citado na página 11.

FIGUEIREDO, A. A. A. et al. Uso de correlações entre temperaturas superficiais e qualificação da técnica de imagens infravermelhas para a detecção e localização de tumores de mama. Universidade Federal de Uberlândia, 2018. Disponível em: <<https://repositorio.ufu.br/handle/123456789/22568>>. Citado 2 vezes nas páginas 14 e 15.

GeeksforGeeks. **DenseNet Explained**. 2024. <<https://www.geeksforgeeks.org/densenet-explained/>>. Citado na página 19.

GFG. **Residual Networks (ResNet) - Deep Learning**. 2025. <<https://www.geeksforgeeks.org/residual-networks-resnet-deep-learning/>>. Citado na página 22.

GONÇALVES, C. B.; SOUZA, J. R.; FERNANDES, H. Classification of static infrared images using pre-trained cnn for breast cancer detection. IEEE, 2021. Disponível em: <<https://ieeexplore.ieee.org/document/9474692>>. Citado 11 vezes nas páginas 11, 13, 18, 19, 26, 30, 33, 34, 37, 41 e 43.

_____. Algoritmos bioinspirados para otimização de cnns aplicados na detecção de câncer de mama utilizando imagens termográficas. Universidade Federal de Uberlândia, 2022. Disponível em: <<https://repositorio.ufu.br/handle/123456789/36092>>. Citado 3 vezes nas páginas 26, 41 e 42.

GOODFELLOW, I.; BENGIO, Y.; COURVILLE, A. **Deep Learning**. MIT Press, 2016. Capítulo 6 – Redes neurais profundas. Disponível em: <<https://www.deeplearningbook.org>>. Citado na página 19.

_____. **Deep Learning - Convolutional Networks**. 2016. Capítulo 9 – Redes Convolucionais. Disponível em: <<https://www.deeplearningbook.org/contents/convnets.html>>. Citado na página 18.

Google Developers. **Curso Intensivo de Machine Learning: Curva ROC e AUC**. 2023. <<https://developers.google.com/machine-learning/crash-course/classification/roc-and-auc?hl=pt-br>>. Acessado em: abr. 2025. Citado 2 vezes nas páginas 5 e 32.

- HE, K.; ZHANG, X.; REN, S.; SUN, J. Deep residual learning for image recognition. **arXiv preprint arXiv:1512.03385**, 2015. Acessado: 2025-03-25. Disponível em: <<https://arxiv.org/abs/1512.03385>>. Citado 3 vezes nas páginas 5, 22 e 23.
- HUANG, G.; LIU, Z.; WEINBERGER, K. Q. Densely connected convolutional networks. **CoRR**, abs/1608.06993, 2016. Disponível em: <<http://arxiv.org/abs/1608.06993>>. Citado 4 vezes nas páginas 5, 19, 20 e 21.
- IBM. **Neural Networks**. 2025. <<https://www.ibm.com/br-pt/topics/neural-networks>>. Acesso em: 11 de abril de 2025. Citado na página 17.
- IBM Data and AI Team. Ai vs. machine learning vs. deep learning vs. neural networks. **IBM**, 2023. Disponível em: <<https://www.ibm.com/think/topics/ai-vs-machine-learning-vs-deep-learning-vs-neural-networks>>. Citado 3 vezes nas páginas 5, 16 e 17.
- INCA, I. N. do C. **Câncer de mama**. 2024. Disponível em: <<https://www.gov.br/inca/pt-br/assuntos/cancer/tipos/mama>>. Citado 2 vezes nas páginas 11 e 14.
- LIMA, M.; RUBIK, E.; MORAIS, R. Introdução ao reconhecimento de imagens. **LAMFO**, 2020. Disponível em: <<https://lamfo-unb.github.io/2020/12/05/Captcha-Break/>>. Citado 2 vezes nas páginas 5 e 18.
- MAMMOOTIL, M.; KULANGARA, L.; CHERIAN, A.; MOHANDAS, P.; HASIKIN, K.; MAHMUD, M. Detection of breast cancer from five-view thermal images using convolutional neural networks. **Journal of Healthcare Engineering**, Hindawi, v. 2022, p. 4295221, 2022. Disponível em: <<https://pmc.ncbi.nlm.nih.gov/articles/PMC8901325/>>. Citado 3 vezes nas páginas 27, 41 e 42.
- MILOSEVIC, M.; JANKOVIC, D.; PEULIC, A. Thermography based breast cancer detection using texture features and minimum variance quantization. **EXCLI Journal**, 2014. Disponível em: <<https://pmc.ncbi.nlm.nih.gov/articles/PMC4464488/>>. Citado 3 vezes nas páginas 11, 27 e 42.
- Ministério da Saúde. **Câncer**. 2024. Disponível em: <[https://www.gov.br/saude/pt-br/assuntos/saude-de-a-a-z/c/cancer#:~:text=CÃncer%20\(ou%20tumor%20maligno\)%20ÃI,origem%20do%20tumor%20\(metÃastases\)](https://www.gov.br/saude/pt-br/assuntos/saude-de-a-a-z/c/cancer#:~:text=CÃncer%20(ou%20tumor%20maligno)%20ÃI,origem%20do%20tumor%20(metÃastases))>. Citado na página 14.
- NIMBUS, M. **EfficientNet: Aumentando a Eficiência em Redes Neurais**. 2025. Acessado em 25 março 2025. Disponível em: <<https://www.monolitonimbus.com.br/efficientnet/>>. Citado na página 25.
- PASTORE, A. R. **Ultrassonografia em Ginecologia e Obstetrícia**. 1. ed. Rio de Janeiro: Editora Revinter, 2003. Informação retirada do Capítulo 64. Citado na página 15.
- SOCIETY, A. C. **About breast cancer**. 2020. Disponível em: <<https://www.cancer.org/content/dam/CRC/PDF/Public/8577.00.pdf>>. Citado 2 vezes nas páginas 11 e 15.
- SÁNCHEZ-CAUCE, R.; PÉREZ-MARTÍN, J.; LUQUE, M. Multi-input convolutional neural network for breast cancer detection using thermal images and clinical data. **Computer Methods and Programs in Biomedicine**, v. 204, p. 106045, 2021. ISSN 0169-2607. Disponível em: <<https://www.sciencedirect.com/science/article/pii/S0169260721001206>>. Citado 2 vezes nas páginas 27 e 42.

TAN, M.; LE, Q. V. Efficientnet: Rethinking model scaling for convolutional neural networks. **arXiv preprint arXiv:1905.11946**, 2019. Disponível em: <https://arxiv.org/abs/1905.11946>. Citado 3 vezes nas páginas 5, 24 e 25.

ZHANG, Z.; YU, D. **ResNet**. 2021. https://pt.d2l.ai/chapter_convolutional-modern/resnet.html. Acessado: 2025-03-25. Citado 2 vezes nas páginas 5 e 22.

ZULUAGA-GOMEZ, J.; MASRY, Z. A.; BENAGGOUNE, K.; MERAGHNI, S.; ZERHOUNI, N. A cnn-based methodology for breast cancer diagnosis using thermal images. 2019. Disponível em: <https://arxiv.org/pdf/1910.13757>. Citado 4 vezes nas páginas 11, 12, 26 e 42.