

# Análise Comparativa de Arquiteturas de Redes Neurais Convolucionais: O Impacto da Remoção de Pelos e Ruídos na Aprendizagem do Algoritmo

Matheus Rodrigues de Oliveira Solon<sup>1</sup>, Vivian Kailany Marques de Queiroz<sup>1</sup>,  
Roney Nogueira Sousa<sup>1</sup>

<sup>1</sup>Instituto Federal de Educação, Ciência e Tecnologia do Ceará (IFCE)  
Av. Parque Central, 1315 - Distrito Industrial I, Maracanaú - CE

{matheus.solon05, vivian.kailany.marques08}@aluno.ifce.edu.br

roney.nogueira@ifce.edu.br

**Abstract.** *This work presents a comparative analysis of different Convolutional Neural Network architectures — ResNet50V2, InceptionV3, EfficientNetB7, and MobileNetV2 — aiming to evaluate the impact of dermoscopic image preprocessing, particularly hair and noise removal, on classification performance. Four experimental scenarios were considered, with and without the use of transfer learning and data augmentation. Results showed that ResNet50V2, when combined with preprocessing and deep learning strategies, achieved the best performance. The study concludes that proper image treatment is crucial to improve accuracy and the discriminative capability of CNN models.*

**Resumo.** *Este trabalho realiza uma análise comparativa entre diferentes arquiteturas de Redes Neurais Convolucionais — ResNet50V2, InceptionV3, EfficientNetB7 e MobileNetV2 — com o objetivo de avaliar o impacto do pré-processamento de imagens dermatoscópicas, especialmente a remoção de pelos e ruídos, no desempenho dos modelos de classificação. Foram considerados quatro cenários experimentais com e sem o uso de transfer learning e data augmentation. Os resultados mostraram que a ResNet50V2, combinada ao pré-processamento e técnicas de aprendizagem profunda, obteve os melhores desempenhos. Conclui-se que o tratamento adequado das imagens é essencial para a melhora na acurácia e na capacidade discriminativa dos modelos.*

## 1. Introdução

O câncer de pele é o tipo mais frequente no Brasil, representando cerca de 30% dos tumores malignos registrados, sendo o melanoma responsável por 1.978 mortes e o não melanoma por 1.488 em 2020 [da Saúde 2024]. Globalmente, estima-se que ocorram entre dois a três milhões de casos de câncer de pele não melanoma e aproximadamente 132 mil casos de melanoma por ano [da Saúde 2025]. A detecção precoce é essencial para o aumento da taxa de sobrevivência, que pode ultrapassar 95% nos estágios iniciais, mas cai drasticamente em estágios avançados [DiMatos and Al. 2009, Esteva et al. 2017].

Apesar dos avanços, o diagnóstico clínico ainda apresenta desafios, já que a acurácia dos dermatologistas na detecção de melanoma varia entre 65% e 80% [G. Argenziano 2001]. Essa dificuldade se deve à semelhança entre lesões malignas e

benignas e à presença de interferências visuais como pelos, marcas ou artefatos. Além disso, fatores como a qualidade das imagens e a variação entre dispositivos de captura podem afetar significativamente a performance de modelos computacionais [Vieira 2022].

Nesse contexto, algoritmos de aprendizado de máquina, especialmente as Redes Neurais Convolucionais (CNNs), têm sido aplicados com sucesso crescente no auxílio à classificação automática de imagens dermatoscópicas. Entretanto, o desempenho desses modelos pode ser impactado negativamente por ruídos visuais. A remoção desses elementos por meio de técnicas de pré-processamento pode ser fundamental para aprimorar a capacidade de aprendizado dos modelos.

Este trabalho investiga o impacto da remoção de pelos e ruídos em imagens dermatoscópicas no desempenho de diferentes arquiteturas de CNNs: *EfficientNetB7* [Tan and Le 2019], *ResNet50V2* [He et al. 2016], *InceptionV3* [Szegedy et al. 2015] e *MobileNetV2* [Sandler et al. 2018]. Os modelos foram avaliados em quatro cenários experimentais, considerando combinações com ou sem *transfer learning* e *data augmentation*, tanto utilizando dados brutos quanto dados pré-processados.

Essa abordagem visa analisar comparativamente como o tratamento das imagens influencia a eficácia da aprendizagem dos modelos, contribuindo para o desenvolvimento de soluções mais precisas e robustas para o apoio ao diagnóstico do câncer de pele.

O restante deste artigo está organizado da seguinte maneira: na Seção 2 é apresentada uma descrição das arquiteturas de CNN analisadas, detalhando suas características e funcionamento. A Seção 3 descreve a base de dados utilizada, o ambiente computacional empregado e os métodos de preparação dos dados. Na Seção 4, os resultados obtidos são apresentados, comparando os impactos da remoção de pelos e ruídos na aprendizagem das diferentes arquiteturas. Finalmente, a Seção 5 traz as conclusões do estudo e sugere possíveis direções para trabalhos futuros.

## 2. Arquiteturas de Rede Neural Convolucional

As CNNs são uma classe especializada de redes neurais artificiais projetadas principalmente para processar dados estruturados em grade, como imagens [Goodfellow et al. 2016]. Desenvolvidas por [LeCun et al. 1998], elas foram inspiradas no funcionamento do córtex visual dos animais, aplicando filtros que realizam operações de convolução nos dados de entrada para extrair características importantes.

### 2.1. *EfficientNet*

A *EfficientNet* é uma família de CNNs desenvolvida para otimizar ao mesmo tempo a acurácia e a eficiência computacional. Sua inovação principal é o escalonamento composto, que equilibra três dimensões da rede: profundidade, largura e resolução [Tan and Le 2019].

Ao contrário de métodos anteriores que escalavam apenas uma dimensão por vez, a *EfficientNet* ajusta simultaneamente as três dimensões por meio de um coeficiente específico ( $\phi$ ). Essa técnica foi obtida por busca automatizada de arquitetura e resultou no bloco MBConv, que utiliza convoluções separáveis em profundidade e conexões residuais [Tan and Le 2019].

## 2.2. ResNet

Antes da *ResNet*, CNNs tinham dificuldades em melhorar seu desempenho ao adicionar muitas camadas devido ao problema do desaparecimento de gradiente [Souza et al. 2020]. A *ResNet* resolveu isso com conexões de atalho, que permitem pular algumas camadas e ajudam o gradiente a fluir melhor durante o treinamento [He et al. 2016].

Essas conexões fazem com que a rede aprenda apenas a diferença entre a entrada e a saída desejada, facilitando o treinamento. Isso permitiu criar redes bem mais profundas, com mais de 150 camadas, que alcançaram desempenho superior em competições importantes como o *ImageNet* [He et al. 2016].

## 2.3. Inception

A família de arquiteturas *Inception* utiliza um módulo básico chamado *Inception*, que processa a entrada por múltiplos caminhos ao mesmo tempo, usando filtros de diferentes tamanhos [Zhong et al. 2015]. A inovação da *InceptionV3* está em substituir convoluções maiores por várias convoluções menores e mais eficientes. Por exemplo, uma convolução  $5 \times 5$  é substituída por duas convoluções  $3 \times 3$  seguidas, reduzindo a complexidade sem perder desempenho [Szegedy et al. 2015]. Além disso, são usadas convoluções assimétricas para reduzir ainda mais o número de parâmetros e o custo computacional [Szegedy et al. 2015].

## 2.4. MobileNet

A família *MobileNet* é feita para dispositivos com recursos limitados, como smartphones [Carvalho 2020]. Sua inovação é usar convoluções separáveis, dividindo a operação padrão em duas partes menores: uma convolução que atua separadamente em cada canal (*depthwise*) e outra que combina esses resultados [Howard et al. 2017]. Isso reduz a complexidade e deixa o modelo mais leve e rápido.

A versão *MobileNetV2* trouxe ainda melhorias, adicionando blocos especiais (gargalo invertido) e conexões residuais, melhorando ainda mais o equilíbrio entre tamanho, velocidade e precisão, sendo ideal para aplicações rápidas em dispositivos limitados [Sandler et al. 2018].

## 3. Metodologia

Esta seção descreve a metodologia adotada, incluindo a base de dados utilizada, o pré-processamento das imagens, os modelos de classificação empregados e o ambiente computacional utilizado nos experimentos.

### 3.1. Base de Dados

Neste estudo, foi utilizada a base de dados *HAM10000*<sup>1</sup>, que contém 10.015 imagens dermatoscópicas de lesões pigmentadas da pele, divididas em sete classes distintas:

- **Queratose actínica** (*akiec*): imagens de queratose solar e Doença de Bowen, representando lesões pré-malignas (327 imagens);

---

<sup>1</sup>Disponível em: <https://dataverse.harvard.edu/dataset.xhtml?persistentId=doi:10.7910/DVN/DBW86T>

- **Queratose benigna** (*bkl*): categoria genérica que inclui queratose seborreica, lentigo solar e líquen plano (1.099 imagens);
- **Carcinoma basocelular** (*bcc*): forma mais comum de câncer de pele, com imagens representativas do carcinoma basocelular (514 imagens);
- **Dermatofibroma** (*df*): lesões benignas da pele que podem surgir como resposta inflamatória a traumas (115 imagens);
- **Nevo melanocítico** (*nv*): neoplasias benignas originadas de melanócitos, sendo a classe com maior número de amostras (6.705 imagens);
- **Melanoma** (*mel*): câncer maligno derivado de melanócitos, com diferentes variantes clínicas (1.113 imagens);
- **Lesões cutâneas vasculares** (*vasc*): englobam angiomas cereja, angioceratomas, granulomas piogênicos e hemorragias (142 imagens).

Na Figura 1, são apresentadas imagens representativas de cada uma das sete classes citadas acima.



**Figura 1. Amostras representativas das classes do dataset HAM10000**

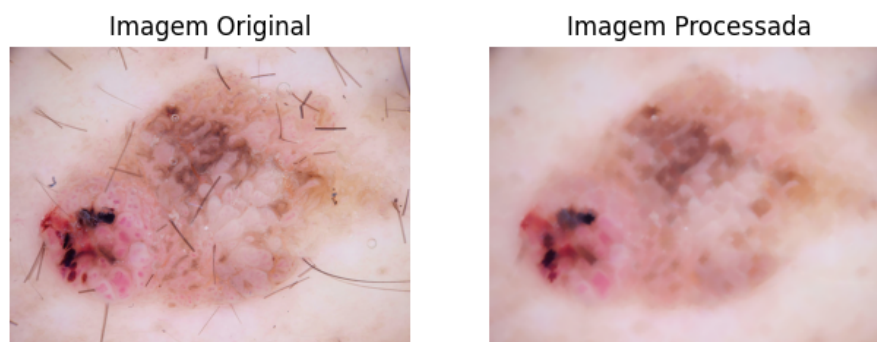
### 3.2. Ambiente Computacional

As redes neurais deste estudo foram implementadas em Python [Van Rossum and Drake 2009], utilizando o *framework* TensorFlow [Abadi et al. 2016] com suporte da API Keras [Chollet et al. 2015]. O pré-processamento das imagens foi realizado com a biblioteca OpenCV [Bradski 2000], e a visualização dos resultados com a Matplotlib [Hunter 2007]. Os experimentos foram conduzidos em um ambiente com processador AMD Ryzen 5 5600X, 32 GB de RAM e GPU NVIDIA GeForce RTX 2060 Super com 8 GB de memória dedicada.

### 3.3. Preparação dos Dados e Classificadores

Para melhorar a qualidade das imagens e reduzir ruídos que poderiam comprometer o treinamento dos modelos, foram aplicadas técnicas de pré-processamento. Inicialmente, realizou-se uma operação de fechamento morfológico para remover artefatos visuais, principalmente pelos nas imagens dermatoscópicas. Essa operação foi implementada utilizando um *kernel* cruzado de 10x10 pixels, aplicado separadamente a cada canal RGB. Em seguida, os canais processados foram recombinaados, resultando em uma imagem colorida com significativa redução de ruídos.

Para suavizar ainda mais as imagens, aplicou-se um filtro gaussiano com *kernel* de 5x5 pixels. A eficácia do pré-processamento foi avaliada por meio da métrica SSIM (*Structural Similarity Index Measure*), que apresentou valor médio de 0,8543, indicando alta preservação da estrutura das imagens originais.



**Figura 2. Exemplo visual do pré-processamento das imagens.**

A Figura 2 ilustra visualmente cada etapa do processo de preparação das imagens.

Após o pré-processamento, foram conduzidos os treinamentos dos classificadores, considerando quatro cenários experimentais distintos, que combinaram o uso ou não de data augmentation e transfer learning.

As imagens foram redimensionadas para 128x128 pixels e normalizadas entre 0 e 1. O conjunto de dados foi dividido em 70% para treinamento, e os 30% restantes foram igualmente distribuídos entre validação e teste.

Nos modelos pré-treinados utilizados, as camadas superiores originais foram removidas e substituídas por uma estrutura adaptada à tarefa de classificação. Essa nova arquitetura foi composta por uma camada Flatten, que transforma a saída convolucional em vetor, seguida por uma camada densa com 256 neurônios e ativação ReLU. Em seguida, adicionou-se uma camada Dropout com taxa de 0,5 para reduzir o overfitting, finalizando com uma camada softmax com sete neurônios, correspondente às classes do problema.

Todos os modelos foram treinados utilizando a função de perda **categorical cross-entropy**, com o otimizador Adam (taxa de aprendizado inicial de 0,001 e decaimento exponencial). Foi adotado também o mecanismo de **early stopping**, com paciência de 15 épocas, para interromper o treinamento caso não houvesse melhora no desempenho. A avaliação dos classificadores foi feita com base nas métricas de acurácia, precisão, recall e AUC (área sob a curva ROC).

## **4. Resultados e Discussões**

Nesta seção, apresentamos os resultados dos modelos em cada cenário previamente discutido com e sem dados pré-processados.

### **4.1. Resultados Com Pré-Processamento**

A presente subseção tem por objetivo, mostrar o comportamento dos classificadores nos cenários descritos no caso de dados com pré-processamento para remoção de ruído das imagens.

A Tabela 1 apresenta os resultados das CNNs utilizando *Data Augmentation* e *Transfer Learning*.

**Tabela 1. Métricas dos modelos com *Data Augmentation* e *Transfer Learning***

Métrica	ResNet50V2	InceptionV3	EfficientNetB7	MobileNetV2
Acurácia	<b>93.14%</b>	92.39%	85.71%	92.69%
Precisão	<b>87.88%</b>	83.38%	67.86%	85.83%
<i>Recall</i>	<b>60.34%</b>	58.41%	29.00%	58.48%
AUC	<b>95.16%</b>	93.39%	86.42%	94.77%

A arquitetura *ResNet50V2* destacou-se com a maior acurácia (93,14%), precisão (87,88%), *recall* (60,34%) e AUC (95,16%), demonstrando melhor capacidade geral de classificação.

*InceptionV3* e *MobileNetV2* tiveram desempenhos similares, com boa acurácia (92%) e precisão (85%), mas *recall* limitado (58%), indicando dificuldade em identificar alguns casos positivos.

A *EfficientNetB7* obteve resultados inferiores, especialmente em *recall* (29%), sugerindo dificuldades significativas na identificação correta de casos positivos sem pré-processamento.

A Tabela 2 apresenta os resultados das CNNs utilizando *Data Augmentation* sem *Transfer Learning*.

**Tabela 2. Métricas dos modelos com *Data Augmentation* e sem *Transfer Learning***

Métrica	ResNet50V2	InceptionV3	EfficientNetB7	MobileNetV2
Acurácia	<b>91.95%</b>	90.55%	90.55%	90.55%
Precisão	<b>80.05%</b>	66.93%	66.93%	66.93%
<i>Recall</i>	58.21%	<b>66.93%</b>	<b>66.93%</b>	<b>66.93%</b>
AUC	<b>92.69%</b>	87.77%	87.40%	87.40%

*ResNet50V2* obteve novamente o melhor desempenho geral, destacando-se em acurácia (91,95%), precisão (80,05%) e AUC (92,69%), porém com *recall* moderado (58,21%).

As demais arquiteturas (*InceptionV3*, *EfficientNetB7*, *MobileNetV2*) apresentaram resultados idênticos entre si, com acurácia de 90,55%, precisão e *recall* iguais (66,93%), e AUC próxima a 87%, indicando limitações na capacidade discriminativa dos modelos quando não é utilizado *transfer learning*.

A Tabela 3 apresenta os resultados das CNNs utilizando *Transfer Learning* sem *Data Augmentation*.

**Tabela 3. Métricas dos modelos sem *Data Augmentation* e com *Transfer Learning***

Métrica	ResNet50V2	InceptionV3	EfficientNetB7	MobileNetV2
Acurácia	<b>93.10%</b>	92.54%	90.55%	92.85%
Precisão	<b>86.49%</b>	83.81%	66.93%	<b>86.49%</b>
<i>Recall</i>	61.34%	59.28%	<b>66.93%</b>	59.21%
AUC	<b>95.30%</b>	93.48%	87.38%	94.92%

Novamente, *ResNet50V2* alcançou o melhor desempenho, com acurácia (93,10%), precisão (86,49%), *recall* (61,34%) e AUC (95,30%), confirmando sua superioridade.

*MobileNetV2* e *InceptionV3* obtiveram resultados próximos, com acurácia ( 92,7%) e AUC ( 94%), indicando boa capacidade classificatória, embora o *recall* esteja um pouco abaixo ( 59%).

A *EfficientNetB7* apresentou desempenho inferior, com menor acurácia (90,55%), precisão reduzida (66,93%) e AUC (87,38%), sugerindo novamente limitações significativas para essa tarefa específica.

A Tabela 4 apresenta os resultados das CNNs sem utilização de *Data Augmentation* e sem *Transfer Learning*.

**Tabela 4. Métricas dos modelos sem *Data Augmentation* e sem *Transfer Learning***

Métrica	ResNet50V2	<i>InceptionV3</i>	<i>EfficientNetB7</i>	<i>MobileNetV2</i>
Acurácia	<b>91.79%</b>	90.55%	90.55%	90.55%
Precisão	<b>80.53%</b>	66.93%	66.93%	66.93%
<i>Recall</i>	56.15%	<b>66.93%</b>	<b>66.93%</b>	<b>66.93%</b>
AUC	<b>92.60%</b>	87.83%	87.37%	87.40%

A *ResNet50V2* novamente obteve os melhores resultados entre as arquiteturas avaliadas, com acurácia (91,79%), precisão (80,53%) e AUC (92,60%), porém com *recall* reduzido (56,15%).

As demais arquiteturas (*InceptionV3*, *EfficientNetB7*, *MobileNetV2*) obtiveram resultados iguais, com acurácia de 90,55%, precisão e *recall* de 66,93% e AUC em torno de 87%, indicando desempenho limitado sem *transfer learning* e sem *data augmentation*.

#### 4.2. Resultados Sem Pré-Processamento

A presente subseção tem por objetivo, mostrar o comportamento dos classificadores nos cenários descritos no caso de dados sem pré-processamento para remoção de ruído das imagens.

A Tabela 5 apresenta os resultados das CNNs utilizando *Data Augmentation* e *Transfer Learning*, sem pré-processamento das imagens.

**Tabela 5. Métricas dos modelos com *Data Augmentation*, com *Transfer Learning* e sem pré-processamento**

Métrica	ResNet50V2	<i>InceptionV3</i>	<i>EfficientNetB7</i>	<i>MobileNetV2</i>
Acurácia	92.96%	92.44%	85.71%	<b>92.99%</b>
Precisão	<b>85.82%</b>	82.00%	67.13%	84.31%
<i>Recall</i>	60.81%	60.34%	15.87%	<b>62.60%</b>
AUC	<b>94.90%</b>	92.94%	87.20%	94.86%

A Tabela 5 apresenta os resultados das CNNs (*ResNet50V2*, *InceptionV3*, *EfficientNetB7*, *MobileNetV2*) utilizando *Data Augmentation* e *Transfer Learning*, porém sem pré-processamento das imagens.

*MobileNetV2* e *ResNet50V2* obtiveram os melhores desempenhos gerais, com acurácias semelhantes ( 93%), precisão alta ( 85%) e bons valores de AUC ( 95%). O *recall* foi relativamente baixo ( 61%), sugerindo espaço para melhoria na identificação correta de casos positivos.

A *InceptionV3* apresentou resultados próximos, com leve queda nas métricas comparada às duas anteriores. Já a *EfficientNetB7* teve o pior desempenho, especialmente no *recall* (15,87%), confirmando dificuldades significativas sem pré-processamento.

A Tabela 6 apresenta os resultados das CNNs utilizando *Data Augmentation*, sem *Transfer Learning* e sem pré-processamento das imagens.

**Tabela 6. Métricas dos modelos com *Data Augmentation*, sem *Transfer Learning* e sem pré-processamento**

Métrica	ResNet50V2	<i>InceptionV3</i>	<i>EfficientNetB7</i>	<i>MobileNetV2</i>
Acurácia	<b>91.67%</b>	90.55%	90.55%	90.55%
Precisão	<b>78.68%</b>	66.93%	66.93%	66.93%
<i>Recall</i>	57.21%	<b>66.93%</b>	<b>66.93%</b>	<b>66.93%</b>
AUC	<b>92.48%</b>	87.26%	87.39%	87.40%

*ResNet50V2* obteve o melhor desempenho geral, com acurácia (91,67%), precisão (78,68%) e AUC (92,48%), embora tenha apresentado um *recall* moderado (57,21%).

As demais arquiteturas (*InceptionV3*, *EfficientNetB7* e *MobileNetV2*) obtiveram resultados idênticos, com acurácia (90,55%), precisão e *recall* iguais (66,93%) e valores de AUC próximos ( 87%). Esses resultados indicam limitações significativas dessas arquiteturas na ausência do *transfer learning* e pré-processamento das imagens.

A Tabela 7 apresenta os resultados das CNNs utilizando *Transfer Learning*, sem *Data Augmentation* e sem pré-processamento das imagens.

**Tabela 7. Métricas dos modelos sem *augmentation*, com *transfer learning* e sem pré-processamento**

Métrica	ResNet50V2	<i>InceptionV3</i>	<i>EfficientNetB7</i>	<i>MobileNetV2</i>
Acurácia	93.02%	92.37%	85.71%	<b>93.14%</b>
Precisão	<b>86.51%</b>	81.89%	11.11%	86.47%
<i>Recall</i>	60.61%	59.99%	15.55%	<b>61.67%</b>
AUC	94.58%	93.40%	87.34%	<b>94.87%</b>

*MobileNetV2* e *ResNet50V2* obtiveram novamente os melhores resultados, com acurácias ( 93%), alta precisão ( 86%) e valores elevados de AUC ( 95%). Entretanto, o *recall* se manteve moderado ( 61%).

A *InceptionV3* apresentou desempenho um pouco inferior, mas ainda competitivo, com acurácia (92,37%) e boa precisão (81,89%). Por outro lado, a *EfficientNetB7* teve baixo desempenho geral, destacando-se negativamente com baixa precisão (11,11%) e *recall* reduzido (15,55%).

A Tabela 8 apresenta os resultados das CNNs sem utilização de *Data Augmentation*, sem *Transfer Learning* e sem pré-processamento das imagens.



**Tabela 8. Métricas dos modelos sem *Data Augmentation*, sem *Transfer Learning* e sem pré-processamento**

Métrica	ResNet50V2	InceptionV3	EfficientNetB7	MobileNetV2
Acurácia	<b>91.60%</b>	90.55%	90.55%	90.55%
Precisão	<b>78.54%</b>	66.93%	66.93%	66.93%
<i>Recall</i>	56.75%	<b>66.93%</b>	<b>66.93%</b>	66.93%
AUC	<b>92.23%</b>	86.45%	87.38%	87.39%

*ResNet50V2* obteve novamente o melhor desempenho, com acurácia (91,60%), precisão (78,54%) e maior AUC (92,23%), apesar do *recall* reduzido (56,75%).

As demais arquiteturas (*InceptionV3*, *EfficientNetB7*, *MobileNetV2*) mostraram desempenho idêntico, com acurácia (90,55%), precisão e *recall* iguais (66,93%) e valores de AUC próximos (87%), evidenciando limitações significativas sem técnicas adicionais de treinamento e pré-processamento.

## 5. Conclusão e Trabalhos Futuros

Este estudo avaliou as CNNs *ResNet50V2*, *InceptionV3*, *EfficientNetB7* e *MobileNetV2*, considerando diferentes combinações de *Data Augmentation*, *Transfer Learning* e pré-processamento de imagens. A arquitetura *ResNet50V2* destacou-se consistentemente em acurácia, precisão e AUC, apesar do *recall* moderado. *InceptionV3* e *MobileNetV2* apresentaram resultados intermediários, enquanto a *EfficientNetB7* exibiu desempenho limitado em todas as métricas avaliadas.

O uso combinado das técnicas de *Transfer Learning*, *Data Augmentation* e pré-processamento mostrou-se eficaz para melhorar significativamente o desempenho dos modelos. Recomenda-se, portanto, a utilização da *ResNet50V2* associada a essas estratégias para problemas similares.

Sugere-se como trabalhos futuros investigar métodos para aumentar o *recall* dos modelos, explorar técnicas avançadas de pré-processamento para melhorar ainda mais a qualidade das imagens, avaliar conjuntos maiores e mais diversificados de dados para melhor generalização, analisar o desempenho de novas arquiteturas emergentes de CNN e considerar abordagens híbridas combinando CNNs com outras técnicas de aprendizado profundo.

## Referências

- Abadi, M., Barham, P., Chen, J., Chen, Z., Davis, A., Dean, J., Devin, M., Ghemawat, S., Irving, G., Isard, M., et al. (2016). Tensorflow: a system for large-scale machine learning. In *12th USENIX symposium on operating systems design and implementation (OSDI 16)*, pages 265–283, Savannah, GA. USENIX Association.
- Bradski, G. (2000). The OpenCV Library. *Dr. Dobb's Journal of Software Tools*.
- Carvalho, M. B. F. d. (2020). Reconhecimento automático de fonemas via rna profunda. *Biblioteca Digital de Teses e Dissertações UFMA*.
- Chollet, F. et al. (2015). Keras. <https://keras.io>.

- da Saúde, M. (2024). Câncer de pele. <https://www.gov.br/saude/pt-br/assuntos/saude-de-a-a-z/c/cancer-de-pele>. [online; acesso em 09 de março de 2025].
- da Saúde, O. M. (2025). Cancer. <https://www.who.int/en/news-room/fact-sheets/detail/cancer>. [online; acesso em 09 de março de 2025].
- DiMatos, D. C. and Al., E. (2009). Melanoma cutâneo no brasil. *Arquivos Catarinenses de Medicina*, 38(Suplemento 01):14.
- Esteva, A., Kuprel, B., et al. (2017). Dermatologist-level classification of skin cancer with deep neural networks. *Nature*, 542(7639):115–118.
- G. Argenziano, H. P. S. (2001). Dermoscopy of pigmented skin lesions—a valuable tool for early diagnosis of melanoma. *The Lancet Oncology*, 2(7):443–449.
- Goodfellow, I., Bengio, Y., and Courville, A. (2016). *Deep learning*. MIT press.
- He, K., Zhang, X., Ren, S., and Sun, J. (2016). Deep residual learning for image recognition. *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 770–778.
- Howard, A. G., Zhu, M., Chen, B., Kalenichenko, D., Wang, W., Weyand, T., Andreetto, M., and Adam, H. (2017). Mobilenets: Efficient convolutional neural networks for mobile vision applications. *arXiv preprint arXiv:1704.04861*.
- Hunter, J. D. (2007). Matplotlib: A 2d graphics environment. *Computing in Science & Engineering*, 9(3):90–95.
- LeCun, Y., Bottou, L., Bengio, Y., and Haffner, P. (1998). Gradient-based learning applied to document recognition. *Proceedings of the IEEE*, 86(11):2278–2324.
- Sandler, M., Howard, A., Zhu, M., Zhmoginov, A., and Chen, L.-C. (2018). Mobilenetv2: Inverted residuals and linear bottlenecks. *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 4510–4520.
- Souza, V., Silva, L., Santos, A., and Araújo, L. (2020). Análise comparativa de redes neurais convolucionais no reconhecimento de cenas. *Anais do Computer on the Beach*, 11(1):419–426.
- Szegedy, C., Liu, W., Jia, Y., Sermanet, P., Reed, S., Anguelov, D., Erhan, D., Vanhoucke, V., and Rabinovich, A. (2015). Going deeper with convolutions. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 1–9.
- Tan, M. and Le, Q. V. (2019). Efficientnet: Rethinking model scaling for convolutional neural networks. *CoRR*, abs/1905.11946.
- Van Rossum, G. and Drake, F. L. (2009). *Python 3 Reference Manual*. CreateSpace, Scotts Valley, CA.
- Vieira, E. Q. (2022). Comparação entre diferentes modelos de redes neurais convolucionais para classificação de melanoma. *Universidade de Brasília*.
- Zhong, Z., Jin, L., and Xie, Z. (2015). High performance offline handwritten chinese character recognition using googlenet and directional feature maps. In *2015 13th International Conference on Document Analysis and Recognition (ICDAR)*, pages 846–850. IEEE.