

# Classificação de Doenças Oculares: Uma Comparação de Modelos CNN

1<sup>st</sup> Alefe Vitor Almeida Gadioli  
Universidade Federal do Espírito Santo - UFES  
Vitória-ES, Brasil  
alefe.gadioli@edu.ufes.br

2<sup>nd</sup> Priscilla Martins Benevides  
Universidade Federal do Espírito Santo - UFES  
Vitória-ES, Brasil  
priscilla.benevides@edu.ufes.br

**Abstract**—A classificação de doenças oculares utilizando aprendizado profundo tem se mostrado promissora na assistência ao diagnóstico médico. Este estudo compara três modelos de redes neurais convolucionais (CNN) – ResNet-18, EfficientNet-B0 e MobileNetV2 – na classificação de catarata, retinopatia diabética, glaucoma e olhos normais. Utilizando o conjunto de dados RFMID, os modelos foram avaliados em termos de acurácia, precisão, recall e F1-score. O MobileNetV2 destacou-se com uma acurácia de 0,92, equilíbrio entre precisão e recall, eficiência computacional e menor uso de memória, tornando-o ideal para aplicações em dispositivos móveis. Futuras melhorias podem incluir a integração de técnicas de aumento de dados e ajustes finos adicionais.

## I. INTRODUÇÃO

A classificação de doenças oculares é uma área de pesquisa dedicada ao desenvolvimento de algoritmos e modelos para identificar com precisão diferentes tipos de doenças oculares a partir de dados de imagem médica. Ela desempenha um papel fundamental ao auxiliar oftalmologistas e profissionais de saúde no diagnóstico e tratamento eficazes das doenças oculares.

A conscientização sobre a saúde ocular, considerando que muitas doenças, como o glaucoma, são assintomáticas nos estágios iniciais é importante para garantir um tratamento adequado e até a cura. [1] O glaucoma pode danificar o nervo óptico e levar à cegueira se não tratado [1]. Mundialmente, 2,2 bilhões de pessoas têm deficiência visual ou cegueira. No Brasil, são mais de 1,5 milhões, com 74,8% dos casos sendo evitáveis com diagnóstico precoce [1].

A cegueira afeta a capacidade laboral, a qualidade de vida e tem impactos econômicos. [1] A catarata, a principal causa de cegueira reversível, afeta 20 milhões de pessoas no mundo. [1] No Brasil, 93% da população conhece a catarata, mas 48% desconhecem que a cirurgia é o único tratamento eficaz [2].

Um estudo do Ibope de 2020 sobre glaucoma revela que, embora 53% dos brasileiros fariam profundamente afetados com a perda da visão, apenas 37% consideram essencial visitar um oftalmologista regularmente. [1]

O objetivo da classificação de doenças oculares é utilizar técnicas de aprendizado de máquina e visão computacional para analisar imagens médicas e identificar quatro condições: catarata, retinopatia diabética, glaucoma e olhos normais.

## A. Objetivo

Este estudo propõe classificar imagens oculares usando aprendizado profundo com os modelos ResNet-18, EfficientNet-B0 e MobileNetV2 no conjunto de dados RFMID, visando diagnósticos iniciais de baixo custo. Os modelos foram escolhidos para integração em sistemas simples e de baixo custo, como triagem e suporte ao diagnóstico. Exemplos de caso de usos: Triagem e Detecção Precoce, Suporte ao Diagnóstico, Planejamento e Monitoramento do Tratamento.

## II. TRABALHO RELACIONADOS

A classificação de doenças oculares com aprendizado profundo tem sido recentemente explorada, com vários estudos desenvolvendo algoritmos para identificar doenças a partir de imagens.

Gulshan et al. (2016) criaram um algoritmo para detecção de retinopatia diabética em fotografias de fundo de olho, demonstrando alta precisão [3]. Ting et al. (2017) desenvolveram um sistema para detectar retinopatia diabética e outras doenças oculares em populações multiétnicas, destacando a importância de dados diversificados [4]. Bismi e Na'am (2023) investigaram a classificação de miopia usando arquiteturas VGG16, VGG19 e InceptionV3 em imagens de fundo de olho [5].

Esses estudos fornecem uma base metodológica valiosa para o presente trabalho. A utilização de arquiteturas robustas como ResNet, Inception e U-Net pode aumentar a precisão dos modelos, apesar de possíveis limitações de eficiência e custo em aplicações práticas.

## III. METODOLOGIA

Usando Aprendizado Profundo para classificar doenças oculares. O método proposto neste estudo pode ser descrito conforme a figura 1. A imagem é um fluxograma que descreve o processo de treinamento e validação de modelos de aprendizado profundo. Aqui está a descrição detalhada de cada etapa do fluxograma:

## A. Dataset

O conjunto de dados utilizado neste estudo foi obtido do Repositório de Imagens de Múltiplas Doenças da Retina (RFMID) [6]. Este trabalho aborda a predição de quatro condições: catarata, retinopatia diabética, glaucoma e olhos

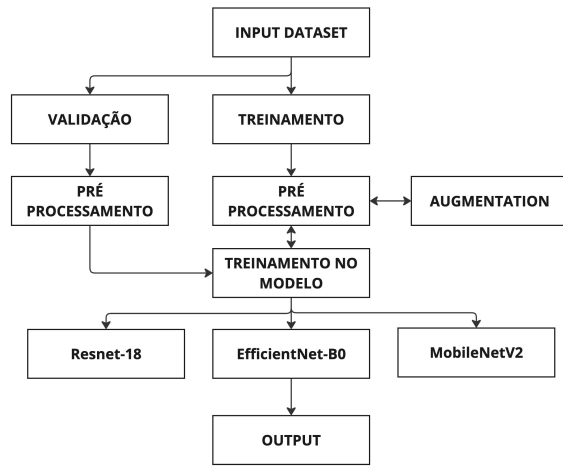


Fig. 1. Método de pesquisa

normais. O conjunto de dados é composto por 4217 amostras. Posteriormente, foi realizada uma seleção entre os conjuntos de dados disponíveis, dividindo-os em quatro classes que explicam a classificação das diferenças entre olhos normais, catarata, retinopatia diabética e glaucoma. Com aproximadamente 1000 imagens em cada classe, a distribuição é uniforme, conforme ilustrado na Figura 2.

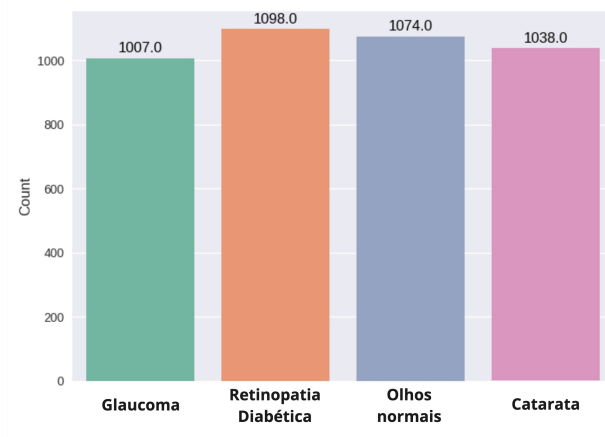


Fig. 2. Distribuição de imagens

O conjunto de dados foi dividido em dois subconjuntos: treinamento e teste, com 85% das imagens destinadas ao treinamento e 15% reservadas para o teste.

### B. Augmentation

*Augmentation* de imagem é uma das estratégias de treinamento mais eficazes para ampliar coleções de imagens para modelos de aprendizado profundo que não possuem dados de imagem suficientes. Aumentação de dados de imagem é necessária porque é essencial para obter melhor precisão nos testes e ajudar na geração de dados equilibrados [7]. As

técnicas de *Augmentation* de imagem utilizadas neste estudo foram as seguintes:

- Espelhamento horizontal aleatório com probabilidade de 50%.
- Espelhamento vertical aleatório com probabilidade de 50%.
- Autocontraste aleatório com probabilidade de 50%.
- Rotação aleatória às imagens dentro de um intervalo de 15 graus.

### C. Pré-processamento

As etapas de pré-processamento realizadas incluem redimensionamento, codificação de rótulos (*labels*) e normalização.

Foi constatado que todos os pixels das imagens não estavam na faixa de [0-255]. Portanto, cada imagem foi normalizada de acordo com o padrão utilizado pelos modelos para garantir consistência e precisão. Todas as imagens foram redimensionadas para (224,224).

Exemplo da imagem utilizada pode ser visto na Figura 3.

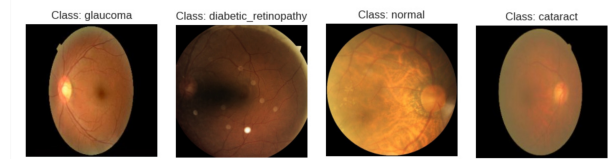


Fig. 3. Imagens das classes escolhidas

### D. Deep-learning

O aprendizado profundo está evoluindo rapidamente e melhorando diversas aplicações médicas, como análise de imagens, detecção de estruturas anatômicas, segmentação de tecidos, diagnóstico auxiliado por computador e prognóstico de doenças [8]. Esse conceito é central em muitos algoritmos, utilizando modelos com múltiplas camadas que transformam dados de entrada (imagens) em saídas (presença/ausência de doença) enquanto aprendem características progressivamente. As redes neurais convolucionais (CNNs) são as mais eficazes para análise de imagens [9], e com o avanço da tecnologia e capacidades computacionais, o aprendizado profundo tem sido amplamente aplicado na classificação de imagens [9].

A seguir iremos tratar de cada arquitetura utilizada.

1) *ResNet-18*: Rede Neural Residual com 18 camadas é uma arquitetura de rede neural convolucional profunda, desenvolvida para superar problemas como o desaparecimento dos gradientes [10]. Faz parte da família de Redes Neurais Residuais (ResNets), conhecidas por suas conexões residuais.

Cada bloco residual permite que o sinal de entrada seja adicionado diretamente ao sinal de saída de algumas camadas subsequentes, utilizando conexões de salto que contornam uma ou mais camadas intermediárias. [10] Essas conexões ajudam a resolver o problema de gradientes desaparecendo e explodindo. [10]. Na figura 4 apresenta a arquitetura usada nesse modelo.

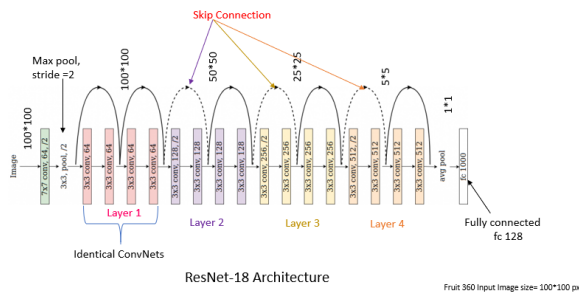


Fig. 4. Arquitetura ResNet-18

2) *EfficientNet-B0*: É uma arquitetura de rede neural convolucional desenvolvida para otimizar a eficiência e a precisão do aprendizado profundo. Introduzida por Mingxing Tan e Quoc V. Le em 2019, a EfficientNet-B0 faz parte da família EfficientNet, que utiliza dimensionamento composto para equilibrar profundidade, largura e resolução da rede, maximizando a performance [11].

A arquitetura EfficientNet-B0 é composta por blocos MB-Conv (Mobile Inverted Bottleneck Convolutional Blocks), que foram introduzidos pela primeira vez na rede MobileNetV2. Esses blocos são eficientes em termos de computação e ajudam a reduzir a quantidade de parâmetros enquanto mantêm a capacidade da rede de aprender representações complexas [11]. Na figura 5 apresenta a arquitetura usada nesse modelo.

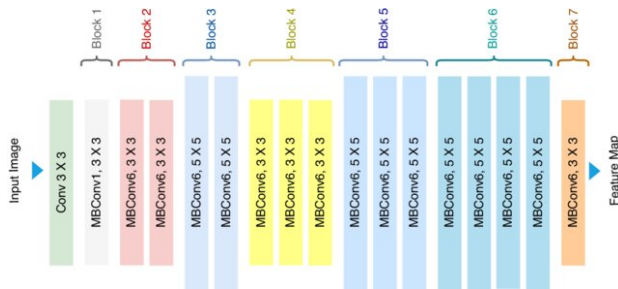


Fig. 5. Arquitetura EfficientNet-B0

3) *MobileNetV2*: Introduzida por Mark Sandler, Andrew Howard e outros pesquisadores do Google em 2018, a MobileNetV2 é uma evolução da MobileNetV1, com melhorias significativas em termos de eficiência e desempenho, projetada especificamente para uso em dispositivos móveis e aplicações com recursos limitados. [12].

A MobileNetV2 introduz o conceito de blocos invertidos de gargalo, que são diferentes dos blocos residuais tradicionais [12]. Esses blocos permitem que as camadas intermediárias da rede sejam expandidas e comprimidas, reduzindo a quantidade de operações computacionais necessárias [12].

Seguindo a MobileNetV1, a MobileNetV2 utiliza convoluções separáveis em profundidade para reduzir drasticamente o número de parâmetros e a complexidade computacional, mantendo a precisão do modelo. [12]. Na figura 6 apresenta a arquitetura usada nesse modelo.

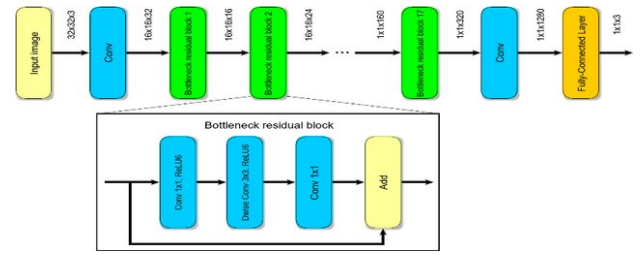


Fig. 6. Arquitetura MobileNetV2

## E. Processo de Treinamento

O processo de treinamento é a etapa na qual a CNN é ajustada para melhorar a precisão da classificação. Este estágio envolve dois processos principais: feedforward e backpropagation [9].

1) *Processo de Feedforward*: É a primeira etapa do treinamento, onde várias camadas são geradas para classificar os dados de imagem, utilizando pesos e vieses atualizados durante o backpropagation. [9]

2) *Processo de Backpropagation*: É a segunda etapa do treinamento. Nesta fase, os resultados do feedforward são analisados para identificar erros da camada de saída até a primeira camada, permitindo ajustes nos pesos e vieses. [9]

3) *Cálculo de Gradiente*: O cálculo do gradiente na rede de convolução é crucial para gerar novos valores de peso e viés necessários durante o treinamento, ajustando a rede para minimizar os erros de previsão [9]

## F. Processo de Teste

O processo de teste envolve a classificação utilizando os pesos e vieses obtidos no treinamento. [9] Este estágio resulta na avaliação da precisão da classificação, identificando os dados que falharam na classificação, o número de imagens que não foram corretamente classificadas e a configuração final da rede a partir do feedforward. A camada de saída está totalmente conectada com os rótulos existentes. [9]

## G. Parâmetros do estudo

Os seguintes parâmetros foram utilizados neste estudo:

- Otimizador: Adam
- Taxa de aprendizado:  $3e-5$
- Tamanho do batch: 128
- Número de épocas: 100

Os parâmetros da camada fully connected são:

- Camada linear: 512 unidades de entrada e 128 unidades de saída
- Ativação: ReLU
- Dropout: Taxa de 0.2
- Camada linear final: 128 unidades de entrada e 4 unidades de saída

## H. Métricas

A seguir, explico como cada uma dessas métricas é calculada e apresento um exemplo de código para gerar essas métricas.

Definições:

**Acurácia (Accuracy):** A proporção de previsões corretas em relação ao total de previsões feitas.

$$\text{Acurácia} = \frac{\text{Número de previsões corretas}}{\text{Número total de amostras}}$$

**Matriz de Confusão (Confusion Matrix):** Uma tabela que é frequentemente usada para descrever o desempenho de um modelo de classificação. A matriz de confusão mostra os valores reais versus os valores preditos, distribuídos por classe.

**Precisão (Precision)**

$$\text{Precisão} = \frac{\text{Verdadeiros Positivos}}{\text{Verdadeiros Positivos} + \text{Falsos Positivos}}$$

**Recall (Recall)**

$$\text{Recall} = \frac{\text{Verdadeiros Positivos}}{\text{Verdadeiros Positivos} + \text{Falsos Negativos}}$$

**F1-Score:** A média harmônica entre a precisão e o recall.

$$\text{F1-Score} = 2 \times \frac{\text{Precisão} \times \text{Recall}}{\text{Precisão} + \text{Recall}}$$

**Support:** O número de ocorrências de cada classe no conjunto de dados.

#### I. Bibliotecas Utilizadas

Para este estudo, utilizamos várias bibliotecas. O Pandas foi empregado para a manipulação e análise de dados estruturados [13]. O NumPy forneceu suporte para arrays e matrizes multidimensionais [14]. Utilizamos o Matplotlib para a criação de gráficos [15] e o Seaborn para gráficos estatísticos [16]. O OpenCV foi utilizado para processamento de imagens [17]. Para treinar modelos de redes neurais profundas, utilizamos o PyTorch [18]. O scikit-learn forneceu ferramentas para distribuição de dataset (treinamento e controle) e avaliação, incluindo matriz de confusão e relatórios de classificação [19].

### IV. RESULTADOS E DISCUSSÃO

#### A. ResNet-18

A classe "glaucoma" tem uma precisão de 0,89, recall de 0,88 e F1-score de 0,89. Isso sugere que o modelo se desempenha bem na identificação correta dos casos de glaucoma, mas pode haver alguns falsos negativos. A classe "normal" tem uma precisão de 0,85, recall de 0,90 e F1-score de 0,88, indicando bom desempenho tanto em precisão quanto em recall para casos normais. A classe "retinopatia diabética" apresenta alta precisão, recall e F1-score de 0,99, indicando excelente desempenho na identificação correta dos casos de retinopatia diabética. A classe "catarata" também apresenta alta precisão, recall e F1-score de 0,95 ou mais, indicando identificação precisa dos casos de catarata. Conforme mostrado na Tabela I.

A análise da matriz de confusão revela que o modelo tem bom desempenho na identificação de retinopatia diabética e catarata, mas apresenta variações na distinção entre glaucoma e olhos normais, conforme a Figura 7.

Classe	Precisão	Recall	F1-Score	Support
Glaucoma	0.89	0.88	0.89	153
Ret. Diabética	0.99	1.00	0.99	163
Normal	0.86	0.88	0.87	151
Catarata	0.97	0.95	0.96	166

TABLE I  
MÉTRICAS DE AVALIAÇÃO POR CLASSE - RESNET-18

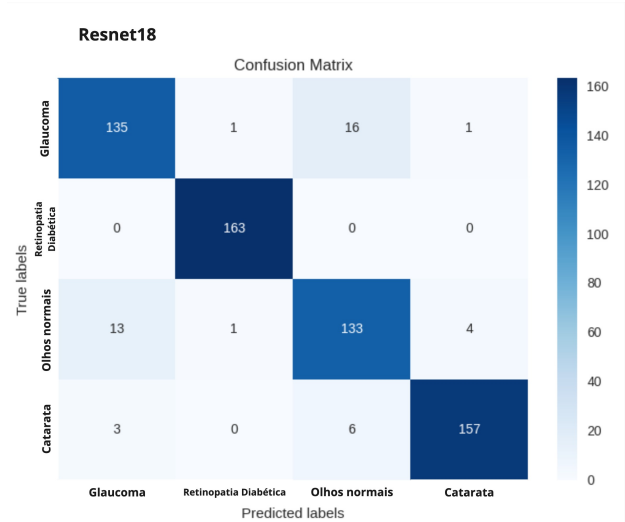


Fig. 7. Matriz Confusão - ResNet-18

Métrica	Precision	Recall	F1-Score	Support
Accuracy	-	-	0.93	633
Macro Avg	0.93	0.93	0.93	633
Weighted Avg	0.93	0.93	0.93	633

TABLE II  
MÉTRICAS DE AVALIAÇÃO GERAL - RESNET-18

A acurácia geral do modelo é de 0,93, indicando a porcentagem de instâncias corretamente previstas em todas as classes, conforme visto na Tabela II.

A perda (loss) de treinamento diminuiu consistentemente, indicando aprendizado, enquanto a perda de validação apresentou alguma variação, sugerindo potencial overfitting, conforme visto na Figura 8.

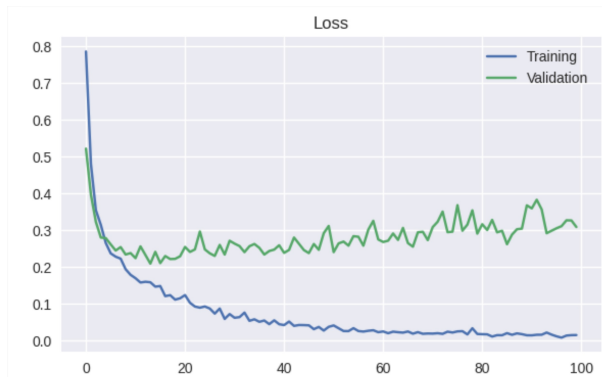


Fig. 8. Loss do Treinamento - ResNet-18

A acurácia de treinamento aumentou de forma constante, aproximando-se de 1.0, enquanto a acurácia de validação variou em torno de 0.90 a 0.95, sugerindo uma boa generalização com algumas flutuações, conforme visto na Figura 9.



Fig. 9. Acurácia do Treinamento - ResNet-18

### B. EfficientNet-B0

A classe "retinopatia diabética" apresenta alta precisão, recall de 0.97 e F1-score de 0.94, demonstrando excelente desempenho na identificação correta dos casos. A classe "glaucoma" apresenta precisão de 0.83, recall de 0.82 e F1-score de 0.83. Embora o modelo se saia razoavelmente bem na identificação de casos de glaucoma, ainda existem falsos negativos e falsos positivos. A classe "normal" tem precisão de 0.86, recall de 0.81 e F1-score de 0.83, indicando um equilíbrio entre precisão e recall. A classe "catarata" também apresenta alta precisão, recall e F1-score de 0.93, indicando uma identificação precisa dos casos de catarata. Conforme mostrado na Tabela III.

Classe	Precision	Recall	F1-Score	Support
Glaucoma	0.83	0.82	0.83	148
Retinopatia Diabética	0.91	0.97	0.94	152
Normal	0.86	0.81	0.83	176
Catarata	0.92	0.94	0.93	157

TABLE III

MÉTRICAS DE AVALIAÇÃO POR CLASSE - EFFICIENTNET-B0

A análise da matriz de confusão revela que o modelo tem bom desempenho na identificação de retinopatia diabética e catarata, mas apresenta dificuldades na distinção entre glaucoma e olhos normais, como visto na figura 10.

A acurácia geral do modelo é de 0.88, indicando a porcentagem de instâncias corretamente previstas em todas as classes, como visto na tabela IV.

Classe	Precision	Recall	F1-Score	Support
Accuracy			0.88	633
Macro Avg	0.88	0.89	0.88	633
Weighted Avg	0.88	0.88	0.88	633

TABLE IV

MÉTRICAS DE AVALIAÇÃO GERAL - EFFICIENTNET-B0

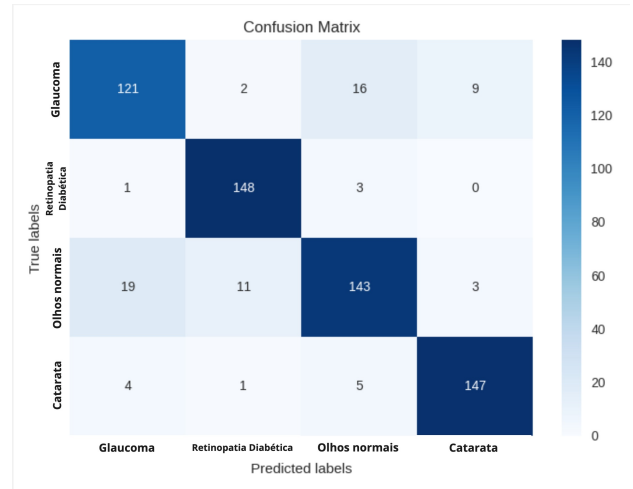


Fig. 10. Matriz de Confusão - EfficientNet-B0

A perda de treinamento (*Loss*) diminuiu consistentemente, indicando aprendizado eficaz, enquanto a perda de validação apresentou alguma variação, sugerindo potencial overfitting, conforme visto na figura 11. A acurácia de treinamento aumentou constantemente, aproximando-se de 0.95, enquanto a acurácia de validação variou entre 0.85 e 0.90, sugerindo uma boa generalização com algumas flutuações, como visto na figura 12.

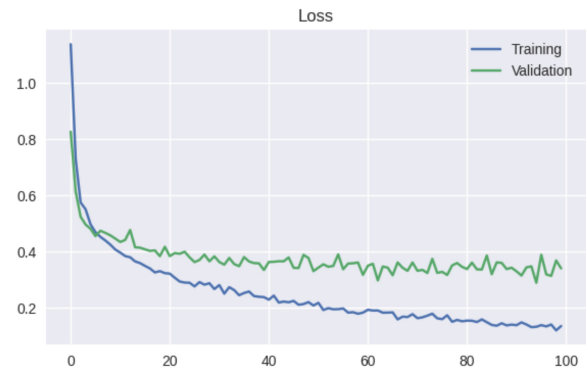


Fig. 11. Loss do Treinamento - EfficientNet-B0

### C. MobileNetV2

A classe "catarata" também apresenta alta precisão, recall e F1-score de 0.95, indicando uma identificação precisa dos casos de catarata. A classe "glaucoma" apresenta precisão de 0.86, recall de 0.87 e F1-score de 0.87. O modelo se desempenha bem na identificação de casos de glaucoma, embora haja falsos negativos e falsos positivos. A classe "normal" tem precisão de 0.87, recall de 0.89 e F1-score de 0.88, indicando um desempenho equilibrado entre precisão e recall. A classe "retinopatia diabética" apresenta alta precisão, recall e F1-score de 0.99, demonstrando excelente desempenho na identificação correta dos casos, conforme a tabela V

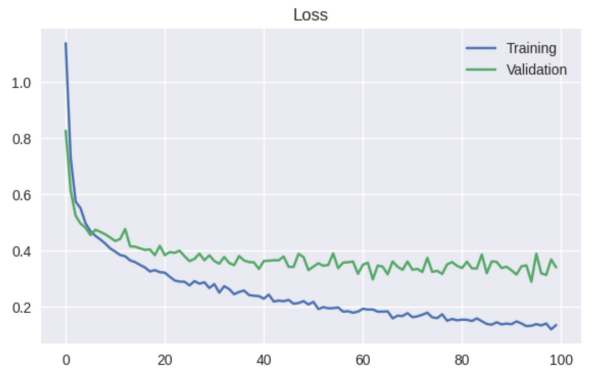


Fig. 12. Acurácia do Treinamento - EfficientNet-B0

Classe	Precision	Recall	F1-Score	Support
Glaucoma	0.86	0.87	0.87	150
Retinopatia Diabética	0.98	0.99	0.99	156
Normal	0.87	0.89	0.88	149
Catarata	0.96	0.93	0.95	178

TABLE V  
MÉTRICAS DE AVALIAÇÃO POR CLASSE - MOBILENETV2

A análise da matriz de confusão revela que o modelo tem bom desempenho na identificação de retinopatia diabética e catarata, mas apresenta algumas dificuldades na distinção entre glaucoma e olhos normais, como visto na figura 13.

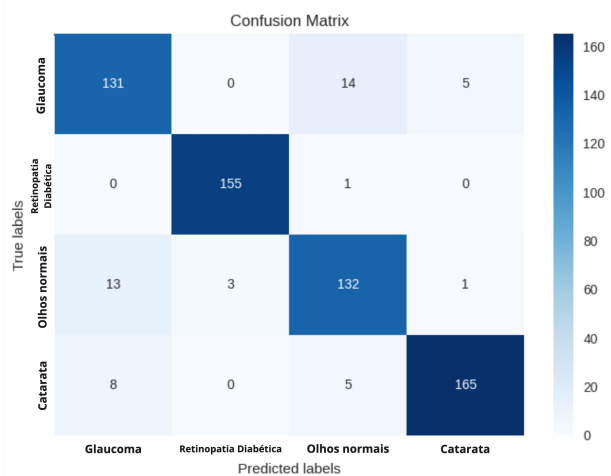


Fig. 13. Matriz de Confusão - MobileNetV2

A acurácia geral do modelo é de 0,92, indicando a porcentagem de instâncias corretamente previstas em todas as classes, como vista na tabela VI

Classe	Precision	Recall	F1-Score	Support
Accuracy			0.92	633
Macro Avg	0.92	0.92	0.92	633
Weighted Avg	0.92	0.92	0.92	633

TABLE VI  
MÉTRICAS DE AVALIAÇÃO GERAL - MOBILENETV2

A perda de treinamento diminuiu consistentemente, indi-

cando aprendizado eficaz, enquanto a perda de validação apresentou alguma variação, mas manteve-se estável, como visto na figura 14. A acurácia de treinamento aumentou constantemente, aproximando-se de 0,95, enquanto a acurácia de validação variou entre 0,90 e 0,93, sugerindo uma boa generalização com algumas flutuações, como visto na figura 15



Fig. 14. Loss do Treinamento - MobileNetV2



Fig. 15. Acurácia do Treinamento - MobileNetV2

#### D. Hardware

Ao analisar os parametros de cada modelos, apresentados na tabela VII, vemos que para aplicações com recursos de hardware limitados, EfficientNet-B0 e MobileNetV2 são mais viáveis devido ao menor número de parâmetros e operações. EfficientNet-B0 se destaca pela eficiência computacional e uso de memória, enquanto MobileNetV2 é ideal para dispositivos com ainda menos capacidade. ResNet-18 é mais robusto em termos de capacidade de modelagem, mas demanda mais recursos de hardware, sendo mais adequado para ambientes com hardware potente, como visto na tabela VIII.

#### V. CONCLUSÃO

Os resultados obtidos neste estudo mostram que a classificação de doenças oculares utilizando modelos de aprendizado profundo, especificamente ResNet-18, EfficientNet-B0



Modelo	Total de Parâmetros	Mult-Adds Totais
EfficientNet-B0	4,007,548	45.99M
ResNet-18	11,176,512	1.84G
MobileNetV2	2,223,872	156.06M

TABLE VII

COMPARAÇÃO DOS PARÂMETROS E MULT-ADDS DOS MODELOS DE REDES NEURAIS

Modelo	Input	F/B Pass	Params	Total
EfficientNet-B0	0.57	7.08	15.29	22.94
ResNet-18	0.57	37.90	42.64	81.11
MobileNetV2	0.57	15.81	8.48	24.87

TABLE VIII

COMPARAÇÃO DOS TAMANHOS EM MEMÓRIA EM MB

e MobileNetV2, é viável e apresenta um bom desempenho, considerando para um diagnóstico inicial.

O modelo ResNet-18 destacou-se com uma acurácia geral de 0,93, especialmente na identificação de retinopatia diabética, com uma precisão e recall de 0,99, e catarata, com uma precisão de 0,97. No entanto, apresentou algumas variações na distinção entre glaucoma e olhos normais, sugerindo possíveis melhorias com maior ajuste e dados adicionais.

O modelo EfficientNet-B0 obteve uma acurácia de 0,88, mostrando eficiência computacional e uso equilibrado de memória. A apresentou uma boa performance na identificação de retinopatia diabética (precisão de 0,91) e catarata (precisão de 0,92), mas com desempenho moderado em glaucoma e olhos normais.

O modelo MobileNetV2 alcançou uma acurácia de 0,92, destacando-se também na detecção de retinopatia diabética (precisão de 0,98) e catarata (precisão de 0,96), mostrando bom equilíbrio entre precisão e recall para todas as classes. Este modelo é vantajoso para aplicações em dispositivos móveis devido ao menor número de parâmetros e eficiência computacional.

#### A. Escolha do Melhor Modelo

Com base nos resultados apresentados e no problema apresentado, o MobileNetV2 é considerado o melhor modelo para esta aplicação específica de classificação de doenças oculares. As principais razões para esta escolha são:

- **Desempenho Balanceado:** O MobileNetV2 apresentou um excelente equilíbrio entre precisão e recall em todas as classes, garantindo uma identificação robusta e confiável das doenças oculares. Por exemplo, a precisão para retinopatia diabética foi de 0,98, e para catarata foi de 0,96.
- **Eficiência Computacional:** Com um menor número de parâmetros (2.223.872) e menor uso de memória (24.87 MB), o MobileNetV2 é altamente eficiente, tornando-o ideal para implementação em dispositivos móveis e ambientes com recursos limitados.
- **Generalização:** O modelo demonstrou boa capacidade de generalização, com acurácia de validação consistente,

o que é crucial para aplicações práticas em contextos clínicos diversos.

Futuras melhorias podem incluir a integração de técnicas de aumento de dados e ajustes finos adicionais para aprimorar ainda mais a precisão e robustez dos modelos.

#### REFERENCES

- [1] Agência Internacional para a Prevenção da Cegueira (IAPB), "Vision atlas," 2020, disponível em: <https://www.iapb.org/learn/vision-atlas/>. [Online]. Available: <https://www.iapb.org/learn/vision-atlas/>
- [2] Ibope Inteligência, "Pesquisa sobre a saúde ocular dos brasileiros," 2020, disponível em: <https://www.ibopeinteligencia.com/>. [Online]. Available: <https://www.ibopeinteligencia.com/>
- [3] V. Gulshan *et al.*, "Development and validation of a deep learning algorithm for detection of diabetic retinopathy in retinal fundus photographs," *JAMA*, vol. 316, no. 22, pp. 2402–2410, 2016. [Online]. Available: <https://jamanetwork.com/journals/jama/fullarticle/2588763>
- [4] D. S. W. Ting *et al.*, "Development and validation of a deep learning system for diabetic retinopathy and related eye diseases using retinal images from multiethnic populations with diabetes," *JAMA*, vol. 318, no. 22, pp. 2211–2223, 2017. [Online]. Available: <https://jamanetwork.com/journals/jama/fullarticle/2665774>
- [5] W. Bismi and J. Na'am, "Classification of myopia levels using deep learning methods on fundus image," *Journal Medical Informatics Technology*, vol. 1, no. 2, pp. 42–48, 2023. [Online]. Available: <https://medinftech.org/index.php/medinftech/article/view/8/10>
- [6] S. Pachade, P. Porwal, D. Thulkar, M. Kokare, G. Deshmukh, V. Sahasrabudhe, L. Giancardo, G. Quéllec, and F. Mériaudeau, "Retinal fundus multi-disease image dataset (rfmid)," 2020. [Online]. Available: <https://dx.doi.org/10.21227/s3g7-st65>
- [7] N. E. Khalifa, M. Loey, and S. Mirjalili, "A comprehensive survey of recent trends in deep learning for digital images augmentation," *Artificial Intelligence Review*, no. 0123456789, 2021.
- [8] Z. Zhang, L. Yang, and Y. Zheng, "Translating and segmenting multimodal medical volumes with cycle- and shape-consistency generative adversarial network," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2018, pp. 9242–9251. [Online]. Available: [https://openaccess.thecvf.com/content\\_cvpr\\_2018/papers/Zhang\\_Translating\\_and\\_Segmenting\\_CVPR\\_2018\\_paper.pdf](https://openaccess.thecvf.com/content_cvpr_2018/papers/Zhang_Translating_and_Segmenting_CVPR_2018_paper.pdf)
- [9] G. Litjens *et al.*, "A survey on deep learning in medical image analysis," *Medical Image Analysis*, vol. 42, pp. 60–88, 2017.
- [10] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," 2015. [Online]. Available: <https://arxiv.org/abs/1512.03385>
- [11] M. Tan and Q. V. Le, "Efficientnet: Rethinking model scaling for convolutional neural networks," 2020. [Online]. Available: <https://arxiv.org/abs/1905.11946>
- [12] M. Sandler, A. Howard, M. Zhu, A. Zhmoginov, and L.-C. Chen, "Mobilenetv2: Inverted residuals and linear bottlenecks," 2019. [Online]. Available: <https://arxiv.org/abs/1801.04381>
- [13] W. McKinney, "Data structures for statistical computing in python," in *Proceedings of the 9th Python in Science Conference*, 2010, pp. 51–56.
- [14] C. R. Harris, K. J. Millman, S. J. van der Walt *et al.*, "Array programming with numpy," *Nature*, vol. 585, pp. 357–362, 2020.
- [15] J. D. Hunter, "Matplotlib: A 2d graphics environment," *Computing in Science & Engineering*, vol. 9, no. 3, pp. 90–95, 2007.
- [16] M. L. Waskom, "Seaborn: Statistical data visualization," *Journal of Open Source Software*, vol. 6, no. 60, p. 3021, 2021.
- [17] G. Bratski, "The opencv library," *Dr. Dobbs Journal of Software Tools*, 2000.
- [18] A. Paszke, S. Gross, F. Massa *et al.*, "Pytorch: An imperative style, high-performance deep learning library," in *Advances in Neural Information Processing Systems*, vol. 32, 2019, pp. 8024–8035.
- [19] F. Pedregosa, G. Varoquaux, A. Gramfort *et al.*, "Scikit-learn: Machine learning in python," *Journal of Machine Learning Research*, vol. 12, pp. 2825–2830, 2011.